Efficiency Effects of Access to Information on Small Scale Agriculture: Empirical Evidence from Uganda using Stochastic Frontier and IRT Models

Yakubu Abdul-Salam and Euan Phimister[1]

Abstract

Low power home electrical items such as radios, mobile phones and televisions are an important source of agricultural information for small scale farmers in developing countries. To empirically test the effects of access to information from these items on efficiency in agriculture, we formulate a stochastic frontier model augmented with a technical efficiency model that controls for an index capturing farmers' 'ability to access information'. The index is constructed with a 2-parameter item response theory (IRT) model based on farmers' access to the electrical items. Using 6 rounds of panel data on small scale farmers in Uganda, we find empirical evidence of a significant and positive relationship between farmer ability to access information and farm efficiency. There is also evidence that the size of these effects is larger for more literate hence better educated farmers. Greater access to information also appears to be associated with increased variance of (in)efficiency and output although the form of the increased variances is underpinned by low risk of lower efficiency and output realisations and high likelihood of higher efficiency and output realisations. Our findings imply that access to limited quantities of electricity needed to power these electrical items can have positive farm efficiency effects, and hence the importance of off-grid electricity (e.g. standalone solar panels) for small-scale farmers in typically isolated communities in developing countries.

---

1. Introduction

In many developing countries where access to electricity is low, e.g. rural sub-Saharan Africa, small scale off-grid solutions are being promoted as an important mechanism in achieving sustainable development goals (UN Sustainable Development, 2015; Abdul-Salam and Phimister, 2016; Deichmann et al, 2011). It is often argued that access to small amounts of electricity, e.g. from a standalone Solar Photo-voltaic (PV) system, can have substantial benefits for rural farm households' welfare by enabling lighting. However, there is little research on whether this limited access to electricity, which cannot provide mechanical power, might have direct farm efficiency benefits.

One mechanism through which limited access to electricity could improve farm efficiency is that it enables farmers to adopt various low power electrical items which might improve their access to agricultural information. This includes information that helps with crop and seed variety choices, reveals demand and market prices, enables planning for transport of inputs and outputs, provides advice on best plant fertilisation and protection measures, as well as other important information such as weather. For example, evidence does exist that mobile phone ownership improves marketing of farm produce. Aker and Fafchamps (2010) find in Niger that use of mobile phones have an impact on price dispersion particularly where travel costs are high. Overå (2006) presents evidence from Ghana of mobile phones helping increase the effectiveness of trade networks, while Jensen (2007) finds that the adoption of mobile phones significantly decreased price dispersion and wastage for Kerari fishermen in India. Recently, Fu and Akter (2016) also find that use of mobile phones enhanced extension service delivery for Indian farmers through greater knowledge and awareness of evolving agricultural best practices.

However, there is limited empirical evidence to support the hypothesis that farmer access to electrical items or services resulting from access to electricity has direct efficiency effects on the farm due to enhanced information access. The World Bank (2012) finds that access to the internet, for example, does raise the efficiency of existing processes and makes new production processes possible. There is also some limited evidence that access to electricity can improve efficiency of maize growing farms in Mozambique (Zavale et al., 2005). However, a survey of World Bank funded off-grid renewable electricity programmes found no evidence of agricultural efficiency effects due to access to electricity (World Bank, 2008).

2

In this paper, we use panel data from farms in Uganda to test whether access to information from radios, mobile phones and televisions have efficiency effects on small-scale agriculture. We estimate a stochastic frontier (SF) model augmented with a technical efficiency model that controls for an index capturing farmers' ability to access information. We introduce a novel way of accounting for farmers' ability to access information using a latent variable approach where the values are constructed drawing on item response theory (IRT). This approach has a long history of use in education testing and psychology (see e.g. Embretson and Reise, 2013) and is increasingly being applied in economics and social science to infer latent traits such as household deprivation, social capital, wealth and labour quality using a range of available information which *a priori* is thought to reflect the underlying trait (Cappellarri and Jenkins, 2007; Filmer and Scott, 2012; Ødegaard and Roos 2014; Vandemoortele, 2014; Pericoli et al 2015). Here, we use the IRT model to infer the farmer's underlying ability to access information using their level of access to a number of electrical items. This captures the potential information effects in a parsimonious but general way, i.e. allowing for the interaction between different information sources.

The remainder of this paper is organised as follows: In Section 2, we discuss our empirical strategy. In particular, we discuss the specification of the full SF model, the IRT model used in the construction of the 'ability to access information' index, and a number of estimation issues. Section 3 presents the data. We discuss our results in Section 4. Section 5 concludes with a discussion of some policy implications.

2. Empirical strategy

2.1 Stochastic frontier (SF) model

A productive unit is said to be fully efficient if it realises the maximal output from the technology represented by its production frontier. Stochastic frontier models presume that no productive unit can achieve efficiencies beyond this frontier and a unit's deviation from this frontier represents its level of technical inefficiency. We index our panel data of Ugandan farm households on farm household i and time period t.[2] Following Dawson et al. (1991), we define the SF model as follows;

---

[2] Productive units in the data are identified at the 'household', land 'parcel' and land 'plot' level. A household may own several parcels of farm land. A parcel of farm land may be segmented into several identifiable

3

$$q_{it} = f\left(x_{k,it}, \beta\right)\exp\left(\varepsilon_{it}\right), \quad i = 1,...,N, \quad t = 1,...,T \tag{1}$$

where q_{it} denotes realised production by farm household i in period t; $f(.)$ denotes the production frontier; $x_{k,it}$ is the vector of k farm inputs, β is a vector of model coefficients and ε_{it} a composite error term that is the sum of two independent elements; a term v_{it} representing idiosyncratic error (e.g. specification and measurement error); and a term u_{it} representing technical inefficiency relative to the technical efficiency frontier, as follows;

$$\varepsilon_{it} = v_{it} - u_{it} \tag{2}$$

The technical inefficiency term u_{it} is fundamental to SF analysis as it represents the inefficiency realised with the technology embodied in the production frontier $f(.)$. Substituting equation (2) into equation (1) and taking natural logarithms of both sides of equation (1) yields:

$$y_{it} = \ln\left\{f\left(x_{k,it}, \beta\right)\right\} + v_{it} - u_{it} \tag{3}$$

$$v_{it} \sim J \tag{4}$$

$$u_{it} \sim F \tag{5}$$

where $y_{it} = \ln\left(q_{it}\right)$, and J and F are the assumed distributions of v_{it} and u_{it} respectively.

2.2 Model Specification

The assumptions about the distributions of the error term v_{it} and inefficiency term u_{it} (i.e. J and F respectively) used to identify model (3)-(5) have received significant attention in the literature and have motivated a wide variety of specifications (e.g. Lee and Schmidt, 1993; Kumbhakar, 1990; Battese and Coelli, 1995; Greene, 2005a/b). The most widely used of

---

plots. Household farm size is the sum of all household parcels and plots. As access to information is a household level variable, this is the unit of analysis used in the paper rather than parcel or plot level. Initial exploratory analysis was undertaken at a parcel level and the main results and conclusions were qualitatively similar to those reported in the paper.

these in the empirical literature is the Battese and Coelli (1995) time varying SF model specification (Kumbhakar et al, 2014). This models technical efficiency by allowing the mean of the inefficiency distribution to be determined by a number of exogenous determinants (e.g. an 'access to information' index). These determinants are usually neither inputs nor outputs of the production process but are factors which are assumed to affect the performance of the productive unit (Belotti et al., 2012). The Battese and Coelli (1995) model assumes homoscedasticity in the variances of the idiosyncratic error and inefficiency term (e.g. Karagiannis and Sarris, 2002). As Kumbhakar et al (2014) note however, this is unnecessarily restrictive and induces potential inconsistency and bias in the estimated coefficients of the production frontier and the technical efficiency model.

We therefore adopt the generalisation of the Battese and Coelli (1995) model introduced by Wang (2002) and used in the recent review SF panel data models by Kumbhakar et al (2014). This model allows for heteroscedasticity in the mean of the inefficiency distribution as well as in the variances of both the inefficiency and idiosyncratic error distributions as follows;

$$
\begin{aligned}
y_{it} &= \ln\left\{ f\left(x_{kit}, \beta\right)\right\} + v_{it} - u_{it} \\
u_{it} &\sim N^{+}\left(\mu_{it}, \sigma_{u,it}^{2}\right) = N^{+}\left(\delta_0 + z_{it}^{'}\delta, \exp\left(\omega_{u0} + z_{u,it}^{'}\omega_{u}\right)\right) \\
v_{it} &\sim N\left(0, \sigma_{v,it}^{2}\right) = N\left(0, \exp\left(\omega_{v0} + z_{v,it}^{'}\omega_{v}\right)\right)
\end{aligned}
\tag{6}
$$

where N^{+} and N denote the truncated-normal and the normal distributions respectively for the technical inefficiency and the idiosyncratic error; μ_{it} is the mean of the technical inefficiency distribution and is modelled as a function of the vector z_{it}; $\sigma_{u,it}^{2}$ and $\sigma_{v,it}^{2}$ are the variances of the inefficiency distribution and the idiosyncratic error respectively and are modelled with the vectors $z_{u,it}$ and $z_{v,it}$. The values $\delta_0, \delta, \omega_{u0}, \omega_{u}, \omega_{v0}$ and ω_{v} capture the impact of the exogenous variables on the mean and variance of technical inefficiency, and the variance of the idiosyncratic error.

Modelling heteroscedasticity in the variance of technical inefficiency limits the bias in the estimates of the production frontier and technical inefficiency parameters, and allows for potential non-monotonic efficiency effects. This can be important in understanding the relationship between technical efficiency and its exogenous determinants (Wang, 2002; Kumbhakar and Lovell, 2003). The heteroscedasticity in the idiosyncratic error variance $\sigma_{v,it}^{2}$

5

is determined by the vector $z_{v,it}$., and has been interpreted as a measure of output variance (Kumbhakar et al., 2014). As in any regression model, there are a range of reasons why the idiosyncratic shocks might vary with exogenous factors including skewness in the distribution of these variables, other types of model misspecification such as omitted variables and inappropriate functional forms.. Hence the empirical modelling should allow for the possibility of variance of the idiosyncratic error as, if ignored, it may also induce bias in the estimates of technical inefficiency as well as the intercept of the production frontier (Kumbhakar and Lovell, 2003).

There are alternative SF models that attempt to exploit panel data in a broader way than the Wang (2002) model adopted here. These include the models by Kumbhakhar and Heshmati (1995), Kumbhakhar and Wang (2005), the Greene (2005a/b) 'true' fixed and random effects models and the Kumbhakar et al (2014) model. These models exploit panel data by disentangling transient (time varying) efficiency from persistent (time invariant) efficiency and/or unit specific (time invariant) heterogeneity. However, there are a range of issues with these models which limit their usefulness for the current study. Firstly, the increased parameter space due to capturing unit specific heterogeneity and/or persistent efficiency presents computational difficulties (Belotti et al., 2012). The second is the so-called incidental parameters problem which arises when the number of units is large relative to the length of the panel. In such instances, intercepts of the unit specific heterogeneities are inconsistently estimated as the number of units increases (i.e. as N →∞) (Lancaster, 2002; Belotti et al., 2012). Due to this problem, the Greene (2005b) fixed effects model for example is only appropriate when the length of panel data is at least 10 periods (i.e. T>=10) (Belotti et and Ilardi., 2012; Belotti et al., 2012).

The most general of models in this class is the recently introduced model by Kumbhakar et al. (2014).[3] This model accounts for unit specific heterogeneity as well as persistent and transient efficiency. However, crucially given the focus of the present paper, this approach does not allow for the incorporation of heterogeneity in mean efficiency or in the variances of efficiency and idiosyncratic error, which prevents the formal testing of information effects

---

[3] Described as Model 6 in Kumbhakar et al (2014).

within this framework.[4] This approach also requires a multi-stage estimation procedure which, as Wang and Schmidt (2002) showed, can lead to severely biased results.

Hence, the Wang (2002) model adopted here is seen as the best compromise between more general panel specifications discussed and the need for a framework in which the information effects can be formally tested. However, as a robustness check, we also present results from the Kumbhakhar et al. (2014) model. As discussed above, this specification does not allow the formal direct estimation of the impacts of access to information on the different types of efficiencies, or on the efficiency and idiosyncratic error variances. However, we can informally explore the impact of information in this setting by disaggregating the results for farmers with low and high access to information and use these as a robustness check on the results of the Wang model. These results are reported in Section 4.4 and our on-line Appendix B.

2.3. Model Parameterization

Michler and Shively (2015) point out that debate exists about the most appropriate functional form for the production frontier, with numerous early studies adopting the less flexible Cobb-Douglas form. However use of flexible functional forms such as the translog have been widely applied in agricultural efficiency studies (Bravo-Ureta et al., 2007) as they can be interpreted as an approximation to any underlying production frontier. Although there are often estimation difficulties due to collinearity between inputs, we also adopt a translog production frontier.[5] The production frontier $f(.)$ in model (6) follows the translog specification in Coelli et al (2005) as follows;

---

[4] Kumbhakar et al. (2014) state that the development of this model to include heterogeneity in mean inefficiency and heteroscedasticity in the variances of inefficiency and idiosyncratic error is a subject for future research.

[5] A likelihood ratio test of the translog frontier specification vs the Cobb Douglas frontier specification was conducted. The associated test statistic was 104.12 (p<0.01), suggesting that the restrictions implied by the Cobb Douglas frontier can be rejected.

$$y_{it} = c + \sum_{k}^{3} \beta_k \ln\left(X_{k,it}\right) + \beta_{ox} Ox_{it} + \frac{1}{2} \sum_{k}^{3} \sum_{l}^{3} \beta_{kl} \ln X_{k,it} \ln X_{l,it} + \frac{1}{2} \sum_{k'}^{3} \beta_{k'} Ox_{it} \ln X_{k',it}$$
$$+ \beta_t \cdot t + \beta_{t^2} \cdot t^2 + v_{it} - u_{it} \tag{7}$$

where y_{it} is the natural log of household farm output per period. There are 3 continuous input variables $X_{k,it}$, namely 'Farm size' measured in acres; 'Labour' measured in the number of days worked on farm and 'Chemical' which is the sum of the real cost of fertiliser (i.e. organic and inorganic) and pesticide applications on a farm. Finally we include one dummy variable 'Ox',[6] indicating whether oxen are used on a farm, and interact this with the other continuous variables. To capture macro level changes over time we include a time trend variable t and its squared term t^2 (t=1,…,6).[7]

Following Coelli and Battese (1996), Wang (2002), Bozoğlu and Ceyhan (2007) and Kumbhakar et al. (2014), we construct a Laspeyres index of real farm output using revenue values in 2009 prices[8] (Ugandan shillings) as a measure of production. The Laspeyres index was constructed using information on quantities and sale prices of crops. The first cropping season of 2009 was used as the reference period for the index. The index was used to deflate production revenues and the monetary cost of chemical inputs. On-line Appendix A describes the construction of deflated production revenues. Labour includes family and hired labour. Following Coelli and Battese (1996), we convert number of days worked to male equivalents according to the rule that female and child number of days worked were equivalent to 0.75 and 0.5 male number of days worked respectively. The 3 continuous input variables $X_{k,it}$

---

[6] 'Ox' ownership is not available for the first year of the data collection. However, after experimentation, ox ownership was observed to be highly persistent. We therefore assume that a household has ownership of an ox in the first year if it has ownership of an ox in the subsequent year. Exploratory estimations were also made excluding this variable from the model. These did not change the results and conclusions in terms of the information effects.

[7] A number of alternative specifications for (7) using time dummies with their interactions, splitting chemicals into its component parts and including information on tractor availability were also explored. The final specification chosen was based on both hypothesis testing and judgement to ensure a parsimonious and robust specification where apparent overfitting of the model to the data was minimized. Hence, for example, available information on tractor use was excluded on the basis of statistical testing while in the final specification the time trend was used in place of individual time dummies to provide a more parsimonious specification which did not affect the nature of the results in terms of the information effects.

[8] As farmers in our data produce multiple crops, we are unable to use physical units (e.g. kilograms) as a measure of production.

were scaled (i.e. divided by their geometric means) prior to their logs being taken so that their first order coefficients can be interpreted as elasticities of output calculated at their means.

As described in model (6), we allow the mean of the inefficiency distribution term to be determined by exogenous determinants thought to affect farm efficiency. We include in the vector z_{it} a number of these determinants for each farm household i in period t. They include age of the household head (i.e. a proxy for experience), household level of literacy,[9] the land tenure status of households (i.e. whether owned or otherwise)[10] and farm location (i.e. rural or urban). These variables have been found to be important determinants of technical efficiency in previous studies (see e.g. Battese and Coelli, 1995; Kumbhakar et al., 2014; Michler and Shively, 2015). To capture information effects, we include in vector z_{it} a dummy variable reflecting whether the household has access to agricultural extension services[11] and a latent variable measuring the household's 'ability to access information' from a range of low power home electrical items (the construction and estimation of this variable is described below in Section 2.4). Further, households with higher literacy levels are more likely to have greater capacity to 'receive, decode and understand information' (Reimers and Stephan, 2013). Hence to allow for the fact that they may be able to use information more effectively to improve efficiency on their managed farms, we allow for possible interactions between the household level of literacy and the latent access to information variable. If better information access improves management practice it ought to improve average household farm efficiency. Hence, we expect that farm efficiency will increase as information access improves.

While we expect mean technical inefficiency μ_{it} to decrease with increased access to information, the effects of improved access to information on the variance of technical inefficiency $\sigma^2_{u,it}$ and idiosyncratic error $\sigma^2_{v,it}$ is ambiguous. Indeed Wang (2002) states that while an exogenous inefficiency variable may be expected to decrease inefficiency, the same

---

[9] Calculated as the proportion of household adults (age >10 years) who can read and write.

[10] A farm household is assumed to own its farm land (i.e. freehold tenure) if the proportion of its managed farm land under freehold tenure is more than 50%.

[11] The underlying questionnaire asks whether farmers 'receive' extension advice. We interpret this to mean 'access' to extension advice and not necessarily the 'use' of it. This makes the 'extension' variable exogenous. Potential endogeneity with 'ability to access information' variable may arise if the 'extension' variable is interpreted as 'use' of extension.

variable may not necessarily reduce inefficiency and/or output variance. In the present instance, improved management practices from improved access to information may alter how farmers respond to changing conditions through the growing season hence potentially reducing inefficiency and/or output variance. However, previous studies have also found that output variance in particularfor farmers who adopted yield enhancing technologies in fact increased (Fafchamps and Pender, 1997), possibly as innovation may increase risk taking.

2.4 Item response theory (IRT) model: Quantifying 'ability to access information'

A novelty in our empirical approach is that we estimate a latent index variable capturing farmers' ability to access information based on their access to 3 electrical items: a radio; a mobile phone; a television.[12] We use the forecast values of this variable for each farmer to capture their level of access to information in model (6). Specifically, we construct the latent 'ability to access information' variable as an index using a 2-parameter logistic IRT model. IRT models have been used to study cognitive abilities, personality traits, attitudes, quality of life, patient satisfaction and other attributes that are latent and so cannot be measured directly (Embretson and Reise, 2013).

More recently, studies in economics and social science have also used this approach to construct indices capturing a variety of latent traits. For example, Cappellarri and Jenkins (2007) construct an index of household deprivation using information on household characteristics, and an individual's perceived ability to afford specific items; Pericoli et al (2015) and Li et al (2005) estimate indices of social capital based on responses to questions indicating the degree to which they were linked to social networks, their civic participation and neighbourhood attachment; Faye et al (2011) construct an index of food insecurity; Filmer and Scott (2012) and Vandemoortele (2014) use wealth indices based on information on the presence or absence of certain assets and household characteristics; Ødegaard and Roos (2014) constructed indices of labour quality using answers to health status and the psychosocial environment questions.

---

[12] We also explored whether access to the internet could be included in the construction of the index. However, less than 1% of the farm households in our data have access to the internet. The lack of variation in this item meant it was not possible to identify the IRT model when it was included.

Within these studies the latent indices produced have been used in a variety of ways. The values of any IRT based latent index are difficult to interpret in isolation but they have been used to compare values of social capital across various categories e.g. income, gender, and education (Li et al, 2005), as a method to compare the performance of other measures of deprivation or as a measure to be explained by other exogenous factors (Cappellarri and Jenkins, 2007). IRT based latent indices have also been used directly to generate thresholds to define various levels of food insecurity (Faye et al., 2011), and as a method of identifying the influence of ownership of different items on wealth status (Vandemoortele, 2014). As in this study, a number of these previous papers have also used the predicted latent indices as an explanatory factor in their subsequent analysis. For example, Pericoli et al (2015) use their index to explain the impact of social capital on consumption insurance and income volatility, while Ødegaard and Roos (2014) use predicted values of latent labour quality in a DEA study explaining the productive efficiency of large firms in Sweden.

The idea underlying IRT models applied in these studies (and here) is that some latent variable or trait exists and that the level of a respondent's trait can be identified by the answers given to a set of (binary) questions which reflect the respondent's level of the trait. We conceptualise the latent trait here as farmers' 'ability to access information' and claim that a farmer's access to radio, mobile phone and television will reflect their latent ability to access information. In an IRT model the relationship between latent trait and the observed binary data is modelled probabilistically. In this case the probability that the farmer has access to a particular item is positively related to the latent "ability to access information" trait, with the exact relationship depending upon how informative having a particular item is in helping identify the latent trait.

More formally in IRT models the relationship between the level of the latent trait and the probability of "success" (or access to a particular item) is given by the item characteristic curve (ICC) (a logistic function). Let $infoIndex_{it}$ represent 'ability to access information' for farm household i in period t. We consider the IRT logistic model by Birnbaum (1968) as follows;

$$\pi_{it,j} = \frac{\exp\left\{a_j\left(infoIndex_{it} - b_j\right)\right\}}{1 + \exp\left\{a_j\left(infoIndex_{it} - b_j\right)\right\}}; \quad infoIndex_{it} \sim N(0,1) \tag{8}$$

11

where $\pi_{it,j}$ is estimated probability that farm household i has access to electrical item j in period t, a_j is the estimated level of item j "discrimination" and b_j is the estimated level of item j "difficulty". The estimated a_j and b_j underpin the shape of the ICCs for the 3 electrical items.

To provide a better intuitive understanding of our IRT model (8),
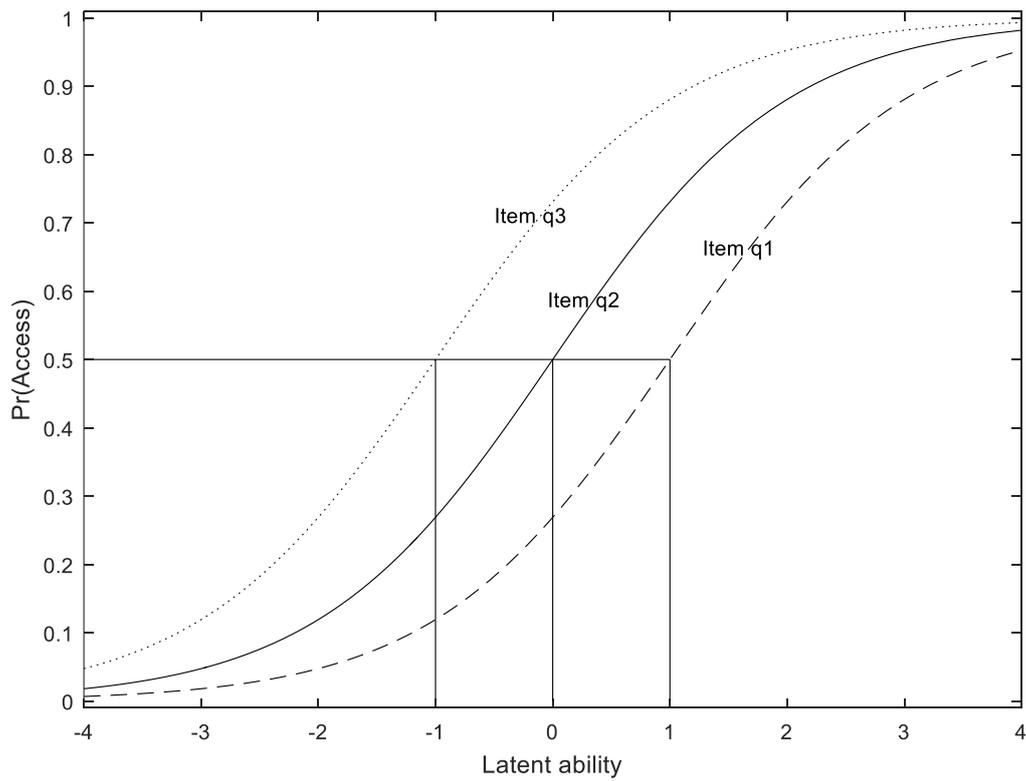
Figure 1 and
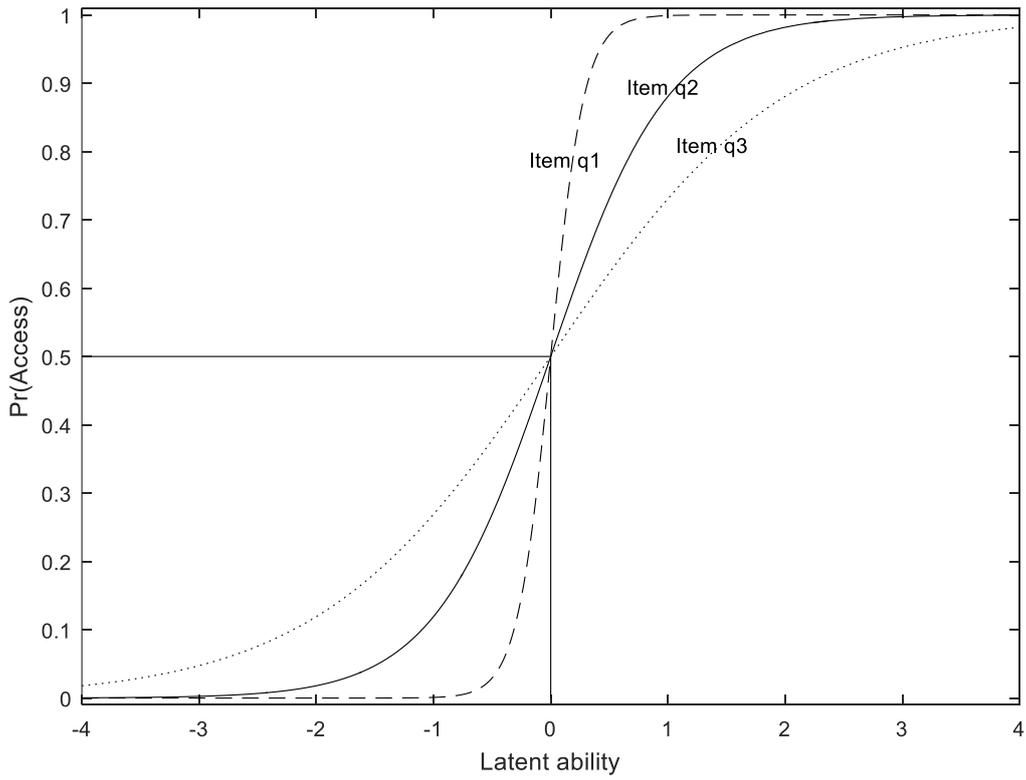
Figure 2 illustrate a hypothetical case of how an item's difficulty and discrimination affects the ICC curve and the information it provides on the latent variable. The *x* axis captures the level of the latent trait. Associated with this trait we assume we have 3 hypothetical items *q1*, *q2* and *q3* (analogous to individuals answering 3 questions about access to electrical items

which   are   assumed   to   provide   information   on   the   latent   trait).
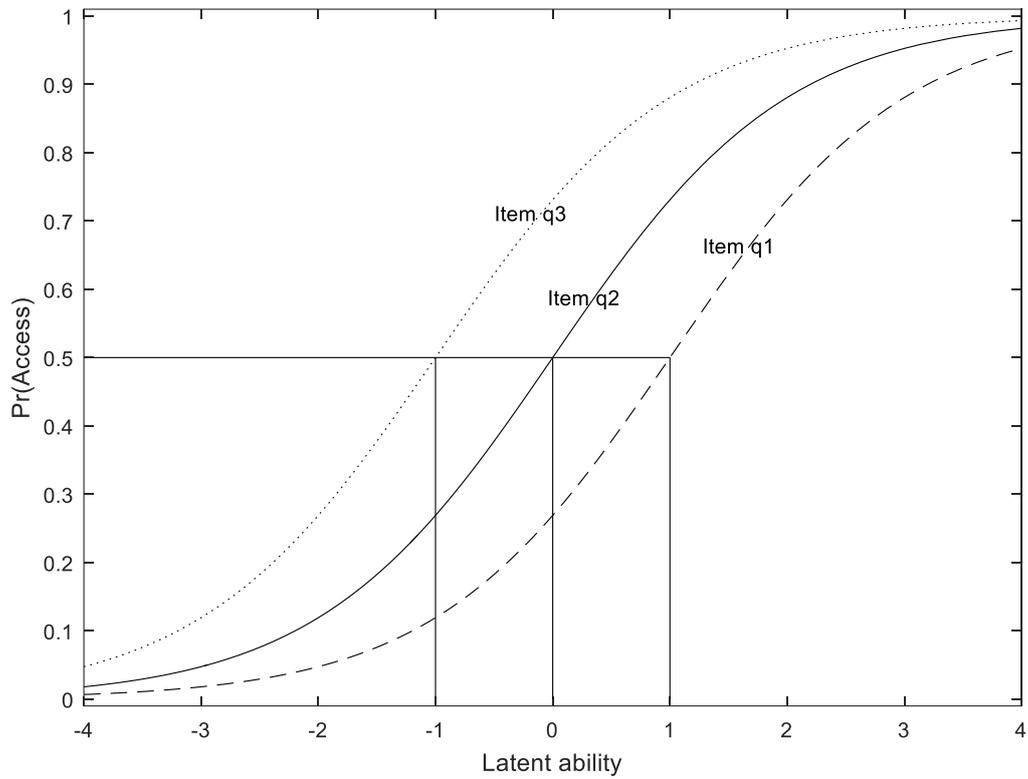
Figure 1 illustrates the 3 ICC curves associated with the 3 hypothetical items and the role of an item's "difficulty". Formally this is defined as the level of the trait such that the

probability   that   individual   has   access   to   the   item   equals   0.5.   In
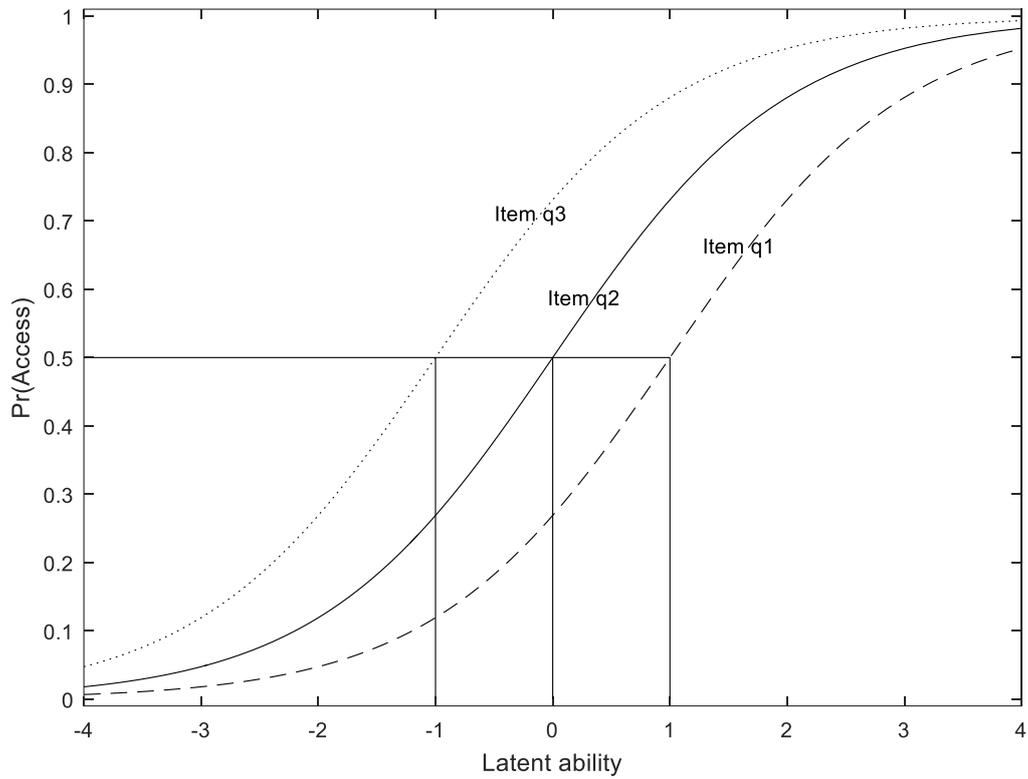
Figure 1 items *q1*, *q2* and *q3* have difficulties of 1, 0 and -1 respectively. Informally this implies that individuals with access to item *q3* only are on average likely to have lower latent ability to access information relative to individuals who have access to items *q1* or *q2*.
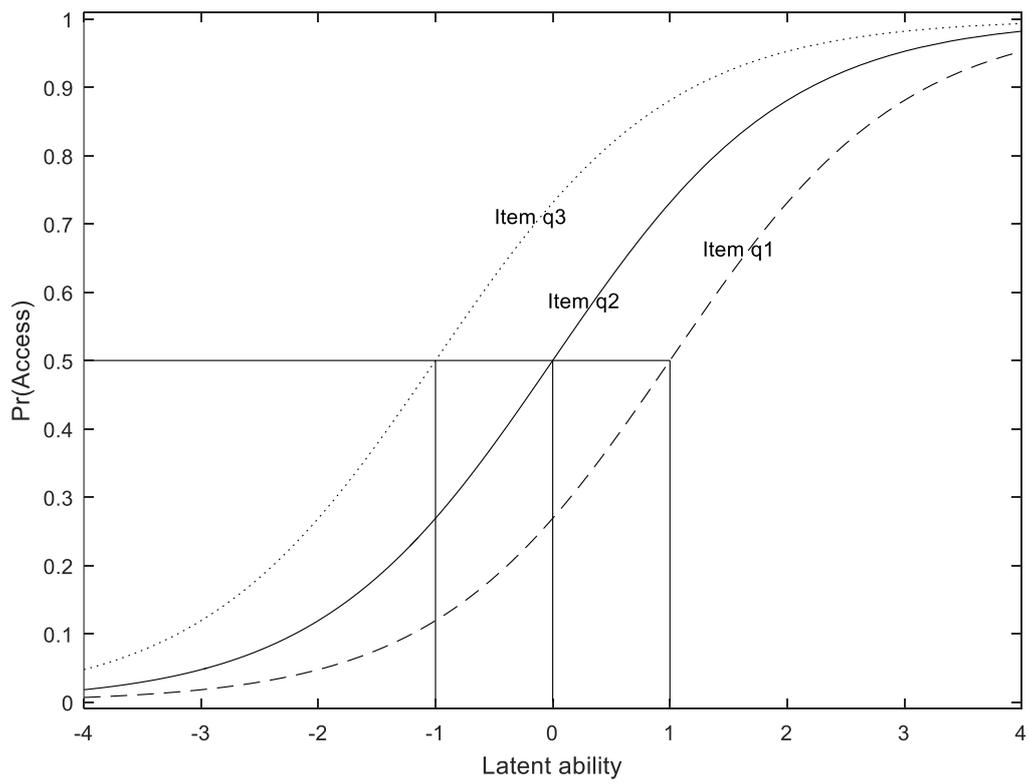
Figure 1: ICCs and item "difficulty".
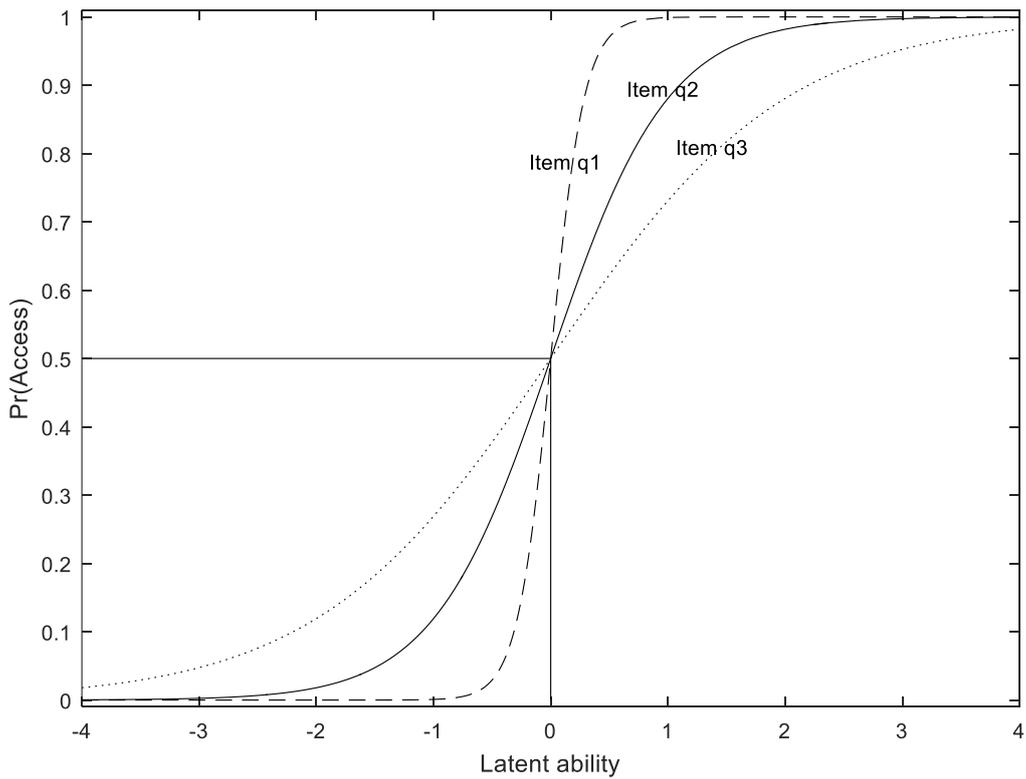
Figure 2: ICCs and item "discrimination".

As Li et al (2005) point out the discrimination (or factor loading) parameter captures how items "vary in the extent to which they measure the underlying factor" (p112, para 6). Hence, if we expect a mobile phone to be more useful to the farmer in accessing agricultural information than say a radio, then item discrimination for a mobile phone should be higher than that of a radio.
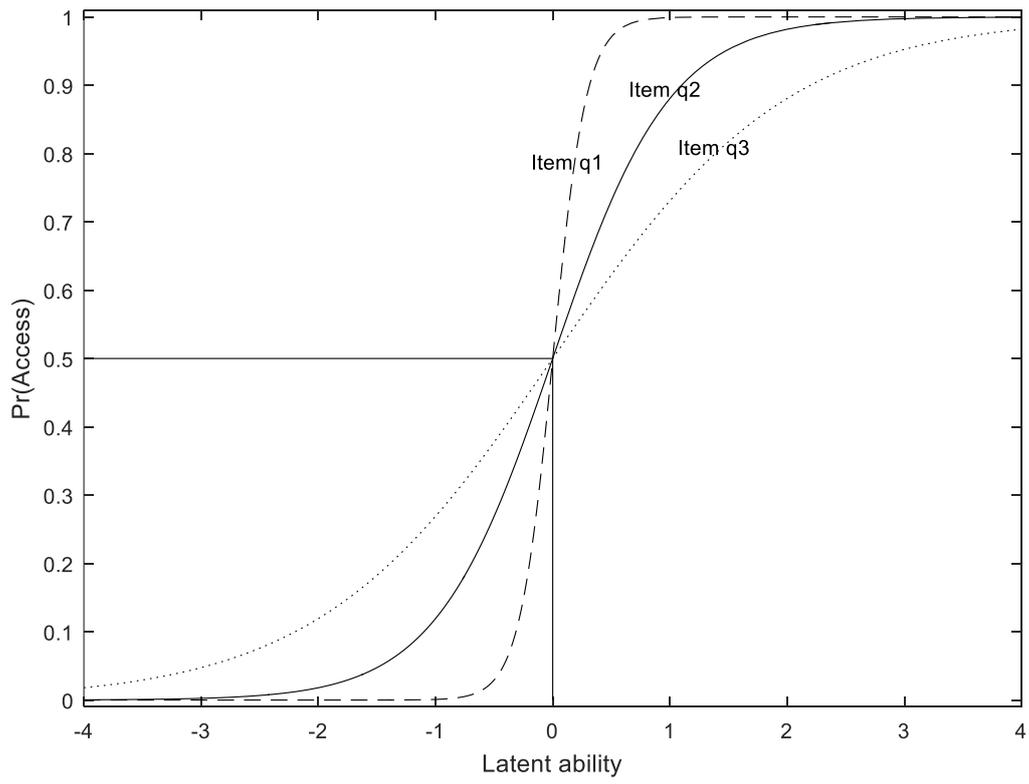
Figure 2 attempts to illustrate why this occurs. In
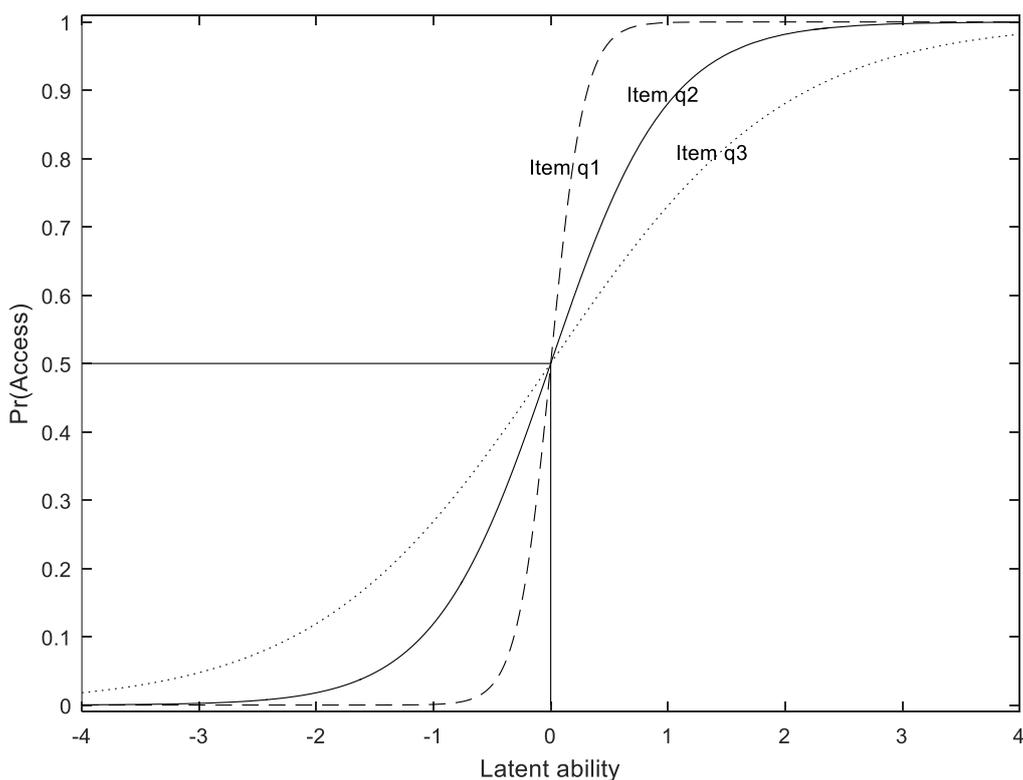
Figure 2, all 3 hypothetical items have the same difficulty i.e. for individuals with ability equal to zero the probability that an individual is able to access an item is 0.5 for all 3 items. However, the 3 items have different degrees of discrimination (i.e. slopes) around zero, with *q3* having the lowest degree of discrimination and *q1* highest. Consider two individuals, one with latent ability to access information just below 0 and the other with a value just above 0. For item *q3* both individuals would have similar probabilities of access to the item. For item *q1* the person with the marginally higher ability to access information would have a significantly higher probability of access to the item. Therefore information on whether an individual has access (or not) to item *q1* will provide more information on the latent trait than item *q3*.

The IRT method has a number of advantages over possible alternative approaches. One alternative used elsewhere is to use the total expenditure on items as a proxy to subjects' ability to access information (e.g. Castiglione, 2012; Shao and Lin, 2001). Another is by

19

simply counting the number of items a farmer has access to (Esselaar et al., 2006).[13] Both methods can be viewed as weighted sums of the items to which the farmer has access. Relative to the IRT approach, the use of weighted sums to generate such indices have two significant limitations. The weightings used are typically arbitrary and tend to imply a priori value judgements, which are not always acknowledged (Vandemoortele, 2014). In addition, unless all the data used is continuous (i.e. not binary), there is no clear link between the index and the underlying latent variable (Cappellarri and Jenkins, 2007).

Alternatively it is possible to model the multiple items using individual dummy variables and their interactions. However, collinearity issues make it difficult to clearly identify separate and joint effects of each of the items within this approach. To allow for all possible combinations of the 3 items in our model would require the inclusion of 7 new dummy variables, whose effects would be extremely difficult to identify separately with any degree of precision. Such an approach would also exacerbate the numerical difficulties often found in the estimation of frontier models due to collinearity (Kumbhakar et al., 2014). Hence, the approach used here provides a parsimonious way of allowing for potential interactions between the 3 information sources.

2.5 Estimation Issues

As argued above the IRT approach provides a sound framework for generating an index from multiple binary data. However, to identify the IRT model, it is necessary to assume that the distribution of ability (i.e. *infoIndex*) is normally distributed with zero mean and unit variance, which may be violated in many empirical situations. The extent to which this assumption induces bias in the IRT approach is an area of active research (Reise and Revicki, 2014). Bartholomew (1988) showed analytically that model bias is not present when the latent trait distribution is symmetrical, while Bock and Aitkin (1981) found that non-normality had minimal implications on model bias. Seong (1990) and Robins et al (2009) provide a method (used below) to check informally whether potential violations of the normality assumption should be considered a potential problem. Seong (1990) showed that

---

[13] We were unable to use the expenditure approach as no information on periodic expenditure on the 3 items was available. We did explore the use of the simple count index based on an item count. Qualitatively, the result compared to the *infoIndex* is similar. This analysis is reported in on-line Appendix B.

bias is not important when model estimates are "moderate" while Robins et al (2009) showed that "moderate" estimates could be interpreted as a range of between -3 to 3 for the difficulty parameter and the latent ability index.

An estimation issue also arises due to the use of the latent index variable *infoIndex* in our SF model (6), as we first estimate the IRT model and then use the predicted *infoIndex* variable value as an independent variable in the SF model. Generally in two step estimation procedures such as this, the asymptotic distribution of the second stage estimation (i.e. SF model) will depend on the first (i.e. IRT model) (Cameron and Trivedi, 2010). To correct for standard errors in the SF model therefore, we implement a sequential two-step non-parametric bootstrapping procedure in which the predicted *infoIndex* variable in the first stage IRT model is used in the second stage SF model.

The traditional approach to estimating the effects of the inefficiency determinants on technical inefficiency uses a two-step procedure. However, Wang and Schmidt (2002) show that this approach, where estimates of inefficiency are obtained in a first regression, while in the second step these are regressed on the exogenous variables, can lead to severely biased results. The Wang (2002) model adopted here, which is a generalisation of the Battese and Coelli (1995) model, is estimated in a one-step procedure using Maximum Likelihood methods hence potentially limiting bias in the results (Belotti et al. 2012).[14]

Finally, recent literature has attempted to account for potential endogeneity both of farm inputs and of factors affecting farm efficiency (e.g. Amsler et al, 2014; Solis et al, 2009). However, identifying appropriate potential instruments is difficult and we therefore follow the approach typically taken in the literature with both distributions assumed independent and identically distributed (Michler, and Shively 2015; Kumbhakar et al, 2014; Rahman, 2010; Bravo-Ureta et al., 2007).

---

[14] The maximum likelihood estimation procedure used in Stata 14 were found to be reliable although convergence issues were encountered when a large number of explanatory variables were included in the equations explaining $\sigma^2_{u,it}$ and $\sigma^2_{v,it}$.

3. Data

We use the Uganda National Panel Survey (UNPS) data which is a product of the Uganda Bureau of Statistics and the World Bank's Living Standards Measurement Survey – Integrated Surveys on Agriculture (LSMS-ISA) project. The panel is an annual survey of a nationally representative sample of Ugandan households (World Bank, 2013). It started in 2009 with data collected for some 3,123 households. Subsequent waves were carried out for 2010/11 and 2011/12 with some 2,716 households surveyed. We use data in the agricultural module of the survey and specifically focus on crop farmers. There are two rounds to each annual wave. These rounds are conducted at 6 month intervals in order to 'better capture agricultural outcomes associated with the two cropping seasons of the country' (World Bank, 2013). To reflect the cropping seasons, we construct a 6 period unbalanced panel from the 3 year data, each period in our panel being a half year.

There was evidence of potential data entry errors, with some farm land parcels or plots recorded as zero although farm output was positive, while some land parcels or plots had extremely large values (although output did not necessarily correspond). To exclude the potential influence of these extremes, we dropped parcels or plots for which output per acre is below the 5th percentile or above the 95th percentile, or where labour variables were unrecorded or recorded as zero. The application of such sample selection rules is common in farm microdata studies to reduce the influence of large outliers on the results (see e.g. Benjamin and Phimister, 2002). Table B3 in on-line Appendix B shows the number of times households occur over the 6 periods of the data after applying the above sample selection rules. While it is possible to conduct our analysis on a balanced panel (i.e. households occurring for all 6 periods only), such an approach would mean a limited portion (i.e. 13.45%) of the available data is used and would likely induce sample selectivity issues. Rather we follow the approach adopted by Kumbhakar et al. (2014) and only exclude those farm households for which there are less than 2 periods of data.[15] This process excluded 372 observations from the final estimation sample of 7524 observations. Table 1 describes key variables in our sample.

Table 1: Descriptive Statistics

---

[15] Panel data estimators are able to exploit panels for which units appear at least twice in the data. Exploratory estimation was also undertaken using all available farms and the main results were unaffected.

	Mean	Standard deviation
SF production frontier variables		
Output (million Ugandan Shilling)	0.41	0.90
Farm size (acres)	14.70	44.53
Labour use (No. of days)	804	1107
Chemical input (U. Sh. thousand)	15.2	179.3
Ox	0.09	0.28
IRT and SF technical efficiency model variables		
Access to radio	0.67	0.47
Access to mobile phone	0.52	0.50
Access to television	0.05	0.22
Access to extension	0.26	0.44
Level of household literacy	0.56	0.31
Age of household heads	46.9	14.7
Male household heads (male = 1, female = 0)	0.73	0.45
Land tenure of farm (owned = 1, otherwise = 0)	0.40	0.49
Urban farm location (urban = 1, rural = 0)	0.11	0.31

All values reported are pooled across all observations and periods. Each household appears at least twice over the 6 periods.

There is clearly significant variation in farm sizes, labour, chemical use and whether oxen were available for use on farms. Most rural people in Uganda have rights to their land through customary tenure arrangements, with a small proportion of the land formally registered. However, the land rental market appears active relative to elsewhere in Sub-Saharan Africa (USAID, 2010; Deininger and Mpuga, 2003). As expected many farms are managed by households with access to radios and to some extent, mobile phones. In the case of radios, this is not surprising given their relative low cost. In the case of mobile phones, the relatively high access rate is consistent with the increasing penetration of this item in the developing world over the past two decades. Access to television is rather low, possibly due to its higher cost, greater power requirement and/or lack of adequate reception.

4. Results

As discussed above the IRT model results are used to construct the latent index which is then used as an explanatory factor in the SF estimation. Hence, we first present the results of the IRT model, and then proceed to the results from the SF estimation.

23

4.1 IRT model results

Table 2 reports the estimated difficulty and discrimination of the 3 items used in the IRT model, with the associated estimated ICCs provided in
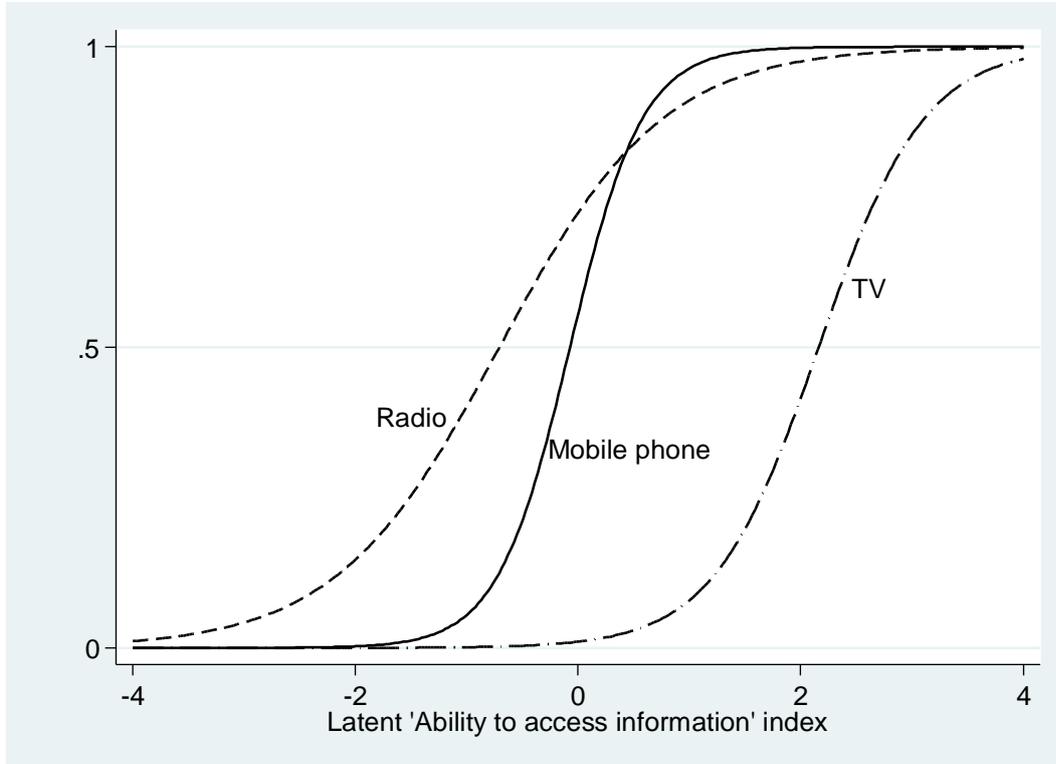
Figure 3. Table 3 reports the predicted values of the underlying latent variable associated with access to each item individually and in combination.

Table 2: IRT model estimation: 'Ability to access information' index.

	Radio	Mobile phone	Television
Difficulty	-0.71***	-0.07***	2.16***
	(0.04)	(0.02)	(0.09)
Discrimination	1.36***	3.07***	2.11***
	(0.10)	(0.47)	(0.19)
Log likelihood	-10,711.81	Observations	7,475
R^2 (Pearson)	0.74	R^2 (Sum of Squares)	0.73
R^2 (McFadden)	0.42	Accuracy	0.95

* Standard errors in parentheses; *** $P<0.01$. Goodness of Fit measures used as suggested by Mair et al (2008), R^2 (Pearson) based on squared correlation between responses and predicted access probabilities, R^2 (Sum of Squares) based on sum of squares measure, R^2 (McFadden) or "pseudo R^2" based on comparison of the fitted model with a null model setting using the mean of all responses as the predicted probability of item access. Accuracy is the proportion of the model predictions which are "correct" (where access to an item is predicted when the predicted probability is 0.5 or above).

24

The estimated parameters reported in Table 2 suggest that all parameters are well determined and that all items are significantly related to the underlying latent variable. The reported goodness of fit measures (Mair et al, 2008) also suggest that the model fit to the data is satisfactory. For example the level of accuracy reported implies that only 5 % of the model predictions were "incorrect" in predicting no access to an item when a household had access or in predicting access when the household did not. The relative values of the estimated discrimination parameters in Table 2 indicate that access to a radio provides the least information on the underlying latent variable while access to a mobile phone provides the most information. In other words, consistent with previous research, access to mobile phones appears important factor in terms of the ability of individuals to access information (Aker and Fafchamps, 2010; Jensen, 2007; Overå, 2006; Fu and Akter, 2016). Although TV has a lower discrimination parameter than mobile phone, its estimated difficulty is significantly higher, reflecting low access to it perhaps due to higher cost.

The   ICC   curves   in
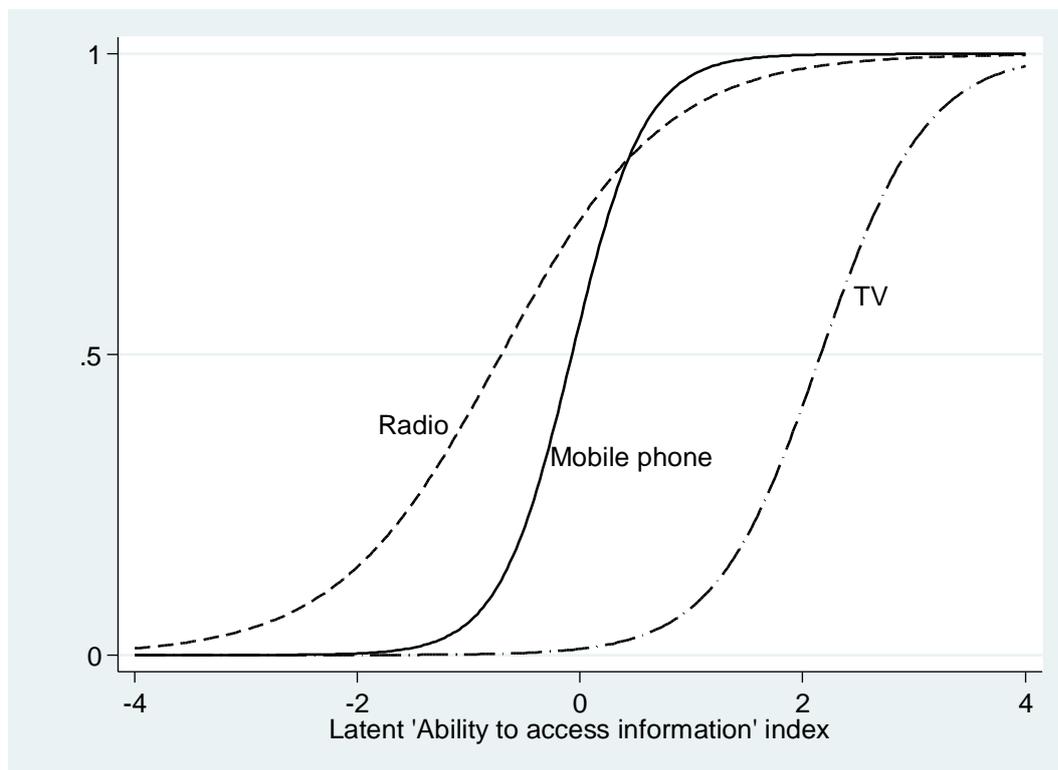
Figure 3 provide a useful alternative representation of the results. The curves show that for a household with a relatively low latent access to information (e.g. Index level = -0.5) the probability of having access to a radio could be relatively high. In comparison, for such a household the probability of access to a mobile phone will be low and the probability of

25

having access to a TV extremely low. In contrast, having a low probability of access to a TV is possible with a relatively high latent value of ability to access information. The steeper slope of the ICC (around zero) for access to a mobile phone underlines that this item captures the underlying latent variable more strongly than the other items, with small differences in underlying latent ability to access information (around zero) implying large differences in the probability of access to a mobile phone.
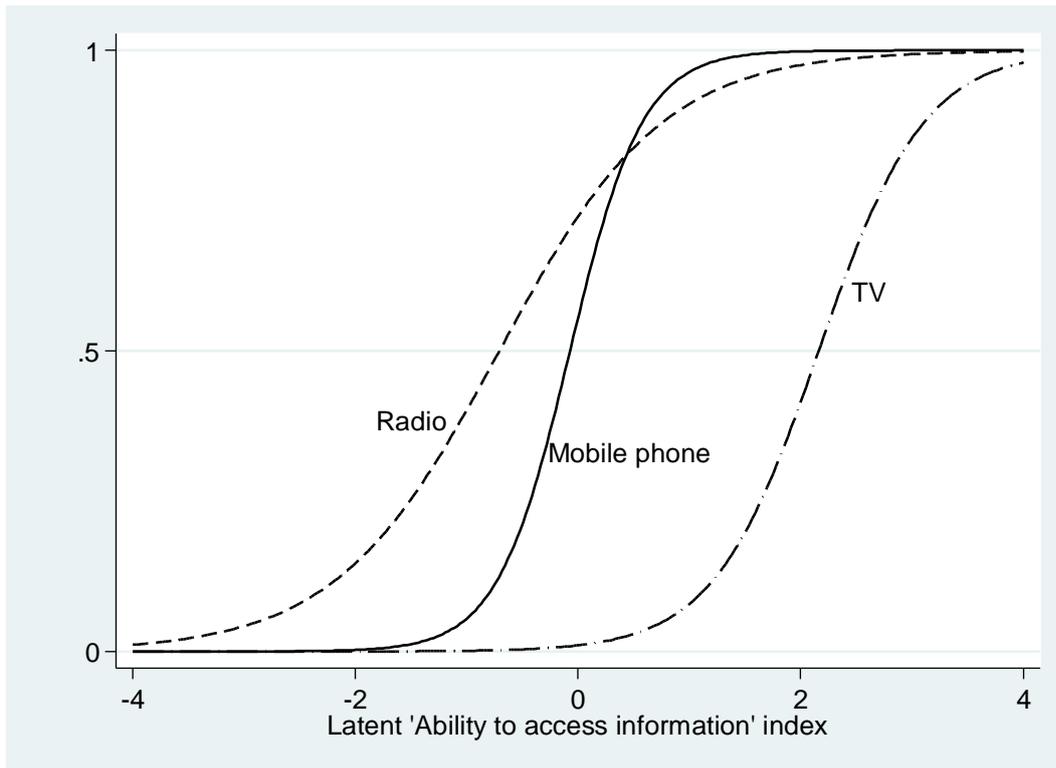
Figure 3: Estimated item characteristic curve for 'ability to access information' index.

Table 3 reports the estimated latent index scores and their standard errors and 90% prediction intervals for different combinations of the electrical items. We conduct a skewness and kurtosis test for normality and reject the hypothesis that the latent information index is normally distributed at the 10% significance level. However, we find that the distribution is symmetric[16] and moderate (i.e. between -3 and 3 for the estimated latent index and difficulty parameters). These results suggest that bias in the IRT model can be considered negligible (Robins et al, 2009; Seong,1990).

---

[16] The absolute value of the skewness is less than twice its standard error, suggesting that the data is symmetric.

Table 3: Predicted Scores for ability to access information (infoIndex) based on access to electrical items.

Household access to item (s)	Predicted Latent index, (*infoIndex*)	Forecast error	90% Prediction Interval
None	-1.01	0.69	(-1.04,-0.98)
Radio only	-0.45	0.60	(-0.47,-0.43)
Television only	-0.22	0.58	(-0.54,0.10)
Mobile phone only	0.17	0.58	(0.13,0.21)
Radio, Television only	0.23	0.59	(-0.06,0.52)
Radio, Mobile phone only	0.67	0.64	(0.65,0.69)
Mobile phone, Television only	0.95	0.67	(0.70,1.20)
All items	1.61	0.71	(1.54,1.68)

While values of the latent index are difficult to interpret in isolation, the relative values provide an informal check on whether the index is consistent with expectations. Firstly, farms managed by households with access to all 3 items score highest on the index (score = 1.61) while farms managed by households with no access to any of the items score lowest (score = -1.01). Secondly, the higher the number of items a farmer has access to, the higher their index score. Thirdly, access to more difficult and discriminatory items lead to higher scoring in the index.

4.2 SF model results

The SF model described by the Wang (2002) model (6) allows for heteroscedasticity both in the inefficiency and idiosyncratic error variances. To ensure a parsimonious specification, we tested 3 nested versions of this model. In Model 1, we allowed for heteroscedasticity in both technical inefficiency and the idiosyncratic error. Model 2, allowed for heteroscedasticity in the inefficiency term only (as heteroscedasticity in the idiosyncratic error term is less important in SF models (Wang, 2002)). Model 3 imposes homoscedasticity in both inefficiency and idiosyncratic error variances (equivalent to the original Battese and Coelli (1995) specification). The results of the likelihood ratio tests suggested that relative to Model 1 (i.e. the general model), the restrictions implied by the nested Model 2 and Model 3 could

be rejected at the 1% significance level.[17] We therefore based our subsequent analysis on Model 1 only.

As discussed above, we implement a sequential two-step non-parametric bootstrapping procedure in which the predicted *infoIndex* variable in the first stage IRT model is used in the second stage SF model. Table 4 reports the results for this model for the specification given in equation (7). For brevity only the production estimates found to be statistically significant at 10% or less are reported in the Table. For example a number of interactions between the presence of an ox on the farm size and use of chemicals were not found to be statistically significant and so these coefficients are not reported.

---

[17] Testing Model 1 vs Model 2 yielded a likelihood ratio statistic of 24.78 (p-value <0.01). For Model 1 vs Model 3, the test statistic was 316.68 (p-value < 0.01). Model 1 is therefore preferable as the restrictions implied by Model 2 and Model 3 can be rejected at the 1% significance level.

Production frontier estimates		Technical efficiency estimates	
Variable	**Coefficient**	**Variable**	**Coefficient**
Production frontier		**Mean technical inefficiency**	
$y_{it} = \ln\left\{ f\left(x_{kit}, \theta \right) \right\} + v_{it} - u_{it}$		$\mu_{it} = \delta_0 + z_{it}^{'}\delta$	
ln (*Farm_size*)	0.24***	*infoIndex*	-2.40***
	(0.02)		(0.84)
ln (*Labour*)	0.24***	Literacy level	-17.45***
	(0.03)		(6.23)
ln (*Chemicals*)	0.16***	*infoIndex**Literacy level	-14.70***
	(0.02)		(5.36)
Ox	0.20*	Access to extension service	-0.61**
	(0.10)		(0.25)
ln (*Farm_size*)^2	0.04***	Age of head farmer	0.44*
	(0.01)		(0.23)
ln (*Farm_size*)*ln(*Labour*)	-0.03***	Gender (Male = 1, Female=0)	-0.17
	(0.01)		(0.15)
ln (*Farm_size*)*ln(*Chemicals*)	0.001**	Land tenure (owned =1, otherwise=0)	-2.06***
	(0.00)		(0.52)
ln (*Labour*)^2	0.04***	Location of farm (Urban=1, Rural=0)	0.56
	(0.01)		(0.35)
ln (*Labour*)*ln(*Chemicals*)	-0.01***	Constant	-3.17**
	(0.00)		(1.61)
ln (*Labour*)*Ox	-0.11***		
	(0.04)	**Heteroscedasticity in technical inefficiency variance**	
ln (*Chemicals*)^2	0.02***	$\sigma_{u,it}^2 = \exp\left(\omega_{u0} + z_{u,it}^{'}\omega_u \right)$	
	(0.00)	*infoIndex*	0.89***
Time, *t*	0.04*		(0.05)
	(0.03)	Literacy level	1.01***
			(0.12)
		Constant	1.28***
			(0.27)
		Heteroscedasticity in idiosyncratic error variance	
		$\sigma_{v,it}^2 = \exp\left(\omega_{v0} + z_{v,it}^{'}\omega_v \right)$	
		infoIndex	0.17***
			(0.04)
		Literacy level	-0.04
			(0.09)
		Constant	-0.43***
			(0.07)
Log likelihood	-10,190		
Bootstrap observations	7353		
Number of replications	100		

29

Table 4: Production frontier and technical efficiency estimates in the SF model.

The first half of Table 4 shows only the parameter estimates of the production frontier component of the SF model which are statistically significant. Coefficients of single variable terms are interpreted as elasticities of output. Consistent with monotonicity requirements for a well behaved production function (Sauer, et al, 2006), the results show that the output elasticities of all the inputs are positive and significantly different from zero, suggesting that increased use of the inputs increases output. [18] After controlling for other factors, the coefficient on the time trend variable suggests positive growth in output over time.

The coefficients of the reported squared terms which were statistically significant show the marginal effect of a change in the level of the associated input on output as the level of output increases. The result suggests that a marginal increase in the use of labour, farm size and chemical input significantly increases output, as the level of output increases. The coefficients of the interaction terms capture the impact of one input on the marginal effect of another. A significant negative interaction suggests that the inputs are substitutes, as is the case for farm size and labour as well as labour and use of chemicals. A significant positive interaction suggests that both inputs are complementary, as is the case for farm size and use of chemicals, although the size of the effect is small in absolute terms.

The second half of Table 4 reports the results of the impact of the various exogenous factors on technical efficiency. A negative coefficient implies the associated variable is linked with higher efficiency. The coefficient on the information index and the level of literacy as well as the interaction of the two variables is negative and statistically significant. This suggests that consistent with our prior expectations, farms managed by households with higher access to information and literacy levels are more efficient and closer to the stochastic production frontier. The interaction also provides some support for the claim that those with higher literacy or education levels have a greater capacity to use information effectively (Reimers and Stephan, 2013). Similarly, access to information through access to agricultural extension officers significantly increases efficiency. As found in previous studies, the security with

---

[18] Quasi-concavity was not tested for nor imposed as this would restrict the flexibility of the production function to allow for possible non-convexities in the production set e.g. due to lumpy inputs (Chapter 10, Ginsburgh and Keyzer, 2002).

which land is held is also important with owned (freehold) farms associated with higher efficiency levels (Michler and Shively, 2015). However, the age of the farm household head is associated with lower efficiency. This is consistent with the evidence suggesting younger farmers are more likely to adopt innovative technologies and best practices hence are more efficient (Bozoğlu and Ceyhan, 2007). Lastly, the results suggest that the gender of the head of household and location of a farm do not significantly affect efficiency.

The results also suggest a positive relationship between the information index and both inefficiency and idiosyncratic error variances. The household level of literacy is also positively related to the variance of inefficiency. This implies increased access to information and higher literacy is associated with increased variance of inefficiency and output. This empirical result is consistent with the findings of Wang (2002) who shows that while a variable may be expected to increase mean efficiency (as is the case for these two variables), the same variable may not necessarily reduce inefficiency and/or output variances.

4.3 Distribution of technical efficiency and idiosyncratic errors by ability to access information

To further understand the nature of the relationship between access to information on the mean and variances of inefficiency and idiosyncratic errors, we split our sample into two groups; one with lower and the other with higher access to information, using the mean information index value as a "natural" threshold to split the sample. In this discussion it is more natural to consider the impact of information on technical efficiency (rather than inefficiency). Hence we derive technical efficiencies from the estimated technical inefficiencies using the Jondrow et al (1982) method. As inefficiency and efficiency variance are equivalent, the variance estimation results translate directly here, i.e. the results suggest

higher   access   to   information   increases   efficiency   variance.
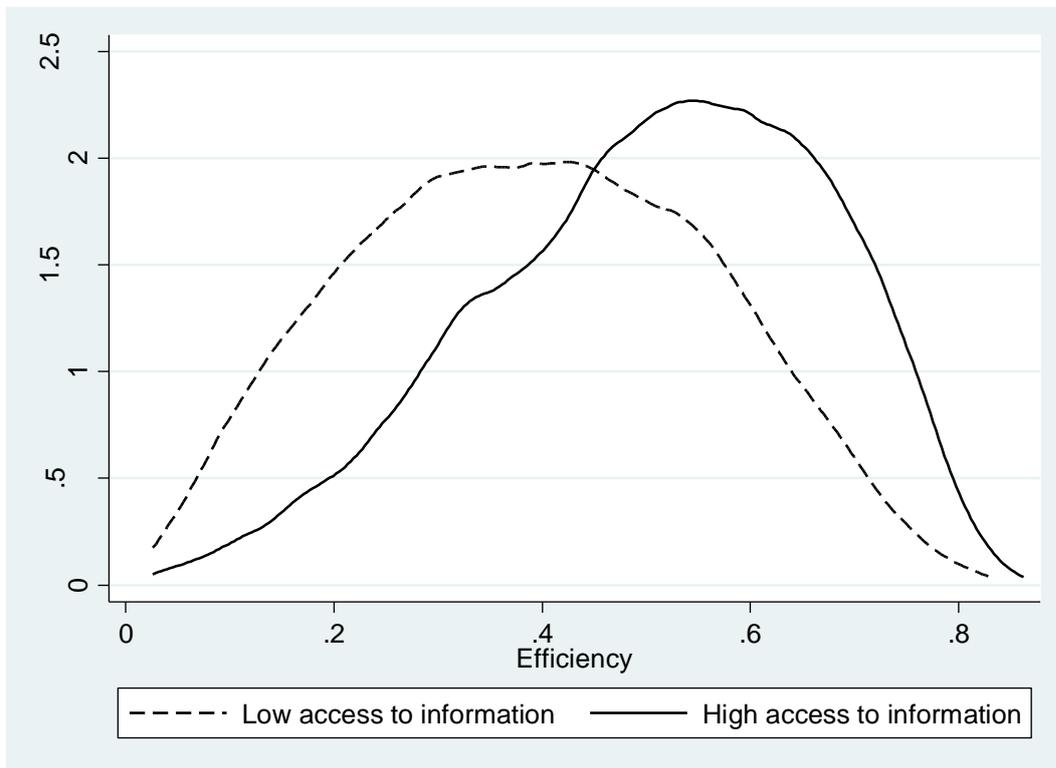
Figure   4   and
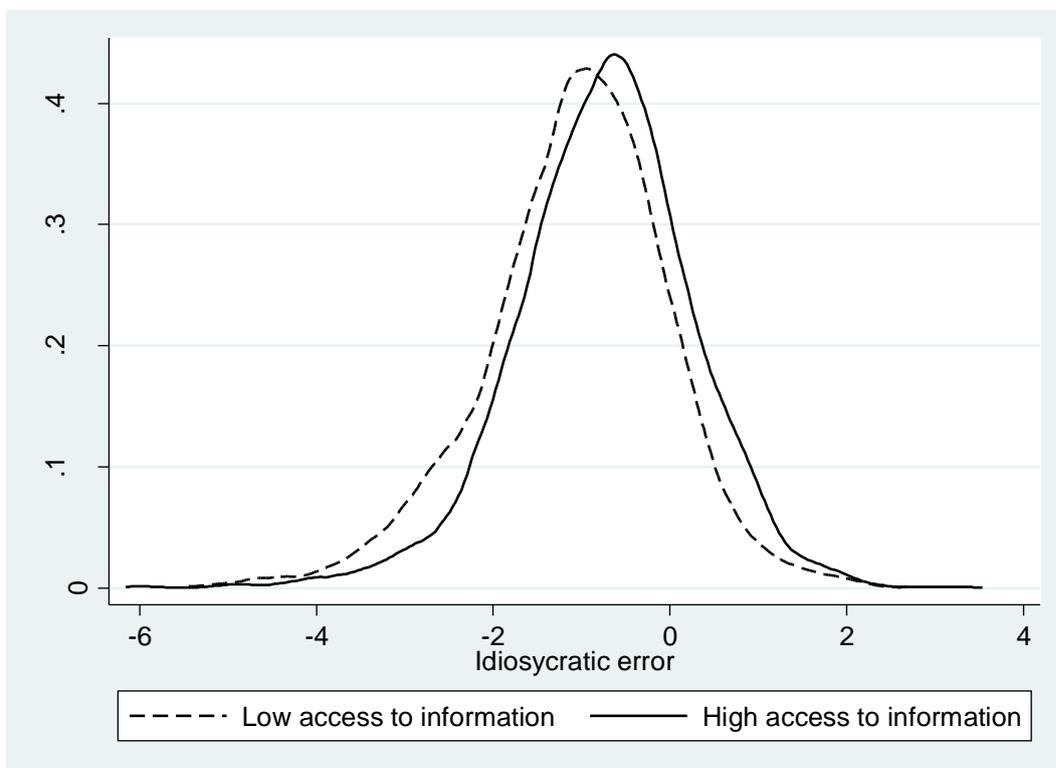
Figure 5 plot the distributions of technical efficiency and idiosyncratic errors for the two groups.
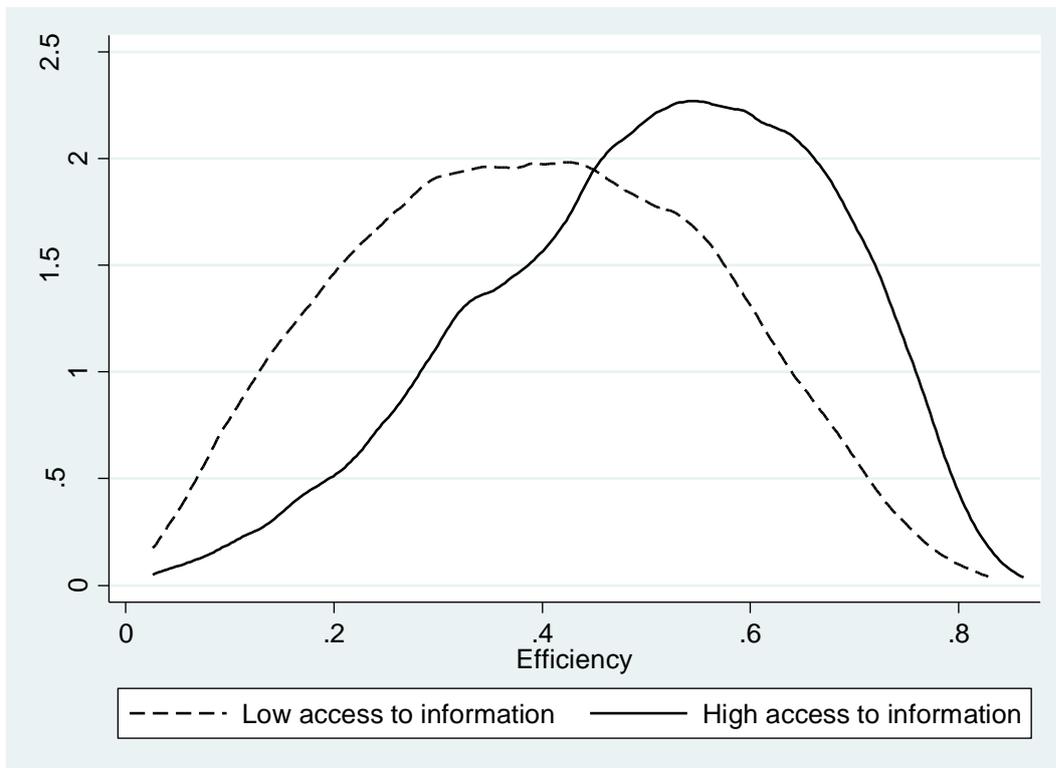
Figure 4: Distribution of efficiency by the access to information index
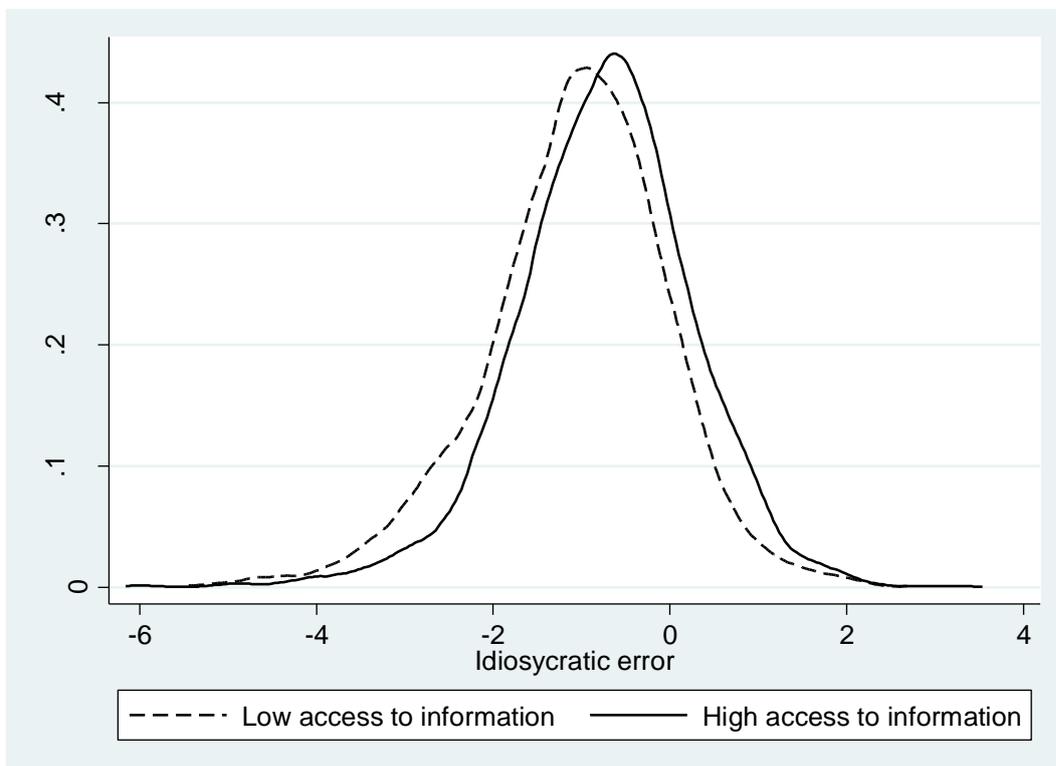
Figure 5: Distribution of idiosyncratic error by access to information index

Based on Kolmogorov-Smirnov tests we find evidence that in both cases the distributions are statistically different.[19]
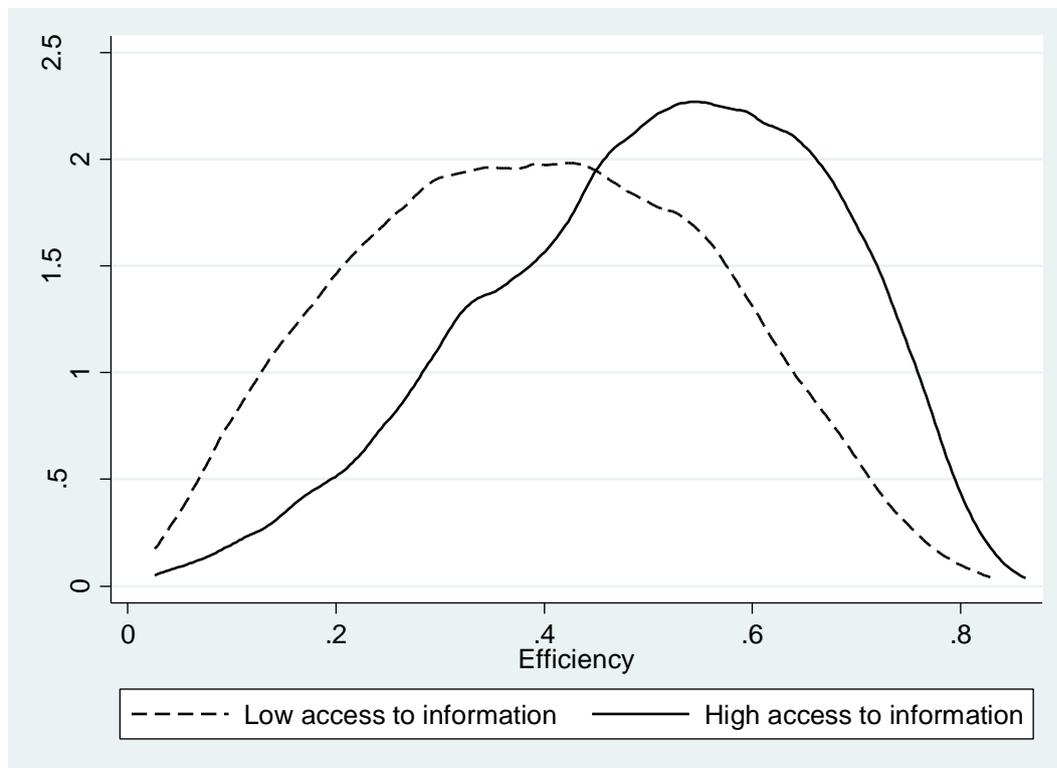
Figure 4 shows that the distribution of efficiency for farms managed by households with higher access to information have greater mean efficiency (51%) than farms managed by households with lower access to information (39%). However, the variance of the efficiency distribution for higher access to information (0.17) is slightly greater than that for lower access to information (0.16) hence corroborating our empirical results (i.e. higher variance of

---

[19] For the efficiency distributions in Figure 4, the combined Kolmogorov-Smirnov test statistic is 0.28 and the p-value < 0.01. For the idiosyncratic error distributions in Figure 5, the combined Kolmogorov-Smirnov test statistic is 0.13 and the p-value < 0.01.

inefficiency   translates   into   higher   variance   of   efficiency).
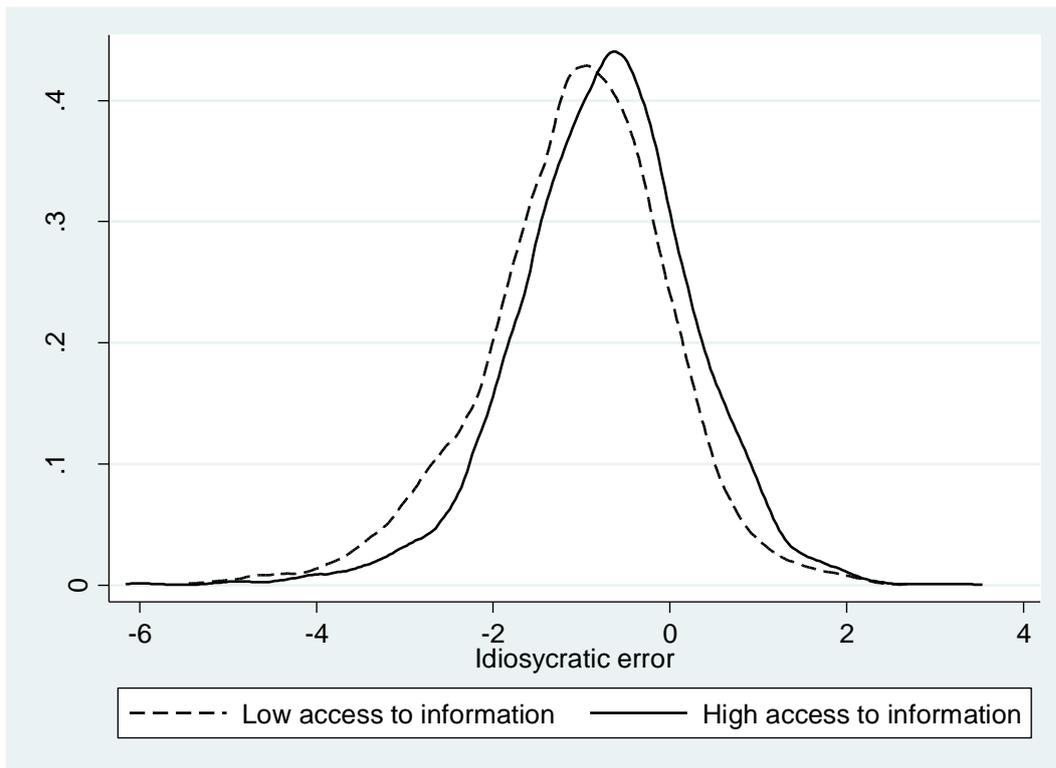
Figure 5 also shows that the distribution of idiosyncratic error for farms managed by households with higher access to information has a higher mean (-0.74) than the mean of the idiosyncratic error distribution for farms managed by households with lower access to

information (-1.06). The results from
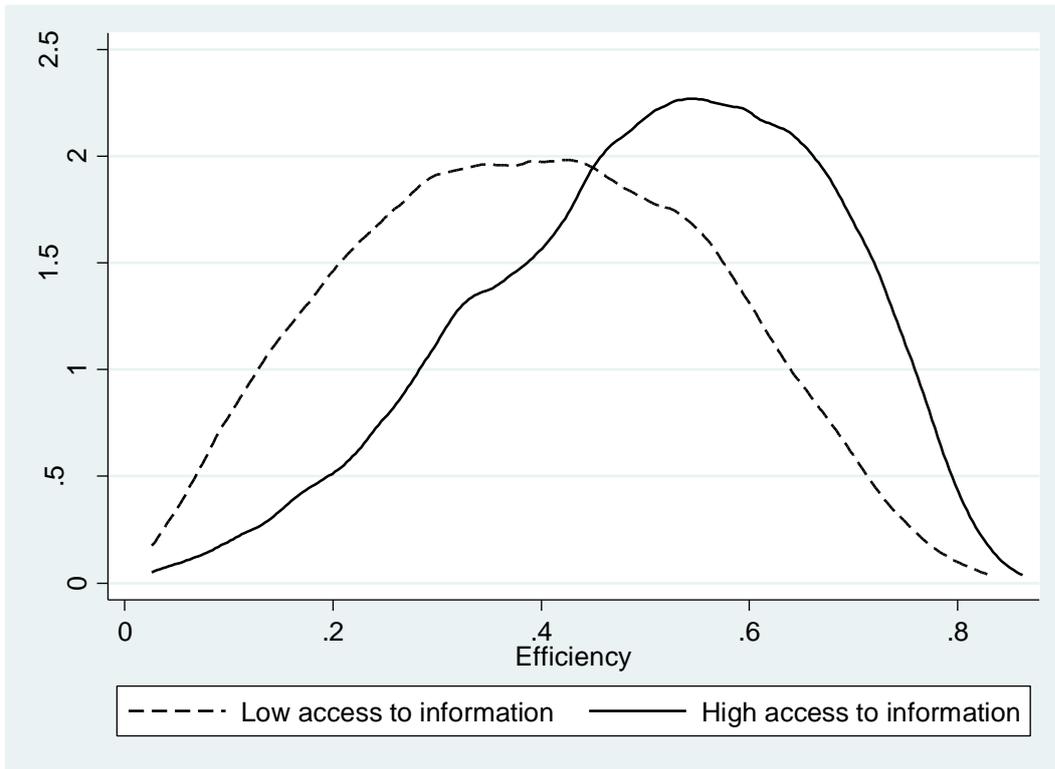
Figure 4 and
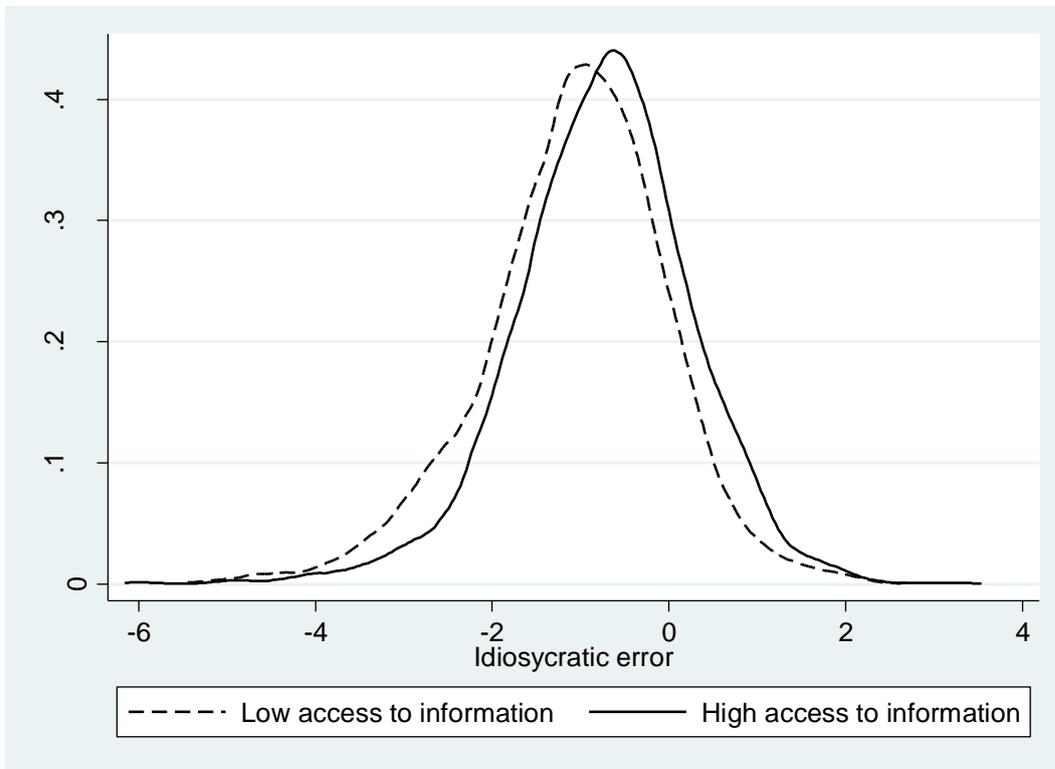
Figure 5 suggest that while improved access to information may increase variances of (in)efficiency and outputas shown in Table 4, it does so by increasing the probability of higher efficiency and output realisations while reducing the likelihood of lower efficiency and output realisations. The higher variances for households with greater access to information may also be consistent with these farmers being more willing to accept risk (Fafchamps and Pender, 1997).

4.4 Robustness checks

We have undertaken a number of supplementary analyses to check the robustness of the results. To explore the impact of the construction of the latent ability to access information variable *infoIndex*, we re-estimated the SF model by replacing this with a simple count of the items to which each household has access. The results of this analysis are qualitatively similar to the results reported in Table 4 and are reported in on-line Appendix B (Table B1).

To explore whether the restrictions placed by the Wang model on the structure of efficiency are driving the key results we also estimate the Kumbhakar et al (2014) model that better exploits the panel data by accounting for unit specific heterogeneity, persistent efficiency and transient efficiency. As previously stated in Section 2.5, we are unable to use this model to directly estimate the impact of access to information on persistent and transient efficiencies, or on variances of (in)efficiencies and/or idiosyncratic errors. However, we are able to estimate persistent and transient efficiencies which we can then calculate for farm households with low and high access to information. Table 5 shows these results. (The full estimation results are reported in Table B2, on-line Appendix B.).

		Number	Mean	Std dev	95% Confidence interval [LB,UB]	
Persistent efficiency	Overall	7475	0.59	0.13	0.59	0.59
	High *infoIndex*	3928	0.61	0.13	0.61	0.62
	Low *infoIndex*	3547	0.56	0.14	0.56	0.57
Transient efficiency	Overall	7475	0.68	0.09	0.67	0.68
	High *infoIndex*	3928	0.68	0.09	0.67	0.68
	Low *infoIndex*	3547	0.67	0.10	0.67	0.68

Table 5: Persistent and transient efficiency by 'ability to access information' index

As we would expect given the results in Table 4, persistent efficiency for households with high access to information is higher than those for households with low information access. This suggests households with greater access to information are inherently more efficient than households with low access to information. However, transient efficiency is only marginally higher in households with greater access to information.

Finally, the conventional approach to SF analysis has been to focus on the effect of exogenous efficiency determinants on production efficiency and not production output. As a further robustness check, we retrieve the predicted production output for households with low and high access to information. In line with expectations, predicted production output for households with high access to information is greater, with mean predicted outputs being 0.31 million U.Sh and 0.48 million U.Sh for households with low and high access to information respectively. This result corroborates our expectations and findings with respect to production efficiency.

5. Conclusions

We tested empirically the relationship between small-scale farmers' ability to access information and farm technical efficiency using a stochastic frontier model. The novelty of our empirical approach is that farmers' ability to access information is constructed as an index using a 2-parameter IRT model. We find evidence of a significant and positive relationship between farm efficiency and farmers' ability to access information. There is also evidence that the size of these effects is larger for literate and hence better educated farmers. Greater access to information also appears to be associated with higher (in)efficiency and output variance although the increased variances are associated with a lower risk of lower efficiency and output realisations and higher likelihood of higher efficiency and output realisations.

While identifying causality remains challenging, the results do suggest a mechanism through which providing limited electricity access to poor farm households can have positive farm efficiency effects. Hence policy makers may be able to improve adoption of small scale off-grid renewables such as Solar PV in sub-Saharan African countries such as Uganda by giving greater emphasis to the possible information and efficiency benefits if the electricity is used to power mobile phones, radios and televisions. It also suggests that there may be benefits associated with improved coordination of extension and electricity access policies so that

improved information access leads to better use of available information and improved farm efficiency.

Clearly simply having access to information does not guarantee its effective use. Hence, exploring ways in which we might better disentangle the impacts of information access and how effectively it is used is clearly an area where future research would be useful. This could also help improve our understanding of the relationship between information access and (in)efficiency and output variances. Finally, cross-country comparisons might also provide useful insights into how different institutional contexts impact on any information effects.

References

Abdul-Salam, Y., Phimister, E. 'The politico-economics of electricity planning in developing countries: A case study of Ghana', *Energy Policy*, Vol 88, (2016b) pp. 299-309.

Aker, J. C., Fafchamps, M. 'Mobile phones and farmers' welfare in Niger'. *Center for the Study of African Economies (CSAE) Conference*, (2010).

Amsler C, Prokhorov, A., Schmidt, P. 'Endogeneity in Stochastic Frontier Models', *The University of Sydney, Working Paper*, (2014).

Bartholomew, D. J. 'The sensitivity of latent trait analysis to choice of prior distribution', *British Journal of Mathematical and Statistical Psychology*, Vol 41, no. 1, (1988) pp. 101-107.

Battese, G. E., Coelli, T. J. 'A model for technical inefficiency effects in a stochastic frontier production function for panel data', *Empirical Economics,* Vol. 20(2), (1995) pp. 325-332.

Belotti, F., Daidone, S., Ilardi, G., Atella, V. 'Stochastic frontier analysis using Stata', *Stata Journals*, Vol ii, (2012) pp. 1-39.

Belotti, F., Ilardi, G. 'Consistent Estimation of the 'True'Fixed-Effects Stochastic Frontier Model.' *CEIS Working Paper No.* 231 (2012)

Benjamin, C., Phimister, E. 'Does capital market structure affect farm investment? A comparison using French and British farm level panel data', American *Journal of Agricultural Economics*, Vol. 84(4), (2002) pp.1115-1129.

Birnbaum, A. 'Some latent trait models and their use in inferring an examinee's ability', *Statistical Theories of Mental Test Scores*, (1968) pp. 395-479

Bock, R. D., Aitkin, M. 'Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm', *Psychometrika* 46, no. 4, (1981) pp. 443-459.

Bozoğlu, M., Ceyhan, V. 'Measuring the technical efficiency and exploring the inefficiency determinants of vegetable farms in Samsun province, Turkey', *Agricultural Systems*, Vol. *94*(3), (2007) pp. 649-656.

Bravo-Ureta B.E, Solis, D., Lo´pez, V. H. M., Maripani, J. F., Thiam, A., Rivas, T. 'Technical efficiency in farming: a meta-regression analysis', *Journal of Productivity Analysis*, Vol 27, (2007) pp. 57–72

Cameron, A.C., Trivedi, P.K. 'Microeconometrics using stata', *College Station TX Stata Press*, Vol. 2, (2010)

Cappellari, L., Jenkins, S. P. 'Summarizing multiple deprivation indicators. In S. P. Jenkins & J. Micklewright (Eds.), Inequality and poverty re-examined', *Oxford University Press*, (2007)

Castiglione, C. 'Technical efficiency and ICT investment in Italian manufacturing firms', *Applied Economics*, Vol. 44(14), (2012) pp. 1749-1763.

Coelli, T.J., Battese, G.E. 'Identification of factors which influence the technical inefficiency of Indian farmers', *Australian Journal of Agricultural Economics,* Vol. 40.2, (1996) pp. 103-128.

Coelli, T.J., Dodla, S., O'Donnell, C.J., Battese, G.E. 'An introduction to efficiency and productivity analysis', *Springer Science & Business Media*, (2005)

Dawson, P. J., Lingard, J., Woodford, C. H. 'A generalized measure of farm-specific technical efficiency', *American Journal of Agricultural Economics,* Vol 73(4), (1991) pp. 1098-1104.

Deichmann, U., Meisner C., Murray, S., Wheeler, D. 'The economics of renewable energy expansion in rural Sub-Saharan Africa', *Energy Policy,* Vol 39, no. 1, (2011) pp. 215-227.

Deininger, K W., Mpuga, P. 'Land Markets In Uganda: Incidence, Impact And Evolution Over Time', *Contributed Paper International Association of Agricultural Economists 2003 Annual Meeting,* (2003). Available at: http://ageconsearch.umn.edu/handle/25809 [Accessed May 2015]

Esselaar, S., Stork, C., Ndiwalana, A., Deen-Swarray, M. 'ICT usage and its impact on profitability of SMEs in 13 African countries', *Information and Communication Technologies and Development,* (2006) pp. 40-47.

Embretson, S. E., Reise, S. P. 'Item response theory'. *Psychology Press*, (2013).

Fafchamps, M., Pender, J. 'Precautionary saving, credit constraints, and irreversible investment: Theory and evidence from serni-arid India', *Journal of Business and Economic Statistics*, Vol. 15(2), (1997) pp. 180-194.

Faye, O., Baschieri, A., Falkingham, J., Muindi, K. 'Hunger and food insecurity in Nairobi's slums: an assessment using IRT models.' *Journal of Urban Health*, Vol *88*(2), (2011) pp. 235-255.

Filmer, D., Scott, K.. 'Assessing asset indices', *Demography*, 49(1), (2012) pp.359–392.

Fu, X., Akter, S. 'The Impact of Mobile Phone Technology on Agricultural Extension Services Delivery: Evidence from India', *The Journal of Development Studies* (2016): 1-16.

Greene, W. 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics,* Vol. 126.2, (2005a) pp. 269-303.

Greene, W. 'Fixed and random effects in stochastic frontier models.' *Journal of productivity analysis,* Vol 23.1 (2005b) pp. 7-32.

Jensen, R. 'The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector', *The Quarterly Journal of Economics*, (2007) pp. 879-924.

Jondrow, J., Lovell, C. K., Materov, I. S., Schmidt, P. 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics*, Vol. 19(2), (1982) pp. 233-238.

41

Karagiannis, G., Sarris, A. 'Direct subsidies and technical efficiency in Greek agriculture', *X^{th} EAAE Congress Zaragoza (Spain),* (2002)

Ginsburgh, V., M. A. Keyzer 'The Structure of Applied General Equilibrium Models' *MIT Press,* (2002).

Kumbhakar, S. C. 'Production frontiers, panel data, and time-varying technical inefficiency', *Journal of Econometrics,* Vol. 46(1), (1990) pp. 201-211.

Kumbhakar S.C, Gudbrand, L., Hardaker, J. B. 'Technical efficiency in competing panel data models: a study of Norwegian grain farming', *Journal of Productivity Analysis*, Vol 41, (2014) pp. 321–337

Kumbhakar, S. C., Heshmati, A. 'Efficiency measurement in Swedish dairy farms: an application of rotating panel data, 1976–88.' *American Journal of Agricultural Economics*, Vol 77(3), (1995) PP. 660-674.

Kumbhakar, S. C., Lovell, C. K. 'Stochastic frontier analysis', *Cambridge University Press,* (2003).

Kumbhakar, S. C., Wang, H. J. 'Estimation of growth convergence using a stochastic production frontier approach.' *Economics Letters*, Vol 88(3), (2005) pp. 300-305.

Lancaster, T. 'The incidental parameter problem since 1948.' *Journal of econometrics*, Vol 95(2), (2000) pp 391-413.

Lee, Y. H., Schmidt, P. 'A production frontier model with flexible temporal variation in technical efficiency', *The Measurement of Productive Efficiency: Techniques and Applications,* (1993), pp. 237-255.

Li, Y., Pickles, A., Savage, M. 'Social capital and social trust in Britain'. *European Sociological Review*, Vol. 21(2), (2005) pp.109–123.

Mair, P., Reise, S.P., Bentler, P.M. 'IRT Goodness-of-Ft Using Approaches From Logistic Regression', Department *of Statistics, UCLA, Los Angeles*, (2008). Available at: http://escholarship.org/uc/item/1m46j62q. [Accessed May 2015]

Michler, J. D., Shively, G. E. 'Land tenure, tenure security and farm efficiency: Panel evidence from the Philippines', *Journal of Agricultural Economics,* Vol. 66(1), (2015) pp. 155-169.

Ødegaard, F., Roos, P. 'Measuring the Contribution of Workers' Health and Psychosocial Work-Environment on Production Efficiency', *Production and Operations Management*, Vol. 23 (12), (2014) pp 1937-5956

Overå, R. 'Networks, distance, and trust. Telecommunications development and changing trading practices in Ghana', *World Development,* Vol. 34(7), (2006) pp. 1301-1315.

Pericoli, F. M., Pierucci, E., Ventura, L. 'The impact of social capital on consumption insurance and income volatility in the UK: evidence from the British Household Panel Survey', *Review of Economics of the Household* , Vol. 13(2), (2015) pp.269-295

Rahman, S. 'Women's Labour Contribution to Productivity and Efficiency in Agriculture: Empirical Evidence From Bangladesh', *Journal of Agricultural Economics*, Vol. 61, No. 2, (2010) pp. 318–342

Reimers, M., Stephan, K. 'Revisiting the role of education for agricultural productivity', *American Journal of Agricultural Economics.* Vol. 95.1: (2013) pp. 131-152.

Reise, S. P., Revicki, D.A. 'Handbook of item response theory modeling: Applications to typical performance assessment'. *Routledge*, (2014).

Robins, R.W., R. Fraley, R.C., Krueger, R.F. 'Handbook of research methods in personality psychology', *Guilford Press*, (2009).

Sauer, J., K. Frohberg, H. Hockmann 'Stochastic efficiency measurement: The curse of theoretical consistency', *Journal of Applied Economics*, Vol. 9, (2006) pp. 139-165.

Seong, T. J. 'Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions', *Applied Psychological Measurement*, Vol. 14(3), (1990) pp. 299-311.

Shao, B. B., Lin, W. T. 'Measuring the value of information technology in technical efficiency with stochastic production frontiers', *Information and Software Technology*, Vol. 43(7), (2001) pp. 447-456.

Solis, D., Bravo-Ureta, B. E., Quiroga, R.E.. 'Technical Efficiency among Peasant Farmers Participating in Natural Resource Management Programmes in Central America', *Journal of Agricultural Economics*, Vol. 60, No. 1, (2009) pp. 202–219

UN Sustainable Development, 2015. Available at: 'https://sustainabledevelopment.un.org/focussdgs.html' [Accessed May 2015]

USAID, 'Country Profile: Property Rights & Resource Governance Uganda', (2010). Available at: http://www.usaidlandtenure.net/uganda [Accessed May 2015]

Vandemoortele, M. 'Measuring Household Wealth with Latent Trait Modelling: An Application to Malawian DHS Data', *Social Indicators Research,* Vol. 118, (2014) pp. 877-891

Wang, H. J. 'Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model', *Journal of Productivity Analysis*, Vol. 18(3), (2002) pp. 241-253.

Wang, H., Schmidt, P. 'One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels', *Journal of Productivity Analysis,* Vol. *18*(2), (2002) pp. 129-144.

World Bank. 'World Development Report, 2008', *The International Bank for Reconstruction and Development* (2008), Washington DC

World Bank. 'World Development Report, 2012', *The International Bank for Reconstruction and Development* (2012), Washington DC

World Bank. 'Living Standards Measurement Survey – Integrated Surveys on Agriculture (LSMS-ISA); The Uganda National Panel Survey', (2013). Available at http://go.worldbank.org/FS2M7AYE00 [Accessed March 2013]

Zavale, H., Mabaya, E., Christy, R. 'Smallholders' cost efficiency in Mozambique: Implications for improved maize seed adoption', *Staff paper No. 4, Cornell University, Department of Applied Economics and Management* (2005)

Efficiency Effects of Access to Information on Small Scale Agriculture: Empirical Evidence from Uganda using Stochastic Frontier and IRT Models

Yakubu Abdul-Salam and Euan Phimister

On-Line Appendices

Appendix A: Calculating a measure of real output

Here, we illustrate the construction of the real output values using a constructed Laspeyres Price Index. Let the following sets represent identities in the data;

i = set of households

t = set of 6 time periods

l = set of land parcels a household farms

p = set of plots in a land parcel

c = set of crops planted on a plot

Table A1 highlights the level of observation and measurement units of two data variables that we use in the construction of deflated revenues.

Variable	Level of observation	Brief Description	Unit of measurement
amount _ harvested	i,t,l,p,c	Amount harvested of crop ID c in period t	Kilograms (kg)
amount _ sold	i,t,l,p,c	Amount sold of crop ID c in period t	Ugandan shilling (U.Sh)

Table A1: Variables used to construct revenues.

Variable $amount\_harvested(i,t,l,p,c)$ is fully described as household i in period t harvests *amount_harvested* kg of crop ID c which was planted on plot ID p of land parcel ID l. We compute prices attained for crop c for household i in period t as follows;

$$price(i,t,c) = \sum_{l,p} \frac{amount\_sold(i,t,l,p,c)}{amount\_harvested(i,t,l,p,c)} \; \forall \; amount\_sold(i,t,l,p,c) > 0$$

where $price(i,t,c)$ is price in Ugandan shillings per kilogram. To prevent outliers, the prevailing market price of crop ID c in period t i.e. $price_{market}(c,t)$ is calculated as the median of prices attained for the crop across all households. Let $rev(i,t)$ (U.Sh) represent the nominal revenue realised by farm household i in period t. Given the prevailing market price for each crop, we can calculate the nominal revenue for a household as follows;

$$rev(i,t) = \sum_{l,p,c} \left(amount\_harvested(i,t,l,p,c) \times price_{market}(c,t) \right)$$

To deflate the revenues to their real values, we construct a Laspeyres price index. Let $quantity(c,t)$ denote the quantity of crop ID c harvested over all households in period t;

$$quantity(c,t) = \sum_{l,i,p} amount\_harvested(i,t,l,p,c)$$

The Laspeyres price index is computed as follows;

$$L(t) = \frac{\sum_c \left(price_{market}(c,t) \cdot quantity(c,t=1) \right)}{\sum_c \left(price_{market}(c,t=1) \cdot quantity(c,t=1) \right)}$$

where $t=1$ is the base period, notionally the first 6 months of the 2009 cropping season in the UNPS data. With this index, we deflate the nominal revenue to find the real output measure $y(i,t)$ as follows;

$$y(i,t) = \frac{rev(i,t)}{L(t)}$$

The natural log of $y(i,t)$ is used as the dependent variable in the SF model.

Appendix B: Results and data

Production frontier estimates	IRT information index	Simple count index
Production frontier		
$y_{it} = \ln\left\{ f\left(x_{kit}, \theta\right)\right\} + v_{it} - u_{it}$		
ln (*Farm_size*)	0.24***	0.23***
	(0.02)	(0.02)
ln (*Labour*)	0.24***	0.24***
	(0.03)	(0.03)
ln (*Chemicals*)	0.16***	0.16***
	(0.02)	(0.02)
Ox	0.20*	0.21**
	(0.10)	(0.10)
ln (*Farm_size*)^2	0.04***	0.04***
	(0.01)	(0.01)
ln (*Farm_size*)*ln(*Labour*)	-0.03***	-0.04***
	(0.01)	(0.01)
ln (*Farm_size*)*ln(*Chemicals*)	0.001**	0.00**
	(0.00)	(0.00)
ln (*Labour*)^2	0.04***	0.04***
	(0.01)	(0.01)
ln (*Labour*)*ln(*Chemicals*)	-0.01***	-0.01***
	(0.00)	(0.00)
ln (*Labour*)*Ox	-0.11***	-0.11***
	(0.04)	(0.04)
ln (*Chemicals*)^2	0.02***	0.02***
	(0.00)	(0.00)
Time, *t*	0.04*	
	(0.03)	
Mean technical inefficiency		
$\mu_{it} = \delta_0 + z_{it}^{'}\delta$		
infoIndex {simpleIndex}	-2.40***	{-1.03***}
	(0.84)	(0.32)
Literacy level	-17.45***	-1.76***
	(6.23)	(0.52)
infoIndex*Literacy level {*simpleIndex**Literacy level}	-14.70***	{-6.36***}
	(5.36)	(1.83)
Access to extension service	-0.61**	-0.54***
	(0.25)	(0.18)
Age of head farmer	0.44*	0.37**
	(0.23)	(0.18)
Gender (Male = 1, Female=0)	-0.17	-0.19

	(0.15)	(0.12)
Land tenure (owned =1, otherwise=0)	-2.06***	-1.65***
	(0.52)	(0.32)
Location of farm (Urban=1, Rural=0)	0.56	0.49*
	(0.35)	(0.26)
Constant	-3.17**	-0.20
	(1.61)	(0.76)

Heteroscedasticity in technical inefficiency variance

$$\sigma_{u,it}^2 = \exp\left(\omega_{u0} + z_{u,it}'\omega_u\right)$$

InfoIndex {simpleIndex}	0.89***	{0.61***}
	(0.05)	(0.05)
Literacy level	1.01***	0.99***
	(0.12)	(0.12)
Constant	1.28***	0.11
	(0.27)	(0.19)

Heteroscedasticity in idiosyncratic error variance

$$\sigma_{v,it}^2 = \exp\left(\omega_{v0} + z_{v,it}'\omega_v\right)$$

infoIndex	0.17***	0.15***
	(0.04)	(0.03)
Literacy level	-0.04	-0.06
	(0.09)	(0.10)
Constant	-0.43***	-0.66***
	(0.07)	(0.07)
Log likelihood	-10,190	-10,230
Bootstrap observations	7353	-
Number of replications	100	-

Note: $^*P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$. The simple count index involves counting the total number of the 3 electrical items a household has access to. For the latent access to information results, the standard errors in parenthesis are bootstrapped. As specified in equation (7), the translog production frontier included all variables, their squares and all interactions plus a time trend and the time trend squared. However, for both sets of results, the table reports significant estimates only. Full estimation results are available from the authors on request.

Table B1: Latent access to information index vs. simple count index

		Steps 2 and 3: Technical efficiency estimates		
Step 1: [a]			**Step 2:** [b]	**Step 3:** [c]
Production frontier		**Production frontier**		
ln (*Farm_size*)	0.24***	Constant	0.44***	0.66***
	(0.02)		(0.03)	(0.03)
ln (*Labour*)	0.23***			
	(0.03)	**Homoscedastic efficiency variance**		
ln (*Chemicals*)	0.13***	Constant	-1.17***	-0.49***
	(0.02)		(0.12)	(0.08)
Ox	0.24**			
	(0.11)	**Homoscedastic error variance**		
ln (*Farm_size*)^2	0.05***	Constant	-1.03***	-0.94***
	(0.01)		(0.04)	(0.05)
ln (*Farm_size*)*ln(*Labour*)	-0.04***			
	(0.01)			
ln (*Farm_size*)*ln(*Chemicals*)	0.00***			
	(0.00)			
ln (*Labour*)^2	0.02*			
	(0.01)			
ln (*Labour*)*ln(*Chemicals*)	-0.01***			
	(0.00)			
ln (*Chemicals*)^2	0.01***			
	(0.00)			
Time, *t*	0.06**			
	(0.03)			
Constant	-0.73***			
	(0.06)			
Sigma_u	0.66			
Sigma_e	0.81			
Rho	0.39			
Log likelihood	-10,313.80		-7,751.26	-8731.28
Bootstrap observations	-		7353	7353
Overall number of replications	-		100	100

Note: $^*P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$; Bootstrapped standard errors in parenthesis. As specified in equation (7), the translog production frontier in the main SF equation included all variables, their squares and all interactions plus a time trend and the time trend squared terms. However, the table reports significant results only. Full estimation results are available on request.

[a] Step 1: Estimation of random effects panel data model. Error component and the random unit effects are saved for use in subsequent steps.
[b] Step 2: Estimation of time varying technical efficiency, achieved by running stochastic frontier model on the predicted idiosyncratic error component from step 1. Production frontier is a contant.
[c] Step 3: Estimation of persistent technical efficiency; achieved by running stochastic frontier model on the predicted random effects from step 1. Production frontier is a constant.

Table B2: Production frontier, transient and persistent technical efficiency estimates in the SF model by Kumbhakar et al (2014).

Number of periods occurring	Number of households	% of households
1	372	4.71
2	832	10.54
3	2,691	34.08
4	1,644	20.82
5	1,295	16.40
6	1,062	13.45
Total	7,896	100

Table B3: Duration of households in the data