# Genebank Impacts

## Working Paper No. 13 | July 2021

# The contribution of the ICRISAT genebank to groundnut improvement and rural poverty in Malawi

**Martin Paul Jr. Tabe-Ojong**
Institute for Food and Resource Economics, University of Bonn, Germany
martin.tabe-ojong@ilr.uni-bonn.de

**Melinda Smale**
Michigan State University, USA
msmale@msu.edu

**Nelissa Jamora**
Global Crop Diversity Trust, Bonn, Germany
nelissa.jamora@croptrust.org

**Surendra S. Manohar**
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) India
m.surendra@cgiar.org

**Patrick Okori**
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) Malawi
p.okori@cgiar.org

**Vania Azevedo**
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) India
v.azevedo@cgiar.org

## Abstract

Genebanks contribute to varietal improvement through the conservation and supply of diverse crop germplasm over vast distribution channels. Tracing their contribution to the welfare and poverty impacts of improved varieties on smallholder farmers presents multiple challenges. We study the role of the ICRISAT genebank in the development of improved groundnut varieties and poverty reduction among groundnut producers in Malawi. First, we apportion the contribution of the genebank to improved varieties using pedigree data. We apply a Tobit model to a three-wave household panel to examine the adoption of improved varieties and test the effect of the genebank contribution. We then estimate adoption impacts on several welfare indicators in an instrumental variables regression. To link the two-stage regression, we use the chain rule and establish the positive role of the ICRISAT genebank in improving household incomes, expanding assets, and reducing income poverty. The main mechanism through which this happened was through the extent of adoption of the improved groundnut varieties, developed with breeding materials from the ICRISAT genebank. Our results are consistent over different specifications. We thus lend support and credence to the targeting and upscaling of improved crop technologies that are developed with access to diverse genebank materials. These have the potential to increase household welfare as well as lift the poorest households out of poverty.

## Suggested citation

*Tabe-Ojong, Martin Paul Jr., Melinda Smale, Nelissa Jamora, Surendra S. Manohar, Patrick Okori and Vania Azevedo. 2021. The contribution of the ICRISAT genebank to groundnut improvement and rural poverty in Malawi. Genebank Impacts. Working Paper No. 13. CGIAR Genebank Platform, IRRI, and the Crop Trust.*

## Acknowledgement

Funding for this research was provided by the CGIAR Genebank Platform, ICRISAT, and the Crop Trust through the 2020 Genebank Impacts Fellowship. We are grateful to Kai Mausch and Michael Hauser for initial discussions and providing access to the dataset.

# Acronyms

| | |
|---|---|
| CARD | Centre for Agricultural Research and Development |
| CRE | correlated random effect |
| ICG | ICRISAT Genebank |
| FAO | Food and Agriculture Organization of the United Nations |
| ICRISAT | International Crops Research Institute for Semi-Arid Tropics |
| IV | instrumental variable |
| ITPGRFA | International Treaty on Plant Genetic Resources for Food and Agriculture |
| LASSO | least absolute shrinkage and selections operator |
| LPM | linear probability model |
| PPP | purchasing power parity |
| RCG | relative genetic contribution |
| RCP | relative contribution of provenance |
| SSP | single plant progenies |
| SDG | sustainable development goals |
| TLU | tropical livestock units |

# Contents

# 1 Introduction and motivation

Despite implementing significant structural and economic reforms to sustain economic growth, Malawi continues to lag behind and is currently ranked as one of the world's poorest economies (World Bank 2020). Further exacerbating the situation is its vulnerability to a plethora of external shocks, particularly climatic shocks and its growing population which is expected to double by 2038 (ibid.). As the economy is dependent on agriculture, the agricultural sector remains the most affected with approximately 55% of farmers cultivating on less than one hectare (FAO 2015). Smallholder production is still highly subsistence oriented and characterized by both low levels of input and output. Against this background, governments, as well as national and international development agencies, have prioritized enhancing the productivity of smallholder farmers as a means of achieving agricultural growth, reducing rural poverty, and achieving the sustainable development goals (SDGs). Varietal improvement is one of the most important ways to enhance crop productivity of smallholder farmers (Evenson and Gollin 2003; Raitzer and Kelley 2008). As highlighted by Walker and Alwang (2015), varietal improvements not only have implications for food and agricultural development but are also one of the pathways to reducing poverty in rural areas (Alwang et al. 2019).

The adoption of improved tropical legumes has been suggested to be both pro-poor and environmentally friendly (Verkaart et al. 2017). Improved legume varieties contribute to poverty reduction by improving market access and income of farmers (Tabe-Ojong et al. 2021) and are key to maintaining environmental sustainability due to their ability to fix atmospheric nitrogen, which in turn can lead to a reduction in the use of inorganic fertilizers (Giller 2001). One particularly propitious legume is the drought-tolerant, disease-resistant, and high-yielding improved groundnut (*Arachis hypogaea* L.) varieties, produced and released by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). The ICRISAT genebank partners with other global and national breeding programs in conserving plants accessions and germplasm with varied traits and characteristics.

# 2 Background and research objectives

In this paper, we trace the impact of the genebank through varietal improvement on the welfare of smallholder farmers in Malawi. By combining pedigree data with econometric methods applied to panel data, we are able to estimate the genebank's contribution to the alleviation of rural poverty.

With regards to the role of the genebank in varietal improvement, we first establish the ancestry of the improved varieties grown by farmers and link their pedigrees to genetic materials maintained in ICRISAT's genebank. To establish the relative genetic contribution (RGC) of the various progenitors from the ICRISAT genebank, we applied the relative contribution of provenance (RCP) based on pedigree data (Bernal-Galeano et al. 2020). From this, we used the Mendelian rule of inheritance to apportion the theoretical genetic components of these varieties to genebank ancestors. This step created the variable of key interest in the first stage of our multi-stage econometric model.

Our research contributes to two strands of literature. First, we move beyond the classical impact evaluation literature that starts at the development of improved groundnut varieties and take a step back to fully understand the role of the genebank in the process. We add empirical evidence and learning on the impacts of improved legume varieties in rural Africa both from a short-term and long-term welfare perspective. Previous research on the impact of improved legumes exists (Tabe-Ojong et al. 2021; Verkaart et al. 2017; Asfaw et al. 2012) but with little insight on the role of genebanks. Our research addresses this lacuna in the literature. We advance recent insights on the role of genebanks in crop varietal improvement and subsequent socio-economic outcomes by analyzing a different crop and context with an original econometric approach. Recent, different econometric approaches to this topic include Villanueva et al. (2020) and Selitti et al. (2020).

Second, we assess the impacts of these genebank materials on both the income and assets of households. For the assets, we consider both productive and non-productive household assets. As poverty in most rural areas is reflected in the lack of assets (Moser 1998; Brockington 2021), we are capturing the longer-term welfare effects as opposed to the short-term effects that we observe when using income as an indicator. Typically, farm income is irregular and lumpy and depends very much on seasonal harvests, making it to be a less forward-looking measure of poverty (Tabe-Ojong et al. 2020).

## 3 ICRISAT genebank and varietal improvement

The success of the green revolution in Asia and Latin America continues to be reflected in the proliferation of improved seeds and varieties in developing nations. In collaboration with national governments, the production of these improved crop varieties has been carried out by many international agricultural research institutions. Genebanks in these institutions are committed to conserving plant species and maintaining their diversity. This is the case of the ICRISAT genebank (https://www.icrisat.org/gene-bank/) which contains about 50,000 accessions of pulses like pigeonpea, chickpea, and groundnuts. The diverse collections of accessions serve the dual purpose of insurance against genetic erosion as well as a source of tolerance to diseases and pests, ecological stresses, higher nutritional quality, and traits related to yield improvement. Most of these collections have been placed in-trust with the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) and the Food and Agriculture Organization (FAO) of the United Nations for global use (ICRISAT 2019).

Groundnut (*Arachis hypogaea* L.*)* is known to have originated in the region that now spans the countries of Bolivia and Argentina and is currently cultivated in 108 countries (ICRISAT 2017). Groundnut is an important cash and food crop, usually grown sole or in combination with other crops. As a cash crop, it is cultivated for its rich source of edible oil and high protein content making it also relevant for food and fodder. Because of the ability of groundnut to synthesize atmospheric nitrogen, the crop can be referred to as pro-poor and environmentally friendly (Verkaart et al. 2017).

Since the 1970s, about 31 improved groundnut varieties have been produced using breeding materials from ICRISAT (ICRISAT 2019). Of these 31 varieties, 12 have been released in Malawi (Appendix Table 1). One of the released varieties, ICG 12991 is an Indian landrace released in 2001 in Malawi as Baka. It was also released in Zambia, Mozambique, and Uganda to help in the fight against rosette virus. Apart from the vagaries of weather, the rosette virus has been one of the greatest constraints affecting the production of groundnuts in sub-Saharan Africa (Naidu et al. 1999).

Other released varieties that have been promoted for production and commercialization in Malawi include ICGV-SM 83708 (CG7), ICGV-SM 90704 (Nsinjiro), JL 24 (Kakoma), Manipintar, and ICGV SM 99568 (Chitala). These varieties were introduced to farmers through participatory varietal selection (PVS), on-farm research trials, and farmer field days (Simtowe et al. 2010). Apart from being resistant to the rosette virus, they have interesting production features like shell size, color, early maturation, branching, and strong marketing potential. For instance, Baka and Chitala mature early (90–110 days to maturation), and are drought and rosette resistant (Deom et al. 2006). However, they have different shelling attributes; while Baka is tiny shelled, Chitala is large seeded, making it a more preferred seed amongst farmers.

# 4 Data and methods

## 4.1 Farm household survey

Our data comes from a joint effort of ICRISAT, the Centre for Agricultural Research and Development (CARD) of the University of Malawi, and National Smallholder Farmer's Association (NASFAM). We employed a multistage sampling technique whereby Balaka and Mchinji districts were purposely selected in the first stage as regions specializing in groundnut production. From each district, two sections were chosen, and three villages were randomly selected from each section. Twelve to thirteen farmers were randomly selected using household lists constructed in these villages. This led to a total of 149 farm households in the base year 2008. These same households were followed in 2010 and 2013 with no attrition. Before the survey, we made a comprehensive list of all the improved and local groundnut varieties and then asked the farmers randomly selected for the survey about their knowledge and use of the varieties.

Preliminary data showed that farming households grew six main varieties in these two study districts represented here by their local names: CG7, Nsinjiro, Kakoma, Manipintar, Baka, and Chitala. From this varietal data at the farm household level, we generated pedigrees by talking to groundnut breeders and several genebank specialists. Apart from this rich varietal information, the survey also garnered information on a range of variables including socio-economic and biographic profiles of households, landed and non-landed farm assets, livestock ownership, membership of household in different village organizations, production, market participation, and household income sources.

As our income sources were recorded in Malawian Kwacha, we deflated them to real values to enable comparison over time. For this, we used the national consumer price index and set 2005 as the base. The 2005 Malawian Kwacha values were then converted to the US dollar ($) purchasing power parity (PPP) values with obtained rates from the International Comparison Program of the World Bank (World Bank 2014).

For the assets, we differentiated between non-productive and productive assets. While the productive assets mostly involved farm and agricultural machinery, the non-productive assets included other assets like houses, televisions, and radios. Since livestock ownership represents wealth in most rural areas (Tabe-Ojong et al. 2020), we considered livestock ownership as a separate productive asset. We convert the ownership of different livestock to the tropical livestock units (TLU) indices using conversion rates obtained from FAO. Here, a cow represents 0.7 units, a sheep and goat, 0.1 units and a chicken 0.01 units. For income poverty, we used the international poverty line of 1.25 US$ PPP and the median poverty line of 2.00 US$ PPP per day per capita to represent the lower and upper bounds of poverty (Ravallion et al. 2009). Based on this line, we differentiate between poor and non-poor households and create a poverty dummy variable.

## 4.2 Measuring genetic contribution

To measure the genetic contribution of the ICRISAT genebank, we used the relative contribution of provenance (RCP) and apportion based on the Mendelian rule of inheritance. The Mendelian rule of inheritance implicitly assumes that every parent in the pedigree of an improved variety contributes equally in each generation (Smale and Jamora 2020). By this, it ignores the effects of any random genetic drift as well as the effect of breeder's selection based on selected traits of interest. RCP[1] is a simplified measurement algorithm that evaluates the contribution of a source entity to a released variety (Bernal-Galeano et al. 2020).

---

[1] This is quite similar to relative genetic contribution (RGC) which is also a method of quantifying the genetic contribution from each of the parents based on the Mendelian rule of inheritance.

We then attribute genetic contribution of the genebank to household welfare outcomes using econometric techniques. From the farm household survey, farmers are mainly using six groundnut varieties (CG7, Nsinjiro, Kakoma, Manipintar, Baka, and Chitala). From this varietal data at the farm household level, we generated pedigrees by talking to groundnut breeders and several genebank specialists. For these six varieties, we have the pedigree information (though limited to one generation) for CG7, Nsinjiro and Chitala. Baka, an Indian landrace released in sub-Saharan Africa, is known to be part of the ICRISAT genebank collection. This makes the apportioning here quite straightforward. The pedigree for Kakoma cannot be easily derived as it originates from an Indian national breeding program. The pedigrees of the varieties are presented in Table 1.

As shown in Table 1, two of the three varieties came directly from parents that were conserved at the ICRISAT genebank (see Appendix Table 2). To enable keen understanding into varietal developments, we explain the pedigree method. In this method, $F_2$ bulks population of a cross, superior single plants were selected depended upon the objective of the cross. It usually varies between 40–50 single plant progenies (SSP). The selected SSP were planted in progeny rows of $F_3$, $F_4$ generations and $F_5$ generation onwards. The single plant progenies were then advanced to the next generation and sown bulk progenies. This is done until it becomes a homogeneous bulk population. Homogeneous bulks were later tested in replicated yield trials in national yield testing system with best controls and superiors test entries were subsequently released in country. Genotypic and phenotypic profiles of all the improved varieties are presented in the appendix.

Following the algorithm of Mendelian inheritance, for CG7 and Chitala, we apportioned 100% of their breeding provenance to the genebank. This is also the case for the Baka variety. For Nsinjiro, we apportioned it with a 50% value as just one of its parents is from the genebank. While RG 1 is from the ICRISAT genebank, the source of Manipintar was not confirmed during the consultation with experts, so we apportioned it a zero value. Finally, for Kakoma, we apportioned a value of 0 since we do not have any information to ascertain its link to the genebank.

In the empirical model, we used the relative provenance as obtained from the apportionment as an indicator or genebank contribution. Additionally, as a measure of robustness, we used a binary variable indicating whether any of the ancestors of the cultivated varieties can be traced to the ICRISAT genebank, based on expert consultation. Here, we take into consideration that the genotype of a cultivated variety has links with a genebank if any of its germplasm was supplied by the genebank collection.

## 4.3 Estimation strategy

We aim to estimate the impact of the ICRISAT genebank on household income, assets, and income poverty. We do this in three steps. First, we estimate the impact of the relative contribution of provenance on the groundnut area in improved varieties, or the extent of adoption. In the second step, we estimate the impact of the extent of adoption on household income, assets and poverty as measured by income. Finally, we apply the chain rule to measure the impact of ICRISAT genebank on our welfare indicators. In what follows, we describe this approach in detail.

To estimate the impact of relative provenance on the extent of adoption, we specify the following panel empirical model as

$$Y_{it} = \beta RP_{it} + \delta \boldsymbol{X}_{it} + c_{1i} + u_{2it}, \qquad (1)$$

where $Y_{it}$ is the extent of adoption for household $i$ in panel year $t$, $RP_{it}$ is the relative provenance of the improved variety. Extent of adoption is defined as the groundnut area planted to the improved variety by the farm household. As explained above, (1) shows the link to the genebank. As a measure of robustness, we replace relative provenance with a dummy variable, genetic ancestry that takes the value of 1 if the improved groundnut variety has any of its parents from the ICRISAT genebank and zero otherwise. $\boldsymbol{X}_{it}$ is a vector of control variables including time fixed effects. While $c_{1i}$ represents time-invariant unobserved heterogeneity, $u_{2it}$ captures time-varying idiosyncratic shocks. As our outcome variable, the extent of adoption has zeros and positive continuous variables, we model it as a corner solution model employing the Tobit model (Tobin 1958) which assumes that $u_{2it}$ is normally distributed, $u_{2it}|\boldsymbol{X}_{it}$~Normal(0, $\sigma^2$).

To control for time-invariant unobserved heterogeneity in the Tobit model, we apply the correlated random effects model (CRE), also known as the Mundlak-Chamberlain Device (Mundlak 1978; Chamberlain 1982). Very similar to the random effects model, it is empirically implemented by adding time averages of all time-varying variables in the regression models. This technique avoids the incidental parameters problem in linear models and reports the coefficients of time-invariant variables while controlling for unobserved household characteristics.

Now, to estimate the impact of extent of adoption on our various outcomes, we again specify a similar empirical model, though with some minor changes as

$$W_{it} = \sigma Y_{it} + \lambda \boldsymbol{V}_{it} + c_{2i} + u_{2it}, \qquad (2)$$

Here, our key interest variable, $Y_{it}$, is the extent of adoption. A positive coefficient of $\sigma$ implies adoption extent has a positive impact on our six outcome variables represented here by $W_{it}$. $\boldsymbol{V}_{it}$ is a vector of control variables while $c_{2i}$ and $u_{2it}$ represent time-invariant heterogeneity and time-varying shocks, respectively.

As our outcome variables have different data properties, we employ different models for their estimation. For the income and asset variables, we use the standard household fixed effect (FE) panel estimator as there is sufficient variation in either variable among our sampled households over the panel period. However for the poverty regressions, the within variation is smaller, so we again use the Mundlak Chamberlain Device which is more efficient for small variations within the outcome variable (Wooldridge 2010).

Despite controlling for unobserved heterogeneity, we had concerns about the endogeneity of the extent of adoption owing to unobserved shocks which are time-varying. To control for such shocks, we follow Bezu et al. (2014) and Verkaart et al. (2017) in using the unconditional expected values of adoption extent as an instrument for the observed extent of adoption in a control function approach (Wooldridge 2015). We begin by estimating the extent of adoption as outlined above. From this, we calculate the unconditional expected values of the extent of adoption using predicted values from the Tobit model. In doing this, we exclude some variables from the original Tobit adoption model in the specification of the outcome models. The exclusion of these variables serves as a source of exogenous variation. Some of these variables include agronomic characteristics like soil characteristics, access to irrigation, and distance to extension agents. We do not expect any of the above variables to be directly correlated with our outcome variables after controlling for the extent of improved groundnut adoption. Thus, our specified variables may provide variation in our instrument over time. To test the validity of these exclusion restrictions, we added them as additional regressors in the welfare outcomes where we obtained statistical insignificant effects (Appendix Tables A3 and A4 ). The results are maintained both for the full sample and when restricted to households that did not use any improved variety. In all the outcome models, the control function residuals are statistically insignificant except for income related outcomes (Appendix Table A9). This suggests that endogeneity may not be an issue in the other outcomes, making us to drop the residuals in these outcomes.

After obtaining the impact of relative provenance on the extent of adoption and the impact of the extent of adoption on our welfare outcomes, we finally apply the chain rule to calculate the effect of relative provenance on our welfare outcomes. This is calculated by taking the unconditional partial effect from the Tobit model and multiplying by the partial effect from the welfare models (Mason and Smale 2013), as follows

$$\frac{\partial W_{it}}{\partial RP_{it}} = \frac{\partial W_{it}}{\partial Y_{it}} \cdot \frac{\partial Y_{it}}{\partial RP_{it}} \qquad (3)$$

# 5 Results and discussion

## 5.1 Descriptive statistics

Figure 1 shows the distribution of the improved varieties used by farmers. As previously mentioned, six improved groundnut varieties were being cultivated in the Balaka and Mchinji districts. In both districts, the most cultivated variety is CG 7, followed by Nsinjiro, Manipintar and Kakoma. Chitala and Baka are cultivated by less than 2% of the farmers. For Baka, this may be due to its very small shells which are undesirable for most farmers who cultivate groundnuts for food. Moreover, since it is one of the earliest released groundnut varieties in Malawi, it has probably lost its relevance over time with the release of varieties with better shelling properties. This is depicted in Figure 2 where the various varieties are represented over the three panel years. Notwithstanding, Baka is now increasingly used as an experimental variety for breeding and advanced line purposes.

While the adoption of the various varieties depicts a varying trend over time, CG 7 is by far the most cultivated variety despite being v-shaped as a result of a drop in its adoption extent in 2010. One noteworthy observation here is the increasing adoption trend of Nsinjiro. Despite having lower adoption rates than CG 7, Nsinjiro maintained an increasing pattern throughout the panel years. Similar to CG 7, Kakoma also depicted a v-shaped extent of adoption while rates for Manipintar and Baka generally decreased over time. A graph of the distribution of the improved varieties over time in the two districts is found in the appendix (A1).

Applying pedigree analysis and with insights from the Mendelian rule of inheritance, we apportion the genetic components of these varieties to individual ancestors. From this, we calculate the percentage of the genetic composition of improved varieties derived from the groundnut accessions housed by the ICRISAT genebank as shown in Table A1 in the appendix. We report the summary statistics of the key genebank attribution variables (Table 2). 43% of households cultivate varieties whose ancestry can be traced to the ICRISAT genebank. While this is based on observing any parent from the genebank, it may not tell a very compelling story as having two parents from the genebank takes on the same value as having just one parent. Among adopters, the ICRISAT genebank contributes 88.6% of the ancestry of improved groundnut varieties cultivated by households in Malawi.

Table 3 shows the summary statistics of some of the socio-economic and farm variables for the pooled sample. Here we also report the differences between farm households who plant improved groundnut varieties and those that do not. Households have an income level of approximately US$2200 which is greater for households that use the improved groundnut varieties with materials from the genebanks. Households also make use of both productive and non-productive assets. There exist significant mean differences in the value of productive assets owned by groundnut adopters and non-adopters. This is also the case of livestock ownership where households report an average of 5.71 livestock units.

Smallholder households in Mchinji and Balaka have a household size of about five members and operate farm sizes of about one hectare. From this, about a third (38%) is under improved groundnut cultivation. About half of the households (46%) grow improved groundnut varieties which as described above have ancestors originating from the ICRISAT genebank.

In terms of socio-economic characteristics, households have an average age of about 46 years and have spent seven years in school which is equivalent to achieving primary education and beginning secondary education. Approximately 80% of the households are male headed and have been involved in groundnut production for about eight years. Coming to institutional variables like access to extension agents and markets, we find that households are an average distance of five km from extension agents. Similarly, households have a mean walking distance of about 12 km to reach markets.

We see significant differences between households that adopted improved groundnut varieties and households that did not. In the subsequent section, we validate the consistency of results by controlling for both observed and unobserved factors that may affect our welfare outcomes in a regression framework.

## 5.2 Effect of genebank contribution on adoption

In this section, we present and discuss the effects of the genebank contribution on adoption using five different specifications. Table 4 shows the estimates of the effect of the relative provenance when and when not controlling for a plethora of confounders. In all the different specifications, our estimate of the effect of relative provenance is the same, implying considerable consistency. A percentage point increase in the access of materials from the genebank is associated with a 0.013 percentage point increase in the area under the adoption of improved groundnut varieties.

Apportioning genebank contribution based on any ancestral link to the genebank (Table 5), we still obtain a positive coefficient, but in this case, the magnitudes are larger. Here, a 10 percentage point increase in any genebank ancestry is associated with a 13.8 percentage point increase in the area under the adoption of improved groundnuts. This increase in magnitude is justified by the fact that unlike in the relative provenance case where we calculate genebank contribution in relative terms, we only apportion contribution based on whether or not any genebank ancestor is present. This is an important finding, strengthening the idea that scales of measurement do matter. That notwithstanding, the results all speak to the significant role of the genebank in driving the adoption of improved groundnut varieties.

## 5.3 Effect of adoption on income, assets, and poverty

In this section, we also present and discuss the effects of adoption on income, assets, and poverty. For all income and assets outcomes, we used fixed-effects instrumental variable regressions (Table 6) while the correlated random effect estimates for income poverty are shown in Table 7.

Results in column (1) of Table 6 suggest that adoption of improved groundnut varieties increase household income by US$766 which is equivalent to an income gain of approximately 48% at the mean. This is quite a substantial effect given that the landholdings are small and the use of modern inputs like improved seeds represents some of the few options to increase food production (Bezu et al. 2014). Similar insights were obtained by Verkaart et al. (2017) for the adoption of improved chickpea varieties in Ethiopia. Improved groundnut varieties, like improved chickpea, may have desirable market traits which make them appealing in output markets (Tabe-Ojong et al. 2021). Adoption also has a statistically significant impact on the values of unproductive assets held by households. Over the years of the panel data, the adoption of improved groundnut varieties increases unproductive assets by US$118. Furthermore, significant effects are obtained for productive assets and the ownership of livestock. Adoption increases the value of productive assets by US$115, equivalent to an asset gain of above 100%. This is also the case for livestock ownership where the gains from adoption are far higher than the current livestock levels. This finding highlights some significant level of diversification. In this case, diversification is favorable as it leads to an increase in livestock ownership. Since livestock represent a significant form of rural wealth and livestock keeping is an important economic activity in such regards, households may invest the proceeds from the farm sector in developing the livestock sector.

The sizeable, positive impacts of adoption on income and assets also contribute reductions in income poverty as shown in Table 8. Here, we estimated a CRE model with and without instrumental variables. However, we only discuss the IV regression but it is noteworthy to mention here that the IV estimates are larger than the OLS estimates as expected. In column (1), adoption reduces income poverty by 7.2 percentage points for households living below the US$1.25 poverty line. Similarly, adoption reduces the probability of households living below the median income poverty line (US$2.00) by 11.7 percentage points. We thus conclude that the adoption of improved groundnut varieties can raise rural households both under the median poverty line as well as the poorest households out of poverty.

Applying the chain rule to the estimates presented above, we derive the contribution of the genebank to our welfare outcomes (Table 8). Since we multiply from the first adoption equation, the magnitudes are also relatively small when compared to the impact of adoption on our household welfare outcomes. On average, a 10 percentage point increase in the availability of materials from genebanks increases household income by US$99.6, unproductive assets by US$15.4 and productive assets by US$15. It also increases livestock ownership by 2.7 livestock units. In terms of reducing income poverty, a 100 percentage point increase in the relative contribution of genebank provenance lifts households below the US$1.25 poverty line by 0.09 percentage points and those below the medium poverty line by 0.15 percentage points. Overall, our findings support recent studies that reported the importance of various genebanks in varietal development as well as its productivity and yield impacts (Sellitti et al. 2020; Villanueva et al. 2020).

### 5.4 Robustness checks

We test an alternative specification where instead of considering the area under improved groundnut adoption, we treat adoption as a dummy. Here, instead of employing probit or logit models, we use a linear probability model (LPM). The LPM has several distinct advantages. First, it avoids identification by functional form which is common when using probit models (Angrist and Pischke 2009). Second, it is easier to interpret the coefficients of the LPM as opposed to the coefficients from the probit model which require parameter transformation into marginal effects for easier interpretation. Despite the advantages of using the LPM model, it has two main shortcomings. Firstly, the LPM approach leads to the generation of heteroscedastic errors, which can be controlled using robust standard errors. Secondly and more importantly, the predicted probabilities from the LPM model can fall out of the strict [0 1] interval. This shortcoming would be crucial if we are predicting the probability of a given outcome. However, as we instead examine the average partial effect of relative provenance on adoption, this should be less of a concern (Wooldridge 2010).

To further confirm this alternative specification, we ran some machine learning regressions. We employed the double selection lasso linear model (Belloni et al. 2014). As shown in Table 9, both the LPM and the double selection lasso linear regressions provided results which are consistent in magnitude and sign with the findings obtained for the extent of adoption. This goes to bolster our claims of the positive role of the ICRISAT genebank in the improvement of varieties which are subsequently adopted.

## 6 Conclusion

In this study, we provide empirical answers to the contributory role of the ICRISAT genebank to the improvement of groundnut varieties and the reduction of poverty in Malawi. We argue that the main mechanism through which this occurs is through the adoption extent of improved groundnut cultivation. Using pedigree data from the improved groundnut varieties, we establish the percentage of genebank materials present in the varieties adopted by farm households. We then use this contribution as a right-hand-side variable to estimate a Tobit model with correlated random effects and use the predicted groundnut area from the Tobit model to instrument for adoption in a mix of both fixed effects and correlated random effect welfare equations. We find substantial impacts of adoption on the income and asset levels of households with accompanying income and asset gains.

Using the chain rule to link first and second stages of the adoption welfare model, we find positive impacts of the ICRISAT genebank to household income increases, as well as the accumulation of both productive farm assets and livestock. Significant reduction in income poverty was also reported based on both the poverty line and the medium poverty line. Our results are robust to different empirical specifications, enabling us to conclude that genebank ancestry in the improved variety contributes to welfare increases.

Given that we observe large effects of adoption on the income and asset levels of households, especially with the ability of adoption to lift the poorest households out of poverty, we lend empirical support to the design, development, and dissemination of improved crop varieties as a significant way out of poverty. We also provide learning to the targeting and upscaling of improved crop varieties as this may have the intended welfare increasing and poverty reduction effects. That said, the fact that we could link these large income and poverty reduction effects to genebanks and still obtain significant effects is even more impressive and relevant for policy. Genebanks matter for both varietal and household welfare improvement.

Significant policy and institutional support should be provided to the genebanks to ensure that their role of conserving crop germplasm and breeding materials is maintained. The consultation with breeders and genebank specialists highlighted the gaps in pedigree information and documentation. Access to germplasm and accession-level data by breeders to develop varieties with desired traits must be improved. This is even more the case in the face of climate change and growing numbers of pests and diseases that are increasingly constraining food production.

# 7 References

Alwang, J., Gotor, E., Thiele, G., Hareau, G., Jaleta, M., & Chamberlin, J. (2019). Pathways from research on improved staple crop germplasm to poverty reduction for smallholder farmers. *Agricultural Systems*, *172*, 16–27. https://doi.org/10.1016/j.agsy.2017.10.005

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Asfaw, S., Shiferaw, B., Simtowe, F., & Lipper, L. (2012). Impact of modern agricultural technologies on smallholder welfare: Evidence from Tanzania and Ethiopia. *Food Policy*, *37*(3), 283–295. https://doi.org/10.1016/j.foodpol.2012.02.013

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*(2), 608–650. https://doi.org/10.1093/restud/rdt044

Bernal-Galeano, V., Norton, G., Ellis, D., Anglin, N. L., Hareau, G., Smale, M., Jamora, N., Alwang, J., & Pradel, W. (2020). Andean potato diversity conserved in the International Potato Center genebank helps develop agriculture in Uganda: the example of the variety 'Victoria.' *Food Security*, *12*(5), 959–973. https://doi.org/10.1007/s12571-020-01037-8

Bezu, S., Kassie, G. T., Shiferaw, B., & Ricker-Gilbert, J. (2014). Impact of improved maize adoption on welfare of farm households in Malawi: A panel data analysis. *World Development*, *59*, 120–131. https://doi.org/10.1016/j.worlddev.2014.01.023

Brockington, D. (2021). Persistent peasant poverty and assets. Exploring dynamics of new forms of wealth and poverty in Tanzania 1999–2018. *The Journal of Peasant Studies*, *48*(1), 201–220. https://doi.org/10.1080/03066150.2019.1658081

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, *18*(1), 5–46. https://doi.org/10.1016/0304-4076(82)90094-X

Crop Trust. (2016). *Securing crop diversity for sustainable development.* Global Crop Diversity Trust. https://cdn.croptrust.org/wp/wp-content/uploads/2017/02/Impact-Paper_16Dec2015_ks.pdf

Deom, C. M., Kapewa, T., Busolo-Bulafu, C. M., Naidu, R. A., Chiyembekeza, A. J., Kimmins, F. M., Subrahmanyam, P., & Merwe, P. J. A. (2006). Registration of ICG 12991 Peanut Germplasm Line. *Crop Science*, *46*(1), 481–481. https://doi.org/10.2135/cropsci2005.010

Evenson, R. E. (2003). Assessing the Impact of the Green Revolution, 1960 to 2000. *Science*, *300*(5620), 758–762. https://doi.org/10.1126/science.1078710

FAO. (2015). *National investment profile: Water for agriculture and energy - Malawi.* http://www.fao.org/fileadmin/user_upload/agwa/docs/NIP_Malawi_Final.pdf

Giller, K. E. (Ed.). (2001). *Nitrogen fixation in tropical cropping systems* (2nd ed.). CABI. https://doi.org/10.1079/9780851994178.0000

ICRISAT. (2017). *Groundnut Dashboard – Patancheru, Genebank ICRISAT.* http://genebank.icrisat.org/IND/Dashboard?Crop=Groundnut

ICRISAT. (2019). *Genebank*. https://www.icrisat.org/gene-bank/

Mason, N. M., & Smale, M. (2013). Impacts of subsidized hybrid seed on indicators of economic well-being among smallholder maize growers in Zambia. *Agricultural Economics*, *44*(6), 659–670. https://doi.org/10.1111/agec.12080

Moser, C. O. N. (1998). The asset vulnerability framework: Reassessing urban poverty reduction strategies. *World Development*, *26*(1), 1–19. https://doi.org/10.1016/S0305-750X(97)10015-8

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69. https://doi.org/10.2307/1913646

Naidu, R. A., Kimmins, F. M., Deom, C. M., Subrahmanyam, P., Chiyembekeza, A. J., & van der Merwe, P. J. A. (1999). Groundnut rosette: A virus disease affecting groundnut production in sub-Saharan Africa. *Plant Disease*, *83*(8), 700–709. https://doi.org/10.1094/PDIS.1999.83.8.700

Raitzer, D. A., & Kelley, T. G. (2008). Benefit–cost meta-analysis of investment in the International Agricultural Research Centers of the CGIAR. *Agricultural Systems*, *96*(1–3), 108–123. https://doi.org/10.1016/j.agsy.2007.06.004

Ravallion, M., Chen, S., & Sangraula, P. (2009). Dollar a day revisited. *The World Bank Economic Review*, *23*(2), 163–184. https://doi.org/10.1093/wber/lhp007

Sellitti, S., Vaiknoras, K., Smale, M., Jamora, N., Andrade, R., Wenzl, P., & Labarta, R. (2020). The contribution of the CIAT genebank to the development of iron-biofortified bean varieties and well-being of farm households in Rwanda. *Food Security*, *12*(5), 975–991. https://doi.org/10.1007/s12571-020-01038-7

Simtowe, F., Asfaw, S., Diagne, A., Shiferaw, B. A., Simtowe, F., Asfaw, S., Diagne, A., & Shiferaw, B. A. (2010). *Determinants of agricultural technology adoption: the case of improved groundnut varieties in Malawi*. https://doi.org/10.22004/AG.ECON.95921

Smale, M., & Jamora, N. (2020). Valuing genebanks. *Food Security*, *12*(5), 905–918. https://doi.org/10.1007/s12571-020-01034-x

Tabe-Ojong, M. P., Hauser, M., & Mausch, K. (2020). *Does agricultural commercialization increase asset and livestock accumulation on smallholder farms in Ethiopia?*

Tabe-Ojong, M. P. J., Mausch, K., Woldeyohanes, T. B., & Heckelei, T. (2021). Three hurdles towards commercialisation: integrating subsistence chickpea producers in the market economy. *European Review of Agricultural Economics*, *jbab023*. https://doi.org/10.1093/erae/jbab023

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*(1), 24. https://doi.org/10.2307/1907382

Verkaart, S., Munyua, B. G., Mausch, K., & Michler, J. D. (2017). Welfare impacts of improved chickpea adoption: A pathway for rural development in Ethiopia? *Food Policy*, *66*, 50–61. https://doi.org/10.1016/j.foodpol.2016.11.007

Villanueva, D., Smale, M., Jamora, N., Capilit, G. L., & Hamilton, R. S. (2020). The contribution of the International Rice Genebank to varietal improvement and crop productivity in Eastern India. *Food Security*, *12*(5), 929–943. https://doi.org/10.1007/s12571-020-01036-9

Walker, T. S., & Alwang, J. R. (Eds.). (2015). *Crop improvement, adoption and impact of improved varieties in food crops in Sub-Saharan Africa*. Published by CGIAR and CAB International.

Wooldridge, J. M. (2003). Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters*, *79*(2), 185–191. https://doi.org/10.1016/S0165-1765(02)00318-X

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed). MIT Press.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, *50*(2), 420–445. https://doi.org/10.3368/jhr.50.2.420

World Bank. (2014). *Purchasing power parities and the real size of world economies: A comprehensive report of the 2011 International Comparison Program*. The World Bank. https://doi.org/10.1596/978-1-4648-0329-1

World Bank. (2020). *The World Bank in Malawi* [Text/HTML]. World Bank. https://www.worldbank.org/en/country/malawi/overview

# 8 Tables

**Table 1.** Pedigree of adopted groundnut varieties in Malawi

| Variety | Pedigrees | Year | Provenance | Ancestry |
|---|---|---|---|---|
| CG 7 | (USA 20 x TMV 10) F2-P3-B1-B1-B1-B1-B1B1-B1-B1 | 1990 | 100% | 1 |
| Nsinjiro | (RG 1 x Manipintar) F2-P23-P59-P59-B1-B1-B13-B1 | 2000 | 50% | 1 |
| JL24/Kakoma | (Unknown source, improved variety) | 2000 | 0% | 0 |
| Baka | (Sourced from genebank, landrace) | 2001 | 100% | 1 |
| Chitala | (ICGV 93437 X ICGV-SM 94586) F2-P10-P4-B1-B1-B1-B1 | 2005 | 100% | 1 |
| Manipintar | (Unknown source, landrace) | 1955 | 0% | 0 |

Notes: F here refers to filial generation, B refers to the bulk selection, and P represents the progenies.

**Table 2.** Contribution of ICRISAT genebank

| | **2008** | **2010** | **2013** | **Pooled** |
|---|---|---|---|---|
| Genebank ancestry (1=Yes) | 0.41<br>(0.49) | 0.36<br>(0.48) | 0.52<br>(0.50) | 0.43<br>(0.49) |
| Relative provenance (%) | 40.26<br>(48.86) | 29.86<br>(42.28) | 44.63<br>(45.81) | 38.25<br>(46.05) |
| Weighted relative provenance (%) | 98.36<br>(8.97) | 82.40<br>(24.10) | 85.25<br>(22.94) | 88.60<br>(21.03) |
| Area under adoption (hectares) | 0.50<br>(0.61) | 0.30<br>(0.78) | 0.34<br>(0.80) | 0.38<br>(0.74) |
| Ratio of area under improved to total area (hectares) | 0.21<br>(0.23) | 0.13<br>(0.19) | 0.22<br>(0.24) | 0.18<br>(0.23) |
| Observations | 149 | 149 | 149 | 447 |

Notes: Genetic ancestry here refers to any parent of the improved variety coming from the genebank. It is treated as a dummy variable here. Relative provenance, on the other hand, is the derived contribution of the genebank based on the Mendelian rule of inheritance. Mean values are presented here with their standard deviation in parentheses.

**Table 3.** Summary statistics of socio-economic and farm characteristics

| Variable | Pooled sample | Improved variety | | Mean difference |
|---|---|---|---|---|
| | | Adopters | Non-adopters | |
| *Dependent variables* | | | | |
| Household income (US$) | 2222.86<br>(5708.63) | 2880.99<br>(593.37) | 1699.54<br>(102.35) | 1181.55*** |
| Productive assets (US$) | 132.39<br>(552.24) | 145.05<br>(36.44) | 122.37<br>(36.91) | 22.73*** |
| Non-productive assets (US$) | 729.58<br>(8801.08) | 1272.07<br>(937.45) | 298.215<br>(51.422) | 973.86 |
| Livestock ownership (TLU) | 5.71<br>(40.36) | 6.89<br>(3.85) | 4.22<br>(1.53) | 2.677** |
| Poor household (<US$1.25) | 0.69<br>(0.46) | 0.66<br>(0.03) | 0.72<br>(0.02) | -0.06* |
| Poor household (<US$2.00) | 0.84<br>(0.36) | 0.81<br>(0.02) | 0.86<br>(0.02) | -0.04* |
| *Key explanatory variable* | | | | |
| Relative provenance (%) | 38.26<br>(46.04) | 86.38<br>(1.77) | 0.00<br>(0.00) | 86.38*** |
| Area under adoption (hectares) | 0.38<br>(0.74) | 0.70<br>(0.06) | 0.13<br>(0.02) | 0.58*** |
| *Covariates* | | | | |
| Age of the household head (years) | 46.33<br>(16.04) | 45.84<br>(1.04) | 46.72<br>(1.08) | -0.87 |
| Educational level of the household head (years) | 6.90<br>(3.62) | 4.39<br>(0.26) | 4.14<br>(0.26) | 0.25*** |
| Household head is male (%) | 0.79<br>(0.40) | 0.82<br>(0.02) | 0.77<br>(0.02) | 0.05* |
| Household size (number) | 5.40<br>(2.14) | 5.52<br>(0.14) | 5.30<br>(0.14) | 0.21 |
| Experience in groundnut cultivation (years) | 7.75<br>(11.49) | 7.65<br>(0.81) | 7.81<br>(0.73) | -0.16 |
| Distance to market (km) | 11.54<br>(8.12) | 11.85<br>(0.56) | 11.29<br>(0.53) | 0.55 |
| Distance to extension agent (km) | 4.79<br>(4.22) | 4.73<br>(0.26) | 4.84<br>(0.28) | -0.11** |
| Farm size (hectares) | 1.04<br>(1.18) | 1.06<br>(0.08) | 1.02<br>(0.07) | 0.04 |
| Irrigation | 0.06<br>(0.25) | 0.07<br>(0.02) | 0.06<br>(0.01) | 0.01 |
| Observations | 447 | 198 | 249 | 447 |

Notes: for all the panel rounds, observations are pooled. Mean values are presented for all variables with their standard deviations in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 4.** Tobit estimates of the relationship between relative provenance and adoption

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Relative provenance | 0.013*** | 0.013*** | 0.013*** | 0.013*** | 0.013*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Pseudo $R^2$ | 0.165 | 0.187 | 0.197 | 0.197 | 0.229 |
| Additional controls | No | No | No | Yes | Yes |
| District dummies | No | Yes | Yes | No | Yes |
| Time dummies | No | No | Yes | No | Yes |
| Observations | 447 | 447 | 447 | 447 | 447 |

Notes: All columns report the relationship between relative provenance and the area under improved groundnuts. Additional controls include the age of the household head, educational level of the household head, household size, sex of the household head, soil characteristics, irrigation access, land ownership, distance to extension agent, and walking distance to the village market. Robust standard errors are in parentheses. The Mundlak-Chamberlain, CRE models include the averages of time-varying, which we do not show here for brevity. All coefficient estimates are reported as average partial effects (APE) obtained by the margins function in STATA *** p<0.01, ** p<0.05, * p<0.1

**Table 5.** Estimates of the relationship between any genebank ancestry and adoption

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Genebank ancestry | 1.380*** | 1.387*** | 1.426*** | 1.370*** | 1.424*** |
| | (0.149) | (0.147) | (0.150) | (0.145) | (0.146) |
| Pseudo $R^2$ | 0.208 | 0.224 | 0.241 | 0.239 | 0.269 |
| Additional controls | No | No | No | Yes | Yes |
| District dummies | No | Yes | Yes | No | Yes |
| Time dummies | No | No | Yes | No | Yes |
| Observations | 447 | 447 | 447 | 447 | 447 |

Notes: All columns report the relationship between genebank ancestry and the area under improved groundnuts. Additional controls include the age of the household head, educational level of the household head, household size, sex of the household head, soil characteristics, irrigation access, land ownership, distance to extension agent, and walking distance to the village market. Robust standard errors are in parentheses. The Mundlak Chamberlain, CRE models include the averages of time-varying, which we do not show here for brevity. All coefficient estimates are reported as average partial effects (APE) obtained by the margins function in STATA *** p<0.01, ** p<0.05, * p<0.1

**Table 6.** Income and asset impacts of groundnut adoption

| | (1) Income | (2) Unproductive assets | (3) Productive assets | (4) Livestock ownership |
|---|---|---|---|---|
| Area under adoption | 766.509** | 118.750*** | 115.827*** | 21.111*** |
| | (198.853) | (45.448) | (37.257) | (3.589) |
| Prob > chi$^2$ | 0.000 | 0.000 | 0.000 | 0.000 |
| R squared | 0.138 | 0.247 | 0.208 | 0.223 |
| Additional controls | Yes | Yes | Yes | Yes |
| District dummies | Yes | Yes | Yes | Yes |
| Time dummies | Yes | Yes | Yes | Yes |
| Observations | 447 | 447 | 447 | 447 |

Notes: All columns report the impact of improved groundnut adoption on the income and asset holdings of households. All models are estimated with fixed effect instrumental variables (FE_IV). Additional controls include the age of the household head, educational level of the household head, household size, sex of the household head, land ownership, experience in groundnut cultivation, and walking distance to the village market. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 7.** Poverty impacts of groundnut adoption

| | (1) Poverty (IV) | (2) Poverty | (3) Poverty (IV) | (4) Poverty |
|---|---|---|---|---|
| Area under adoption | -0.072** | -0.044 | -0.117*** | -0.082*** |
| | (0.032) | (0.028) | (0.030) | (0.027) |
| Wald chi2 | 51.30 | 48.31 | 21.13 | 24.62 |
| Additional controls | Yes | Yes | Yes | Yes |
| District dummies | Yes | Yes | Yes | Yes |
| Time dummies | Yes | No | Yes | No |
| Observations | 447 | 447 | 447 | 447 |

Notes: All columns report the impact of improved groundnut adoption on income poverty. In columns (1) and (3), we employ IV regressions. All models are estimated with the Mundlak-Chamberlain device. The Mundlak-Chamberlain, CRE models include the averages of time-varying, which we do not show here for brevity. All coefficient estimates are reported as average partial effects (APE) obtained by the function of the margin in STATA. Additional controls include the age of the household head, educational level of the household head, household size, sex of the household head, land ownership, experience in groundnut cultivation, and walking distance to the village market. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 8.** Estimates from the chain rule

| **Outcomes** | **Adoption impacts** | **Genebank ancestry** |
|---|---|---|
| Income | 766.509*** | 9.9646*** |
| Unproductive assets | 118.750*** | 1.5437*** |
| Productive assets | 115.827*** | 1.5057*** |
| Livestock ownership | 21.111*** | 0.2744*** |
| Poverty | -0.072** | -0.0009** |
| Medium poverty | -0.117*** | -0.0015*** |

Notes: *** p<0.01, ** p<0.05, * p<0.1

Table 9. LPM estimates of the relationship between provenance and adoption

| | **(1)** **Adoption (LPM)** | **(2)** **Adoption (LPM)** | **(3)** **Adoption (LASSO)** | **(4)** **Adoption (LASSO)** |
|---|---|---|---|---|
| Relative provenance | 0.010*** | 0.010*** | 0.010*** | 0.010*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| F test | 2965.58*** | 278.01*** | | |
| Pseudo $R^2$ | 0.870 | 0.875 | | |
| Prob ($chi^2$) | | | 0.000 | 0.000 |
| Additional controls | No | Yes | No | Yes |
| District dummies | No | Yes | No | Yes |
| Time dummies | No | Yes | No | Yes |
| Observations | 447 | 447 | 447 | 447 |

Notes: Columns (1) and (2) report the relationship between relative and adoption. Here adoption is treated as a dummy and estimated with the linear probability model (LPM) with the household fixed effect estimator. In models (3) and (4), we report the LASSO regressions which are also quite similar to the LPM model. Additional controls include the age of the household head, educational level of the household head, household size, sex of the household head, soil characteristics, irrigation access, land ownership, distance to extension agent, and walking distance to the village market. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

# 9 Figures

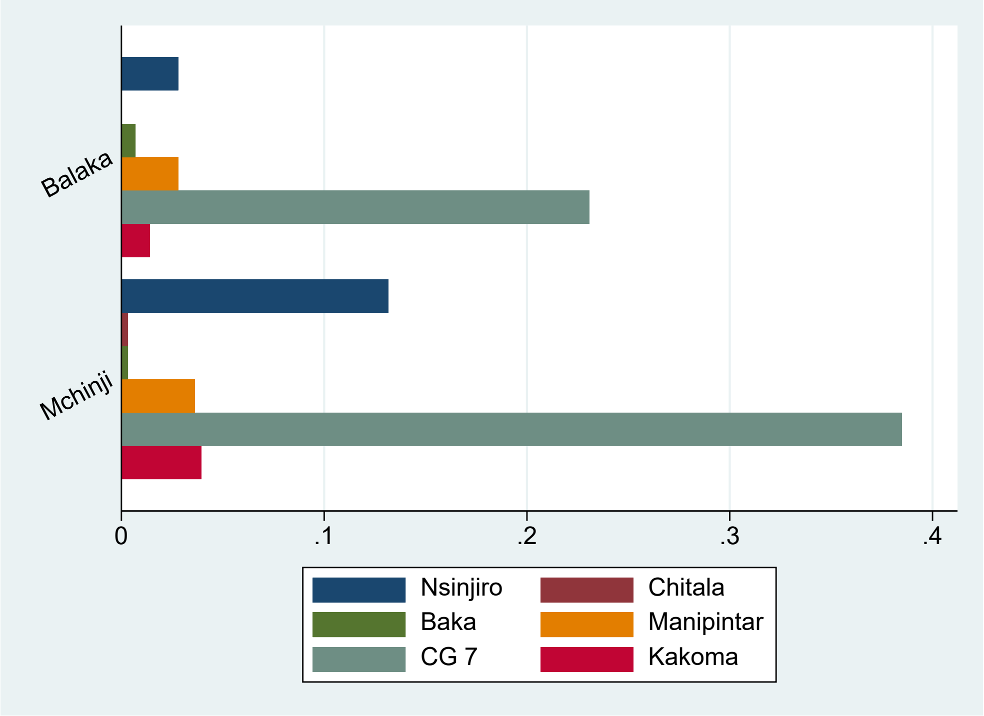

**Figure 1.** Percentage distribution of groundnut area by improved variety and district

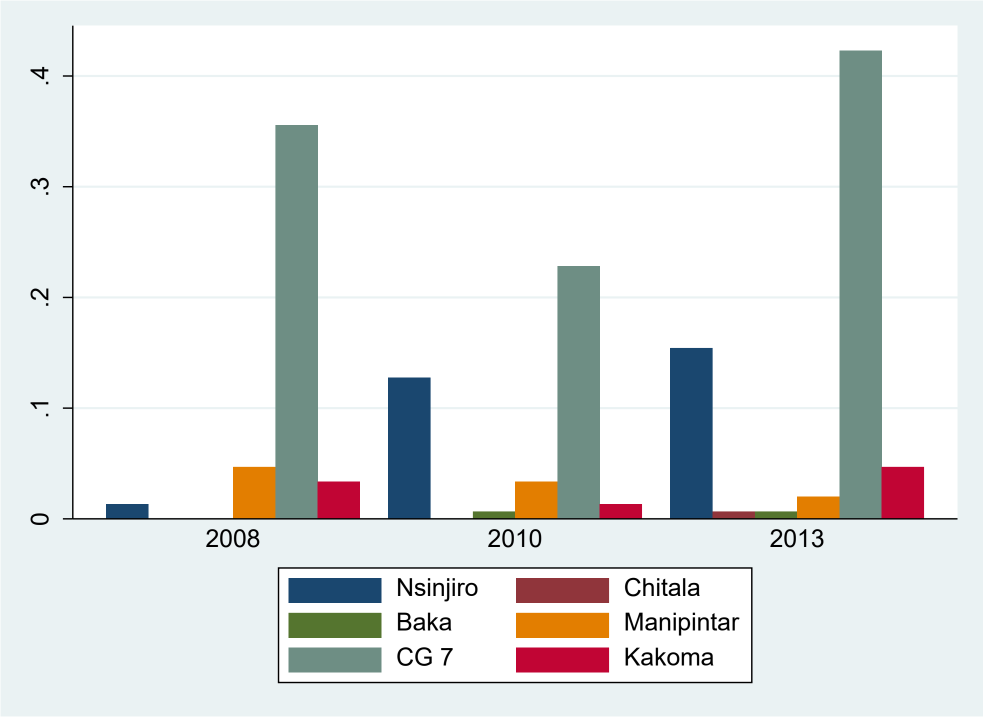

**Figure 2.** Percentage distribution of groundnut area by improved variety and survey year

# 10 Appendix

## 10.1 Varietal information

1. **CG 7**

CG 7, also known as ICGMS 42 or ICGV-SM 83708, is a high-yielding Virginia bunch variety released in 1990, jointly developed by ICRISAT and the Department of Agricultural Research and Technical Services (DARTS). Recommended for cultivation in all groundnut-growing areas of Malawi, it is suitable for confectionery use and oil extraction (Subrahmanyam et al. 2000). It is more tolerant of drought and much easier to harvest than Chalimbana, currently the most widely grown variety in Malawi. Potential seed yields of CG 7 can reach 2 t/ha. Apart from being a high-yielding medium-duration variety, it is resistant to the groundnut rosette virus. It is also well adapted to the central plateau of Lilongwe and Kasungu. It is uniform in size, red, and blanches easily. It is usually described as not early maturing as it can take about 150 days to mature.

2. **Nsinjiro**

Nsinjiro released as ICGV-SM 90704 is high-yielding medium-duration groundnut germplasm that was developed at ICRISAT Malawi. In collaboration with the National Agriculture Research Systems (NARS), it was evaluated in the Eastern and Southern Africa (ESA) region. It was then released in 1999 in Uganda as serenut 2 and a year later in Malawi as ICGV-SM 90704. It was also later released in Zambia. In the ESA region, it has been widely used as it is resistant to the rosette virus, although susceptible to the aphid vector. It results from a cross between RG1 and Manipintar. It was developed following a series of bulk selections for rosette disease reaction using the inferior row technique (Freeman et al. 2002). It is very high yielding with an average seed yield of 1.04 tonnes/ha as compared to 0.52 tonnes/ha and 0.84 tonnes/ha for Chalimbana and CG7 respectively. It has a low resistance to the rosette virus of 2% as against 81% and 83% for Chalimbana and CG7 respectively. It shells at about 67% and it has readily available seeds at ICRISAT Malawi.

3. **Kakoma**

JL 24 (Phule Pragati), a pure line selected variety from the exotic germplasm ‘EC 94943’, has been released for commercial cultivation. It became a national variety due to its wide adaptability and superior yields and is still popular among farmers across India (Ahire and Khalache 2007). JL 24 was developed in Jalgaon, India and released for commercial cultivation in Maharashtra and Gujarat. Jalgaon is one of the important oilseed research stations in Maharashtra (Ingale and Shrivastava 2011). It was released for cultivation in 1979. Kakoma is very susceptible to diseases like rosette and aflatoxin. It is also an early-maturing variety taking about 90–120 days to mature. It is usually advisable to grow Kakoma when the rains taper off early.

### 4. Baka

ICG 12991 is a short duration (90–110 days to maturation), drought-tolerant, Spanish-type peanut with field resistance to groundnut rosette disease. ICG 12991 was originally collected from a farmer's field in south India in 1988. In 1994, ICRISAT introduced ICG 12991 into Malawi for evaluation during a germplasm screening program for resistance to groundnut rosette disease and early leaf spot disease. ICG 12991 was released in Malawi as 'Baka' in 2001. Baka can be referred to as a groundnut landrace from India which was released in sub-Saharan Africa. It is high yielding and very resistant to the groundnut rosette virus. It branches sequentially with about 4.5 and 2.5 primary and secondary branches, respectively. It has moderate oil and protein content and a shelling percentage of about 75%. It is also highly resistant to aphids.

### 5. Chitala

The ICGV-SM 99568 (Chitala) is a short-duration (100-110 day), medium seed size with good tolerance to rosette disease. It is a popular groundnut variety that was released in Malawi in 2005. It is extensively used as a parental line in the development of high oleate groundnut. ICGV-SM 99568 is a short-duration rosette-resistant variety. It takes about 90–105 to mature and sow. Its seed coat has a tan color with a 100 seed mass of 40 g and 46% oil content. It has no fresh seed dormancy. It is moderately resistant to groundnut rosette disease (Deom et al. 2006).

### 6. Manipintar

Manipintar (or Mani Pintar) is a long-duration groundnut variety with both white and red variegated seed color, reported to have originated from the Bolivian strain of groundnuts (Smartt 1978). It was obtained from the Queensland Department of Agriculture and Stock, Australia, in the early part of 1955 (Smartt 1960). It was further developed by the Department of Research and Specialized services in Zambia. It is one of the parents from which Nsinjiro was bred. One of its low points is its susceptibility to the rosette virus. Apart from that, it is a high-yielding variety (McEwen 1961). It is generally large (length, width, and thickness) as compared to other varieties. It is late maturing (140–150 days) and fairly resistant to *Cercospora* leafspots. It produces oil with high kernel content (Smartt 1960). It remains one of the varieties that has been extensively used for academic purposes and it adapts quite well in local conditions.

## 7. References

Ahire, M. C., & Khalache, P. G. (2007). *Knowledge level of groundnut growers (Phule Pragati) from Hatkangale Tahsil of Kolhapur, India*. *3*(2), 286–289.

Deom, C. M., Kapewa, T., Busolo-Bulafu, C. M., Naidu, R. A., Chiyembekeza, A. J., Kimmins, F. M., Subrahmanyam, P., & Merwe, P. J. A. (2006). Registration of ICG 12991 Peanut Germplasm Line. *Crop Science*, *46*(1), 481–481. https://doi.org/10.2135/cropsci2005.010

Freeman, H. A., van der Merwe, P. J. A., Subrahmanyam, P., Chiyembekeza, A. J., & Kaguongo, W. (2002). Assessing adoption potential of new groundnut varieties in Malawi. *Experimental Agriculture*, *38*(2), 211–221. https://doi.org/10.1017/S0014479702000261

Ingale, S., & Shrivastava, S. K. (2011). In vivo study of nutritive value of oil seeds from new varieties of sunflower, safflower and groundnut. *IIOAB Letters*, *1*(1). https://doi.org/10.5195/IIOABLETT.2011.2

McEwen, J. (1961). Groundnut variety, Mani pintar. *Nature*, *192*(4797), 92–92. https://doi.org/10.1038/192092b0

Smartt, J. (1960). Genetic instability and outcrossing in the groundnut variety Mani Pintar. *Nature*, *186*(4730), 1070–1071. https://doi.org/10.1038/1861070a0

Smartt, J. (1978). Makulu red—A 'Green Revolution' groundnut variety? *Euphytica*, *27*(2), 605–608. https://doi.org/10.1007/BF00043190

Subrahmanyam, P., Van der Merwe, P. J. A., Chiyembekeza, A. J., & Freeman, H. A. (2000). *Groundnut variety CG 7: a boost to Malawian Agriculture*. *20*, 33–35.

**Table A1.** Groundnut varieties released by ICRISAT in Malawi

| ICRISAT Name | Release Name | Year |
|---|---|---|
| ICGV-SM 83708 (ICGMS 42) | CG 7 | 1990 |
| ICGV-SM 90704 | Nsinjiro | 2000 |
| JL 24 | Kakoma | 2000 |
| ICG 12991 | Baka | 2001 |
| ICGV SM 99568 | Chitala | 2005 |
| ICGV-SM 08501 | CG8 | 2014 |
| ICGV-SM 8503 | CG9 | 2014 |
| ICGV-SM 01731 | CG10 | 2014 |
| ICGV-SM 01724 | CG11 | 2014 |
| ICGV-SM 01514 | CG12 | 2014 |
| ICGV-SM 99551 | CG13 | 2014 |
| ICGV-SM 99556 | CG14 | 2014 |

**Table A2**. Source of parents

| SN | Name of Accessions | ICG No. |
|---|---|---|
| 1. | USA 20 | ICG 983 |
| 2. | RG 1 | ICG 12938 |
| 3. | TMV 10 | ICG 618 |
| 4. | JL 24 | ICG 7827 |
| 5. | Manipintar | Not Available |

**Table A3.** Falsification test for income and assets

|  | (1) Income | (2) Unproductive assets | (3) Productive assets | (4) Livestock ownership |
|---|---|---|---|---|
| Sandy soil (1=Yes) | 984.026 | 1889.74 | 145.025 | -7.290 |
|  | (2885.38) | (4521.93) | (298.487) | (26.466) |
| Irrigation (1=Yes) | 2104.68 | 6129.94 | 61.181 | -8.310 |
|  | (1081.04) | (1694.19) | (129.963) | (11.528) |
| Distance to extension agent (km) | -104.085 | -94.886 | 0.403 | 0.219 |
|  | (65.836) | (103.178) | (9.216) | (0.817) |
| Prob > chi$^2$ | 0.000 | 0.000 | 0.000 | 0.000 |
| R squared | 0.208 | 0.02 | 0.102 | 0.221 |
| Additional controls | Yes | Yes | Yes | Yes |
| District dummies | Yes | Yes | Yes | Yes |
| Time dummies | Yes | Yes | Yes | Yes |
| Observations | 447 | 447 | 237 | 237 |

**Table A4.** Falsification test for income poverty

| | (1)<br>Income Poverty ($1.25) | (2)<br>Income Poverty ($2.00) |
|---|---|---|
| Sandy soil (1=Yes) | -0.242 | 0.356 |
| | (0.222) | (0.329) |
| Irrigation (1=Yes) | 0.020 | 0.433 |
| | (0.075) | (0.485) |
| Distance to extension agent (km) | -0.001 | 0.045 |
| | (0.005) | (0.040) |
| Prob > chi$^2$ | 0.000 | 0.000 |
| R squared | 0.21 | 0.31 |
| Additional controls | Yes | Yes |
| District dummies | Yes | Yes |
| Time dummies | Yes | Yes |
| Observations | 447 | 237 |

**Table A5**. Tobit estimates of the relationship between relative provenance and adoption (full results)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Relative provenance | 0.013***<br>(0.001) | 0.013***<br>(0.001) | 0.013***<br>(0.001) | 0.013***<br>(0.001) | 0.013***<br>(0.001) |
| Age of the household head (years) | | | | -0.005<br>(0.005) | 0.001<br>(0.005) |
| Educational level (years) | | | | 0.025<br>(0.023) | 0.007<br>(0.019) |
| Household head is male (%) | | | | 0.180<br>(0.285) | 0.169<br>(0.281) |
| Household size (number) | | | | -0.006<br>(0.047) | 0.006<br>(0.046) |
| Experience (years) | | | | -0.003<br>(0.006) | -0.008<br>(0.007) |
| Distance to market (km) | | | | -0.019**<br>(0.008) | -0.014<br>(0.008) |
| Distance to extension agent (km) | | | | 0.017<br>(0.018) | 0.007<br>(0.019) |
| Farm size (hectares) | | | | -0.101*<br>(0.061) | -0.101*<br>(0.059) |
| Irrigation | | | | -0.038<br>(0.265) | 0.148<br>(0.261) |
| Sandy soil | | | | -0.548<br>(0.697) | -0.670<br>(0.677) |
| Constant | -0.930***<br>(0.125) | -1.306***<br>(0.158) | -1.070***<br>(0.168) | -0.107<br>(1.057) | -1.354<br>(1.010) |
| Mchinji | | 0.640***<br>(0.136) | 0.657***<br>(0.136) | | 0.740***<br>(0.151) |
| 2010 | | | -0.393***<br>(0.139) | | -0.373***<br>(0.143) |
| 2013 | | | -0.365***<br>(0.134) | | -0.441***<br>(0.168) |
| Pseudo $R^2$ | 0.165 | 0.187 | 0.197 | 0.197 | 0.229 |
| Observations | 447 | 447 | 447 | 447 | 447 |

Notes: All columns report the relationship between relative provenance and the area under improved groundnuts. Robust standard errors are in parentheses. The Mundlak-Chamberlain, CRE models include the averages of time-varying, which we do not show here for brevity. All coefficient estimates are reported as average partial effects (APE) obtained by the margins function in STATA *** p<0.01, ** p<0.05, * p<0.1

**Table A6.** Estimates of the relationship between any genebank ancestry and adoption (full results)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Genebank ancestry | 1.380***<br>(0.149) | 1.387***<br>(0.147) | 1.426***<br>(0.150) | 1.370***<br>(0.145) | 1.424***<br>(0.146) |
| Age of the household head (years) | | | | -0.006<br>(0.005) | 0.001<br>(0.005) |
| Educational level (years) | | | | 0.030<br>(0.022) | 0.027<br>(0.022) |
| Household head is male (%) | | | | 0.204<br>(0.281) | 0.197<br>(0.277) |
| Household size (number) | | | | -0.005<br>(0.047) | 0.010<br>(0.045) |
| Experience years) | | | | -0.001<br>(0.005) | -0.008<br>(0.007) |
| Distance to market (km) | | | | -0.018**<br>(0.008) | -0.012<br>(0.008) |
| Distance to extension agent (km) | | | | 0.023<br>(0.018) | 0.008<br>(0.019) |
| Farm size (hectares) | | | | -0.116**<br>(0.060) | -0.119**<br>(0.059) |
| Irrigation | | | | 0.002<br>(0.259) | 0.246<br>(0.258) |
| Sandy soil | | | | -0.748<br>(0.680) | -0.910<br>(0.664) |
| Constant | -1.131***<br>(0.132) | -1.438***<br>(0.163) | -1.158***<br>(0.169) | -0.096<br>(1.036) | -1.039<br>(1.053) |
| Mchinji | | 0.543***<br>(0.135) | 0.571***<br>(0.136) | | 0.603***<br>(0.151) |
| 2010 | | | -0.492***<br>(0.138) | | -0.484***<br>(0.142) |
| 2013 | | | -0.463***<br>(0.133) | | -0.554<br>(0.168) |
| Pseudo $R^2$ | 0.208 | 0.224 | 0.241 | 0.239 | 0.269 |
| Observations | 447 | 447 | 447 | 447 | 447 |

Notes: All columns report the relationship between genebank ancestry and the area under improved groundnuts. Robust standard errors are in parentheses. The Mundlak-Chamberlain, CRE models include the averages of time-varying, which we do not show here for brevity. All coefficient estimates are reported as average partial effects (APE) obtained by the margins function in STATA *** p<0.01, ** p<0.05, * p<0.1

**Table A7.** Income and asset impacts of groundnut adoption (full results)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | **Income** | **Unproductive assets** | **Productive assets** | **Livestock ownership** |
| Area under adoption | 766.509*** | 118.750*** | 115.827*** | 21.111*** |
| | (198.852) | (48.361) | (37.257) | (3.589) |
| Age of the household head (years) | -3.699 | -2.949 | -2.076 | 0.184 |
| | (12.845) | (3.119) | (2.557) | (0.246) |
| Educational level (years) | -21.418 | 1.172 | 6.561 | -1.272 |
| | (52.232) | (12.635) | (10.358) | (0.997) |
| Household head is male (%) | 293.874 | -0.499 | 14.349 | -0.121 |
| | (571.816) | (137.675) | (112.861) | (10.872) |
| Household size (number) | 157.473* | 68.860*** | 37.557** | 0.417 |
| | (92.335) | (22.278) | (18.263) | (1.759) |
| Experience (years) | 8.484 | 2.581 | 0.460 | 0.117 |
| | (15.316) | (3.720) | (3.049) | (0.293) |
| Distance to market (km) | -11.423 | 1.214 | 5.097 | -0.497* |
| | (19.401) | (4.654) | (3.815) | (0.367) |
| Farm size (hectares) | 153.554 | 39.913 | 16.321 | 1.952 |
| | (179.774) | (43.319) | (35.511) | (3.420) |
| Constant | 1521.887 | 273.099 | -217.991 | -3.470 |
| | (1052.248) | (249.978) | (204.923) | (19.740) |
| Mchinji | 453.564 | -221.226 | -113.057 | -11.294 |
| | (779.085) | (188.219) | (154.295) | (14.863) |
| 2010 | -222.655 | 118.974 | -1.784 | 12.039 |
| | (449.387) | (106.745) | (87.505) | (8.429) |
| 2013 | -332.053 | 143.506 | 63.261 | 9.636 |
| | (781.783) | (186.522) | (152.904) | (14.729) |
| R squared | 0.138 | 0.247 | 0.208 | 0.223 |
| Observations | 447 | 447 | 447 | 447 |

Notes: All columns report the impact of improved groundnut adoption on the income and asset holdings of households. All models are estimated with fixed effect instrumental variables (FE_IV). Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table A8.** Poverty impacts of groundnut adoption (full results)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | **Poverty (IV)** | **Poverty** | **Poverty (IV)** | **Poverty** |
| Area under adoption | -0.064** | -0.044 | -0.117*** | -0.082*** |
| | (0.032) | (0.028) | (0.030) | (0.027) |
| Age of the household head (years) | -0.001 | -0.001 | 0.001 | 0.008 |
| | (0.001) | (0.001) | (0.001) | (0.019) |
| Educational level (years) | 0.004 | 0.004 | -0.006 | -0.006 |
| | (0.008) | (0.008) | (0.007) | (0.007) |
| Household head is male (%) | -0.111 | -0.095 | -0.146 | -0.124 |
| | (0.106) | (0.106) | (0.107) | (0.108) |
| Household size (number) | 0.051*** | 0.051*** | 0.036** | 0.035 |
| | (0.016) | (0.016) | (0.016) | (0.017) |
| Experience years) | -0.003 | -0.003 | 0.003 | 0.003 |
| | (0.002) | (0.002) | (0.003) | (0.003) |
| Distance to market (km) | 0.006 | -0.001 | 0.008*** | 0.007** |
| | (0.029) | (0.002) | (0.003) | (0.003) |
| Farm size (hectares) | -0.008 | -0.014 | 0.024 | 0.016 |
| | (0.022) | (0.022) | (0.017) | (0.017) |
| Mchinji | -0.213*** | -0.179*** | -0.868 | -0.337 |
| | (0.048) | (0.044) | (0.536) | (0.421) |
| 2010 | 0.055 | 0.038 | -0.533 | -0.677* |
| | (0.042) | (0.040) | (0.379) | (0.360) |
| 2013 | -0.279*** | -0.284*** | -1.701*** | -1.555*** |
| | (0.060) | (0.059) | (0.581) | (0.502) |
| Wald chi$^2$ | 50.77 | 48.31 | 21.68 | 24.62 |
| Observations | 447 | 447 | 447 | 447 |

Notes: All columns report the impact of improved groundnut adoption on income poverty. In columns (1) and (3), we employ IV regressions. All models are estimated with the Mundlak-Chamberlain device. The Mundlak-Chamberlain, CRE models include the averages of time-varying, which we do not show here for brevity. All coefficient estimates are reported as average partial effects (APE) obtained by the function of the margin in STATA. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table A9.** Endogeneity tests

| | **Income** | **Unproductive assets** | **Productive assets** | **Livestock ownership** | **Poverty (\$1.25)** | **Poverty (\$2.00)** |
|---|---|---|---|---|---|---|
| Residual | 1072.871*** | 84.891 | 61.789 | 12.522 | -0.421* | -1.0.36** |
| | (403.522 | (98.075) | (80.432) | (7.706 | (0.215) | (0.443) |

Notes: Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

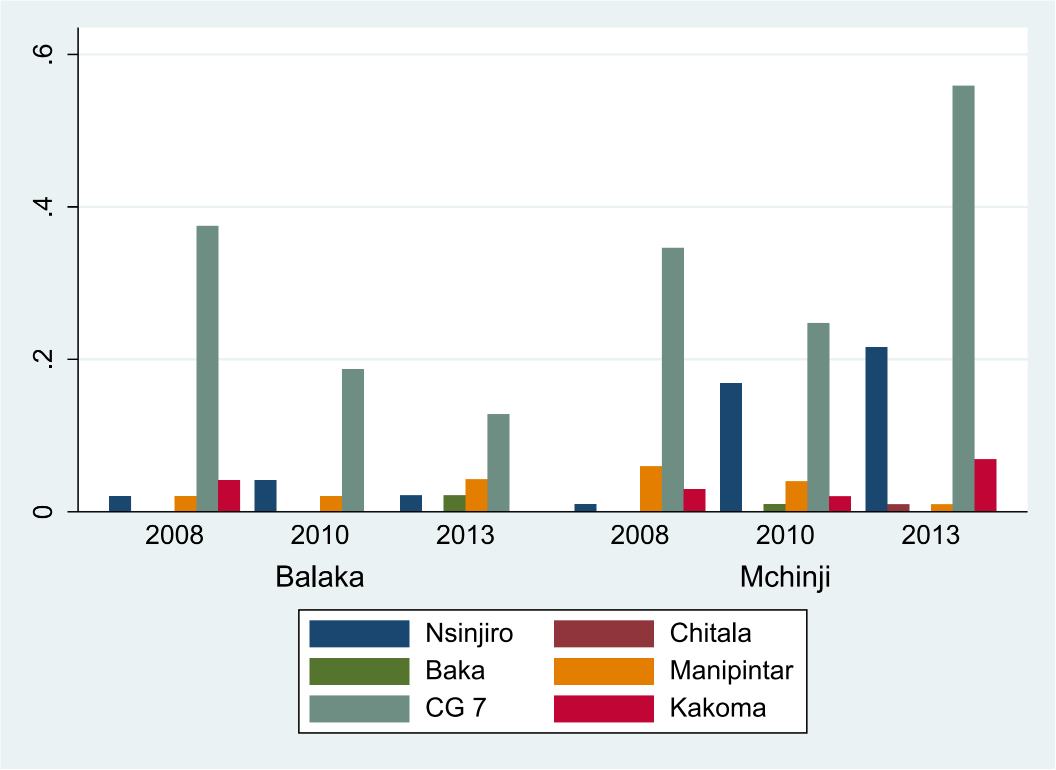

**Figure A1.** Adoption of improved groundnut varieties over time, by districts

## Author bios and pictures

Martin Paul Jr. Tabe-Ojong a doctoral researcher with the Institute for Food and Resource Economics at the University of Bonn and currently a consultant with the Social Protection and Jobs Global Practice at the World Bank. He is currently completing his PhD in Agricultural Economics and holds an MSc in Agricultural and Food Economics from the University of Bonn, Germany, and a BSc in Agricultural Economics from the University of Buea, Cameroon. Passionate about international development, poverty reduction and shared prosperity, his research focuses on selected strands in development, agricultural and behavioral economics using econometric impact analysis. Specifically, he does empirical research on aspirations and economic mobility, technology adoption, social protection and welfare as well as productivity and efficiency analyses. He is highly skilled at data collection through household surveys with field work and research stays in Cameroon, Kenya, Tanzania, Ethiopia, Ghana, and Cote d'Ivoire.

## Author contributions

The first author contributed to the research conceptualization and design, data gathering and cleaning, data analysis, writing, and editing. The second, and third authors contributed to the research conceptualization and design, writing, editing, supervision, and funding acquisition. The fourth, fifth and last authors contributed to pedigree data provision, writing, and editing.