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

Kibrom A. Abay*, Gashaw T. Abate, Christopher B. Barrett, and Tanguy Bernard

February 2019 revision

Abstract:

We show that non-classical measurement errors (NCME) on both sides of a regression can bias the parameter estimate of interest in either direction. Furthermore, if these NCME are correlated, correcting for either one alone can aggravate bias relative to ignoring mismeasurement in both variables, a 'second best' result with implications for a broad class of economic phenomena of policy interest. We then use a unique Ethiopian dataset of matched farmer self-reported and precise ground-based measures for both plot size and agricultural output to re-investigate the long-debated relationship between plot size and crop productivity. Both self-reported variables contain substantial NCME that are negatively correlated with the true variable values, and positively correlated with one another, consistent with prior studies. Eliminating both sources of NCME eliminates the estimated inverse size-productivity relationship. But correcting neither variable generates a parameter estimate not statistically significantly different from that generated using two improved measures, while correcting for just one source of NCME significantly aggravates the bias in the parameter estimate. Numerical simulations demonstrate that over a relatively large parameter space, expensive collection of objective measures of only one variable or correcting only one variable's NCME may be inadvisable when NCME are large and correlated. This has practical implications for survey design as well as for estimation using existing survey data.

Keywords: Agricultural development, bias, Ethiopia, measurement, smallholder agriculture.

Acknowledgements:*Corresponding author at k.abay@cgiar.org.Affiliations: Abay: International Livestock Research Institute; Abate: International Food Policy Research Institute (IFPRI); Barrett: Cornell University; Bernard: IFPRI and University of Bordeaux. This paper benefited from comments by Marc Bellemare, Leah Bevis, Chris Boone, Brian Dillon, John Gibson, Kalle Hirvonen, Berk Ozler, Alemayehu Seyoum Taffesse, Chris Udry, and an anonymous reviewer, conference participants at the African Development Bank, International Conference of Agricultural Economists, and Agricultural and Applied Economics Association annual meetings, and seminar participant at Cornell University, Ethiopian Development Research Institute, and University of Bordeaux. Any remaining errors are the authors' sole responsibility.

1. Introduction

Measurement drives analysis. In recent years, empirical researchers have begun to devote considerable effort to more carefully measure 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). Many papers show the widespread prevalence of NCME and its relevance for policy inference, especially in 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. 2019), and agriculture (De Groote and Traoré, 2005; Carletto et al., 2013; Carletto et al., 2015; Gourlay et al., 2017). That literature sensibly suggests we employ better measurement methods so as to reduce error. Hence 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.

Yet in many domains, multiple variables fall prey to NCME and mismeasurements may correlate among variables. For example, survey respondents might systematically underreport or overreport multiple variables so as to reduce prospective tax liabilities or to increase the likelihood of eligibility for some benefit. Or unconscious error may arise from regression-to-the-mean or rounding (also known as 'focal point bunching') of multiple variables of interest. Or respondents might use one mismeasured variable to generate an optimal prediction of another variable (Hyslop and Imbens, 2001). Mismeasured attitudes may be correlated because respondents unconsciously respond to questions in a way that is coherent with answers they gave to previous questions, or because of their concern to give socially acceptable answers (Bertrand and Mullainathan, 2001).

What happens when the dependent and one of the independent variables in a regression suffer from correlated NCME, and what should an applied researcher do about it? If multiple variables are measured with error but only some are amenable to correction, does correction for just one mismeasured variable necessarily reduce bias and improve inference with respect to a parameter estimate of interest? Correlated NCMEs matter for the same reason that omitted relevant

[1] See in particular the special issue of the *Journal of Development Economics* on measurement and survey design, introduced by McKenzie and Rosenzweig (2012), and Ozler (2013)'s Development Impact blog entry.

variables matter because each NCME is, by definition, correlated with a relevant variable. With multiple NCME, if positively correlated measurement errors generate biases of opposing signs, or if negatively correlated errors each produce bias of the same sign, then correcting for only one source of NCME could inadvertently increase bias. In such cases, if one cannot correct for both sources of measurement error, a 'second best' estimate based on multiple NCME may, ironically, 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 hours worked, 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 children's ages used to construct standard anthropometric indicators such as height-for-age, can significantly bias estimates of the determinants of child health (Larsen et al., 2019).

We explore this issue analytically and then empirically as it relates to 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 or transactions costs (e.g., Sen, 1966; Feder, 1985; Barrett, 1996; Lamb, 2003; Foster and Rosenzweig 2017), omitted land attributes, including soil quality (Benjamin, 1995; Lamb, 2003; Assuncao and Braido, 2007; Barrett et al., 2010), or 'edge effects' arising for biophysical or farmer behavioral reasons (Bevis and Barrett, 2018).

Lamb (2003) speculated that measurement error in farm size might account for the observed inverse relationship. Recently, improvements in agricultural data collection have allowed researchers to explore the implication of measurement errors in both self-reported production and farm or plot size.[2] Some papers have examined the implication of improved, GPS measurement of the surface area of plots for estimation of the SPR (Carletto et al., 2013; Holden and Fisher, 2013;

[2] We use the terms 'output' and 'production' synonymously, and similarly 'area' and 'size'.

Carletto et al., 2015). A few recent 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). These papers find that NCME in self-reported production drives a spurious inverse relationship, conditional on GPS plot size measures. The inverse relationship disappears upon using crop-cuts output measures instead.

While these recent studies explore the implication of NCME in either area or production, no study has yet considered both measurement problems in a unified framework, much less generalized them beyond the specific SPR case. By studying correlated NCME in a more general setting, we can reconcile prior findings in the SPR literature and tease out far broader lessons.

In what follows, we first set up a general framework that allows for potentially correlated NCME in both output and area. We analytically characterize the implication of alternative features of NCME in output and area on the estimated SPR. We then empirically demonstrate our analytical findings, employing both self-reported and objective measures of output and area from an agricultural household survey in Ethiopia. We use crop-cut output data and area based on compass-and-rope method,[3] each widely considered the gold standard measures (Schoning et al., 2005; Keita and Carfagna, 2009; Fermont and Benson, 2011; Carletto et al., 2015; Carletto et al., 2016). By employing these four different measures of farm size and production, we illustrate empirically the patterns predicted by our analytical results 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 suffer from correlated NCME, the effect of these measurement errors on the estimated parameter of interest is analytically ambiguous. This appears to be the first paper to provide a general analytical framework for understanding the inferential implications of multiple correlated NCMEs, and of their incomplete correction. We discuss the very general implications of these findings for survey design as well as econometric analysis.

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 under certain

[3] Also known as polygon or traverse measurement, this involves measuring the length of each side and the angle of each corner using a rope and a compass to calculate the plot surface area using trigonometry (De Groote and Traoré 2005; Schoning et al., 2005; Casley and Kumar, 1988). Fermont and Benson (2011) show that GPS estimates significantly underestimate smaller farm sizes while both methods perform comparably for larger plots (those greater than 0.5 ha). Thus, although time consuming, compass-and-rope remains the preferred approach for specialized data collection (Diskin, 1997; Schoning et al., 2005; Keita and Carfagna, 2009; Carletto et al., 2015; Carletto et al., 2016). Lobell et al. (2015) and Gourlay et al. (2017) also use high-resolution remote sensing-based measurements for estimating crop production, which shows promise.

conditions, 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' (Lancaster and Lipsey, 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 dataset the core findings of recent studies (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 generalize these prior findings, which might be easily misinterpreted to suggest that one should correct for NCME whenever possible, although the authors never make that assertion. We show conditions under which (incomplete) correction of measurement error might aggravate, rather than correct for, bias in SPR estimates.

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 show that, when both variables suffer from similar measurement errors, error in output measures 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).

2. Analytical Framework

Due to cost and logistical considerations, most survey data are collected through single household visits 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. The recall and aggregation of information often generates substantial errors. 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, farmer illiteracy or innumeracy may lead to significant, but random and symmetric (around the true value) measurement error (De Groote and Traoré, 2005). 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 holdings or crop

production so as to conceal wealth or income and thereby avoid taxes or be found eligible for proxy means tested benefits of various types (Diskin, 1997). Second, farmers may *unintentionally* introduce errors in the reported information, in a way that is related to the level of the true value. For example, 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.[4] These NCME mechanisms can vary systematically with true area cultivated, true output, or both. And the same mechanisms might easily generate correlated measurement error across multiple variables. Other examples relate to the use of traditional measurement units which are commonplace in many rural settings. Because these can vary between locations and farming systems (an in particular larger versus smaller farm size or outputs), they may introduce systematic errors upon conversion into more standardized units.

In this section we work through the analytics for the simple case of a bivariate regression, in order to make clear the core intuition of our analysis.[5] 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, X^*. Instead of true measures of production and land area 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:[6]

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

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

[4] These biases may affect other measures besides area and output, including labor use (Arthi et al., 2018). We abstract from that possibility here.
[5] More formal extension to the multivariate case is left to future work.
[6] This specification implies that measurement errors are assumed to be additive in their logarithmic transformed values and hence multiplicative in their original, untransformed state.

Let $Y^*$and X^*measure true (log) production and land area, respectively. The standard estimated relationship between partial productivity, expressed as crop yield (production/area), and area cultivated is (typically augmented by a vector of controls omitted here for the sake of focus):

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

Where $\beta \equiv \theta - 1$. We now analyze the effects of alternative forms of measurement errors in either production or area on the β estimate of interest, the SPR.

One can use this basic framework to examine four cases of NCME: (i) measurement error in the dependent variable, *u*, is correlated with the true outcome; (ii) *u* is correlated with the true explanatory variable; (iii) the measurement error in the explanatory variable, *v*, is correlated with its true value; and/or (iv) the measurement errors *u* and *v* are correlated. Readers interested in the technical details should review the derivations found in the online appendix for each of these four special cases. Here we focus just on the most general case, which happens also to be consistent with the data we employ below.

As shown in the online appendix, if we allow for both output and area measurement error to potentially be correlated with true plot size,

$$u = \lambda X^* + \zeta \tag{4}$$

$$v = \alpha X^* + \ell \tag{5}$$

*w*here ℓ and ζ are uncorrelated with both X^* and the error term in equation 1, then OLS estimation of the size-productivity relationship using self-reported area and output measures yields

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

where ρx_*^2 is the variance of X^*, $\rho\ell^2$is the variance of the random component of the measurement error in *X*, and π is the covariance of the measurement errors, *u* and *v*. Equation 6 provides a very general representation encompassing various types of classical and non-classical measurement errors as well as those affecting the dependent and independent variables. For example, the textbook attenuation bias associated with only classical measurement error in the explanatory variable of interest can be shown by setting $\lambda = \alpha = \rho\ell^2 = \pi = 0$. Similarly, we can show the special case of 'division bias' that may arise as area enters both the right and left-hand side of the

equation (3). Accordingly, area measurement error affects both the dependent and independent variables, generating a 'division bias' (Borjas, 1980). This bias arises in the case of classical measurement error (α=0) even in the absence of either measurement error in the dependent variable level (i.e., self-reported production, λ=0) or correlation between both types of measurement errors ($\pi = 0$) .This is reflected in the third term of equation 6: the simple existence of measurement error in X (i.e., $\rho\ell^2$>0) generates downward bias. In sum, even for classical measurement error, the appearance of X^* on both sides of the regression generates attenuation plus division bias. That problem goes largely unnoticed in the literature, although this issue will affect any dependent variable constructed as a ratio with a mismeasured denominator, as in the case of crop yields (e.g., tons/hectare), anthropometric indicators (e.g., weight/height, height/age, or weight/age), earnings per unit time measures, even per capita expenditure measures where one wants to address the equivalence scales across households of different sizes with misreported composition.

The first thing to recognize in equation 6 is that we cannot in general sign the bias of the OLS estimator of β, the true SPR parameter. Some components lead to attenuation bias, others to downward or upward bias depending on the specific values of the different structural parameters describing the NCME mechanisms. This refines and qualifies prior claims that measurement error in size or production spuriously generates the standard inverse SPR (Lamb, 2003; Carletto et al., 2015; Holden and Fisher, 2013; Carletto et al., 2015; Gourlay et al., 2017; Desiere and Jolliffe, 2018). 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.[7]

Second, even in the absence of correlation between measurement errors ($\pi = 0$) , if both area and output suffer from non-classical measurement error ($\alpha \neq 0, \lambda \neq 0$) then correcting for measurement error in one of the variables does not ensure an unbiased estimator of the SPR. Relatedly, even if we correct for measurement error in one measure, the bias in the OLS estimator could grow rather than shrink. For example, using GPS measures to eliminate NCME in area (i.e., let $\alpha = \rho\ell^2 = 0$), then the fourth term in equation 6 would still imply a biased estimate with downward bias for $\lambda<0$, as we, Gourlay et al. (2017), and Desiere and Jolliffe (2018) all find. That bias is no longer counterbalanced by the attenuation bias that comes from the random component

[7] Adding more covariates to equation 1 further complicates the prediction of the direction of the bias in the OLS estimator of the SPR parameter, 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, or soil quality.

of X eliminated by the improved area measure. As we discuss in section 5, for a range of reasonable parameter values, cleaning up one key measure might inadvertently aggravate bias in the inference that is a central objective of improved measurement, especially if the measurement errors are strongly positively correlated. This our 'second best' policy inference result; a SPR parameter estimate based on multiple NCME – i.e., on two uncorrected measures – may, ironically, be preferable in the sense of reduced bias to an estimate based on just one corrected measure.

Third, equation 6 indicates that we can assess the relative effects of the different types of measurement errors on bias in the parameter estimate. Such information can inform decisions as to whether to use subjective, error-ridden measures in econometric inference or whether to invest in more expensive collection of accurate, objective measures, as we discuss in section 5. Meanwhile, the first step in understanding the net effects of correcting one or both measures is to study the patterns of measurement error evident in the data.

3. Characterizing measurement error among wheat producers in Ethiopia

Based on the above analysis, we now empirically investigate how NCME affect the SPR amongst wheat farmers in Ethiopia. Our original sample consists of 504 farmers, randomly selected from 36 villages (*kebeles*) spanning 18 districts (*woredas*) of the Ethiopian wheat belt.[8] Farmers were interviewed in February and March 2014, a few months after the *meher* season harvest.[9] 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 these data arises from 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 harvest of one random subplot (4 meters × 4 meters), and weighing the cut output. Crop-cut wheat production was successfully measured on 365 of the sampled plots.[10]

[8] See Abate et al. (2015) for detailed discussion on the sampling design.
[9] *Meher* is the long (main) rainy and production season in Ethiopia.
[10] Among our original sample of 504 plots we miss some plot level information on 16 of the plots, effectively reducing our sample to 488 plots. Crop-cuts could not be measured for the remaining 123 plots for three reasons. First, 7 farmers had no wheat plot during the 2013 *meher* season. Second, 5 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. This latter reason

Table 1 presents the summary statistics of main household and plot level characteristics. The first four rows provide alternative measures of plot size and production for the subsample for which both crop-cut and self-reported production measures are available. 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 panel of the table, there are significant discrepancies between self-reported and objective measures of land area and production. The third panel of Table 1 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.[11]

In the remainder of this section, we explore the extent and nature of output and area measurement errors. We note that while crop-cut and compass-and-rope area measurements are considered the most reliable and objective measures of production and plot size, respectively, they may not be completely free of errors. For instance, the choice of subplots and associated extrapolation process in crop-cutting exercises may introduce sampling error due mainly to variations in the productivity of plot parts (e.g., interior vs. periphery or edge). However, these types of errors are expected to be uncorrelated with true measures of production, plot size, or relevant explanatory variables, implying that these inaccuracies are generally less systematic relative to self-reported measures.

3.1 Plot size

While measuring plot size through farmers' self-report in household surveys is least costly, self-reported measures can be subject to considerable measurement error. 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 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'

could introduce attrition bias in either direction. If harvest was early because the farmer chose a shorter cycle variety, yields would typically be lower. If early harvest arises, however, due to unusually robust growth, the missing yields would be above average. In Table A2, we show that these nonresponses are not systematic and appear uncorrelated with the household- and plot-level characteristics.

[11] 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).

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. Third, farmers may strategically report lower landholding, a means to avoid state taxes or involuntary redistribution or to qualify for social programs.

Figure 1 reports the error in self-reported plot size (blue columns), by categories of plot size as measured by the compass-and-rope (CR) gold standard. 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.[12]

Next, we explore potential sources of mismeasurement in self-reported plot sizes. Figure 2, panel (a) plots the self-reported plot size against actual plot size measured by CR. Most observations lie above the 45° line, indicating a clear tendency for farmers to over-estimate their plot size. Self-reported plot sizes also exhibit rounding error around values that correspond to the conversion factor between the common local unit and hectare (e.g. ½ oxen day=0.125 ha; 1 oxen day=0.25 ha). Rounding necessarily has larger proportional consequences for smaller plots. Measurement error could also arise 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., 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 2 reports correlates of measurement error in self-reported farm size, expressed as differences in logarithmic values of self-reported and CR plot size, i.e., ln(self-reported)-ln(CR).[13] To facilitate comparison of estimates, we restrict our sample to those plots for which crop-cuts are available.[14] Table 2 provides only estimates from our most exhaustive (i.e.,

[12] Appendix Table A3 provides further details on this distribution.
[13] Due to the skewed distribution of some of our variables (e.g., plot size), we also re-estimate all regressions using the inverse hyperbolic sine transformation of our main variables of interest (Burbidge et al., 1988). Results based on inverse hyperbolic sine transformation are similar in sign, significance and magnitude as those presented throughout this paper.
[14] Results based on the full sample are almost identical and available upon request.

full set of controls) regressions.[15] In all regressions, we control for village (*kebele*) fixed effects to capture village-level misreporting that might arise from, for example, local measurement units or enumerator-specific bias, since local units vary across villages and different villages were surveyed by different enumerators.[16] We consistently find a negative correlation between measurement error and true plot size. The magnitude of this correlation is larger than those reported by Carletto et al. (2013) and Carletto et al. (2015). That might arise because the GPS-based area measures they use are least accurate on the smallest plots (Schoning et al., 2005; Fermont and Benson 2011), where the measurement errors appear largest. 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.

3.2 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 that this extrapolation may introduce errors due mainly to variations in the productivity of plot parts (e.g., interior vs. periphery or edge).[17] However, one may account for crop-cut distance to edges to minimize such problems, as we do in all regressions.

Figure 1 reports the error in self-reported production (red columns), by categories of (CR) plot size. Farmers tend to over-estimate (or at least over-report) their wheat harvest, although the bias appears much more pronounced for smaller plots. Here also, we find average over-estimation of 250% for plots smaller than 0.125 hectares, and 150% for plots between 0.125 and 0.250

[15] Table A5 in the Appendix provides full set of results, including unconditional and conditional correlations.

[16] About 40 and 60 percent of the sample households used standard units for reporting their plot size and production, respectively. The use of standard units (e.g., kilograms (kg) or hectares (ha)) may introduce measurement error among those who do not commonly use these measures. For this reason, we control for these measurement units in our empirical work.

[17] For example, previous agronomic studies indicate that the periphery of a plot is often more productive than its interior (Little and Hills, 1978; Barchia and Cooper, 1996; Ward et al., 2016). More recently, Bevis and Barrett (2018) argue that this could be one explanation for the inverse size-productivity relationship. We explore that hypothesis below. Gourlay et al. (2017) show using crop-cuts of entire plots versus cuts on only samples of plots that sampling error in crop cuts does not affect results.

hectares. These biases significantly decrease as plots become larger, albeit remaining positive and statistically significant.[18]

Figure 2 panel (b) further confirms farmers' general tendency to over-estimate their production, with most observations lying above the 45° line. We do find clear evidence of rounding error (as was the case in panel (a)), mainly due to the fact that farmers often report their production estimates in bags of 50 to 100 kg each.

In column 2 of Table 2, we present correlates of measurement error in production, conditional on plot size.[19] Results confirm that measurement error in self-reported production is negatively correlated with true (CR) plot size measure. These correlations are much higher than those reported in Gourlay et al. (2017) and Desiere and Jolliffe (2018), which could likewise be due to their use of GPS-measures that are less accurate than the Compass-and-Rope method for smaller plots. We do not however uncover significant correlation between measurement error in production and the introduced household, farm and plot characteristics.

In Table 2, column 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 repeatedly documented in earnings (Bound and Krueger, 1991) and consumption (Gibson et al., 2015).[20] Measurement error in self-reported production is also correlated with soil quality once one controls for actual production.

3.3 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 both plot size and actual production. Table 2 column 4 reports estimates of the correlation 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 0.5 percent increase in measurement error in production.[21]

[18] See Appendix Table A4 for further details on this distribution.
[19] The full set of estimation results, including unconditional and conditional correlations, are given in Table A6 in the Appendix.
[20] Table A7 in the Appendix provides full regression results.
[21] For the sake of parsimony, we report here the results including all covariates, including true plot size, in order to net out the effects of other sources of errors. Point estimates are larger in magnitude when we do not include these covariates, and in particular the true plot size variable. We report the full results in Appendix Table A8.

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 posed in analytical section 2: (i) NCME in self-reported production manifest in a 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, reflected in a negative correlation between the measurement error and true plot size; 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.

4. 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, now augmented by control variables in vector Z:

$$Y^* - X^* = \beta_1 X^* + \tau_1' Z + \varepsilon_1 \tag{7}$$

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 2 and ε_1 is a mean zero error term. Equation 7 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:

$$Y - X^* = \beta_2 X^* + \tau_2' Z + \varepsilon_2 \tag{8}$$

$$Y^* - X = \beta_3 X + \tau_3' Z + \varepsilon_3 \tag{9}$$

$$Y - X = \beta_4 X + \tau_4' Z + \varepsilon_4 \tag{10}$$

Before estimating the parametric regressions in equations 7-10 we run unconditional non-parametric regressions corresponding to these specifications. Figure 3 suggests stark differences in the estimated SPR based on which measures one uses, with the weakest inverse relationship arising when using both error free variables or both error-ridden measures and the strongest arising when one uses just one corrected measure. This suggests the 'second best' inference result we raised earlier.

The conditional parametric results from estimating equations 7-10 are presented in Table 3.[22] For sake of comparability across estimations, we limit the sample to those plots with crop-cut estimates, although similar results are obtained using full sample (e.g., for regression 10) or the sub-sample with GPS measured plot size (e.g., for regression 8). Column 1 reports the benchmark estimates associated with equation 7. The estimated β_1 parameter is negative, 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 (2018) 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 8-10 against the null of $\beta_1 = 0$ and the benchmark $\widehat{\beta_1}$ estimate based on the gold standard values of wheat output and plot size.

Results in column 2 (corresponding to equation 8) show a large, negative and statistically significant β_2 estimate. Measurement error in production appears to substantially exaggerate the estimated inverse size productivity 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.

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, we showed that measurement error may lead to over- or under-estimation of the SPR depending on the relative magnitudes of the structural parameters δ and λ. In our data, the effect of production mismeasurement correlated with true plot size appears to dominate that of its correlation with true production. Overall, our

[22] Appendix Tables A9-A12 provide corresponding full regression results.

results suggest that using self-reported production leads to substantial overestimation of the magnitude of an inverse relationship, if it exists, and can lead to a spurious negative finding if there truly is no size-productivity relationship.

Column 3 presents the estimation results of equation 9, where production is correctly measured but plot size is not. NCME 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 effects on the SPR estimate. 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 plot area. Our descriptive statistics indicate that variance of the self-reported plot size is smaller than that of the true area (Table 1), implying a negative correlation between measurement error in plot size and true land area measure. Thus, OLS estimates using self-reported plot size will also tend to overestimate the inverse relationship. The results in column 3 support this prediction. 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 contrasts with the results in Carletto et al. (2013), likely because the consequences of plot size measurement error may vary across contexts, sources and empirical features of measurement errors, as we established analytically in section 2.

Finally, column 4 reports estimation results of equation 10, using self-reported plot size and production. The estimated β_4 parameter suggests a statistically significant, but relatively mild, inverse relationship between plot size and productivity. The magnitude of this point estimate is statistically significantly less than – and less than one-half or one-third as large as – 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. Furthermore, the point estimate in column 4 is not statistically significantly different from the parameter estimate that uses only corrected values (column 1). This is consistent with our analytical finding in equation 6, 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 other variable. In such a situation, ignoring both types of measurement errors appears 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 4 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 important differences among columns 2-4 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 output, 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 Table 3, columns 1 and 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; Lamb, 2003; Assuncao 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). 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 (2018) and Abay, Bevis, and Barrett (2019). 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.

5. So what to do about NCME in econometric analysis and data collection?

Having established that NCME is pervasive and commonly correlated among variables, and that partial correction of NCME in just one variable may aggravate rather than ameliorate bias in parameter estimates of interest, the crucial question is what to do. If one has available both self-reported and objectively-measured values of only some variables, should one use the corrected or uncorrected values? When might it be preferable to employ a "second-best" estimation strategy and ignore known measurement error rather than correcting just one variable's measurement error, assuming it is infeasible to correct the second variable's measurement error? And in designing data collection, under what conditions is it worthwhile to invest in (presumably more expensive) collection of more accurate measures?

We tackle the first question by numerical simulation, using our analytical findings from equation 6 and drawing on estimates for the various measurement error correlation coefficients

found across the literature. Table 5 reports various values found in relevant studies to provide a sense of the range established in roughly similar data sets. Extensions to related questions – e.g., the age profile of changes in height-for-age or weight-for-age z-scores, the hourly wage elasticity of labor supply, equivalence scales in per capita expenditures, etc. – are reasonably straightforward; analysts can either find similar estimates in existing data or make educated guesses.

For illustrative purposes, in the numerical simulations that follow, we use equation 6, assuming that there is no statistically significant relationship between farm size and productivity $(\beta = 0)$, so that the attenuation bias component (i.e., the first term) disappears. One could use some non-zero value instead, with some increase in complexity. We illustrate numerically the parameter space over which the option of ignoring known measurement error is likely to reduce bias compared to the seemingly-better 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 unit variance of the mismeasured and correctly measured explanatory variables $(\rho x_*^2 = \rho x^2 = \rho v^2 = \rho u^2 = 1, \rho \ell^2 = 0)$. We then consider the most common values of correlations between measurement errors and true measures, values established by previous studies and shown in Table 5.

Figure 4 depicts RB as a function of the correlation between measurement errors. We compare the consequences of ignoring both mismeasurements versus correcting measurement error in the explanatory variable only (and leaving mismeasurement in the dependent variable). We consider four values of correlations between measurement errors and true values $(\alpha$ and $\lambda)$, ranging from -0.2 to -0.8. All the four graphs under Figure 4 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 graphs reveals three important points. First, the second-best approach can dominate no matter the sign of the correlation between the measurement errors.

Second, as the correlation between measurement error and true values (α and λ) increases, so does the region where RB<0, suggesting the superiority of the second-best approach over incomplete correction. Third, even if the measurement errors are perfectly correlated, as in the case of $corr(u, v) = 1$, with α=λ=-0.8 in the lower-right panel, correcting the one measurement error might still not be superior.

Figure 5 replicates the exercise but now for RB when the econometrician has the option to correct measurement error in the dependent variable only. The same basic pattern emerges but with some important differences: the parameter space over which the second-best approach of not correcting for measurement error dominates is appreciably smaller. Similarly, unlike the patterns in Figure 4, the region RB<0 does not meaningfully increase with the size of correlations between measurement error and true values (α and λ). These pieces of evidence reinforce one of our key points, 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 A1-A2 (in the appendix) offer three-dimensional representations of the same phenomena, relaxing some key assumptions imposed in Figures 4-5. In Figure A1, 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 A2 offers a similar comparison when the option is to correct measurement error in output rather than area. Beyond reinforcing the findings from Figures 4 and 5, these plots add important nuance, such as the importance of the correlation between the measurement error and the true value of the independent variable. If α≥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 of correlation 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.

Overall, our findings can offer imperfect, but hopefully-useful, rules of thumbs to the applied econometrician confronted with the decision of whether or not to use objective data for either the dependent or the independent variable – assuming that objective data are not available on both variables – or to rely on error-ridden subjective reports by respondents for both variables

in inferential analysis.[23] First, partial correction, using an improved measure of just one variable, is more likely preferable to no correction when the corrected variable is the dependent one (in our case, production). Note that under purely classical measurement error one reaches precisely the opposite conclusion. Second, if one suspects a high magnitude of correlations between the measurement error in the dependent or independent variables and the true value of the independent variable (the parameters λ and α, respectively), then the use of uncorrected measures will typically be preferable to partial correction when the covariance between the errors π is expected to be high, as is evident in the data we use in this study.

When considering the implications of our finding for the design of data collection, potential improvements in inference (and description) should be weighed against costs. If collection of more accurate data were costless, it would always be first best. But on the assumption that collection of more accurate measures adds costs, and that the opportunity cost of such investments is positive, an affirmative finding to the prior question, 'should we use improved measures over self-reported ones?', is necessary but not sufficient to justify the extra data collection. The cost-benefit analysis here would depend on the marginal value to the analyst – or her client(s) – of reduced bias in the parameter estimate of interest and in the descriptive evidence on variables of interest, weighed against the marginal cost of the extra data collection. The high cost of crop cutting unfortunately matches the greater benefit of improved measures of output. This underscores the prospective gains from accurate, low cost, remotely sensed measures of crop production presently being studied by various research groups.

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 might be correlated. We show that the signs and magnitude of resulting biases in estimates of a key parameter are analytically ambiguous and depend on several parameters characterizing measurement errors in these variables as well as the relationship under

[23] Of course, for descriptive purposes, the more accurate, objectively measured data are always better. The issue we address concerns inference about key parameters of interest.

investigation. We also show that correcting 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 in either or both of the variables (e.g., through GPS data for area cultivated, and/or crop-cuts for production), none has investigated the relationship more generally, by addressing measurement issues in both variables in a more general way, 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 wheat output and area cultivated in Ethiopia. These data enable us to empirically validate our analytical 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. Indeed, it may substantially aggravate bias in the parameter estimate of interest as compared to not correcting measurement error in either variable.

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] Toward that end, we also use numerical simulation to offer some rough rules of thumb as to when it does, or does not, make sense to correct for measurement error in just one among several potentially mismeasured variables.

[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., de Brauw, A., Minot, N., Bernard, T., 2015. The impact of the use of new technologies on farmers' wheat yield in Ethiopia: Evidence from a randomized controlled trial. IFPRI Discussion Paper 1462. Washington, D.C.

Abay, K., 2018. Measurement Errors in Agricultural Data and their Implications on Marginal Returns to Modern Agricultural Inputs. Working paper.

Abay, K., Bevis, L.E.M., Barrett, C.B. 2019. Measurement Error Mechanisms Matter: Agricultural intensification with farmer misperceptions and misreporting. Working paper.

Arthi, V., Beegle, K., De Weerdt, J., Palacios-López, A., 2018. Not your average job: measuring farm labor in Tanzania. *Journal of Development Economics* 130, 160-172.

Assuncao, J. J., Braido, L.H., 2007. Testing household-specific explanations for the inverse productivity relationship. *American Journal of Agricultural Economics* 89(4), 980–990.

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

Barchia, I., Cooper, J.L., 1996. Edge row effects on yields in an oat trial. *Animal Production Science* 36(4), 459–461.

Barrett C.B., Bellemare, M., Hou, J., 2010. Reconsidering Conventional Explanations of the Inverse Productivity–Size Relationship. *World Development* 38 (1), 88–97.

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

Beegle, K., De Weerdt, J., Friedman, J., Gibson, J., 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.

Bertrand, M., Mullainathan, S., 2001. Do people mean what they say? Implications for subjective survey data. *American Economic Review*, *91*(2), pp.67-72.

Bevis, L., Barrett, C. B., 2018. Close to the edge: High productivity at plot peripheries and the inverse size-productivity relationship. Working paper.

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

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

Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data, in Heckman J. and Leamer E. (eds), *Handbook of Econometrics*, Vol. 5, Amsterdam: Elsevier, pp. 3705–3843.

Bound, J., Krueger, A. B., 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., Magee L., Robb L., 1988. Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association* 83, 123-127.

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

Carletto, C., Savastano, S., Zezza A., 2013. Fact or Artefact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship. *Journal of Development Economics*, 103: 254–61.

Carletto, G., Gourlay, S., Murray, S., Zezza, A., 2016. Cheaper, faster, and more than good enough: is GPS the new gold standard in land area measurement? World Bank Policy Research Working Paper 7759.

Casley, D. J., Kumar, K., 1988. *The Collection, Analysis and Use of Monitoring and Evaluation Data*. Baltimore, MD: Johns Hopkins University Press.

Chao, L.W., Szrek, H., Peltzer, K., Ramlagan, S., Fleming, P., Leite, R., Magerman, J., Ngwenya, G.B., Pereira, N.S., Behrman, J., 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., Hammer, J., Sánchez-Paramo, C., 2012. The impact of recall periods on reported morbidity and health seeking behavior. *Journal of Development Economics*, *98*(1), 76-88.

De Groote H., Traoré O., 2005. The Cost of Accuracy in Crop Area Estimation. *Agricultural Systems*, 84 (1), 21–38.

De Janvry, A., Sadoulet, E., 2000. Rural poverty in Latin America: dterminants and exit paths. *Food Policy* 25 (2000) 389–409.

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

Diewert, W.E., Fox, K.J., 1999. Can measurement error explain the productivity paradox? *Canadian Journal of Economics*, *32*(2), 251-280.

Diskin, P., 1997. Agricultural Productivity Indicators Measurement Guide. Food and Nutrition Technical Assistance Project. Washington: US Agency for International Development.

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., Benson, T., 2011. Estimating Yield of Food Crops Grown by Smallholder Farmers, A Review in the Uganda Context. IFPRI Discussion Paper 01097, Washington, D.C.

Foster, A.D., Rosenzweig, M.R., 2017. Are there too many farms in the world? Labor-market transaction costs, machine capacities and optimal farm size. NBER working paper 23909.

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

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

Gibson J., Beegle K., De Weerdt J., Friedman J., 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.

Gourlay, S., Kilic, T., Lobell, D. 2017. Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for Inverse Scale - Productivity Relationship in Uganda. Paper Presented at the 2017 World Bank Land and Poverty Conference, March 23, 2017, Washington, DC.

Gottschalk, P., Huynh, M., 2010. Are Earnings Inequality and Mobility Overstated? The Impact of Non-Classical Measurement Error. *Review of Economics and Statistics* 2(2): 302–315.

Holden, S. T., Fisher, M., 2013. Can Area Measurement Error Explain the Inverse Farm Size Productivity Relationship? CLTS Working Paper No. 12, Norwegian University of Life Sciences, Ås, Norway.

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

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

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

Lamb, R.L., 2003. Inverse productivity: land quality, labor markets, and measurement error. *Journal of Development Economics* 71(1): 71– 95.

Larsen, A.F., Headey, D., Masters, W.A. 2019. Misreporting Month of Birth: Diagnosis and Implications for Research on Nutrition and Early Childhood in Developing Countries. *Demography*, in press.

Lipsey, R. G., Lancaster, K., 1956. The General Theory of Second Best. *Review of Economic Studies* 24 (1), 11–32.

Little, T.M., Hills, F.J., 1978. *Agricultural experimentations– Design and analysis*. John Wiley and Sons. New York, NY, USA.

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

McKenzie, D., Rosenzweig, M., 2012. Preface for symposium on measurement and survey design. *Journal of Development Economics* 98(1):1-2.

Özler, B., 2013. Economists have experiments figured out. What's next? (Hint: It's Measurement). World Bank Development Impact blog. https://blogs.worldbank.org/impactevaluations/economists-have-experiments-figured-out-what-s-next-hint-it-s-measurement.

Schoning P., Apuuli, E., Muwanga-Zake, E., 2005. Handheld GPS Equipment for Agricultural Statistic Surveys. Experiments on Area Measurements Done During Fieldwork for the Uganda Pilot Census of Agriculture, 2003. Oslo: Statistics Norway.

Sen A. K., 1966. Peasants and Dualism with or Without Surplus Labor. *Journal of Political Economy* 74 (5), 425–50.

Ward, B., Roe, B. E., Batte, M.T., 2016. Potential Profitability of Strip Intercropping with Corn and Soybeans. *Journal of the ASFMRA*.

Figures and Tables

Figure 1. Measurement error in plot size and production, as a function of accurate plot size

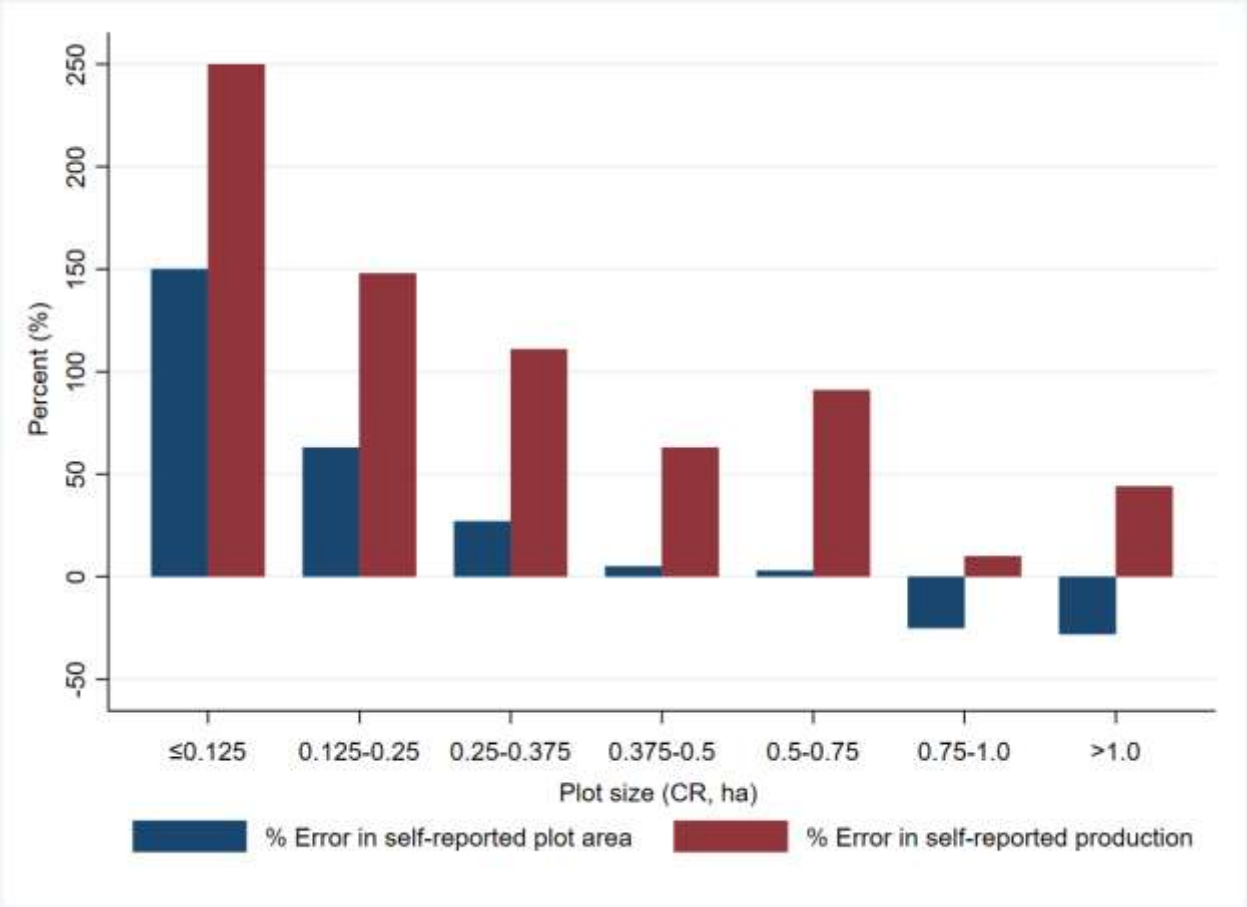

Figure 2. True vs self-reported plot size and production

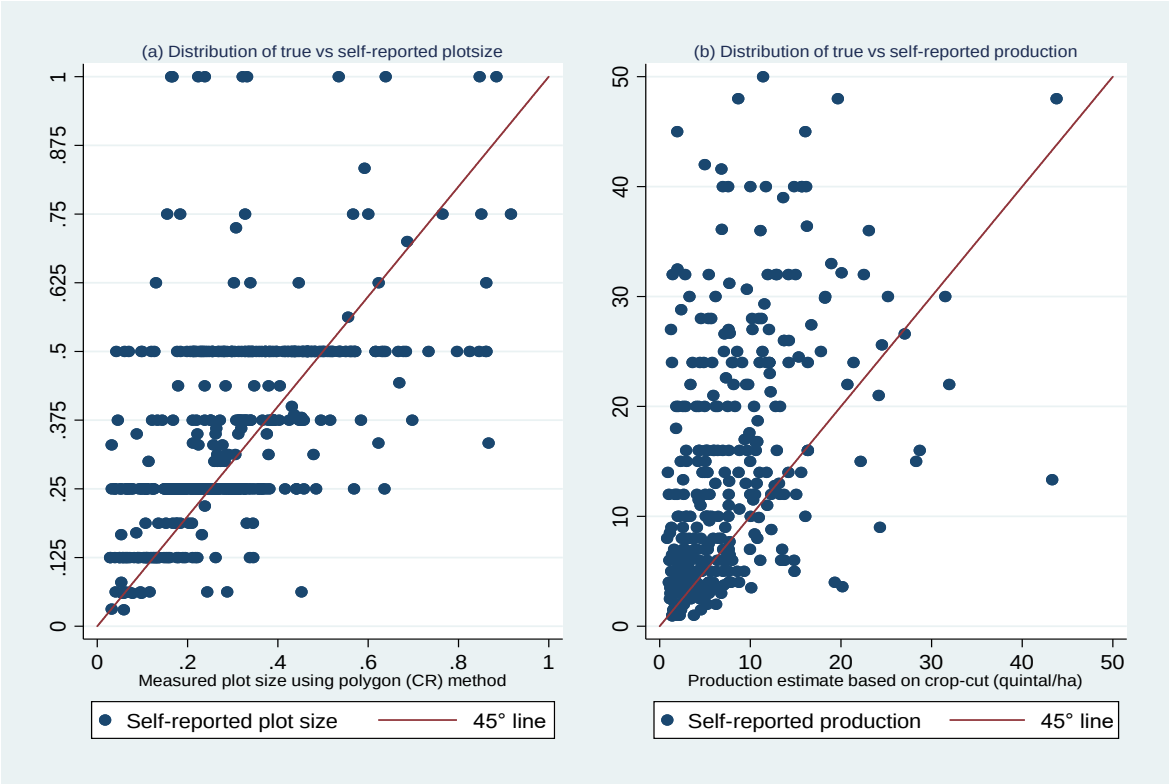

Figure 3: Non-parametric relationships between alternative measures of productivity and land area

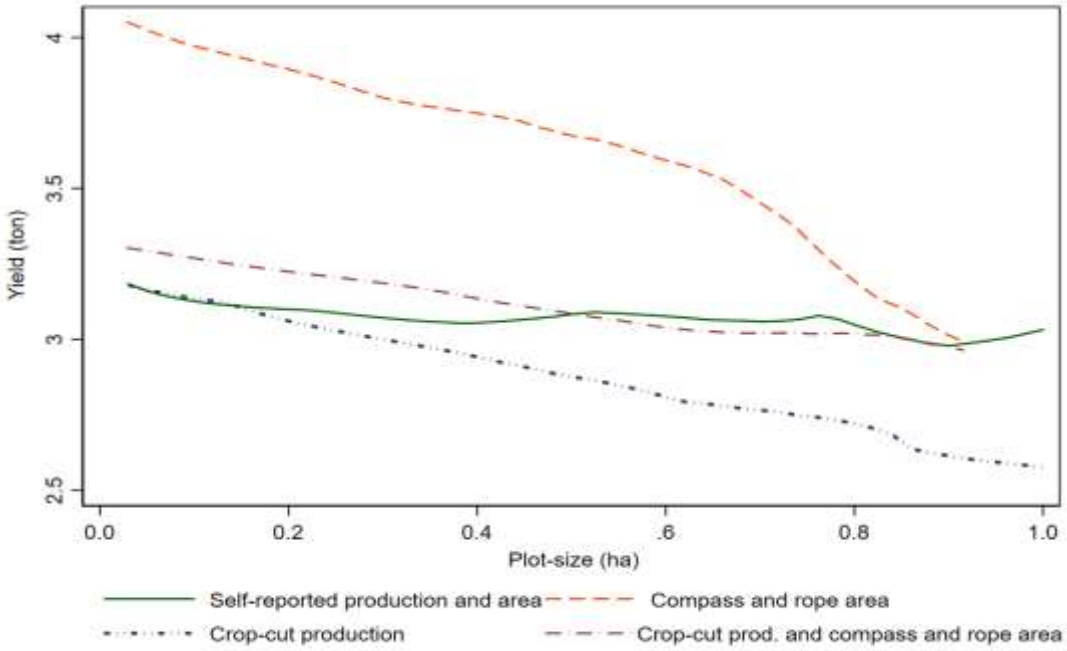

Figure 4: Relative bias in SPR, where RB>(<)0 implies correcting the mismeasured explanatory variable reduces (increases) bias when measurement error remains in the dependent variable.

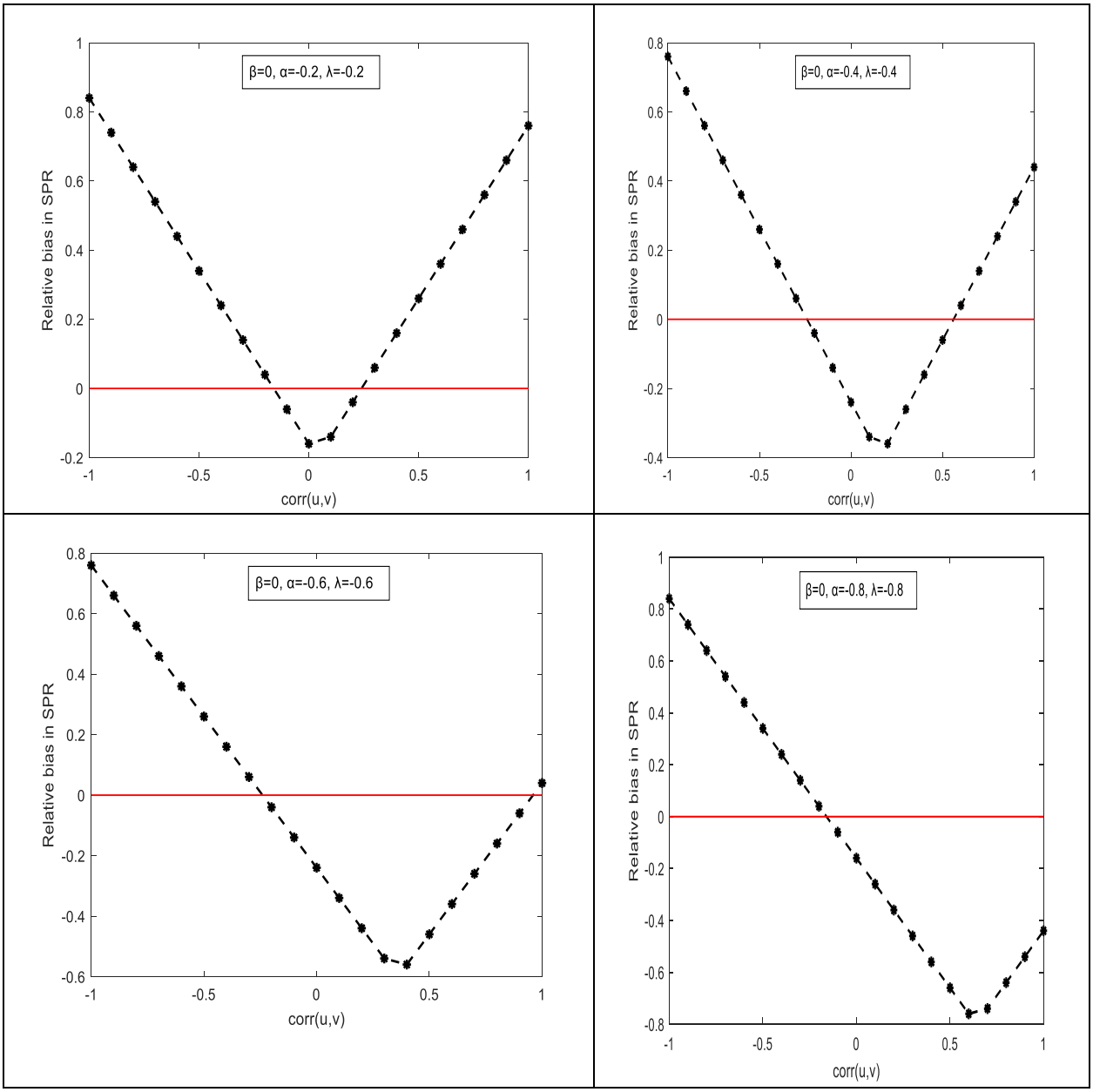

Figure 5: Relative bias in SPR, where RB>(<)0 implies correcting the mismeasured dependent variable reduces (increases) bias when measurement error remains in the explanatory variable.

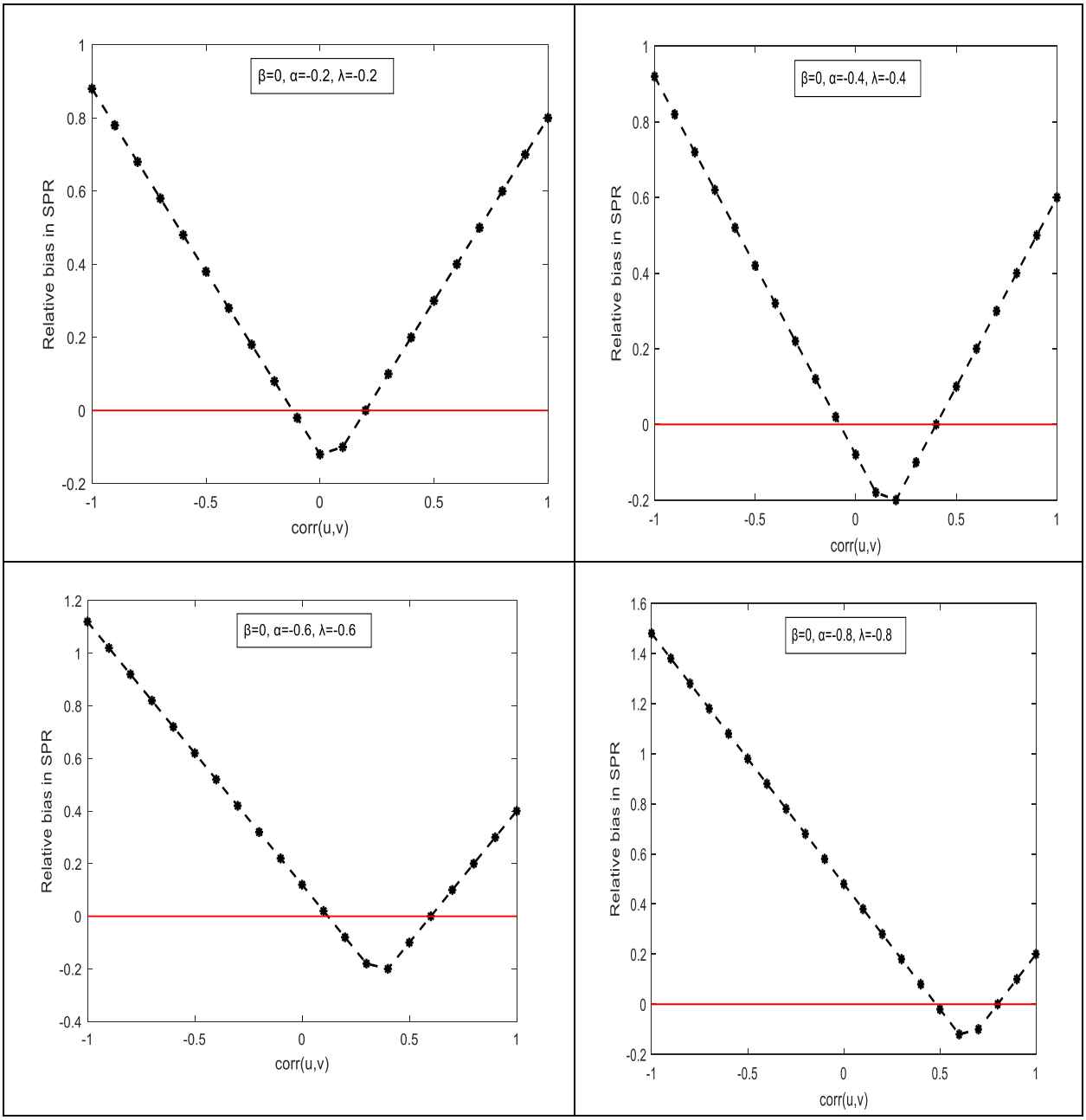

Table 1: Summary statistics

Variable	Description	Mean	Std. Dev.	Min	Max	Obs.
Area SR	Subjective self-reported area size (ha)	0.42	0.39	0.03	4.00	365
Area measured	Objectively measured area size during crop-cut (ha)	0.37	0.41	0.03	3.80	365
Production SR	Subjective self-reported production for reference plot (qt.)	17.53	17.03	0.94	120.00	365
Production measured	Objectively estimated production based on crop-cut (qt.)	8.98	9.91	0.81	101.5	365
Yield SR	Subjectively self-reported (production/area), (qt./ha)	27.32	15.98	1.00	96.00	365
Yield measured	Objectively measured (production/area), (qt./ha)	28.22	15.07	2.78	95.38	365
Age of HH head	Age of the household head in completed years	45.70	11.03	20.00	75.00	365
Gender of head of HH	Gender of head of household =1 if male	0.84	0.35	0.00	1.00	365
HH size	Number of household members	6.73	2.30	1.00	16.00	365
Literacy of HH head	=1 if the household head is literate	0.60	0.49	0.00	1.00	365
No. of corners	Number of corners of the reference plot	8.71	4.97	4.00	32.00	365
Closure error	Closure error in plot area measurement	1.10	0.91	0.02	4.50	365
Area unit†	=1 if farmers used ha for SR area measurement	0.44	0.49	0.00	1.00	365
Total owned area†	Total farm land owned by sample farmers	2.38	2.28	0.00	20.00	365
Crop-cut to edge	Distance between the crop-cut and shortest or closest plot edge (meters)	25.92	18.89	1.40	148.00	360
Production unit	=1 if farmers used kg for SR production measurement	0.61	0.48	0.00	1.00	365
Total wheat produced†	Total wheat production during 2013/14 *meher*	39.91	63.97	0.95	700.00	365
Soil fertility†						
High	=1 if the fertility of the reference plot is high	0.42	0.49	0.00	1.00	365
Medium	=1 if the fertility of the reference plot is medium	0.49	0.50	0.00	1.00	365
Poor	=1 if the fertility of the reference plot is poor	0.08	0.27	0.00	1.00	365
Soil color†						
Red	=1 if the color of the reference plot is red	0.28	0.45	0.00	1.00	365
Black	=1 if the color of the reference plot is black	0.52	0.50	0.00	1.00	365
Grey/sand	=1 if the color of the reference plot is grey or sandy	0.19	0.39	0.00	1.00	365
Distance to plot†	Walking distance between the dwelling and the plot (in minutes)	30.73	10.28	0.00	120.00	365
Plot ownership	=1 if the reference plot owned by the HH	0.82	0.37	0.00	1.00	365

Notes: † denotes that the values are self-reported by farmers during the household survey. HH refers to household.

Table 2. Correlates of measurement errors

Explanatory variables	ln (self-reported plot size/compass-and-rope plot size) (1)	ln (self-reported production/crop-cut production) (2)	ln (self-reported production/crop-cut production) (3)	ln (self-reported production/crop-cut production) (4)
ln (compass-and-rope plot size)	-0.532***	-0.558***		-0.246*
	(0.042)	(0.080)		(0.126)
ln (crop-cut production)			-0.670***	
			(0.055)	
ln (land area bias)				0.492***
				(0.114)
Age of HH head	0.010	0.009	0.016	0.010
	(0.015)	(0.022)	(0.024)	(0.021)
Age square	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Male HH head	-0.047	-0.020	0.032	-0.026
	(0.090)	(0.101)	(0.098)	(0.094)
Size of HH	-0.013	-0.013	-0.007	-0.013
	(0.011)	(0.013)	(0.014)	(0.015)
Education of HH head	0.007	0.022	-0.001	0.030
	(0.068)	(0.115)	(0.104)	(0.114)
Total landholding size	0.059***	0.034	0.041	0.007
	(0.017)	(0.027)	(0.024)	(0.029)
Soil fertility				
Medium	-0.098*	-0.059	-0.150**	-0.022
	(0.055)	(0.089)	(0.072)	(0.081)
Poor	-0.262***	-0.051	-0.289***	0.068
	(0.084)	(0.103)	(0.082)	(0.127)
Soil color				
Black	-0.137	0.121	-0.035	0.181
	(0.087)	(0.128)	(0.102)	(0.122)
Grey or sandy	-0.003	0.260*	0.234	0.275*
	(0.082)	(0.151)	(0.146)	(0.152)
Distance from home	0.004	0.003	0.003	0.001
	(0.003)	(0.004)	(0.004)	(0.004)
Distance to the edge	0.001	0.000	0.000	-0.004
	(0.002)	(0.003)	(0.003)	(0.003)
Number of corners	-0.002	-0.014	-0.002	-0.029**
	(0.008)	(0.010)	(0.010)	(0.012)
Own plot (1=yes)	-0.007	-0.130	-0.101	-0.125
	(0.075)	(0.090)	(0.081)	(0.097)
Village (*kebele*) dummies	Yes	Yes	Yes	Yes
Constant	-0.981**	-1.379**	0.739	-0.076
	(0.415)	(0.608)	(0.552)	(0.499)
Observations	360	360	360	360
R-squared	0.518	0.516	0.635	0.521

Notes: Standard errors are clustered at *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table 3. Plot size – productivity relationship

Explanatory variables	ln (crop-cut production/compass-and-rope plot size) (1)	ln (self-reported production/compass-and-rope plot size) (2)	ln (crop-cut production/self-reported plot size) (3)	ln (self-reported production/self-reported plot size) (4)
ln (compass-and-rope plot size)	-0.104	-0.662***		
	(0.063)	(0.074)		
ln (self-reported plot size)			-0.578***	-0.204***
			(0.077)	(0.073)
Age of HH head	0.012	0.021	-0.013	0.022
	(0.012)	(0.025)	(0.013)	(0.022)
Age square	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Male HH head	0.069	0.049	0.187*	0.067
	(0.085)	(0.109)	(0.109)	(0.084)
Size of HH	0.007	-0.006	0.037**	-0.002
	(0.010)	(0.014)	(0.018)	(0.018)
Education of HH head	-0.028	-0.006	-0.095	-0.033
	(0.073)	(0.105)	(0.078)	(0.084)
Total landholding size	0.011	0.045	-0.027	0.005
	(0.020)	(0.027)	(0.026)	(0.029)
Soil fertility				
Medium	-0.137***	-0.196***	-0.091	-0.063
	(0.049)	(0.067)	(0.085)	(0.063)
Poor	-0.363***	-0.414***	-0.162	-0.184
	(0.118)	(0.089)	(0.137)	(0.111)
Soil color				
Black	-0.237***	-0.116	-0.142	-0.062
	(0.069)	(0.092)	(0.092)	(0.089)
Grey or sandy	-0.043	0.217	0.009	0.127
	(0.074)	(0.151)	(0.126)	(0.121)
Distance from home	-0.001	0.002	-0.002	0.001
	(0.002)	(0.004)	(0.004)	(0.003)
Distance to the edge	-0.001	-0.001	0.011***	0.002
	(0.002)	(0.003)	(0.003)	(0.002)
Number of corners	0.012	-0.002	0.072***	0.014
	(0.012)	(0.010)	(0.012)	(0.010)
Own plot (1=yes)	0.045	-0.085	0.028	-0.065
	(0.065)	(0.085)	(0.094)	(0.071)
Village (*kebele*) dummies	Yes	Yes	Yes	Yes
Constant	3.351***	1.972***	1.682***	7.060***
	(0.426)	(0.608)	(0.443)	(0.566)
Observations	360	360	360	360
R-squared	0.562	0.607	0.535	0.476

Notes: Standard errors are clustered at *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table 4. Summary of estimation results

Source of NCME	Key NCME parameter estimates δ (1)	λ (2)	α (3)	π (4)	Estimated SPR (5)	Relative implication for estimates of the SPR
No error	NA	NA	NA	NA	-0.104 (0.063)	Weak, insignificant inverse SPR
Error in production	-0.670*** (0.055)	-0.558*** (0.080)	NA	NA	-0.662*** (0.074)	Strongest inverse SPR
Error in plot size	NA	NA	-0.532*** (0.042)	NA	-0.578*** (0.077)	Strong inverse SPR
Error in both	-0.670*** (0.055)	-0.558*** (0.080)	-0.532*** (0.042)	0.250	-0.204*** (0.073)	Weakest inverse SPR

Notes: we extracted the above estimates and standard errors (given in parenthesis) from our conditional regressions associated with equations (7)-(10). NA indicates that these parameters are either not relevant or not estimated.

Table 5. Summary of previously established correlations between measurement errors and true values

Source	Key NCME parameter estimates α (1)	λ (2)	π (3)
This study	-0.5	-0.6	0.25
Abay, Bevis and Barrett (2019)	[-0.3, -0.4, -0.5]	-	-
Abay (2018)	-0.3	-0.4	0.20
Carletto et al. (2015)	[-0.4, -0.6, -0.7, -0.8]	-	-
Gourlay et al. (2017)	-	[-0.2, -0.3]	-
Desiere and Jolliffe (2018)	-	-0.2	
Gibson et al. (2010)	[-0.3, -0.4, -0.7]	-	-
Gibson et al. (2015)	[-0.3, -0.4, -0.6]	-	-
Gottschalk and Huynh (2010)	-0.3	-	-

Notes: we extracted the above estimates from the list of studies given in column 1. We extract multiple parameter values from those studies involving multiple countries or multiple cases of mismeasurements.

Online Appendix to

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

Kibrom A. Abay,* Gashaw T. Abate, Christopher B. Barrett, and Tanguy Bernard

February 2019

* Corresponding author at k.abay@cgiar.org. Affiliations: Abay: International Livestock Research Institute; Abate: International Food Policy Research Institute (IFPRI); Barrett: Cornell University; Bernard: IFPRI and University of Bordeaux.

The first part of this appendix walks through the detailed derivations of the biases that arise under each of the four special cases of non-classical measurement error (NCME) nested within the general form discussed in the main text. The remainder of the appendix provides tables and figures that support the findings presented in the main text.

As a reminder, we are interested in the 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 (here we repeat equation numbers from the main text, supplementing them with new appendix equations, the latter carrying the prefix A):

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

We assume that the regression error term, ε, is mean zero and uncorrelated with the explanatory variable, X^*. Instead of true measures of production and land area 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:[1]

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

Now let $Y^*$and X^*measure true (log) production and land area, respectively. The standard estimated relationship between partial productivity, expressed as crop yield (production/area), and area cultivated is (typically augmented by a vector of controls omitted here for the sake of focus):

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

where $\beta \equiv \theta - 1$. We now analyze the effects of alternative forms of measurement errors in either production or area on the β estimate of interest – the SPR – under four distinct, realistic NCME mechanisms, wherein measurement error in the dependent variable numerator, *u*, is correlated with (i) the true outcome, and/or (ii) the true explanatory variable, (iii) the measurement error in the explanatory variable, *v*, is correlated with its true value, and/or (iv) the measurement errors *u* and *v* are correlated.

[1] This specification implies that measurement errors are assumed to be additive in their logarithmic transformed values and hence multiplicative in their original, untransformed state.

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

This case assumes that

$$u = \delta Y^* + \omega \tag{A1}$$

Where δ is any real number, ω is a random term uncorrelated with the explanatory variable, X^*, and the error term in equation (1). Such a mechanism would underpin, for example, a regression-to-the-mean process.[2] This implies that,

$$Y = (1+\delta)Y^* + \omega \tag{A2}$$

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

$$\beta^{OLS} = \frac{\text{cov}(Y - X^*, X^*)}{\text{var}(X^*)} = \frac{\text{cov}((1+\delta)(\beta+1)X^* + (1+\delta)\varepsilon + \omega - X^*, X^*))}{\text{var}(X^*)} = (1+\delta)\beta \tag{A3}$$

In the context of SPR, Case 1 implies measurement error in production that is correlated with the true output. If we assume, as we find empirically and as is typical under the regression-to-the-mean phenomenon, 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 the β^{OLS} estimate using self-reported production suffers attenuation bias for 0>δ>-1, takes the opposite-of-true sign for δ<-1, and is unbiased only in the case that $\delta = 0$ or -1. In most cases, 0>δ>-1, thus β^{OLS} attenuates more the true parameter estimate as the (negative) correlation between measurement error and true production, δ , increases (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 the dependent variable numerator, u, and the true explanatory variable value, X^*.[3] This type of measurement error may be more consequential in our context, at least in generating correlation across measurement errors in production and area, the case we analyze next.

[2] Focal point bunching is just a localized version of regression-to-the-mean, so could be represented by a generalized version of equation A1,$u = f(\delta Y^* + \omega)$ where the $f()$ mapping reflects a sequence across the conditioning domain.
[3] This is always the case if production is a deterministic function of land area.

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

This case assumes that

$$u = \lambda X^* + \zeta \tag{4}$$

where ζ is random noise uncorrelated with X^* and the error term in equation (1).[4] Using similar substitutions, one obtains the following expression:

$$\beta^{OLS} = \frac{\text{cov}(Y - X^*, X^*)}{\text{var}(X^*)} = \frac{\text{cov}((\beta+1)X^* + \varepsilon + \lambda X^* + \xi - X^*, X^*)}{\text{var}(X^*)} = \beta + \lambda \tag{A4}$$

Applied to SPR, Case 2 implies that output mismeasurement is correlated with plot size, as appears to be the case in our data and in Desiere and Jolliffe (2018) and Gourlay et al. (2017). Given a regression-to-the-mean and/or focal point bunching process behind NCME such that $\lambda<0$, then equation (A5) implies that β^{OLS} suffers downward bias, exaggerating the – or conjuring up a spurious – inverse size-productivity relationship. Following this reasoning, Desiere and Jolliffe (2018) and Gourlay et al. (2017) show that self-reported production measures can generate a spurious inverse relationship even when productivity appears invariant with respect to area when one uses more accurate, objective measures of both area and output.

So far, we have considered two cases of measurement errors in production that carry different implications for estimates of the SPR. If there truly is an inverse relationship (i.e., if $\beta < 0$), then the first case attenuates the inverse relationship while the second case amplifies it. If both cases 1 and 2 hold, then the overall net effect depends on the relative sizes and sources of the measurement error. Considering similar levels of correlations, the level differencing effect caused by Case 2 will generally dominate the fractional attenuation effect 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.

To this point, we have treated the explanatory variable as if it were not measured with error. When that is not true, we have a still-different complication due to NCME.

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

[4] The assumed linear relationship between measurement error and the true value of plot size is imposed for simplicity, but also reasonably consistent with our data.

$$v = \alpha X^* + \ell \tag{5}$$

where ℓ is uncorrelated with the true explanatory variable and the error term in equation (1). This, by substitution, implies that,

$$X = (1+\alpha)X^* + \ell \tag{A5}$$

Letting the variance of $X^* = \rho x_*^2$ and variance of $\ell = \rho\ell^2$, OLS estimation of the relationship in equation (3) using self-reported land area results in the following parameter estimate:[5]

$$\beta^{OLS} = \frac{\beta(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} - \frac{\alpha(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} - \frac{\rho\ell^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} \tag{A6}$$

This is the case of measurement error in land area assuming that we have an error-free measure of production.

Equation A6 offers 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, anthropometric indicators, earnings per unit time measures, etc. The first term in equation A6 reflects the special cases where the explanatory variable only appears on the right-hand side of equation 3. This expression simplifies further to the usual attenuation bias if the measurement error associated with the explanatory variable is classical ($\alpha = 0$).

The second and third terms in equation A6 arise 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. The second term again disappears if measurement error behaves classically ($\alpha = 0$). But the last term remains (and can generate inverse SPR) even for classical measurement error. Borjas (1980) called attention to this issue of 'division bias' in discussing estimation of the wage elasticity of labor

[5] To see this, consider $\beta^{OLS} = \frac{\text{cov}(Y^* - X, X)}{\text{var}(X)} = \frac{\text{cov}((\beta+1)X^* + \varepsilon - ((1+\alpha)X^* + \ell), ((1+\alpha)X^* + \ell))}{\text{var}((1+\alpha)X^* + \ell)}$ such that $\beta^{OLS} = \frac{(\beta-\alpha)(1+\alpha)\rho x_*^2 - \rho\ell^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2}$ which leads to equation A6.

supply. This makes even classical measurement error in explanatory variables of greater concern than is often recognized.

Importantly, we cannot know *a priori* the direction of bias associated with using *X*, self-reported area, in this case. 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.[6] 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 remaining terms in equation A6 render ambiguous the overall effect of inaccurate land area measurement.

Case 4*: Measurement errors in both dependent and independent variables are correlated,* $\mathrm{cov}(u,v) = \pi \neq 0$.

Using analogous substitutions and reformulation, we can show that OLS estimation of the size-productivity relationship using both self-reported measures yields the following estimator:[7]

$$\beta^{OLS} = \frac{\beta(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} - \frac{\alpha(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} - \frac{\rho\ell^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} + \frac{\lambda\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} + \frac{\pi}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2} \quad (6)$$

Equation 6 is our most general representation of the OLS estimator of β, 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

[6] In our data (Table 2), the variance of the self-reported area measure is smaller than that of the true area measurement, as is also reflected in the negative correlation between measurement error associated with farm size and true farm size.

[7] To see this, consider

$$\beta^{OLS} = \frac{\mathrm{cov}(Y-X,X)}{\mathrm{var}(X)} = \frac{\mathrm{cov}((\beta+1)X^* + \varepsilon + u - (X^* + v), X^* + v)}{\mathrm{var}((1+\alpha)X^* + \ell)} = \frac{(\beta-\alpha)(1+\alpha)\rho x_*^2 - \rho\ell^2 + \lambda\rho x_*^2 + \pi}{(1+\alpha)^2\rho x_*^2 + \rho\ell^2}$$

by setting $\lambda = \alpha = \rho\ell^2 = \pi = 0$. Similarly, we can show that ignoring measurement error in self-reported production $(\lambda = 0)$ and correlations between both types of measurement errors ($\pi = 0$) results in the special case of equation A6.

Table A1 summarizes the findings of these four cases.

Table A1: Summary of Analytical Results

Source of NCME	Key Parameters				OLS estimator of SPR	Bias (assuming $\beta \leq 0$)
	δ	λ	α	π		
No error	0	0	0	0	β	No bias
Error in production	<0	0	0	0	$(1+\delta)\beta$	Positive/attenuation*
Error in production	≤0	≤0	0	0	$\beta + \lambda$	Negative/exaggeration
Error in plot size	≤0	0	≤0	0	$\frac{\beta(1+\alpha)\rho x_*^2 - \alpha(1+\alpha)\rho x_*^2 - \rho\ell^2}{(1+\alpha)^2 \rho x_*^2 + \rho\ell^2}$	Ambiguous
Error in both	≤0	≤0	≤0	0	$\frac{\beta(1+\alpha)\rho x_*^2 - \alpha(1+\alpha)\rho x_*^2 - \rho\ell^2 + \lambda\rho x_*^2}{(1+\alpha)^2 \rho x_*^2 + \rho\ell^2}$	Ambiguous
Correlated error in both	≤0	≤0	≤0	>0	$\frac{\beta(1+\alpha)\rho x_*^2 - \alpha(1+\alpha)\rho x_*^2 - \rho\ell^2 + \lambda\rho x_*^2 + \pi}{(1+\alpha)^2 \rho x_*^2 + \rho\ell^2}$	Ambiguous

Notes: we rely on our data and the existing literature (cited in the main paper) to suggest the likely sign of the key parameters of interest. The parameter δ reflects the association between output measurement error and true output, per case 1, equation A1. The parameter λ reflects the association between output measurement error and true plot size, per case 2, equation 4. The parameter α reflects the association between area measurement error and true plot size, per case 3, equation 5. And the parameter π reflects the covariance between output and area measurement error, per case 4, equation 6.

* Assuming 0>δ>-1, as is generally true in measurement error studies.

Table A2: Characterizing non-responses in crop-cut production

Explanatory variables	Dependent variable: Crop-cuts (1=yes)
	(1)
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

Note: Standard errors are clustered at *kebele* level and given in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A3. Discrepancy between compass-and-rope and self-reported plot size

Plot size group (CR)	Number of obs.	Self-Reported (SR) (1)	Compass-and-Rope (CR) (2)	Bias (SR) – (CR)		Difference in mean (p-value) (5)
				Bias=(1)-(2) (3)	%Bias=(3)/(2) (4)	
≤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 A4. Discrepancy between crop-cut and self-reported production

Plot size group (CR)	Number of obs.	Self-reported (SR) (1)	Crop-cut (CC) (2)	Bias (SR) – (CC)		Difference in mean (p-value) (5)
				Bias=(1)-(2) (3)	%Bias=(3)/(2) (4)	
≤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.

Table A5: Correlates of measurement error in land area

Explanatory variables	Dependent variable: ln (self-reported area/compass-and-rope plot size)		
	(1)	(2)	(3)
ln (compass-and-rope plot size)	-0.550***	-0.540***	-0.532***
	(0.045)	(0.044)	(0.042)
Age of HH head		0.014	0.010
		(0.015)	(0.015)
Age squared		-0.000	-0.000
		(0.000)	(0.000)
Gender of HH head		0.005	-0.047
		(0.088)	(0.090)
Size of HH		-0.020*	-0.013
		(0.011)	(0.011)
Education of HH head		0.017	0.007
		(0.066)	(0.068)
Total landholding size		0.061***	0.059***
		(0.016)	(0.017)
Soil fertility			
Medium			-0.098*
			(0.055)
Poor			-0.262***
			(0.084)
Soil color			
Black			-0.137
			(0.087)
Grey or sandy			-0.003
			(0.082)
Distance from home			0.004
			(0.003)
Distance to the edge			0.001
			(0.002)
Number of corners			-0.002
			(0.008)
Own plot (1=yes)			-0.007
			(0.075)
Village (*kebele*) dummies	Yes	Yes	Yes
Constant	-0.889***	-1.162***	-0.981**
	(0.090)	(0.341)	(0.415)
Observations	365	365	360
R-squared	0.463	0.494	0.518

Notes: Standard errors are clustered at *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A6: Correlates of measurement error in production

Explanatory variables	Dependent variable: ln (self-reported production/crop-cut production)		
	(1)	(2)	(3)
ln (compass-and-rope plot size)	-0.596***	-0.590***	-0.558***
	(0.073)	(0.058)	(0.080)
Age of HH head		0.020	0.009
		(0.024)	(0.022)
Age square		-0.000	-0.000
		(0.000)	(0.000)
Gender of HH head		0.004	-0.020
		(0.105)	(0.101)
Size of HH		-0.018	-0.013
		(0.019)	(0.013)
Education of HH head		0.018	0.022
		(0.087)	(0.115)
Total landholding size		0.024	0.034
		(0.022)	(0.027)
Soil fertility			
Medium			-0.059
			(0.089)
Poor			-0.051
			(0.103)
Soil color			
Black			0.121
			(0.128)
Grey or sandy			0.260*
			(0.151)
Distance from home			0.003
			(0.004)
Distance to the edge			0.000
			(0.003)
Number of corners			-0.014
			(0.010)
Own plot (1=yes)			-0.130
			(0.090)
Village (*kebele*) dummies	Yes	Yes	Yes
Constant	-1.327***	-1.793***	-1.379**
	(0.147)	(0.574)	(0.608)
Observations	365	365	360
R-squared	0.495	0.501	0.516

Notes: Standard errors are clustered at *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A7: Correlates of measurement error in production

Explanatory variables	Dependent variable: ln (self-reported production/crop-cut production)		
	(1)	(2)	(3)
ln (crop-cut production)	-0.656***	-0.658***	-0.670***
	(0.055)	(0.043)	(0.055)
Age of HH head		0.028	0.016
		(0.021)	(0.024)
Age square		-0.000	-0.000
		(0.000)	(0.000)
Gender of HH head		0.072	0.032
		(0.092)	(0.098)
Size of HH		-0.016	-0.007
		(0.017)	(0.014)
Education of HH head		0.013	-0.001
		(0.076)	(0.104)
Total landholding size		0.038*	0.041
		(0.020)	(0.024)
Soil fertility			
Medium			-0.150**
			(0.072)
Poor			-0.289***
			(0.082)
Soil color			
Black			-0.035
			(0.102)
Grey or sandy			0.234
			(0.146)
Distance from home			0.003
			(0.004)
Distance to the edge			0.000
			(0.003)
Number of corners			-0.002
			(0.010)
Own plot (1=yes)			-0.101
			(0.081)
Village (*kebele*) dummies	Yes	Yes	Yes
Constant	0.987***	0.324	0.739
	(0.093)	(0.502)	(0.552)
Observations	365	365	360
R-squared	0.609	0.617	0.635

Notes: Standard errors are clustered at *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A8: Correlation between both types of measurement errors

Explanatory variables	Dependent variable: ln (self-reported production/crop-cut production)		
	(1)	(2)	(3)
ln (land area bias)	0.623***	0.542***	0.492***
	(0.091)	(0.105)	(0.114)
Ln (CR Plot size)		-0.310**	-0.246*
		(0.125)	(0.126)
Age of HH head			0.010
			(0.021)
Age square			-0.000
			(0.000)
Gender of HH head			-0.026
			(0.094)
Size of HH			-0.013
			(0.015)
Education of HH head			0.030
			(0.114)
Total landholding size			0.007
			(0.029)
Soil fertility			
Medium			-0.022
			(0.081)
Poor			0.068
			(0.127)
Soil color			
Black			0.181
			(0.122)
Grey or sandy			0.275*
			(0.152)
Distance from home			0.001
			(0.004)
Distance to the edge			-0.004
			(0.003)
Number of corners			-0.029**
			(0.012)
Own plot (1=yes)			-0.125
			(0.097)
Village (*kebele*) dummies	Yes	Yes	Yes
Constant	-0.263***	-0.194***	-0.076
	(0.020)	(0.039)	(0.499)
Observations	365	365	360
R-squared	0.481	0.494	0.521

Notes: Standard errors are clustered at *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A9: Benchmark results: plot size-productivity relationship

(correcting for both area and production measurement errors)

Explanatory variables	Dependent variable: ln (**crop-cut production/compass-and-rope plot size)**		
	(1)	(2)	(3)
ln (compass-and-rope plot size)	-0.083**	-0.086*	-0.104
	(0.040)	(0.042)	(0.063)
Age of HH head		0.012	0.012
		(0.012)	(0.012)
Age square		-0.000	-0.000
		(0.000)	(0.000)
Gender of HH head		0.098	0.069
		(0.082)	(0.085)
Size of HH		0.002	0.007
		(0.009)	(0.010)
Education of HH head		-0.005	-0.028
		(0.068)	(0.073)
Total landholding size		0.021	0.011
		(0.024)	(0.020)
Soil fertility			
Medium			-0.137***
			(0.049)
Poor			-0.363***
			(0.118)
Soil color			
Black			-0.237***
			(0.069)
Grey or sandy			-0.043
			(0.074)
Distance from home			-0.001
			(0.002)
Distance to the edge			-0.001
			(0.002)
Number of corners			0.012
			(0.012)
Own plot (1=yes)			0.045
			(0.065)
Village (*kebele*) dummies	Yes	Yes	Yes
Constant	3.542***	3.247***	3.351***
	(0.082)	(0.263)	(0.426)
Observations	365	365	360
R-squared	0.518	0.525	0.562

Notes: Standard errors are clustered at the *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A10: Plot size-productivity relationship *(correcting for area measurement only)*

Explanatory variables	Dependent variable: ln (self-reported production/**compass-and-rope plot size**)		
	(1)	(2)	(3)
ln (compass-and-rope plot size)	-0.679***	-0.675***	-0.662***
	(0.079)	(0.083)	(0.074)
Age of HH head		0.033	0.021
		(0.025)	(0.025)
Age square		-0.000	-0.000
		(0.000)	(0.000)
Gender of HH head		0.102	0.049
		(0.113)	(0.109)
Size of HH		-0.015	-0.006
		(0.014)	(0.014)
Education of HH head		0.013	-0.006
		(0.103)	(0.105)
Total landholding size		0.045*	0.045
		(0.027)	(0.027)
Soil fertility			
Medium			-0.196***
			(0.067)
Poor			-0.414***
			(0.089)
Soil color			
Black			-0.116
			(0.092)
Grey or sandy			0.217
			(0.151)
Distance from home			0.002
			(0.004)
Distance to the edge			-0.001
			(0.003)
Number of corners			-0.002
			(0.010)
Own plot (1=yes)			-0.085
			(0.085)
Village (*kebele*) dummies	Yes	Yes	Yes
Constant	2.215***	1.454**	1.972***
	(0.160)	(0.572)	(0.608)
Observations	365	365	360
R-squared	0.576	0.587	0.607

Notes: Standard errors are clustered at *kebele* level and given in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A11: Plot size-productivity relationship *(correcting for production measurement only)*

Explanatory variables	Dependent variable: ln (**crop-cut production**/self-reported plot size)		
	(1)	(2)	(3)
ln (self-reported plot size)	-0.410***	-0.404***	-0.578***
	(0.067)	(0.070)	(0.077)
Age of HH head		-0.019	-0.013
		(0.017)	(0.013)
Age square		0.000	0.000
		(0.000)	(0.000)
Gender of HH head		0.252**	0.187*
		(0.115)	(0.109)
Size of HH		0.029*	0.037**
		(0.017)	(0.018)
Education of HH head		-0.072	-0.095
		(0.091)	(0.078)
Total landholding size		-0.029	-0.027
		(0.031)	(0.026)
Soil fertility			
Medium			-0.091
			(0.085)
Poor			-0.162
			(0.137)
Soil color			
Black			-0.142
			(0.092)
Grey or sandy			0.009
			(0.126)
Distance from home			-0.002
			(0.004)
Distance to the edge			0.011***
			(0.003)
Number of corners			0.072***
			(0.012)
Own plot (1=yes)			0.028
			(0.094)
Village level dummies	Yes	Yes	Yes
Constant	2.752***	2.921***	1.682***
	(0.121)	(0.389)	(0.443)
Observations	365	365	360
R-squared	0.403	0.424	0.535

Notes: Standard errors are clustered at *kebele* level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Table A12: Plot size—productivity relationship *(with no correction of measurement errors)*

Explanatory variables	Dependent variable: ln (self-reported production/self-reported plot size)		
	(1)	(2)	(3)
ln (self-reported plot size)	-0.154**	-0.155**	-0.204***
	(0.062)	(0.061)	(0.073)
Age of HH head		0.027	0.022
		(0.022)	(0.022)
Age square		-0.000	-0.000
		(0.000)	(0.000)
Gender of HH head		0.108	0.067
		(0.086)	(0.084)
Size of HH		-0.009	-0.002
		(0.017)	(0.018)
Education of HH head		-0.024	-0.033
		(0.082)	(0.084)
Total landholding size		0.002	0.005
		(0.030)	(0.029)
Soil fertility			
Medium			-0.063
			(0.063)
Poor			-0.184
			(0.111)
Soil color			
Black			-0.062
			(0.089)
Grey or sandy			0.127
			(0.121)
Distance from home			0.001
			(0.003)
Distance to the edge			0.002
			(0.002)
Number of corners			0.014
			(0.010)
Own plot (1=yes)			-0.065
			(0.071)
Village level dummies	Yes	Yes	Yes
Constant	7.693***	7.128***	7.060***
	(0.111)	(0.476)	(0.566)
Observations	365	365	360
R-squared	0.459	0.465	0.476

Notes: Standard errors are clustered at *kebele* level and given in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. High and red are reference categories for soil fertility and color, respectively.

Figure A1: Relative bias in SPR, where RB>(<)0 implies correcting the mismeasured explanatory variable reduces (increases) bias when measurement error remains in the dependent variable.

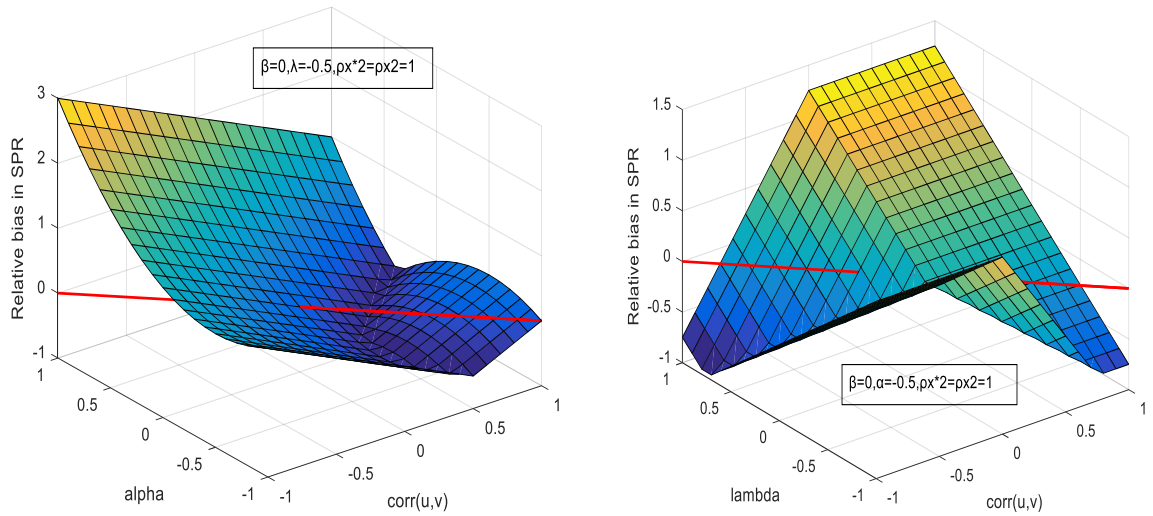

Figure A2: Relative bias in SPR, where RB>(<)0 implies correcting the mismeasured dependent variable reduces (increases) bias when measurement error remains in the explanatory variable.