Impact of improved cassava varieties in Nigeria

Final Report

submitted to:

The CGIAR Research Program on Roots, Tubers and Bananas (RTB)

By

International Institute of Tropical Agriculture
Oyo Road, PMB 5320, Ibadan
November 2017

Tesfamicheal Wossen, Tahirou Abdoulaye, Arega Alene, Shiferaw Feleke, Ismail Rabbi, G. Asumugha, P. Kulakow and Victor Manyong.

Contents

List of Tables

List of Figures

Acknowledgement

Funding for this study came from the Consultative Group for International Agricultural Research (CGIAR) research program on Roots, Tubers and Bananas (RTB), The Bill and Melinda Gates Foundation (BMGF) and the Standing Panel on Impact Assessment (SPIA) of the CGIAR Independent Science and Partnership Council (ISPC). The contributions of Dr. Olanrewaju Adetunji, Dr. Mark Tokula, Abass Adebayo and Mr Henry Musa Kpaka of IITA and the all the participants to the Tanzania CMS workshop, to the conceptualization and design of this project is gratefully acknowledged. In addition, we wish to acknowledge Dr. Gezahegn Tessema for undertaking the DNA-fingerprinting process in this project. The contribution of the National Roots Crops Research Institute, Umudike, Nigeria and other local partners to the success of this project is also acknowledged. The enumerators, extension agents, and IITA staff from socioeconomics, bioscience, and cassava breeding units are recognized and thanked for their contributions. A special thanks to all the farmers and other people who generously gave their time for the interviews.

Executive summary

This report answers the following key research question: Does adoption of improved cassava varieties have any significant causal effects on productivity and poverty? The results of this project showed that about 60% of the farmers growing cassava have adopted improved varieties. However, when adoption was measured using DNA-fingerprinting approach, it was found that about 66% of the farmers have adopted improved cassava varieties. Despite higher adoption rates, the intensification rate of improved cassava varieties was found to be about 38%, which is quite modest. The productivity effect of adoption of improved cassava varieties was estimated using alternative measures of adoption (using self-reported adoption data from household surveys and DNA-fingerprinted adoption data) as well as specifications (OLS and IV estimation strategies). Using OLS estimation strategy, we found that the effect of adoption of improved cassava varieties on cassava yield is about 55%. Further, IV estimation results suggest a 64% productivity gain as a result of adoption of improved cassava varieties. Using a poverty line of $1.25 per person per day, adoption has led to a 4.7% and 4.02% poverty reduction in the closed economy and small open economy case, respectively. This poverty reduction role of adoption at $1.25 per person per day poverty line implies that 6.2%-7.15% of the rural poor cassava producers have escaped poverty in the current year due to adoption of improved cassava varieties. Similarly, at the poverty line of $1.9 per person per day poverty line, adoption has led to a 2.06% to 2.92% poverty reduction in a small open economy and closed economy, respectively. These changes correspond to a 2.9%-4% poverty reduction among rural poor cassava producers.

1. General Introduction

It is widely recognized that improved agricultural technologies play a critical role in agricultural transformation and economic growth in developing countries. Applied correctly, adoption of improved agricultural technologies should, ceteris paribus, increase overall productivity and provide additional income for farmers. In doing so, technology adoption can accelerate economic growth, create marketing opportunities and help millions of farmers to move out of poverty traps (Wossen *et al*., 2017). In this regard, the dissemination and diffusion of improved crop varieties has been cited as the primary pathway through which technological change in the agricultural sector can bring about productivity gains (Gollin et al., 2002). Understanding how and why households adopt improved varieties and their subsequent effects on poverty reduction and productivity gains is, therefore, important to disseminate technologies that are appropriate to the conditions of smallholder farmers.

The poverty impact of adoption of improved varieties can be direct or indirect. While the former impact operates at the household level, the latter impact works through regional, national or economy wide growth effects. In other words, the direct (micro-level) effects are realized via rising yields per cultivated area, lowering the risk of crop failure and generating year-round employment while the indirect (market-level) effects are materialize through ameliorating economic growth (Zeng et al., 2015). This report focuses on the adoption of improved cassava varieties in Nigeria—the largest cassava producer in the world. Cassava is the most widely cultivated root crop in terms of area allocation and the number of growers in Nigeria (Abdoulaye et al., 2013). The importance of cassava is increasing in recent years and is fast replacing yam and other traditional staple foods as a famine reserve and insurance crop against hunger (Wossen et al., 2017). The crop is important not only as a food but also as a major source of income for rural households. As a cash crop, cassava generates income for the largest number of households compared to other staples (Wossen et al., 2017), which justifies our focus on the crop. Improving agricultural productivity—in particular, cassava productivity-through efficient dissemination of improved varieties is therefore central for poverty reduction efforts in Nigeria. Cognizant of this fact, the International Institute of Tropical Agriculture (IITA) initiated cassava research in the early 1970s with a focus on developing varieties with resistance to major diseases such as cassava mosaic virus disease (CMD) and cassava bacterial blight (CBB). Consequently, IITA has developed and released more than 46 cassava varieties with multiple disease resistance and high yield potentials. In addition, IITA has developed good agronomic practices and biological control and integrated pest management options to reduce losses due to insect pests. Despite these major efforts made by IITA and partners to develop and disseminate a growing number of improved cassava varieties, there is still a lack of comprehensive and rigorous evidence of adoption and impacts of these varieties on poverty reduction. Without documenting adoption rates, it will therefore be very difficult to justify any investment for further development and dissemination efforts of improved cassava varieties.

Against this backdrop, this report entitled "impact of improved cassava varieties in Nigeria*"* answers the following policy relevant research questions in Nigeria.

i) What is the extent of adoption of individual improved varieties as well as improved varieties of cassava as a whole?
ii) What are the determinants of uptake and spread of improved varieties of cassava?
iii) Does adoption of improved cassava varieties have any significant causal effects on crop yields and poverty? If so, what are the aggregate impacts of adoption of improved cassava varieties on poverty reduction in Nigeria?

The rest of the report is organized as follows: The second section provides an overview the data collection process, the sampling strategy and some descriptive results from the household survey. The third section elaborates the process of DNA-based varietal identification. The fourth section presents the empirical econometric strategy employed for estimating the productivity and poverty reduction effects of adoption of improved cassava varieties. Section five then presents the main results of our analysis, focusing on the effect of adoption on productivity and poverty. The last section concludes with implications for policy and provides a list of open questions for further research.

1.1 Overview of the project and Household level results

1.1.1 Developing the sampling frame

The list of enumeration areas (EAs) for conducting the national census in Nigeria was obtained from the National Population Commission of Nigeria (NPCN). The EA list was obtained for the 16 states that contribute at least 80% of the total production of cassava in Nigeria. The states cut across four geopolitical regions (Table 1).

Table 1. Study regions and states in Nigeria.

s/n	Region	States
1	Southwest	Ogun
2	Southwest	Ondo
3	Southwest	Oyo
4	Southwest	Ekiti
5	Southwest	Osun
6	North	Kaduna
7	North	Nasarawa
8	North	Taraba
9	North	Benue
10	North	Kogi
11	Southeast	Enugu
12	Southeast	Imo
13	Southeast	Anambra
14	South-South	Cross River
15	South-South	Akwa Ibom
16	South-South	Delta

Relying on agricultural development programs (ADPs) in the targeted states, a prior visit was made to each of the selected EAs to develop the lists of all cassava growing households. This list provided a sampling frame for the selection of at least 50 cassava growing households in each EA, of which five household heads and two spouses were interviewed. This exercise facilitated the unbiased selection of samples for the final interviews.

1.1.2 Sample selection

In this project, a multistage stratified sampling design was employed to select the sample households. First, the list of Enumeration Areas (EAs) for conducting national census in Nigeria was obtained from the National Population Commission (NPC). The list of EAs by Local Government Areas (LGA) was obtained for the 17 states that together account for 80% of the total cassava production in Nigeria. These states were grouped into four geopolitical zones in a stratified sampling design (Table 1). From each region 100 EAs were selected using probability proportional to size (PPS) sampling approach. Finally, from each EA, random samples of 5 cassava growing households were selected for interview.

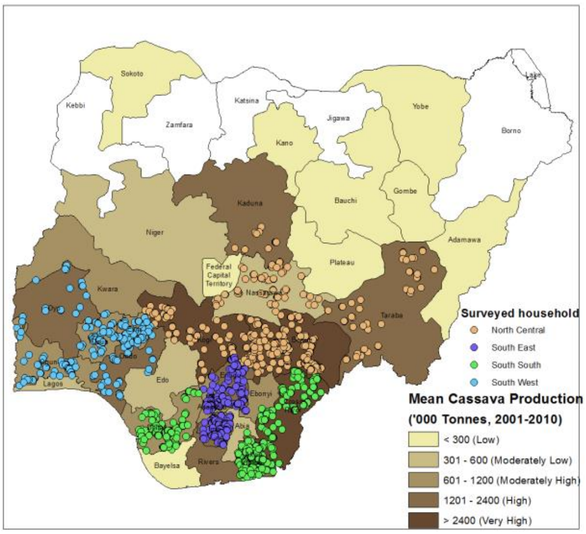

Figure 1: Study areas of CMS. The points on the map represent the distribution of the HHs

This gave a total of 625 households per region and a total of 2,500 farming households. For each surveyed household, information on self-reported treatment status (adoption of improved cassava varieties) was collected. This was done at the variety and plot level as many of the households own more than one plot of cassava and grow many different varieties within the same plot. In addition, from each identified variety in the farm plot, samples of cassava leaves were collected for DNA-fingerprinting analysis. In addition to treatment status, data on socio-economic/demographic characteristics of the households as well as other outcomes of interest such as production, expenditure on food and non-food items were collected. The DNA-fingerprinting process is one of the most novel aspects of this project. To date, several varietal identification methods for tracking adoption of improved varieties have been conducted. However, most of these methods have inherent uncertainty levels. Compared to other conventional varietal identification methods, the DNA-fingerprinting technique offers a reliable method to accurately identify varieties grown by farmers, thereby allowing credible measurement of adoption of improved varieties by farmers. Unlike phenotype-based methods, DNA-based varietal identification is independent of environmental conditions or plant growth stage. However, undertaking a credible DNA-based varietal identification is not trivial. It requires establishing a reference library and collecting samples from farm plots for DNA extraction and genotyping-by-sequencing (Rabbi et al., 2015). The detail procedure used for DNA-fingerprinting for this study is explained in section 3.

1.2 Household survey results

1.2.1 Adoption of improved cassava varieties

Fig. 2 presents adoption rates of improved cassava varieties at country and regional levels based on farmers self-reported adoption status. On average, about 60% of farmers have adopted improved cassava varieties. Adoption rates also show a large spatial heterogeneity. In particular, adoption rates reach as high as 79% in the Southwest region of the country while it is only about 31% in the Southeast region of the country.

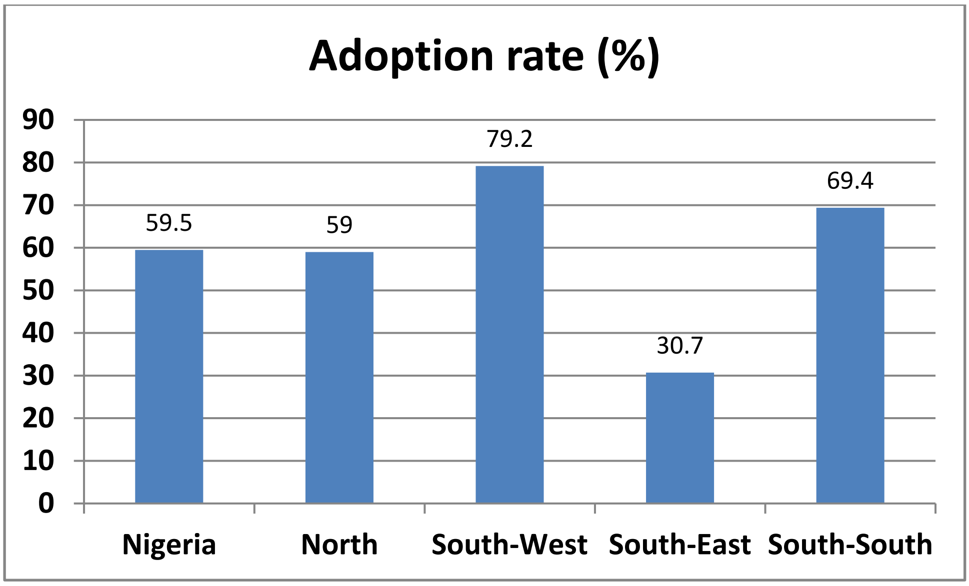

Figure 2: Adoption level based on farmers-self reported adoption status

In what follows, we also report the adoption rate based on intensity of adoption at the plot level (Fig. 3). Intensity of adoption is calculated by considering the area under improved cassava varieties out of the total cassava area. The result shows that despite high rates of adoption the intensity of adoption is very low. The current adoption rate, based on intensity of adoption, stands at 38% while using farmers' self-reported data. Regional distribution of adoption rates further reveals that the intensity of adoption rate is the highest in the South-West region. The lowest intensity rate is reported in the South-East. These results are not surprising as IITA is located and has been operating in the south-western part of the country for the last 50 years. The South-South region of the country has the second highest adoption rate. This might be due to the presence of national research centers in the region.

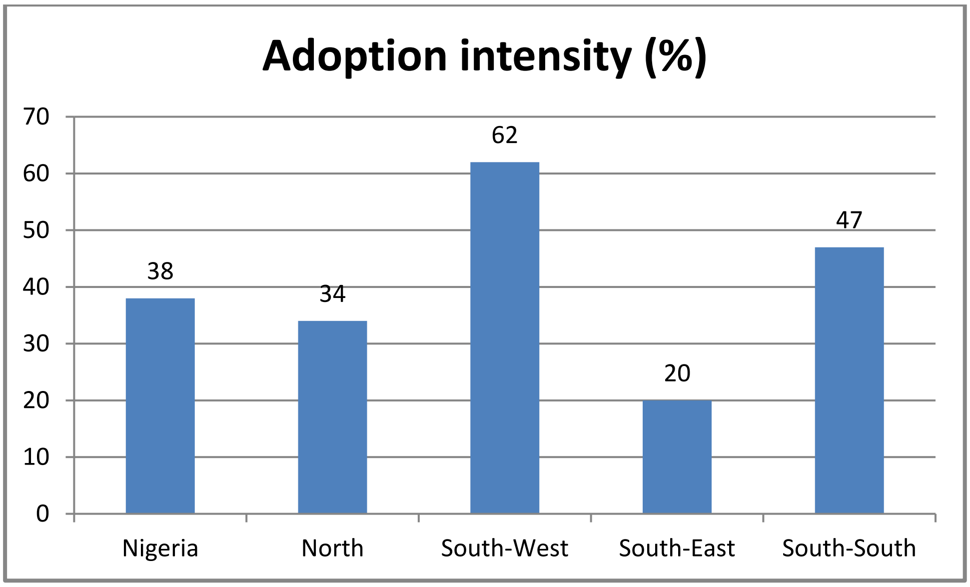

Figure 3: Adoption level based on farmers-self reported adoption status

1.2.2 Access to improved planting materials

Access to planting materials is an important issue for policy makers as it is the most important entry point for promoting improved cassava varieties. In this regard, this project collected data on access to planting materials from both formal and informal sources. What is really striking is that more than 70% of the farmers reported that their primary source of improved planting material is social networks (friends, relatives and neighbours). This shows that social norms, like the norm of reciprocity, play a prominent role in the distribution of planting material. This highlights the importance of social networks in contexts where farmers face limited access to credit and formal seed markets. Other important sources included extension and government sources (13%). Few farmers reported that they obtained planting material through nongovernmental organizations, processors, research institutes, the cassava market, and farmers' associations. Each of these sources accounted for less than 6% as a source of improved planting material. Disaggregated results over the different regions of the country further suggest the same trend where planting materials are being distributed by informal social networks. The contribution of local markets is rather insignificant. In addition, the role of private processor networks seems as insignificant as that of local markets in terms of the distribution of planting material.

Table 2. Source of planting material for improved cassava varieties.

	Full sample	North	Southwest	Southeast	South-South
Family/Friends/Relatives/Farmers/Neighbors	70.4	67.8	79.8	63.1	66
Extension/Government	12.6	13.2	8.1	14.8	16.0
Cassava market	5.7	6.3	2.5	12.8	5.2
Research institutes	4.6	5.1	4.8	3.0	4.7
NGO	3.7	3.5	2.5	4.4	4.7
Processors	2.4	3.3	2.1	1.5	2.3
Farmer associations	0.3	0.3	0.2	0.5	0.5
Others	0.3	0.5	0.0	0.0	0.7

In addition, the replenishment rate of planting material seems to be very low. Our survey results reveal that only 6% of adopters managed to replenish their planting material and about 94.1% have never done so (Table 3). This result is directly linked to our previous finding about the lack of access to planting material from formal sources. Developing a formal seed system is therefore crucial since access to planting material is a requisite for the adoption of productive and yield-enhancing varieties.

Table 3. Planting material replenishment rate

	Full sample	North	Southwest	Southeast	South-South
Never renewed (%)	94.08	94.81	94.54	88.4	95.3

1.2.3 Distribution of yield based on self-reported adoption status

Using GPS-based area measurement and self-reported adoption status as a bench mark, we calculated yield (output per unit of area). According to our data, average cassava yield stands at 14.7 t/ha. However, average cassava yield among adopters (16.1 t/ha) is significantly higher than for non-adopters (11.3 t/ha) and this difference is statistically significant at 1% significance level. However, this difference in cassava yield cannot simply be attributed to adoption by looking at mean differences between adopters and non-adopters. In particular, this observed yield difference between adopters and non-adopters is only an indicative of correlations and cannot be used to make causal inferences regarding the impacts of adoption on cassava yields without controlling for confounding factors. Further Figure 4 below shows the distribution of cassava yield for adopters and non-adopters. The left tail of the distribution suggests that a significant number of non-adopters have lower yield compared to adopters. Further the Kolmogorov–Smirnov equality-of-distributions test suggests that the two distributions are different (Equity of the two distributions is rejected at 1% as the p-value on Combined K-S is less than 0.01).

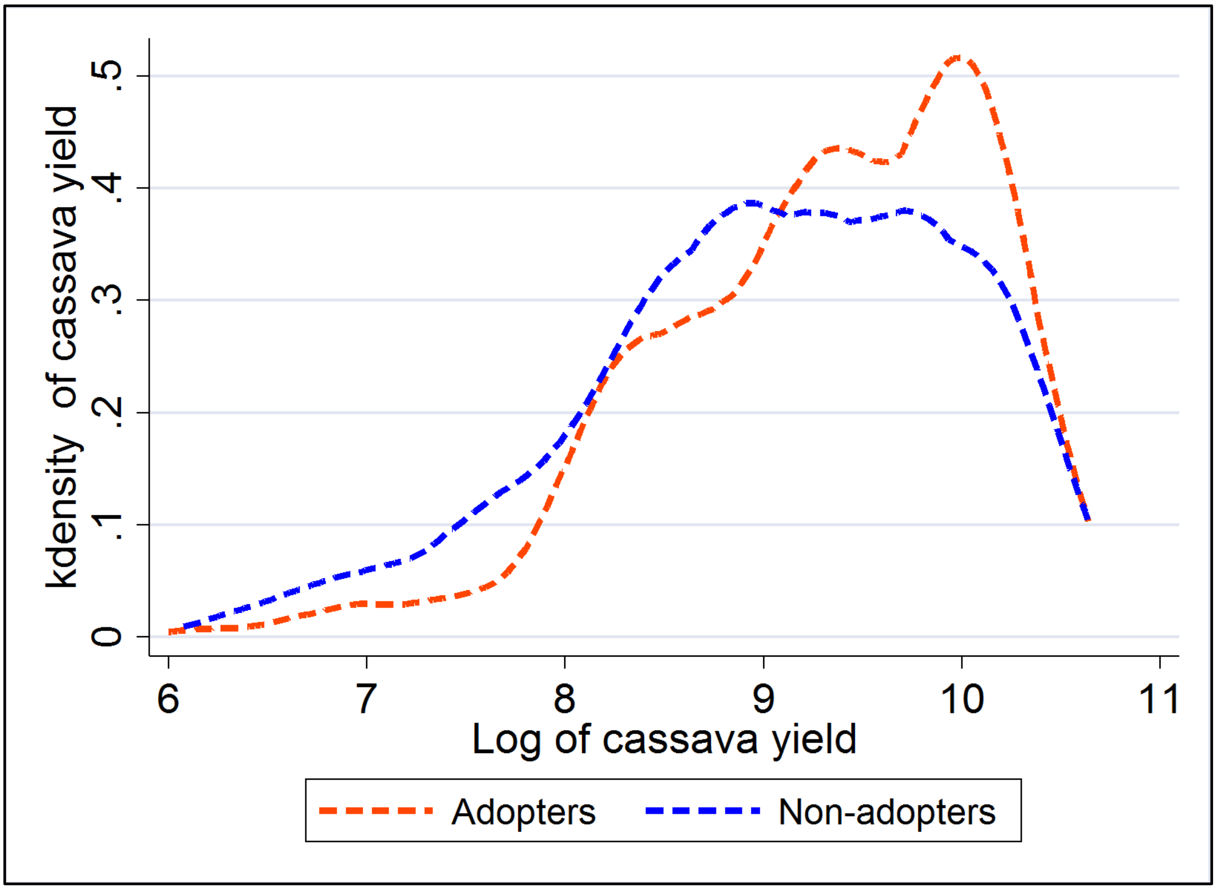

Figure 4: Distribution of cassava yield

2 The process of DNA-fingerprinting

For this project, DNA was extracted following the DNA extraction protocol (Dellaporta *et al*., 1983) from a total of 7376 genotypes collected from 2500 household's including 89 samples for quality control (clones genotyped in duplicate). In house modified protocol (Rabbi *et al*., 2014) that enables to extract up to 10 plates of 96 samples each per day was implemented. All the extracted DNA samples were quantified using spectrophotometer and agarose gel electrophoresis for quality and quantity assessments. Furthermore, test digestion with restriction enzyme was performed for 10% of the samples extracted as suggested by Genomic Diversity Facility (GDF) at Cornell University for standard Genotyping by Sequencing (GBS) library preparation. DNA samples with high concentration were diluted to 1000ng/μl. All extracted samples that pass the minimum quantity requirement (300ng/μl) were shipped to GDF for genotyping by sequencing (GBS). The ApeKI restriction enzyme (recognition site: G|CWCG) that produces less variable distributions of read depth was used for the GBS library preparation and therefore a larger number of scorable SNPs in cassava were used. Eighty 96-plex GBS libraries were constructed following the standard procedure (Elshire et al. 2011) and sequenced at the GDF using the Illumina HiSeq2500.

The raw read sequences obtained in the current study including accessions in the reference library (Rabbi et al 2015) and duplicate of 89 samples for quality control were processed through a TASSEL-GBS discovery pipeline developed using TASSEL 5.0 (Glaubitz et al. 2014). SNP calling was performed based on TASSEL-GBS production pipeline by aligning the tags to the most recent cassava reference genome version 6.0. The 89 randomly selected and genotyped in duplicates were used to determine a distance threshold between genotypes that can help to declare a distance at which two or a set of genotypes are similar or distinct. A frequency distribution of distance (IBS) was plotted and resulted in bimodal distribution of pairwise genetic distance. The bimodal distribution shows the frequency distribution of the data where one of the curves shows artefact that could occur due to genotyping error. The point between the bimodal distributions was therefore declared as a distance threshold where any pair of genotypes or set of genotypes below the point are identical.

Once the distance threshold is determined, the distance-based hierarchical clustering, a pairwise genetic distance (identity-by-state, IBS) matrix were computed for all the genotypes including the intentional duplicates and accessions from the reference library (Rabbi et al, 2005). A Ward's minimum variance hierarchical cluster dendrogram were built from the IBS matrix. The critical distance threshold determined was applied for the whole data and individuals belonging to the same cluster group below the threshold were considered as the same genotypes, i.e. if any of the genotypes from the reference library fall in the cluster of different individuals representing the same variety then it will be identified based on the variety from the reference (Details about the whole DNA-finger printing process is documented in Wossen et al., 2017)

2.1 Key results of the DNA-finger printing analysis

Of the total 7376 farmers' genotypes collected in the current study only 4822 matched genotypes in the reference library, whereas 2554 did not match any of the genotypes there. On the other hand, 1663 of the 3891 genotypes in the reference library did not match any of the varieties collected from the farmer's field. Altogether a total of 114 different varieties were identified. Among these, 46 varieties matched the genetic gain cluster group (improved variety group) whereas 68 matched landrace groups. In our analysis only 18 matched officially released varieties, and 15 matched varieties that are improved and released. Among the officially released varieties only 14 were from the genetic gain group whereas the remaining 4 cultivars represented landrace collections evaluated on experimental plots and officially released.

In addition to the level of genetic matching of the farmers' varieties to the genotypes in the reference library it is also important to have variety release information. This will help to have clear figures on further analysis on the adoption of improved varieties based on DNA fingerprinting. In the current study, more than 35% of the total number of samples collected on the farmers' fields represented "improved varieties" which are categorized in three different groups: improved and officially released, improved but not officially released, and released but matching those accessions from the landrace. In addition, about 33% of the varieties matched the genetic gain (GG) cluster group. However, only 12.50% matched improved and officially released varieties. In addition, 12.56% of the cultivars from farmers' fields matched those landrace collections that went through field evaluation and official released. A significant percentage (10.28%) of the farmers' varieties also represent improved varieties in the GG group that were not officially released but made their way to farmers' fields (Table 4).

Table 4. Percentage of improved and/or released varieties on the farmer's field based on DNA fingerprinting.

Varieties	**GG**	**LR**	**Total**	**%**	**Cumulative %**
Improved AND released	919	3	922	12.50	12.50
Improved NOT released	267	491	758	10.28	22.78
Single accessions in GG cluster	110		110	1.49	24.27
Not in Library, in GG cluster	598		598	8.11	32.38
Matching LR, in GG cluster	55		55	0.75	33.12
Released but TMEB1, TMEB2.		926	926	12.55	45.68
Local varieties		4007	4007	54.32	54.32
Total	1949	5427	7376	100.00	

GG =genetic gain; LR = landrace;

Singular varieties not matching any of the genotypes but found in the GG group represent 1.49% of the collection. Similarly, about 8.11% represent varieties matching the GG cluster group but none were in the reference library. The larger proportion (54.32%) represents local varieties.

2.2 Farmers' variety matching released varieties

IITA, in collaboration with national agricultural research systems (NARS), has officially released 46 cultivars. However, in the current study conducted in the main cassava growing regions of Nigeria a large number (28) of the 46 cultivars were not observed on farmers' fields, suggesting that these were either dis-adopted or no initial dissemination effort had been made. Among the 28 varieties not found on farmers' fields, 11 were not in the reference library, whereas none of the varieties matched the remaining 17 which are in the library. Of the officially released varieties only 18 were encountered on farmers' fields and five were the dominant varieties growing in the study regions with a frequency greater than 100. These are TMS30572, MS-3 (Odongbo), MS-6 (Antiota), TMS 50395, and TME 419 in descending order of their frequencies on farmers' fields.

2.3 Combining self-reported and DNA-finger printed adoption data

After identifying varieties using DNA-finger printing, we then matched farmers own self-reported and DNA-based varietal identification to identify the extent of adoption of improved cassava varieties. As mentioned before, adoption rate stands at 60% while using self-reported adoption status from the household survey. Herein, we calculated adoption rate of improved cassava varieties using our DNA-finger printed adoption data. However, classifying varieties into improved and landraces is not straightforward even after DNA fingerprinting due to issues of measurement and library matching. We therefore develop two scenarios based on commonly used criteria for the identification of varieties into improved and non-improved. Table 5 reports the varietal groups as well as definitions for each group as obtained from DNA fingerprinted results.

Table 5. categories used for defining improved varieties based on DNA fingerprinting results.

Serial no.	Variety groups	Definition
1	Improved and Released	This category includes varieties that match improved and released varieties in the reference library.
2	Improved not Released	This group of varieties matches improved varieties but are not in the officially released list.
3	Single accessions in GG cluster	These are accessions that genetically do not match any specific clone in the collection but cluster with improved varieties in the reference library. This can happen when cross-pollinated (or even self-pollinated) seeds germinate in farmers' fields and the farmers eventually propagate these as a variety.
4	Not in Library, in GG cluster	No library can completely encompass every possible genotype found in household farms. As this is a clonal crop, the breeding program cannot maintain every genotype that it produced or even disseminated because of natural attrition over time. So where we found clones that did not match any specific genotype in the library, the next best thing was to use the cluster analysis method and find where they belong. In this case, these accessions clustered with the improved variety collection we have in the library.
5	Matching LR, in GG cluster	The "landraces" (LR) are accessions whose name start with prefix "TMEB" and were gathered during germplasm collection expeditions in Nigeria and other African countries over the years. Many of these are obviously landraces but some, in reality, are improved varieties which had lost their original names, were collected again, and brought back to IITA. Because of their unknown identities, these accessions are usually placed in the landrace collection, even though some are improved.
6	Released but TMEB1, TMEB2	Among the officially released varieties are some genotypes that were not developed through formal breeding processes (i.e., not from breeder crosses). After several years of purification and testing, these landraces were found to have superior characteristics and were recommended for release. In some instances these genotypes were transferred from one country/region to another which would not have occurred without the intervention of formal breeding programs like IITA. A good case is TMEB419 which was brought to IITA cassava breeder brought from Togo and is now grown in Nigeria and other countries. It is an officially released variety.
7	Local varieties	These are landrace varieties

When DNA-based varietal identification is used, the adoption rate is 66.

3 Methodology

Impact evaluation is crucial for attributing observed impacts (changes in poverty, income, productivity, etc) to an intervention, in our case to adoption of improved cassava varieties. However, attributing observed impacts to an intervention is not trial in the absence of random assignment of adoption status. In the context of randomized control trial (RCT) impacts can easily be attributed to an intervention (adoption) as households are assigned to control and treatment groups randomly. As such, adopters and non-adopters will be similar in both observed and unobserved characteristics except that adopters have received the intervention (in this case improved cassava varieties) and non-adopters did not. Therefore, we can be certain that the observed impacts are indeed the result of adopting improved cassava varieties, and not some other mediating factors. However, identification of the causal effect of adoption in observational (non-experimental) cases is not trivial due to self-selection/endogeneity bias. As such, accurate measurement of impacts requires controlling for both observable and unobservable characteristics between adopters and non-adopters. In other words, identifying the counterfactual (what would have happened to adopters had they not adopted improved cassava varieties) would be crucial. However, constructing a reliable counter factual is challenging since one cannot observe the outcome of adopters had they not been an adopter. As such, the issue of counterfactual becomes effectively a missing data. The best way of tackling this missing data problem is to identify a group of non-adopters (controls) who mimics the behaviour of adopters with one key difference: the control households differ from adopters only in adoption status. Controlling for all observed and unobserved differences between adopters and non-adopters is important as such factors may affect observed impacts of adoption. In the context of non-experimental data, a wide variety approaches have been utilized for constructing counterfactual groups, the most common approaches being matching techniques, difference in difference (especially with fixed effects in panel data) and instrumental variable (IV) regression approaches. In this section, we present the main methodological approaches employed in this report for establishing the causal impact of adoption of improved cassava varieties on productivity, poverty and food security. The first sub-section- where we addressed household level treatment effects- focuses on matching and instrumental variable regression approaches. The implication of misclassifying adoption status is also discussed in this section. In the second sub-section, we then present the methodological approach used for estimating the aggregate (market-level) effects of adoption on poverty.

3.1 Approaches for estimating household level treatment effects

3.1.1 Matching approaches

The most common matching techniques in the impact evaluation literature are the propensity score matching (PSM) and inverse probability weighted adjusted regression (IPWRA) approaches. The basic idea behind PSM is to match each adopter with a similar non-adopter and then measure the average difference in the outcome variable between adopters and non-adopters. In other words, we are interested in the question, "*How would the outcome of adopters (in terms of productivity, income, poverty etc) have changed had adopters chosen not to adopt improved cassava varieties*?" In this case, the average treatment effect on the treated (ATT) is defined as:

$$ATT = E[Y(1) - Y(0)|T = 1] \qquad 1.$$

Where $Y(1)$ and $Y(0)$ are outcome indicators (in our case, productivity and welfare level of households with and without adoption respectively). T is a treatment indicator that takes a value of 1 if a household is an adopter and 0 otherwise. However, we can only observe $E[Y(1)|T = 1]$ in our data set and $E[Y(0)|T = 1]$ is missing. In essence, we cannot observe the productivity and welfare level of adopters had they been non-adopters, after they become adopters. Simple comparison of productivity and welfare level of adopters and non-adopters introduces bias in estimated impacts due to self-selection. The magnitude of self-selection bias is formally presented as:

$$E[Y(1) - Y(0)|T = 1] = \text{ATT} + \underbrace{E[Y(0)|T = 1 - Y(0)|T = 0]}_{\text{Self-selection bias}} \qquad 2.$$

The right hand side term after ATT represents the magnitude of the selection bias. By creating comparable counterfactual households for adopters, PSM reduces the part of the bias due to observables. Once households are matched with observables, PSM assumes that there are no systematic differences in unobservable characteristics between adopters and non-adopters. Given this conditional independence assumption and the overlap conditions, ATT is then computed as follows:

$$ATT = E[Y(1)|T = 1, p(x)] - E[Y(0)|T = 0, p(x)] \qquad 3.$$

The above equation states that ATT is a propensity score weighted mean difference between adopters and non-adopters over the common support area. However, ATT from PSM can still produce biased results in the presence of misspecification in the propensity score model (Robins *et al*., 2007; Wooldridge, 2007; Wooldridge, 2010). A potential remedy for such misspecification bias is to use IPWRA. According to Wooldridge (2010), ATT will be consistent despite misspecification of either the treatment or the outcome model, but not both. As a result, the IPWRA estimator has the double-robust property that ensures consistent results as it allows the outcome and the treatment model to account for misspecification. Following Imbens *et al*., (2009), ATT in the IPWRA model is estimated in two steps. Suppose that the outcome model is represented by a linear regression function of the form $Y_i = \alpha_i + \varphi_i x_i + \varepsilon_i$ for $i = [0\ 1]$ and the propensity scores are given by $p(x; \gamma)$. In the first step, we estimate the propensity scores as $p(x; \hat{\gamma})$. In the second step, we then employ linear regression to estimate (α_0, φ_0) and (α_1, φ_1) using inverse probability weighted least squares as

$$\min_{\alpha_0,\varphi_0} \sum_{i}^{N} (Y_i - \alpha_0 - \varphi_0\, x_i)/p(x,\hat{\gamma}) \;\; if\; T_i = 0 \qquad 4.$$

$$\min_{\alpha_1,\varphi_1} \sum_{i}^{N} (Y_i - \alpha_1 - \varphi_1\, x_i)/p(x,\hat{\gamma}) \;\; if\;\; T_i = 1 \qquad 5.$$

The ATT is then computed as the difference between Eq. (4) and Eq. (5).

$$ATT = \frac{1}{\mathrm{N_w}} \sum_{i}^{N_w} [(\hat{\alpha}_1 - \hat{\alpha}_0) - (\hat{\varphi}_1 - \hat{\varphi}_0)\, x_i\,] \qquad 6.$$

where, ($\hat{\alpha}_1$, $\hat{\varphi}_1$) are estimated inverse probability weighted parameters for adopters while ($\hat{\alpha}_0$, $\hat{\varphi}_0$) are estimated inverse probability weighted parameters for non-adopters. Finally, $\mathrm{N_w}$ stands for the total number of adopters.

Yet, matching techniques—regardless of adjustments for misspecification bias— can only overcome the selection bias that arises from observable characteristics. When the cause of selection bias is unobservable heterogeneity, such as farmer's inherent skill, results based on matching techniques will be biased. As such, proper causal identification requires controlling for both observable and unobservable factors that influence adoption decision. Hence, estimates of both PSM and IPWRA can yield biased estimates due to biases stemming from unobservable factors that affect adoption decision and the outcome indicators simultaneously. A method that takes into account both observed and unobserved sources of heterogeneity between adopters and non-adopters is an IV regression approach. However, finding an instrument that satisfies the orthogonality condition, a variable that is strongly correlated with adoption decision but that does not directly affect outcome indicators (productivity and welfare etc), is a challenge.

3.1.2 Instrumental variable regression approach

Instrumental variable (IV) methods are widely used to identify causal effects in the presence of endogeneity problems in key variables of interest. Herein, we presented the IV regression approach employed in this report for estimating the causal effect of adoption (the endogenous treatment variable) on productivity, poverty and food security. Following Zeng et al., (2015) and Suri (2011), we assume that a particular farm household adopts improved cassava varieties based on expected benefits. Assume that the net befit a given farmer derives from adoption and non-adoption of improved cassava varieties is given by π^a and π^n respectively. Adoption implies that the utility of expected net-return from adoption is higher than from non-adoption:

$$\{\boldsymbol{E}[\boldsymbol{u}(\boldsymbol{\pi}^{\boldsymbol{a}})] > E[\boldsymbol{u}(\boldsymbol{\pi}^{\boldsymbol{n}})]\} \quad \textbf{7.}$$

The net return from adoption depends on the structure of the return and the cost. On the revenue side we assume that the final output price of improved and traditional cassava varieties will be the same. However, productivity is expected to be higher with adoption. Therefore, on the revenue side, adoption decision depends on expected yield. On the cost side, adoption and non-adoption entail different cost structures in terms of labor, cash and information. For example, adoption may require more labor, cash, information and knowledge and some fixed costs of acquiring seeds and planting material. Hence, we assume that adoption of improved cassava varieties is costly. The return from adoption is then specified as follows:

$$\boldsymbol{E}[(\boldsymbol{\pi}^{\boldsymbol{a}})] = \boldsymbol{E}[(\boldsymbol{P}\boldsymbol{Y}^{\boldsymbol{a}})] - [\boldsymbol{C}^{\boldsymbol{a}}] \quad \textbf{8.}$$

Where, P is price of cassava, Y^a, yield with adoption, C^a includes all production costs incurred with adoption (this includes for example, transport cost, acquiring knowledge and information about new cassava varieties, costs of fertilizer and pesticides etc). The above equation shows that costs are incurred ex-ante based on expected revenue. Therefore, the ex-ante variable and fixed costs play a significance role in the decision to adopt improved cassava varieties. Similarly return without adoption is given as

$$\boldsymbol{E}[(\boldsymbol{\pi}^{\boldsymbol{n}})] = \boldsymbol{E}[(\boldsymbol{P}\boldsymbol{Y}^{\boldsymbol{n}})] - \boldsymbol{C}^{\boldsymbol{n}} \quad \textbf{9.}$$

Where Y^n & C^n are the expected yield and costs without adoption. Given that the farmer adopts improved cassava varieties when the utility of expected benefit from adoption is higher than the utility of expected returns without adoption, adoption decision implies,

$$\boldsymbol{E}[\boldsymbol{u}(\boldsymbol{\pi}^{\boldsymbol{a}})] - \boldsymbol{E}[\boldsymbol{u}(\boldsymbol{\pi}^{\boldsymbol{n}})] > \mathbf{0} \qquad \mathbf{10.}$$

With some algebraic manipulation and taking price as a *numeraire*, it can be shown that adoption decision depends on yield and cost differences between improved and traditional cassava varieties.

$$\boldsymbol{E}[(\boldsymbol{Y}^{\boldsymbol{a}} - \boldsymbol{Y}^{\boldsymbol{n}})] > (\boldsymbol{C}^{\boldsymbol{a}} - \boldsymbol{C}^{\boldsymbol{n}}) \qquad \mathbf{11.}$$

Equation 11 above implies that improved cassava varieties will be adopted if the yield gain from adoption is higher than the cost of adoption. As such by capturing yield differences between adopters and non-adopters through a production function and cost differences through a cost function, one can then capture the benefits from adoption. Following Zeng et al., (2015) and Suri (2011), the production function for cassava production can be specified in the following way:

$$\begin{aligned} \mathbf{Y^a} &= \left(\boldsymbol{\alpha}_{\mathbf{a}} + \boldsymbol{\vartheta} + \boldsymbol{\beta}_{\mathbf{a}}\mathbf{X} + \boldsymbol{\mu}_{\mathbf{ay}}\right) \\ \mathbf{Y^n} &= \left(\boldsymbol{\alpha}_{\mathbf{n}} + \boldsymbol{\beta}_{\mathbf{n}}\mathbf{X} + \boldsymbol{\mu}_{\mathbf{ny}}\right) \end{aligned} \qquad \mathbf{12.}$$

where ϑ is the plot-specific percentage yield gain with adoption; X is the input vector with coefficients β and μ_y is the idiosyncratic error term. In the potential outcome framework proposed by Rubin (1974), the above production function can further be expressed as:

$$\mathbf{Y} = \mathbf{T}\mathbf{Y^a} + (\mathbf{1} - \mathbf{T})\mathbf{Y^n} \qquad \mathbf{13.}$$

Where T is a treatment status which takes a value of one if a given farmer is an adopter of improved cassava varieties and zero otherwise. Given Eq. (12 &13) above, the production function can then be expressed as follows:

$$\boldsymbol{Y} = \boldsymbol{\alpha}_{\mathbf{a}} + \mathbf{T}(\boldsymbol{\alpha}_{\mathbf{a}} - \boldsymbol{\alpha}_{\mathbf{n}}) + \mathbf{T}\boldsymbol{\vartheta} + \boldsymbol{\beta}_{\mathbf{n}}\boldsymbol{X} + \mathbf{TX}(\boldsymbol{\beta}_{\mathbf{a}} - \boldsymbol{\beta}_{\mathbf{n}}) + \boldsymbol{\mu} \qquad \mathbf{14.}$$

Parameter estimates of the above production function (ϑ) measures the yield advantage of improved cassava varieties over traditional cassava varieties. Similarly, change in the distribution of costs as a result of adoption can also be estimated using a cost function in the following manner:

$$\begin{aligned} \mathbf{C^a} &= (\boldsymbol{\delta}_\mathbf{a} + \boldsymbol{\sigma} + \mathbf{P}\boldsymbol{\gamma}_\mathbf{a} + \boldsymbol{\mu}_\mathbf{ac}) \\ \mathbf{C^n} &= (\boldsymbol{\delta}_\mathbf{n} + \mathbf{P}\boldsymbol{\gamma}_\mathbf{n} + \boldsymbol{\mu}_\mathbf{nc}) \end{aligned} \qquad \mathbf{15.}$$

Where P is a vector that includes input prices and other key socio-economic indicators. μ_c denotes the idiosyncratic error term. Following the same potential framework approach, the treatment effect (effect of adoption on cost changes) can then presented as follows:

$$\boldsymbol{C} = \boldsymbol{\delta}_\mathbf{a} + \mathbf{T}(\boldsymbol{\delta}_\mathbf{a} - \boldsymbol{\delta}_\mathbf{n}) + \mathbf{T}\boldsymbol{\sigma} + \boldsymbol{\gamma}_\mathbf{n}\boldsymbol{P} + \mathbf{TP}(\boldsymbol{\gamma}_\mathbf{a} - \boldsymbol{\gamma}_\mathbf{n}) + \boldsymbol{\mu} \qquad \mathbf{16.}$$

The parameter σ is interpreted as the plot-specific treatment effect in terms of percentage cost increase due to adoption. In both the yield and cost treatment effect models, the treatment (the decision to adopt improved cassava varieties) is endogenous (as farmers self-select into adoption). In fact, there are several reasons for the adoption decision to be endogenous. First, governments may target households that are more/less productive. Hence, it is likely that adoption decision is correlated with initial productivity levels, poverty status, household income, or underlying features that influence these outcome variables. Second, there is a possibility that adopters share common intrinsic characteristics, such as poor/better farming skills and management abilities, which are likely to be related to poverty status and productivity levels. As such, causal identification of adoption impacts requires an instrument that satisfies the orthogonality condition (a variable that is strongly correlated with adoption decision but that does not directly affect productivity and welfare outcome indicators). In our case, the IV approach relies on a two-stage estimation strategy. In the first stage, a probit/logit model is used to predict the probability of adoption. In the second stage, predicted probabilities from the first stage are used as instruments in the outcome equation. This procedure is very efficient and it is preferred to other IV methods when the endogenous variable is binary as it explicitly considers the binary nature of the endogenous variable (Wooldridge, 2007).

3.2 Capturing market level effects of adoption

To measure overall welfare effects of technology adoption, indirect effects of adoption need to be accounted for. To capture indirect effects, we first need to look into the different pathways through which adoption may affect welfare of adopters and non-adopters. Generally, there are three pathways through which an exogenous change in agricultural productivity (such as adoption) may affect the distribution of outcomes such as productivity, income and poverty. These include:

i. Effects through output price changes: If adoption increases productivity, it affects local supply and hence reduces local prices. However, such changes in food price benefits only net-food buyers

ii. Effects through farm profits: If outputs expand faster than price fall, then adoption increases the income level of net-food sellers

iii. Effects through rural wage—general equilibrium effect

For now, we focus on the first two cases, ignoring the effect of wage adjustments. Estimating the aggregate effect of adoption in the above two cases requires the following steps:

i. Estimating treatment effects in terms of yield and cost changes due to adoption

ii. Estimating income effects (producer and consumer surplus changes) based on yield and cost treatment effects and allocate the resulting income changes (producer and consumer surplus) to households

iii. Estimating the counterfactual distribution based on changes in producer and consumer surplus

From the above discussion two points are apparently crucial: Accurate estimation of the treatment effects and allocation of effects (producer and consumer surplus) to appropriate farm households. Estimation of treatment effects was discussed in the previous section. In this section, we focus on how the allocation of changes into appropriate households is done. Allocation of adoption induced income changes to farm households largely depends on the nature of market the farm-households face (open/closed economy) and the market position of a given farmer (net buyer/seller) of the product under consideration. Following Zeng et al., (2015), we considered two scenarios for allocating adoption induced income changes to appropriate households: The small open economy and closed economy case. In the small open economy case, the price that prevails at the local market would be the same as the prevailing world market price for cassava. Therefore, any productivity shock (increased in production of cassava due to adoption) would not affect the price that the consumers and producers face in the local market. In this case, welfare effects will only be accrued to producers (simply due to productivity gains). In the closed economy case, local supply shocks will necessarily affect the local market price. As such, the price of cassava would undoubtedly decline as a result of adoption due to supply shifts; leading to potential benefits (loses to both producers and consumers (Zeng et al., 2015). These benefits from adoption can however be different for consumers and producers (adopters). For producers (net-sellers), the effects can only be positive if the per-unit production cost reduction is larger than the price fall. However, consumers (net-buyers) will always benefit due to lower prices (higher purchasing power).

3.2.1 Estimation of aggregate effects in a closed-economy

Estimating the poverty impacts of adoption in a closed-economy requires understanding:

i. Treatment effects (effects on productivity and costs) to determine the per-unit cost reduction as a result of adoption
ii. Demand and supply elasticities

In the case of closed economy, estimating a composite demand function which takes into account demand structures facing cassava produces is crucial to determine the aggregate economic benefits. One approach that captures aggregate benefits of adoption through aggregation of farm-level effects is the economic surplus model (ESM, hereafter). The ESM captures adoption induced supply responses (yield gain, per unit cost reductions etc) through a simple shift in the supply function that faces producers (Alston et al., 1995). Assuming a downward-sloping demand curve, such a shift reduces the price received by cassava producers/price paid by consumers. Fig. 5 below shows the ESM in a closed economy. Point **b** shows observed cassava production at price level of **(P^{obs})**. Point **b** reflects observed output level after adoption of improved cassava varieties (i.e, the supply curve shifts from **S^*** to **S^{obs}** as a results of adoption). Assuming a downward-sloping demand curve (**D**), a supply shift from **S^*** to **S^{obs}** leads to a decline in price from **P^{ct}** to **P^{obs}** and the corresponding price changes from **Q^{ct} to Q^{obs}**. However, counterfactual price and quantity (**P^{ct}** and **Q^{ct})** cannot be observed and can be calculated algebraically based on **P^{obs}, Q^{obs},** magnitude of the supply shift and the size of supply (**ε**) and demand elasticity (**η**) estimates.

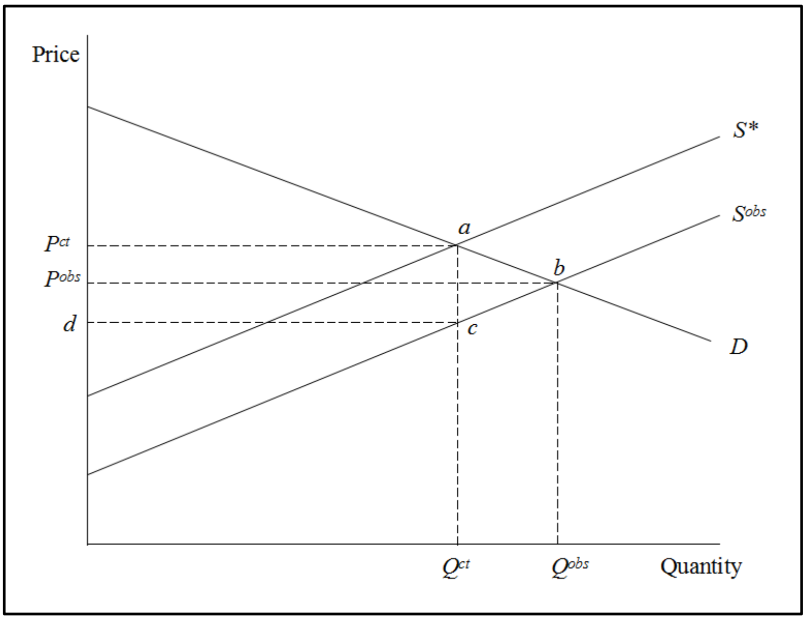

Figure 5: Ex-post economic surplus changes in a closed economy.

In the above figure, yield and cost treatment effects are used to derive the cost reduction per unit of output due to adoption (the *k*-shift). In particular, following Alston et al. (1995) the k-shift parameter can be calculated as follows:

$$\mathbf{k} = \left(\frac{\boldsymbol{\vartheta}}{\boldsymbol{\varepsilon}} - \frac{\boldsymbol{\sigma}}{\mathbf{1} + \boldsymbol{\vartheta}}\right) \times \text{adoption rate} \qquad \textbf{17.}$$

Where $\boldsymbol{\vartheta}$ is the effect of adoption on cassava yield (Average treatment effect of adoption on yield, **Eq. 14**) and $\boldsymbol{\sigma}$ is the cost increase due to adoption (Average treatment effect of adoption on cost of production for cassava, **Eq. 16**). Using the estimated *k*-shift parameter, observed prices ($\boldsymbol{P^{obs}}$), supply elasticity ($\boldsymbol{\varepsilon}$) and demand elasticity ($\boldsymbol{\eta}$), the counterfactual price level ($\boldsymbol{P^{ct}}$) that would have existed without adoption of improved cassava varieties is calculated as follows (c.f. Zeng et al., 2015)

$$\boldsymbol{P^{ct}} = \boldsymbol{P^{obs}}\left(\frac{\boldsymbol{\varepsilon} + \boldsymbol{\eta}}{\boldsymbol{\varepsilon} + \boldsymbol{\eta} - \mathbf{k}\boldsymbol{\varepsilon}}\right) \qquad \textbf{18.}$$

Output level of cassava in the absence of adoption (counterfactual production, $\mathbf{Q^{ct}}$) is calculated by subtracting aggregate production gains from adoption (i.e. treatment effects on yield aggregated over the adopted area) from observed cassava production (**$\mathbf{Q^{obs}}$).** Following Alston et al. (1995), the changes in producer and consumer surplus as a result of adoption of improved cassava varieties are calculated as follows:

$$\Delta\mathbf{PS} = \boldsymbol{P^{ct}Q^{ct}}(\boldsymbol{k} - \boldsymbol{Z})(\mathbf{1} + \mathbf{0.5Z}\boldsymbol{\eta}) \qquad \textbf{19.}$$

$$\Delta\mathbf{CS} = \boldsymbol{P^{ct}Q^{ct}Z}(\mathbf{1} + \mathbf{0.5Z}\boldsymbol{\eta}) \qquad \textbf{20.}$$

Where $\boldsymbol{Z}$ equals the proportional reduction of market price, **($\mathbf{P^{ct}}$ – $\mathbf{P^{obs}}$)/ $\mathbf{P^{ct}}$.** After calculating producer and consumer surplus changes, the next step involves allocating these surplus changes to appropriate farm household to establish household level effects of adoption. The allocation of surplus changes depends on the market position of farmers: net-seller/net buyers of cassava products. For net-buyers (consumers), consumer surplus gains due to a lower out price will be allocated using their cassava purchased quantities from total purchase as a weight. Similarly, for net-sellers (producers), producer surplus is allocated to households based on their adoption status using their sales quantities from total production as a weight. Following the approach of Zeng et al. (2015), we decompose the aggregate producer surplus gains into adoption and price effects ($\mathbf{\Delta PS} = \mathbf{\Delta PS}_{\boldsymbol{adoption}} + \mathbf{\Delta PS}_{\boldsymbol{price}}$). Producer surplus changes as a result of price effect can be calculated as follows:

$$\mathbf{\Delta PS}_{\boldsymbol{price}} = \frac{\boldsymbol{k\varepsilon P^{obs}Q^{ct}}}{\boldsymbol{\varepsilon} + \boldsymbol{\eta} - \mathbf{k\varepsilon}}\left(\frac{\boldsymbol{k\varepsilon P^{obs}}}{\mathbf{2}\boldsymbol{P^{ct}}(\boldsymbol{\varepsilon} + \boldsymbol{\eta} - \mathbf{k\varepsilon})} - \mathbf{1}\right) \qquad \mathbf{21.}$$

Then it follows that:

$$\mathbf{\Delta PS}_{\boldsymbol{adoption}} = \mathbf{\Delta PS} - \mathbf{\Delta PS}_{\boldsymbol{price}} \qquad \mathbf{22.}$$

3.2.2 Estimation of aggregate effects in a small open-economy

In the case of a small open economy, market prices don't change as a result of domestic supply shocks (changes in cassava productivity as a result of adoption of improved cassava varieties) as producers and consumers face the world market price for cassava. The idea is that, in the presence of trade, any domestic shock (as far as the country is small, in terms of trade volume), don't affect market supply and hence price. In this context, the impact of adoption on poverty can easily be computed using yield and cost treatment effect estimates (by calculating the distribution of observed and counterfactual income).

$$\mathbf{\Delta}\boldsymbol{I_{ij}} = \mathbf{P}(\boldsymbol{Y_{ij}^{obs}} - \boldsymbol{Y_{ij}^{ct}}) - (\boldsymbol{C_{ij}^{obs}} - \boldsymbol{C_{ij}^{ct}}) \qquad \mathbf{23.}$$

Where **P** is the unchanging cassava of price; $(Y_{ij}^{obs}, Y_{ij}^{ct})$ as well as $(C_{ij}^{obs} - C_{ij}^{ct})$ are observed and counterfactual yield and cost pairs of plot $\boldsymbol{j}$ for household $\boldsymbol{i}$. In both small open and closed economy case, poverty is calculated using the Foster, Greer, Thorbecke (1984), poverty indices. The poverty effect of adoption of improved cassava varieties is calculated by comparing counterfactual and observed poverty rates (based on counterfactual and observed income levels).

4 Results

4.1 Key descriptive statistics

Table 6 presents key socio-economic and plot- level variables. Household characteristics such as age, household size and education, membership in different social groups as well as wealth indicators such as livestock ownership measured in terms of total livestock units (TLU) are included to control for possible heterogeneities between adopters and non-adopters. We hypothesize that these household characteristics affect farmers' adoption decisions as well as their productivity levels

Table 6: Descriptive statistics of socio-economic characteristics

	Full sample (N=5123)	Adopters (N=2836)	Non-adopters (N=2287)	Mean diff
Household Size	4.58	4.77	4.35	0.42***
Education (years of schooling)	8.8	9.1	8.4	0.71***
Marital status (1=married, 0=otherwise)	0.88	0.90	0.85	0.05***
Age (measured in years)	51.7	51.1	52.6	-1.5***
Sex (1= female, 0= otherwise)	0.88	0.91	0.85	0.06***
Livestock ownership (TLU)	0.75	0.89	0.56	0.33***
Access to extension (1= village has access, 0= otherwise)	0.49	0.61	0.35	0.26***
Access to credit (1=village has access, 0= otherwise)	0.66	0.72	0.58	0.14***
Mobile phone ownership (1=owns, 0 otherwise)	0.97	0.98	0.96	0.02***
Television ownership (1=owns, 0 otherwise)	0.75	0.74	0.75	-0.006
Membership in credit and saving groups (1=yes, 0=no)	0.34	0.37	0.30	0.07***
Membership in cooperatives (1=yes, 0=otherwise)	0.25	0.29	0.19	0.1***
Membership in cassava growers association (1=yes, 0=no)	0.20	0.26	0.14	0.12***
Good soil (1=good, 0 otherwise)	0.74	0.77	0.69	0.08***
Medium soil (1=medium, 0 otherwise)	0.24	0.21	0.28	0.07***
Poor soil (1=poor, 0 otherwise)	0.02	0.02	0.03	-0.01
Labour use (MD/ha)	103.0	82.6	119.6	37
Fertilizer use (1=use, 0=no)	0.33	0.31	0.35	0.004***
Friend/neighbor is adopter (1=yes, 0=no)	0.55	0.58	0.53	0.05***
Advice on cassava production (1= yes, 0=no)	0.41	0.50	0.33	0.17***

Note that, in the data the rate of extension and credit access at the household level is quite low. However, both extension and credit access can easily be endogenous in the adoption decision model as household level access to extension and credit and adoption decision of households are simultaneously determined based on specific farm characteristics. This could lead to a reverse causation between the access variables (extension, credit) and the individual’s decision to adopt improved cassava varieties. We deal with such possible endogeneity of extension access by aggregating it at village level, with the presumption that village level extension and credit access would be exogenous to individual household characteristics. We constructed an indicator to distinguish villages with relatively easy access to extension and credit services to those without access. We set a threshold of 25% to distinguish *"extension and credit villages"* from *"non-extension and non-credit villages"*. For instance, at a threshold of 25%, a village is considered to be an *"extension villages"* if more than 25% of the households in a given village respond to have access to extension. In this case, all households in that village will be considered as having access to extension. A similar approach was also used by Di Falco & Bulte (2013). We also included plot-level variables to control for plot level heterogeneity. Our plot level controls are mainly for soil fertility. We found statistically significant differences between adopters and non-adopters for most of the socio-economic and plot level variables. In general, adopters tend to be more educated, wealthier (have more livestock), and younger. Moreover, adopters have better extension and credit access. Adopters are also significantly different from non-adopters in terms of membership in social networks. However, there is no statistical difference between adopters and non-adopters in terms of labor application.

4.2 Effect of adoption on cassava yield: OLS Estimation Results

In this section, we present OLS results using both household survey and DNA-fingerprinted adoption data. The first column presents results based on farmers-self reported adoption status from the household survey. The second column then presents the benchmark parameter estimates based on DNA-fingerprinted data. Both models are estimated at the plot level to account for plot level heterogeneities between adopters and non-adopters. The results in Table 7 suggest a yield advantage of 39%–54% for improved cassava varieties over traditional varieties.

Table 7: Ordinary Least Squares estimates (dependent variable: ln (yields)

	Self-reported adoption data	DNA finger-printed adoption data
Improved cassava variety	0.391***	0.539***
	(0.034)	(0.031)
Fertilizer use	0.008	0.018
	(0.037)	(0.034)
Labour use	0.116***	0.130***
	(0.014)	(0.013)
Intercropping	0.098**	0.112***
	(0.039)	(0.039)
Good soil	0.074	0.105
	(0.112)	(0.106)
Medium soil	0.041	0.045
	(0.114)	(0.111)
Household size	0.015*	0.012*
	(0.008)	(0.007)
Education	0.004	-0.001
	(0.004)	(0.004)
Married	-0.018	-0.032
	(0.065)	(0.058)
Age	0.001	0.006
	(0.009)	(0.009)
Age2	-0.000	-0.000
	(0.000)	(0.000)
Sex	-0.009	0.087
	(0.070)	(0.069)
Livestock ownership (TLU)	0.002***	0.002***
	(0.000)	(0.000)
Extension access	0.009	-0.011
	(0.042)	(0.038)
Access to credit	0.035	0.041
	(0.038)	(0.036)
Mobile phone ownership	-0.120	-0.135
	(0.103)	(0.101)
Television ownership	0.036	0.033
	(0.042)	(0.040)
Membership in credit and saving association	-0.054	-0.027
	(0.039)	(0.037)
Membership in cooperatives	0.049	0.040
	(0.046)	(0.046)
Membership in cassava association	-0.079*	-0.062
	(0.045)	(0.045)
Distance from output market	-0.004	-0.003
	(0.003)	(0.004)
Distance from input market	-0.000	0.000
	(0.000)	0.539***
Regional fixed effects	Yes	Yes
Joint F-statistic	17.23***	38.45***
R2	0.076	0.115
N	5694	5694

Standard errors clustered at the local government level are reported in parenthesis. ***, ** and * refers to significant at 1%, 5% and 10% respectively

As expected, misclassification results in attenuation bias as parameter estimates from the DNA-based adoption data are 15 percentage points higher than estimated from the self-reported adoption data. The results are consistent with the findings of past studies showing that attenuation bias in the presence of misreporting (Nguimkeu et al., 2016; Aigner, 1973; Black et al., 2000; Frazis and Loewenstein, 2003; Hausman et al. 1998; Kane et al. 1999; Lewbel, 2007; Mahajan, 2006; Hu and Schennach, 2008). Note that, in the absence of misclassification, parameter estimates of the two models should be exactly the same. Therefore, the large discrepancy between the two results suggests that measurement error is consequential. Note that OLS estimates could still be biased due to the endogeneity of adoption decision. The next section presents IV estimation results.

4.3 Effect of adoption on cassava yield: IV Estimation Results

In this section, we present IV regression results where we control for the endogeneity of the adoption decision. The first two columns present parameter estimates based on self-reported and DNA-fingerprinted adoption status, respectively. The last column presents results using a consistent sub-sample. We refer to the last column as a "matched sub-sample" as it involves households whose self-reported adoption status matches their DNA-based adoption status. In the matched sub-sample, adoption status is measured by a dummy variable which takes on a value of one if the matched self-reported and DNA-fingerprinted adoption data confirm adoption of improved varieties and zero if the matched household survey and DNA-finger printed adoption data confirm non-adoption of improved varieties. If adoption status from self-reported and DNA-fingerprinted adoption data do not match, the dependent variable is recorded as missing (this explains the difference in the sample size in the first two and the last column). In doing so, we exclude both "false negatives" and "false positives" in the last column. As such, parameter estimates will capture both technological and behavioural dimensions of farmers consistently.

Table 8: Instrumental variable regression estimates (dependent variable: ln (yields))

Variable	Based on self-reported adoption data	Based on DNA-finger printed data	Matched sample
Improved cassava variety	0.461**	0.578**	0.645***
	(0.196)	(0.227)	(0.179)
Fertilizer use	0.006	0.018	-0.046
	(0.026)	(0.025)	(0.030)
Labour use	0.116***	0.131***	0.121***
	(0.010)	(0.011)	(0.011)
Intercropping	0.098***	0.113***	0.075**
	(0.028)	(0.028)	(0.034)
Good soil	0.074	0.107	0.103
	(0.088)	(0.086)	(0.107)
Medium soil	0.042	0.045	0.104
	(0.089)	(0.086)	(0.109)
Household size	0.015***	0.012**	0.022***
	(0.005)	(0.006)	(0.007)
Education	0.004	-0.001	0.003
	(0.003)	(0.003)	(0.003)
Married	-0.014	-0.032	-0.021
	(0.050)	(0.049)	(0.052)
Age	0.001	0.006	-0.005
	(0.006)	(0.006)	(0.006)
Age2	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Sex	-0.017	0.090	0.084
	(0.058)	(0.057)	(0.059)
Livestock ownership (TLU)	0.002***	0.002***	0.001***
	(0.001)	(0.000)	(0.000)
Extension access	0.007	-0.013	-0.043
	(0.029)	(0.031)	(0.034)
Access to credit	0.034	0.041	0.014
	(0.027)	(0.026)	(0.030)
Mobile phone ownership	-0.133*	-0.141*	-0.063
	(0.079)	(0.082)	(0.104)
Television ownership	0.037	0.033	0.055
	(0.030)	(0.029)	(0.036)
Credit and saving	-0.058**	-0.026	-0.027
	(0.029)	(0.027)	(0.031)
Cooperatives	0.049*	0.039	0.035
	(0.028)	(0.028)	(0.031)
Membership in cassava ass	-0.083***	-0.062**	-0.039
	(0.032)	(0.030)	(0.036)
Distance from output market	-0.004*	-0.003	0.005
	(0.002)	(0.002)	(0.003)
Distance from input market	0.000	0.000	0.001**
	(0.000)	(0.000)	(0.000)
Regional fixed effects	Yes	Yes	Yes
Prob > F	0.0000	0.0000	0.0000
R2	0.078	0.116	0.197
N	5964	5964	2816

Standard errors clustered at the local government level are reported in parenthesis. ***, ** and * refers to significant at 1%, 5% and 10% respectively

IV estimation results suggest that the effect of adoption on productivity ranges from 46%-64%, which are higher than OLS estimates. The results indicate that failure to account for endogeneity of treatment status causes attenuation bias. For example, when controlling for the endogeneity of the true treatment status in the absence of misclassification in the second column, parameter estimates of the treatment variable increased from 53.9% (Table 7) to 57.8%. Hence failure to control for the endogeneity of treatment status in the absence of misclassification reduces the anticipated impact on productivity by about 3.9 percentage points. When controlling for the endogeneity of the true treatment status in the presence of misclassification in the first column, parameter estimates of the treatment variable increased from 39.1% (Table 7) to 46.1% (Table 8). Further, results from the second and third columns suggest that considering behavioural adjustments is important. Note that while using DNA-fingerprinted adoption data, the pure technological effects are consistently estimated. However, the effects of the inherent behavioural adjustment of farmers based of their own subjective self-assessment of treatment status are mixed-up with technological effects, leading to parameter estimates that are biased towards zero. Since the third column (the consistent sub-sample) captures both technological and behavioural adjustments consistently, parameter estimates will be consistent in the presence of unobserved dimensions of behavioural adjustment. Our results suggest that when both technological effects and unobserved behavioural adjustments of farmers are considered, adoption increase productivity by 64.5% (column 3).

4.4 Effect of adoption on cassava yield using matching techniques

Table 9 presents robustness cheeks using PSM and IPWRA approaches for cassava yield. As mentioned before, our treatment variable (adoption status) which takes a value of one if a farmer adopts improved cassava varieties and zero otherwise is measured using farmers self-reported adoption data and DNA-finger printed adoption data. We find a positive and statistically significant effect of adoption on yield in both PSM and IPWRA specifications, suggesting the robustness of our reported results. The results show that adoption (based on farmers self-reported adoption status) increased cassava yield by 36.7% and 25.8% in PSM and IPWRA, respectively. While measuring adoption status using DNA-finger printed adoption status, we found the effects on cassava yield are 53.7% and 53.8% in PSM and IPWRA, respectively.

Table 9: PSM and IPWRA estimation results

	PSM	IPWRA
Cassava yield with self-reported adoption data	0.367***	0.258***
	(0.039)	(0.028)
Cassava yield with DNA-finger printed adoption data	0.537***	0.538***
	(0.035)	(0.027)

Robust standard errors in bracket, *** p<0.01, ** p<0.05, * p<0.1

4.5 Adoption effects on cost of production

Table 10 presents the plot-specific treatment effect in terms of percentage cost increase due to adoption. In estimating the treatment effects in the cost function, the treatment (the decision to adopt improved cassava varieties) is endogenous (as farmers self-select into adoption). We therefore use instrumental variable regression (probit-2SLS) to control for the endogeneity of adoption status. Like before, the first two columns present parameter estimates based on self-reported and DNA-fingerprinted adoption status, respectively. The last column presents results based on a "matched sub-sample" as it involves households whose self-reported adoption status matches their DNA-based adoption status. The results presented in Table 10 suggest that, adoption has a positive and statistically significant effect on the cost of production. Using self-reported adoption data, our estimates in Table 10 suggests that the cost of producing cassava has increased by 33.7% as a result of adoption of improved cassava varieties. The effects on costs of production are a bit higher when using DNA-finger printed adoption data.

Table 10: Effect of adoption on cost of production (Dep. variable, log of cost per ha)

Variable	Based on self-reported adoption data	Based on DNA-finger printed data	Matched sample
Improved cassava variety	0.337*	0.381*	0.415*
	(0.204)	(0.226)	(0.252)
Fertilizer use	0.130***	0.126***	0.141***
	(0.028)	(0.026)	(0.030)
Labour use	0.234***	0.243***	0.240***
	(0.015)	(0.015)	(0.015)
Intercropping	0.003**	0.004**	0.004**
	(0.001)	(0.001)	(0.001)
Good soil	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
Medium soil	0.048	0.082	0.080
	(0.075)	(0.073)	(0.073)
Household size	0.064	0.062	0.073
	(0.079)	(0.073)	(0.075)
Education	-0.003	-0.002	-0.003
	(0.006)	(0.005)	(0.006)
Married	0.004	0.001	0.003
	(0.003)	(0.004)	(0.003)
Age	0.010	0.010*	0.008
	(0.006)	(0.006)	(0.006)
Age2	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Sex	-0.085**	-0.013	-0.041
	(0.043)	(0.047)	(0.040)
Livestock ownership (TLU)	-0.061*	-0.058*	-0.054
	(0.037)	(0.034)	(0.033)
Extension access	-0.116***	-0.103***	-0.105***
	(0.030)	(0.027)	(0.028)
Access to credit	-0.199**	-0.180**	-0.198**
	(0.083)	(0.076)	(0.084)
Mobile phone ownership	0.117***	0.094***	0.099***
	(0.032)	(0.030)	(0.031)
Television ownership	0.026	0.034	0.034
	(0.030)	(0.028)	(0.029)
Credit and saving	-0.035	-0.043	-0.049
	(0.032)	(0.032)	(0.034)
Cooperatives	-0.046	-0.025	-0.039
	(0.038)	(0.032)	(0.035)
Membership in cassava associations	-0.003	-0.002	-0.002
	(0.003)	(0.003)	(0.003)
Distance from output market	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Distance from input market	0.337*	0.381*	0.415
	(0.204)	(0.226)	(0.253)
Prob > F	0.0000	0.0000	0.0000
N	4191	4191	4191

4.6 Effects on poverty reduction

As discussed in section 3.2, estimating the aggregate poverty reduction effects of adoption of improved cassava requires estimating the counterfactual distribution –the income level of adopters had they not adopted improved cassava varieties. We followed the approach of Zeng et al (2015) and used our yield and cost treatment effects from Table 8 &10 to estimate the distribution of counterfactual income for estimating the poverty reduction effects of adoption of improved cassava varieties. One of the most innovative aspects of this project was the use of DNA finger printing to credibly identify adoption status. As such we estimated the poverty reduction effects for alternative measures of treatment (adoption status). These include:

i. Adoption status based on DNA-finger printing and assuming a small open economy &closed economy
ii. Adoption status based on self-reported adoption status and assuming a small open economy &closed economy

4.6.1 Poverty reduction effects of adoption using DNA-fingerprinted adoption data

We used treatment effects on yield (a 57.8% increase in cassava yield) and treatment effect on cost of production (a 38% increase in cost of production as a result of adoption) along with supply and demand elasticities to estimate the poverty reduction effects of adoption. However, estimating supply and demand elasticities is not feasible given the data available to us. In the absence of any empirical evidence from the literature, we therefore assumed a supply elasticity of 0.5 and a demand elasticity of -1 for cassava. According to FAOSTAT (2014), the average observed price ($\boldsymbol{P^{obs}}$) per kilogram of cassava is \$0.16. The corresponding year production of cassava at the national level is about 54.83 million metric tons. With an observed price of \$0.16/kg, adoption rate of 66% and the above-mentioned treatment effects on yield and cost of production, the **k**-shift is computed as a 39.6% cost reduction per kilogram of maize. Given, the above k-shift parameter and observed price of \$0.16, the counterfactual price ($\boldsymbol{P^{ct}}$) becomes

$0.185/kg. Given the above information, we then estimated the total changes in producer and consumer surplus that could be allocated for surveyed households. For allocating producer surplus changes to appropriate households, we used their sales quantities from total production as a weight. We then used observed per-capita expenditure to calculate observed poverty rates. For calculating counterfactual per-capita expenditure, we subtracted observed income gains from adoption (through producer surplus) from observed per-capita expenditure. We then used four alternative poverty lines ($1.25, $1.45 and $1.90 per person per day) to estimate the poverty reduction effects of adoption. Table 11 below presents the results. For all alternative poverty lines, poverty has reduced as a result of adoption.

Table 11: Poverty impacts of adoption of improved cassava varieties

Poverty line($per person per day)	FGT-poverty index	Observed	Closed economy	Poverty impact	Small open economy	Poverty impact
1.25	Headcount	0.60861	0.6555	-0.04689	0.6488	-0.04019
	Depth	0.39312	0.48679	-0.09367	0.47021	-0.07709
	Severity	0.29420	0.41522	-0.12102	0.39076	-0.09656
1.45	Headcount	0.65072	0.68708	-0.03636	0.67943	-0.02871
	Depth	0.42568	0.51248	-0.0868	0.49714	-0.07146
	Severity	0.32396	0.43703	-0.11307	0.41478	-0.09082
1.90	Headcount	0.69665	0.72584	-0.02919	0.71722	-0.02057
	Depth	0.48544	0.55884	-0.0734	0.54496	-0.05952
	Severity	0.38018	0.47914	-0.09896	0.46015	-0.07997

The two most utilized poverty lines in the literature are $1.25 and $1.90 per person per day. At the poverty line of $1.25 per person per day, adoption has led to a 4.7% and 4.02% poverty reduction in the closed economy and small open economy case, respectively. This poverty reduction role of adoption at $1.25 per person per day poverty line implies that 6.2%-7.15% of the rural poor cassava producers have escaped poverty in the current year due to adoption of improved cassava varieties. Similarly, at the poverty line of $1.9 per person per day poverty line, adoption has led to a 2.06% to 2.92% poverty reduction in a small open economy and closed economy, respectively. These changes correspond to a 2.9%-4% poverty reduction among rural poor cassava producers[1]. With regard to the depth and severity of poverty, we again found significant reductions at all poverty lines considered in our analysis.

Next, we estimated the poverty reduction effects of adoption using self-reported adoption rates. As mentioned in the introduction, one of the main objectives of this report was to improve the measurement of adoption status using DNA-finger printing approaches thereby improving the measurement of adoption status and subsequent impacts of adoption on poverty reduction and productivity. Herein, we used self-reported adoption status to show how misclassification of adoption status from self-reported adoption status my bias estimated impacts of adoption. Our theoretical model in section 3.1.3 suggests that misclassifying adoption status may attenuate impact estimates depending on the pervasiveness of false positives and false negatives (false negatives leading to downward bias and false positives leading to upward bias). Our results are reported in Table 12 using the case of closed economy-the more likely scenario in the current condition of Nigeria.

Table 12: Poverty reduction effects using self-reported adoption status

Poverty line($per person per day)	FGT-poverty index	Observed	Closed economy	Poverty impact	Small open economy	Poverty impact
1.25	Headcount	0.60861	0.63876	-0.03015	0.66411	-0.0555
	Depth	0.39312	0.45821	-0.06509	0.50106	-0.10794
	Severity	0.29420	0.37925	-0.08505	0.41674	-0.12254
1.45	Headcount	0.65072	0.6756	-0.02488	0.68804	-0.03732
	Depth	0.42568	0.48578	-0.0601	0.52501	-0.09933
	Severity	0.32396	0.40321	-0.07925	0.44163	-0.11767
1.90	Headcount	0.69665	0.71866	-0.02201	0.7244	-0.02775
	Depth	0.48544	0.53602	-0.05058	0.56826	-0.08282
	Severity	0.38018	0.44919	-0.06901	0.48634	-0.10616

[1] Note that, for calculating these values we followed the approach of Zeng et al., 2015. For example, in the small open economy, the counterfactual poverty headcount ratio and poverty impact under the $1.25 poverty line are 0.6555 and 0.04689, respectively. Thus, the percentage of the originally poor who have escaped poverty is 0.04689/0.6555 = 7.15%

Table 12 clearly suggests that while using self-reported adoption status, the poverty reduction effects of adoption are smaller. Using a poverty line of $1.25 and $1.90 per person per day, we found a 3% to 5.5% reduction in headcount poverty ratio in a closed and small open economy, respectively.

5 References

Abdoulaye, T., Abass, A., Maziya-Dixon, B., Tarawali, G., Okechukwu, R., Rusike, J., Alene, A., Manyong, V., Ayedun, B., 2013. Awareness and adoption of improved cassava varieties and processing technologies in Nigeria. J. Dev. Agric. Econ. 7(4).

Aigner, D. J. 1973. Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics* 1(1), 49-59.

Black, D. A., Berger, M. C. & Scott, F. A. 2000. Bounding parameter estimates with non-classical measurement error. *Journal of the American Statistical Association* 95(451), 739-748.

Frazis, H. & Loewenstein, M. A. 2003. Estimating linear regressions with mis-measured, possibly endogenous, binary explanatory variables. *Journal of Econometrics* 117(1), 151-178.

Gollin, D., Stephen, P., Richard, R. 2002. The Role of Agriculture in Development. *American Economic Review* 92 (2): 160–64.

Hausman, J. A., Abrevaya, J. & Scott-Morton, F. M. 1998. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, 87(2), 239-269

Hu, Y., and S. Schennach. 2008. Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica*, 76 (1): 195–216

Kane, T. J., Rouse, C. E. & Staiger, D. 1999. Estimating returns to schooling when schooling is misreported, Technical report, National bureau of economic research.

Lewbel, A. 2007. Estimation of average treatment effects with misclassification. *Econometrica*, 75(2), 537-551.

Mahajan, A. 2006. Identification and estimation of regression models with misclassification. *Econometrica* 74(3), 631-665

Nguimkeu, P., Denteh, A & Tchernis, R. 2016. On the Estimation of Treatment Effects with Endogenous Misreporting. Working paper

Rabbi, I.Y., Kulakow P.A., Manu-aduening, J.A., Dankyi, A.A., Asibuo, J.Y., Parkes, E.Y., Abdoulaye, T., Girma, G., Gedil, M.A., Ramu P, Reyes B. & Maredia M.K. 2015. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (Manihot esculenta Crantz). *BMC Genetics*, 16(115): 1–11.

Suri T. 2011. Selection and Comparative Advantage in Technology Adoption. *Econometrica* 79: 159–209

Wossen, T., Tahirou, A. Alene, A., S. Feleke, M. Haile, Olanrewaju, A. and V.M. Manyong. 2017. Impacts of extension access and cooperative membership on technology adoption and household welfare. Journal of rural studies (54), 223-233

Wossen, T., Tahirou, A. Alene, A., S. Feleke, M. Haile, and V.M. Manyong. 2017. Estimating the Productivity Impacts of Technology Adoption in the Presence of Misclassification. Forthcoming, American Journal of Agricultural Economics.

Wossen, T., Girma, G., Tahirou, A., Rabbi, I., Olanrewaju, A., Alene, A., S. Feleke., Kulakow, P., Asumugha, G., Abass, A,, Manyong, V. 2017. The Cassava Monitoring Survey in Nigeria. Avillable online at: http://www.iita.org/wp-content/uploads/2017/The_Cassava_Monitoring_Survey_in_Nigeria.pdf

Wooldridge, J. M. 2010. Econometric Analysis of Cross Section and Panel Data. The MIT Press.

Zeng, D., J. Alwang, G.W. Norton, B. Shiferaw, M. Jaleta, and C. Yirga. 2015. *Ex post* impacts of improved maize varieties on poverty in rural Ethiopia. Agricultural Economics 46(4): 515-526.