Feed the Future Innovation Lab for Food Security Policy

Research Paper 72 October 2017

Nigeria Agricultural Policy Project

# THE ROLE OF THE LOCATIONS OF PUBLIC SECTOR VARIETAL DEVELOPMENT ACTIVITIES ON AGRICULTURAL PRODUCTIVITY: EVIDENCE FROM NORTHERN NIGERIA

By

Hiroyuki Takeshima and Abdullahi Mohammed Nasir

FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

USAID
FROM THE AMERICAN PEOPLE

MICHIGAN STATE UNIVERSITY

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

FMARD
FEDERAL MINISTRY OF AGRICULTURE AND RURAL DEVELOPMENT

**Food Security Policy *Research Papers***

This *Research Paper* series is designed to timely disseminate research and policy analytical outputs generated by the USAID funded Feed the Future Innovation Lab for Food Security Policy (FSP) and its Associate Awards. The FSP project is managed by the Food Security Group (FSG) of the Department of Agricultural, Food, and Resource Economics (AFRE) at Michigan State University (MSU), and implemented in partnership with the International Food Policy Research Institute (IFPRI) and the University of Pretoria (UP). Together, the MSU-IFPRI-UP consortium works with governments, researchers and private sector stakeholders in Feed the Future focus countries in Africa and Asia to increase agricultural productivity, improve dietary diversity and build greater resilience to challenges like climate change that affect livelihoods.

The papers are aimed at researchers, policy makers, donor agencies, educators, and international development practitioners. Selected papers will be translated into French, Portuguese, or other languages.

Copies of all FSP Research Papers and Policy Briefs are freely downloadable in pdf format from the following Web site: http://foodsecuritypolicy.msu.edu/

Copies of all FSP papers and briefs are also submitted to the USAID Development Experience Clearing House (DEC) at: http://dec.usaid.gov/

## AUTHORS

**Hiroyuki Takeshima** is a Research Fellow at the International Food Policy Research Institute (IFPRI), based in Washington, DC.

**Abdullahi Mohammed Nasir** is a Senior Technical Officer at the Agricultural Research Council of Nigeria, based in Abuja.

## AUTHORS' ACKNOWLEDGMENT

This paper has been prepared as part of the Feed the Future Nigeria Agricultural Policy Project through the Nigeria Strategy Support Program (NSSP) managed by the International Food Policy Research Institute (IFPRI). The research presented here was conducted as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM) which is led by IFPRI. The authors thank Hiroshi Kodama, Oladere Bakare, Johnson Onyibe, Keijiro Otsuka, and participants at the Policy Dialogue held in Abuja in April 2017 for constructive comments on the background issues for this paper.

This Research Paper was prepared for USAID/Nigeria by Michigan State University (MSU), Federal Ministry of Agriculture and Rural Development (Nigeria), and the International Food Policy Research Institute (IFPRI) under the USAID/Nigeria funded Food Security Policy Innovation Lab Associate Award, contract number AID-620-LA-15-00001.

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID). This publication has not been independently peer reviewed and the contents are the responsibility of Michigan State University and the International Food Policy Research Institute. Any opinions expressed here belong to the authors and do not necessarily reflect those of MSU, IFPRI, PIM, CGIAR, USAID or the United States Government.

This Food Security Policy Research Paper has also been published as Nigeria Strategic Support Program Working Paper No. 42 in July 2017.

*This study is made possible by the generous support of the American people through the United States Agency for International Development (USAID) under the Feed the Future initiative. The contents are the responsibility of the study authors and do not necessarily reflect the views of USAID or the United States Government*



**Published by the Department of Agricultural, Food, and Resource Economics, Michigan State University, Justin S. Morrill Hall of Agriculture, 446 West Circle Dr., Room 202, East Lansing,**

**ABSTRACT**
Despite the importance of location-specific adaptive crop breeding research, past reforms of breeding systems in Nigeria have focused more on centralizing the breeding activities into fewer locations. This has been based partly on the premise that such research systems can still effectively meet the need for a diverse set of varietal technologies that are suitable for different agroecological conditions through the use of numerous outstations and multilocational trials, regardless of the locations of the headquarters or the outstations where breeders are located. However, little empirical evidence exists to support this premise. Using panel data for agricultural households in northern Nigeria, as well as spatial data on agroecological factors, this study fills this knowledge gap. Specifically, it empirically shows that agricultural productivity and technical efficiency at farm household level is significantly and positively affected by similarity between the agroecological conditions of the locations of these households and where major crop breeding institutes are headquartered in Nigeria, namely Maiduguri, Kano, Zaria, Badeggi, Ibadan, and Umudike, after controlling for the agroecological conditions and various relevant household characteristics of these households. These findings suggest that where improved varieties are developed or evaluated affects agricultural productivity and technical efficiency in different locations. Overall agricultural productivity in Nigeria can be significantly increased not simply by increasing support for public sector varietal development, but by doing so in a manner that increases the similarity in agroecological conditions between areas where crop breeding is conducted and the areas where farm households produce those crops.

**Keywords**: varietal technologies, crop breeding institutes, agroecological similarity, agricultural productivity, Generalized Propensity Score-Inverse Probability Weighting, northern Nigeria

## TABLE OF CONTENTS

## LIST OF TABLES



## LIST OF FIGURES

# 1. BACKGROUND

Agricultural productivity growth has been an important contributor to poverty reduction, improvement of food security, and overall economic development around the world, including Nigeria. Improved agricultural production technologies, including improved crop varieties, are important tools to raise agricultural productivity. One of the unique characteristics of many agricultural production technologies, particularly varietal technologies, is their location-specificity. The performance of particular varieties can vary considerably across space depending on climate and soil conditions. The similarity of agroecological conditions between the location where the improved technologies are developed and the areas where the technologies are used by farmers (agroecological similarity hereafter), may shorten "technological distance" by providing technologies that are suitable for the particular locations (Griliches 1991; Evenson & Westphal 1995), raising the performance of these technologies and enhancing agricultural productivity in these locations. The importance of such similarity and its effect on spillover potentials is recognized broadly in the literature, particularly in developed countries (Brennan et al. 1997; Alston 2002; Alston et al. 2010).

In our analysis, an indicator of agroecological similarity is constructed for each household based on the locations of these households and the locations of crop breeding stations (headquarters of major crop breeding institutes and their outstations). The technical definitions of agroecological similarity are provided in section 4. Essentially, agroecological similarity can be defined as the similarity between locations of interests – where a particular household is located and where crop breeding stations are located – in terms of climate, soil, and topographic conditions. An indicator of agroecological similarity for a particular household is constructed as a particular function of the similarity of each climate, soil, and topographic variable – such as rainfall, soil organic matter, or slope – between the location of that household and the location or locations relevant for the household, i.e., the location(s) of each crop breeding station. The agroecological similarity between each pair of locations is calculated as a relative value, i.e., relative to the average similarity of each of these variables between all pairs of locations in northern Nigeria. This way, similarities of different agroecological variables become comparable across each agroecological variable and across each pair of locations. From these, an indicator of agroecological similarity for a particular household can be constructed as an aggregate using a particular function, as described in section 4. The indicator of agroecological similarity thus reflects similarity over multiple dimensions and pairs of locations. For example, even if rainfall levels are similar between the location where a particular household is and locations where crop breeding stations are, if soil conditions are very dissimilar between these locations, the calculated agroecological similarity for this household may be only moderate. Similarly, even if rainfall levels are similar between a pair of locations, if they are very dissimilar to rainfall conditions at other locations, the agroecological similarity for this household may be only moderate.

Whether agroecological similarity is a significant determinant of agricultural productivity has not been widely tested empirically in developing countries, including countries in Africa south of the Sahara (SSA), such as Nigeria. Investigating the effect of agroecological similarity on the performance of improved crop variety technologies is particularly important because the public sector has historically led the advancement of the development of improved crop varieties, in contrast to other improved agricultural technologies for which the private sector has often played the leading role (Evenson & Gollin 2003; Walker & Alwang 2015). The resulting distributions of agroecological similarity often remain unchanged if left to the private sector initiatives alone. These distributions change only with public sector interventions. In addition, overall agricultural research and development support generally has remained low in SSA countries. Attempts to raise efficiency in agricultural research and development activities have included centralization of crop breeding systems, including in countries like Nigeria. However, the efficiency benefits of centralized crop breeding potentially comes at the cost of a loss in agroecological similarity for substantial parts of the country. While such centralization is based on the premises that agroecological diversity can be effectively overcome through effective research and intensive evaluations of potentially improved crop varieties at various outstations spread across the country, little empirical evidence exists to support this premise.

In this paper, we partly fill this knowledge gap using the example of northern Nigeria, using the Living Standard Measurement Study-Integrated Survey on Agriculture (LSMS-ISA), a panel dataset of households in Nigeria, collected in three waves between 2010 and 2016, as well as various spatial data on agroecological conditions. We construct indicators of agroecological similarity using the locations of crop breeding institutes as well as outstations, estimate the indicators of agricultural productivity and technical efficiency at the farm household level through standard production function estimations and through simple Data Envelopment Analyses (DEA), and assess the effects of agroecological similarity on crop productivity and efficiency indicators. We do so by addressing the potential endogeneity of input variables in the production function estimations and of agroecological similarity with respect to productivity and efficiency.

Investigating the effect of agroecological similarity is potentially important for countries like Nigeria. It is a country with one of the largest areas of arable land in the world[1], with considerable heterogeneity in agroecological conditions. Yet its crop breeding activities are concentrated at only a handful of agricultural research institutes[2]. Consequently, there may be significant variation in technological distances for specific improved crop varieties across locations. In addition, over the past two decades, yields of many crops in Nigeria have stagnated at one of the lowest levels in the world. Improved design of crop breeding systems can have potentially significant effects on overall crop yield growth.

Our analyses focus on farm households in the northern part of Nigeria (specifically, the North Central, North East and the North West geopolitical zones). The crops produced in northern Nigeria are mostly annual crops so that the estimation of production function is less complicated relative to the southern part of the country where perennial crops are widely grown and the estimation of productivity is more complicated. The northern part of Nigeria accounts for more than two-thirds of the total area in Nigeria. Consequently, in assessing agroecological similarity, we include all the crop breeding institutes and outstations across the country in our analysis.

This paper contributes to various strands of the literature. First, it builds on earlier studies investigating the effects of agricultural research and development on agricultural productivity around the world (Fan & Pardey 1997; Craig et al. 1997), as well as SSA (Alene 2010; Block 2014; Benin 2016), by providing related evidence from Nigeria. Our paper also contributes to the literature on agricultural research and development in Nigeria (Beintema & Ayoola 2004; Alene et al. 2009a), by providing evidence from the angle of agroecological similarity. Second, our paper applies broad assessments of the linkages between agroecological similarity and agricultural productivity (Griliches 1991; Evenson & Westphal 1995; Brennan et al. 1997; Maredia & Byerlee 1999; Byerlee & Traxler 2001; Alston et al. 2010; Johnson et al. 2014; Bazzi et al. 2016) to the context of Nigeria. Few studies have investigated the impact of agroecological similarity of crop breeding institutes on agricultural productivity at farm household level within a particularly country. Third, our paper contributes to the literature on research spillovers in the agricultural sector (Alston 2002; Maredia et al. 1996; You & Johnson

[1] Nigeria's arable land of 34 million ha is the tenth largest in the world. Among developing countries, only India, Russia, China, Brazil and Argentina have more arable land than Nigeria.

[2] For example, rice breeding institutes in many Asian countries and in the US are more decentralized and are greater in numbers given the rice area, compared to Nigeria (Takeshima & Maji 2016 Table 5). In addition, the organization of Nigeria's crop breeding efforts stand in stark contrast to countries like Japan or China where crop breeding is organized and conducted at administrative levels as low as the prefecture, comparable to the Local Government Area (LGA) in Nigeria.

2010), by providing evidence on the nature of agroecological similarity and its effects on technology adoption and productivity growth. Lastly, our paper contributes to the literature on impact evaluation, by extending the Generalized Propensity Score-Inverse Probability Weighting (GPS-IPW) method (Imbens 2000; Flores & Mitnik 2013) to the case where, as is shown, the assumption of normality fails, whereas other distributions, like gamma distributions, are found more appropriate.

The paper proceeds in the following way. Section 2 briefly describes the crop development systems in Nigeria. Section 3 illustrates the conceptual framework. Section 4 describes the empirical methodologies. Section 5 describes the data and descriptive statistics. Section 6 discusses the empirical results. Finally, section 7 concludes the paper.

## 2. CROP VARIETAL DEVELOPMENT IN NIGERIA

Institutionalized crop breeding in Nigeria started in the early 20th century, and evolved thereafter. Varietal development in Nigeria has been primarily conducted by the public sector, particularly the National Agricultural Research Institutes (NARIs), with other institutions, including higher education institutes like universities, occasionally being involved. Since the 1970s, agricultural research systems in Nigeria have shifted toward centralization. The Agricultural Research Institutes Decree in 1973 is considered the impetus for such centralization. It provided the authority to the federal government to establish agricultural research and training institutes, and take over existing state research stations, leading to reduced incentives for states to fund agricultural research (Roseboom et al. 1994). Most NARIs in Nigeria were established under the 1975 Research Institutes (Establishment) Order. Federal control over regional universities and their agricultural research institutes have strengthened since then (Roseboom et al. 1994). Over time, each NARI established has been given the mandate to develop improved varieties of specific crops (Roseboom et al. 1994; Beintema & Ayoola, 2004; Alene et al. 2009a, 2009b; Flaherty et al. 2010).

**Table 1. Frequencies of locations appearing as the development sources of released improved varieties in Nigeria**

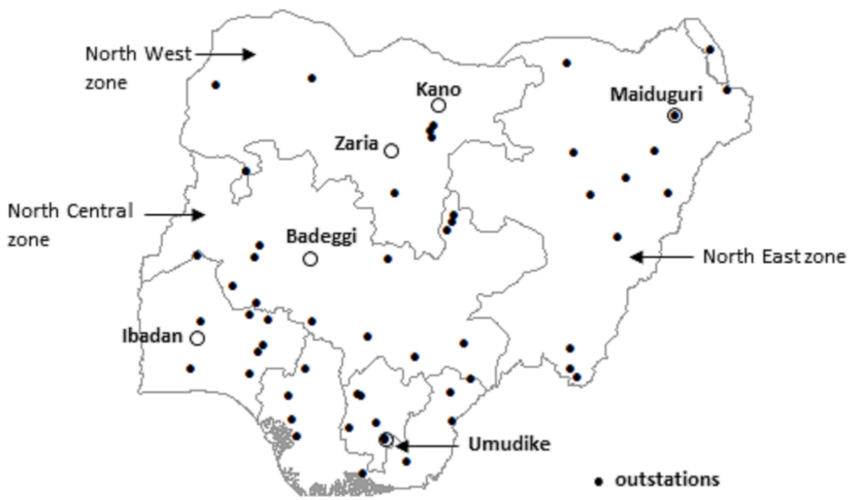
| | Locations of major breeding institutes | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Crops | Maiduguri (LCRI) | Kano (IITA, etc.) | Zaria (IAR, private seed companies) | Badeggi (NCRI) | Ibadan (IITA, other CGIAR centers) | Umudike (NRCRI) | Others | TOTAL |
| Cassava | | | | | 23 | 28 | 3 | 54 |
| Cotton | | | 13 | | | | | 13 |
| Cowpea | | 1 | 16 | 3 | 21 | | 3 | 44 |
| Groundnut | | 3 | 26 | | | | 3 | 32 |
| Maize | | | 31 | 5 | 110 | | 9 | 155 |
| Pearl millet | 5 | 1 | 7 | | | | | 13 |
| Rice | | | | 59 | 49 | | 2 | 110 |
| Sorghum | | 8 | 42 | | | | 7 | 57 |
| Soybean | | | 11 | 1 | 18 | | 1 | 31 |
| Sugarcane | | | | 9 | | | 11 | 20 |
| Tomato | | | 11 | | 2 | | 4 | 17 |
| Wheat | 8 | | 9 | | | | | 17 |
| Yam | | | | | 7 | 19 | | 26 |
| Others[a] | | | 11 | 5 | 74 | 5 | 28 | 123 |
| TOTAL | 13 | 13 | 177 | 82 | 304 | 52 | 71 | 712 |

Source: Author's compilations based on the catalogue of released varieties by NACGRAB (2017). LCRI-Lake Chad Research Institute; IITA-International Institute of Tropical Agriculture; IAR-Institute of Agricultural Research; NCRI-National Cereals Research Institute; NRCRI-National Root Crop Research Institute. [a]Others include forage legume, rubber, sesame, *amaranthus*, *sokoyokoto*, *corchorus*, okra, *solanum*, pepper, melon, cocoa, cashew, kola, coffee, oil palm, coconut, date palm, raphia palm, sweet potato, Irish potato, potato, sweet orange, tangelo, kenaf, sunflower, and cabbage.

One of the noticeable outcomes of this pattern is that most improved varieties in Nigeria have been released by a relatively small number of institutes. **Error! Reference source not found.** summarizes the frequencies of locations appearing as the development source of officially released improved crop varieties in Nigeria, based on the list of improved varieties in Nigeria in NACGRAB (2017). Note that the total frequencies are greater than the number of varieties released, because multiple locations have been listed for some varieties. In terms of locations, Maiduguri, Kano, Zaria, Badeggi (Niger state), Ibadan, and Umudike (Abia state) account for 90 percent of all improved varieties released in Nigeria so far. (See Figure 1.)

Consequently, the number of breeding institutes releasing improved varieties is also low for each crop compared to other countries. For example, Badeggi and Ibadan account for 98 percent of all variety-location combinations of rice released in Nigeria. This is considerably lower than in other countries in Asia where rice is bred at multiple national institutes, in addition to International Agricultural Research Centers, or the United States where rice is developed at seven institutes despite the US having a rice area that is less than half that of Nigeria (Takeshima & Maji 2016). Similarly, Zaria, Badeggi, and Ibadan account for 94 percent of all variety-location combinations of maize released in Nigeria. This contrasts with other Asian countries, such as China, where as much as three-quarters of maize varieties are developed at prefectural institutes, which are below provincial institutes (Jin et al. 2005, Table 1), or India, where maize is developed at 24 public breeding institutes (in addition to 25 private institutes) across country (Morris et al. 1998), even though India's maize area is not more than double that of Nigeria.

**Figure 1. Locations of the major crop breeding institutes in Nigeria listed in** Error! Reference source not found. **and their outstations**




Source: Author.

Generally, these crop breeding institutes are responsible for cross-breeding, as well as primary selections of varieties of mandated crops to be evaluated further in the country. Each NARI has several outstations are spread across the country (refer to Appendix B for a list of outstations), where candidate varieties that were selected at headquarters are evaluated. Such cross-location evaluations contribute to the selection of varieties suited for different agroecological conditions. However, the specific agroecological conditions of the locations where the headquarters of these NARIs are located are likely to considerably influence the characteristics of the improved varieties that are eventually released. This importance of the location of the headquarters may be further strengthened due to a general scarcity of crop breeders in Nigeria. Nigeria has less than two full-time-equivalent (FTE) breeders for rice (Takeshima & Maji 2016). Even for maize with relatively many breeders, the 11 FTE breeders are considerably fewer than are found in other countries, like India where there are about 70 FTE maize breeders in the public sector (Morris et al. 1998). Most of the crop breeders in Nigeria are positioned in the headquarters of each NARI, which further raises the costs of conducting evaluations at outstations.

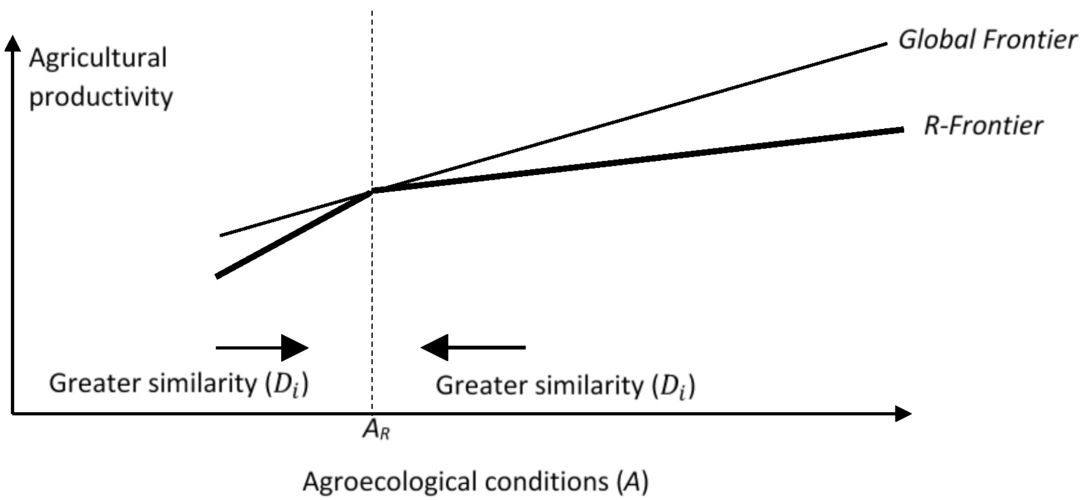
This locational concentration of the improved varieties released, as is shown in **Error! Reference source not found.**, is likely to magnify spatial variation in productivity depending on agroecological (dis)similarity. This is the focus of this paper. In our empirical model, we show that, while the agroecological similarity between each household and all breeding stations – both headquarters and the outstations of major crop breeding institutes – is significant, the agroecological similarity between the household and headquarters of major crop breeding institutes alone is also a significant determinant of the productivity.

## 3. CONCEPTUAL FRAMEWORK

In this section, we first illustrate in detail how agroecological similarities across breeding institutes can affect agricultural productivity or technical efficiency in the crop production of farm households in different locations. However, to motivate our hypotheses testing, we also discuss briefly the counter-mechanisms which may offset these effects of agroecological similarity.

**Error! Reference source not found.** illustrates how realized crop productivity is affected by agroecological conditions and the spatial distributions of breeding institutes. The productivity of crops selected at breeding institute *R* depends on both agroecological conditions and a similarity index. First, the overall productivity frontier achievable with all varieties of a crop (*Global Frontier* hereafter) naturally responds to agroecological conditions and, therefore, will not be constant across locations $A$.[3] Second, the loss of productivity due to the dissimilarity of agroecological conditions in $A$ with those of the location of R ($A_R$) leads to downward deviations from the *Global Frontier*. Consequently, as in **Error! Reference source not found.**, the productivity frontier that is specific to crop varieties developed at location R (*R-Frontier* hereafter) not only depends on the agroecological conditions in $A$ which are commonly understood, but also the agroecological conditions at location R, $A_R$ ,where the crop varieties were selected.

**Figure 2 Agroecological similarity and productivity on farm of a variety developed at a specific crop breeding institute**



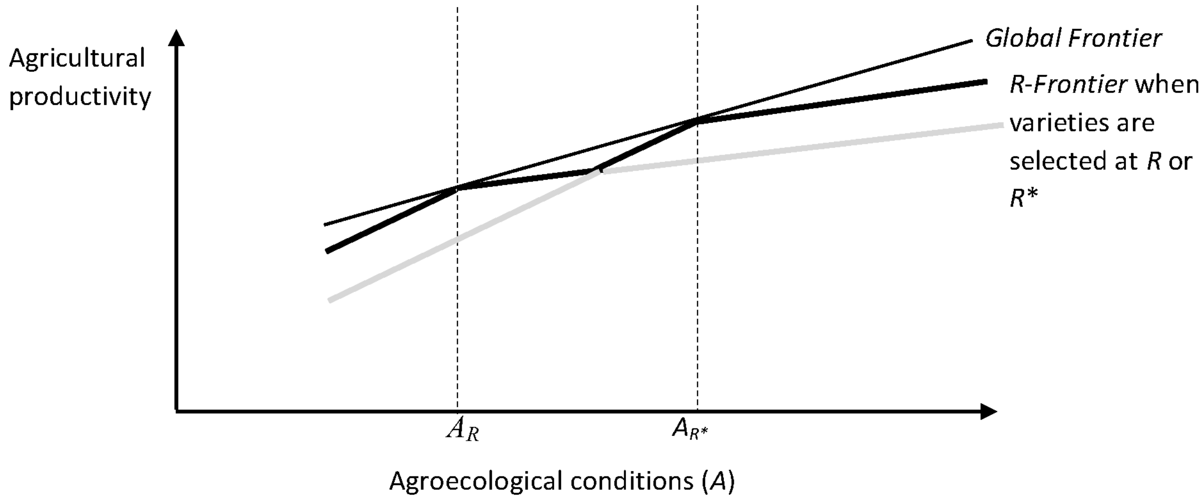

Source: Authors

In this setting, the presence of additional breeding institutes changes the *R-Frontier* curve. **Error! Reference source not found.** illustrates how an additional breeding institute *R** at $A_{R^*}$ shifts up the *R-Frontier* curve. This curve is the black solid line in Figure 3 which is the maximum of the two gray lines indicating the two *R-Frontier* curves for *R* and *R**. The assumption is that realized crop productivity at a particular location depends on *R* or *R**, whichever has the more similar agroecological condition to that location. In either case, these

[3]Because of the non-constant *Global Frontier*, productivity can still be higher in some locations than at R. This is consistent with the patterns for rice productivity in Nigeria. Rice varieties that have been largely selected in Badeggi, Niger state in the North Central zone, often exhibit higher yields in northern Nigeria. This may be due to greater solar radiation in the northern Nigeria, and, thus, a higher *Global-Frontier* there. An implication of such patterns is that, if there are additional rice breeding institutes in northern Nigeria, further yield increases are possible through the selection of rice varieties that are more suitable for the northern environments.

examples illustrate how the increased density of *R** along *A* helps in moving the *R-Frontier* curve closer to the *Global Frontier*.

**Figure 3 Agroecological similarity and productivity on farm of a variety developed at more than one crop breeding institute – maximum frontier curve**

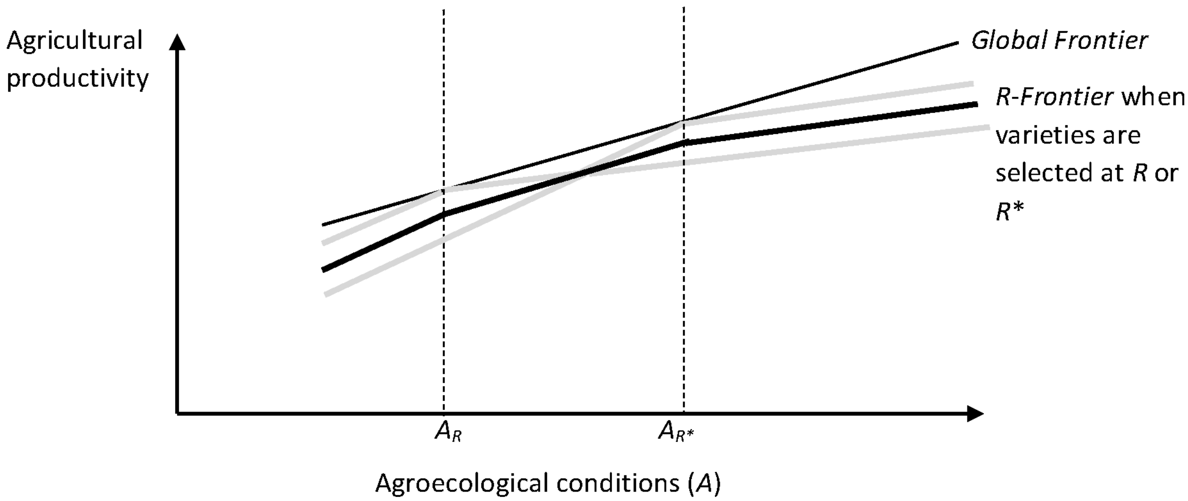



Source: Authors.

[a]Black solid line is the R-Frontier curve, which is the maximum of gray lines indicating two separate R-Frontier curves for R and R*

Figure 4 illustrates a similar example, but where realized productivity depends on the average agroecological similarity across *R* and *R**. Similar to **Error! Reference source not found.**, *R-Frontier* is closer to the *Global Frontier* where breeding institutes are concentrated. While the effect of adding more breeding institutes is less straightforward in this case, the overall effect is positive if the original breeding institutes are clustered away from the center of the distribution of *A*.

**Figure 4. Agroecological similarity and productivity on farm of a variety developed at more than one crop breeding institute – average frontier curve**




Source: Authors.
[a]Black solid line is the R-Frontier curve, which is the average of gray lines indicating two separate R-Frontier curves for R and R*

Importantly, while **Error! Reference source not found.** through Figure 4 draw the *Global Frontier* as linear lines for illustrative purpose, these need not be the case. We show in our empirical analyses that such results are generally robust to deviations from this assumption, such as when the *Global Frontier* is a function of the natural log of *A*, instead of a linear function of *A*.

The effect of agroecological similarity on the gap between the *R-Frontier* and the *Global Frontier*, conceptualized above, is important also because of how the private sector may or may not respond to potential returns from investments to narrow these gaps. Generally, the private sector may enable realized crop productivity to be close to the similarity-determined productivity (bold lines in **Error! Reference source not found.** and **Error! Reference source not found.**), while generally failing to fill the gap between the similarity-determined productivity and the *Global Frontier*. Bazzi et al. (2016) suggest that migrants can often efficiently transfer their farm production skills to new locations if the agro-climatic conditions in those new locations are similar to those in their origin regions.

On the other hand, development of most improved crop varietal technologies has been led by the public sector, at least until recently. The private sector invests to exploit economic rents arising due to such productivity gaps, including in the selections of local varieties. However, the speed of such innovations has been generally slow, because it is often difficult to recover investment costs once improved seeds are developed. Hybrid technologies have filled some of the gap, but hybrid crop varietal technologies have been limited to certain crops like maize. Private investments in other technologies, like mechanization, may not fully fill the productivity gap either, unless those technologies can perfectly substitute for varietal technologies. However, these hypotheses may not always hold due to various other factors – counter-mechanisms – that might confirm the null hypotheses that agroecological similarity among breeding institutes does not affect crop productivity over space. It is possible that productivity may not be affected by agroecological similarities in breeding locations once other factors are controlled for, including agroecological conditions in levels. If this is the case, public sector breeding and the locations in which it is done cannot greatly overcome the effects of agroecological diversity on crop productivity.

Alternatively, there may be a sufficient pool of improved crop varieties developed by the public sector for the prevailing agroecological conditions farmers face, but socioeconomic factors, including physical distance from breeding institutes, are what constrains the diffusion of these improved varieties. Even if varieties are bred and tested in some locations, if they are tested in controlled environments that are very different from farmers' field, then the agroecological environments in testing stations may not matter. If induced innovation in the private sector can overcome agroecological diversity sufficiently, but the capacity for induced innovation to occur depends more crucially on socioeconomic factors, then improvements in productivity depend solely on the nature of those socio-economic factors.

While the underlying causes of these null hypotheses are different from each other, they make the testing of our hypotheses highly relevant for agricultural research and development policies in Nigeria.

## 4. EMPIRICAL METHODS

Our empirical approach will estimate indicators of crop productivity across locations and assess how productivity levels they are affected by agroecological similarity with the breeding institutes.

**Agroecological similarity index with breeding institutes**

Modifying Bazzi et al. (2016), we define the raw similarity index for household $i$ with respect to the breeding institute $R$ ($d_{i.R}$) as,

$$D_{i.R} = -\sum_{k} w_k(|A_i^k - A_R^k|) \tag{1}$$

where $A_i^k$ and $A_R^k$ are the values of key agroecological parameters $k$ in areas where farm household $i$ and breeding institute R is located, respectively. $|A_i^k - A_R^k|$ is the absolute deviations, as in Bazzi et al. (2016). $w_k$ is the weight assigned to each $k$, which captures the effect of the similarity of $k$ for the overall similarity with breeding institute R. $D_{i.R}$ is therefore the weighted sum of the absolute differences in the values of parameter $k$ between $i$ with respect to $R$. The negative sign "$-$" is added in front of the summation operator in (1) so that an increase in $D_{i.R}$ indicates an increase in agroecological similarity.

The overall similarity index for household $i$ ($D_i$) is

$$D_i = f(D_{i,R}) \tag{2}$$

In which *f* denotes various functions that translate $D_{i.R}$ to $D_i$. We primarily present the case where *f* is the average so that $D_i = \sum_R D_{i.R} / N_R$ in which $N_R$ is the number of reference breeding institutes or stations. We then present the robustness of the results using different *f*s, such as the maximum, average weighted by the number of improved varieties released. (More details are provided in the results.)

$D_i$ is then standardized so that it is distributed between 0 and 1, with 0 being the least similar and 1 the most similar. This is simply for ease of interpreting $D_i$.

Our primary specifications use the reference locations of the key breeding institutes – Maiduguri, Kano, Zaria, Badeggi, Ibadan, and Umudike – in view of the concentration of released improved crop varieties to the institutes headquartered in these locations (**Error! Reference source not found.**). However, we also try different $D_i$, by incorporating not only these major breeding institutes, but also the locations of all the research outstations that belong to each breeding institute and the locations of other NARIs focusing on other research than crop breeding.

The key agroecological parameters k consists of three types – (1) climate related (annual rainfall, wind speed, solar radiation), (2) soil related (cation exchange, acidity, proportion of sand, proportion of silt, organic carbon content, bulk density), and (3) topography related (terrain ruggedness, slope). These are expanded from Bazzi et al. (2016) by adding wind speed and solar radiation to account for potentially important agroecological conditions in the Nigerian context. Wind is an important yield-limiting factor for many crops. Wind erosion is also an important cause of soil erosion (Tittonell & Giller 2013). Solar radiation can vary considerably within Nigeria, with a substantial effect on the yield of many crops, including rice (Takeshima & Bakare 2016). We also originally included other parameters, but found that they were highly correlated with the above-mentioned parameters. We therefore focus on the aforementioned set of agroecological parameters.

Focusing on these agroecological parameters is based on the assumption that some of the other productivity-limiting factors, such as pests, viruses, and diseases, are also largely determined by these parameters. Therefore, while it is difficult to find information on the geographical distribution of pests and diseases, using these agroecological parameters can indirectly account for their distributions. Pest incidence and its evolution is correlated with agricultural intensification processes, such as an increase in crop homogeneity, switching from mixed cropping to monocropping, and increased production intensity (Thottappilly 1992; McMillan & Meltzer 1996). All are likely to be associated with the greater productivity that is affected by the agroecological parameters.[4] Additionally, certain pests, viruses, and diseases are more likely to occur in specific agroecological environments – for example, the parasitic weed on cereals, *Striga*, is mostly found in the Northern Guinea Savanna agroecological zone in Nigeria, (Olanya et al. 1993; Iken & Amusa 2004)). However, in other cases, viruses and diseases occur relatively randomly across all agroecological zones in Nigeria – for example, maize mottle/chlorotic stunt virus (Thottappilly 1992).

### Productivity / efficiency measurements

Indicators of agricultural productivity are obtained through both parametric and nonparametric regressions of production function.

#### Parametric model

The parametric production function is estimated through the following Cobb-Douglas production function:

$$y_{it} = \alpha^y + c_i + X_{it}\beta + W_{it}\gamma + \varepsilon_{it}^y \qquad (3)$$

in which $y_{it}$ is the agricultural output expressed as the natural log of real production revenue for household *i* at time *t*, $\alpha^y$ is the intercept, $c_i$ are unobserved individual fixed effects, the main parameter that is assumed

[4] For example, yellow mottle virus, one of the major viruses for rice, spread faster when twice-a-year rice production became possible under irrigation (Thottappilly 1992).

to capture variations in agricultural productivity,[5] $X_{it}$ is a vector of inputs expressed as natural log, $W_{it}$ is the vector of other time-variant factors, and $\varepsilon_{it}^{y}$ is the idiosyncratic errors that are assumed uncorrelated with $c_i$ and $W_{it}$. We estimate (3) through both a standard fixed effects (FE) model and a fixed effects-generalized method of moments (FE-GMM), by using other time-variant instrumental variables (IV) $\tilde{Z}_{it}$, to account for the possibility that some $X_{it}$ are endogenous even after $c_i$ is separated from $\varepsilon_{it}^{y}$. GMM is more efficient than other IV estimators, like two-stage least squares, when $\varepsilon_{it}$ is heteroskedastic or serially correlated across space, which may be common in agricultural sector due to spatial correlations of unobserved climatic or biotic shocks. A FE-GMM estimation of (3) is done using the within-transformed IVs (Baltagi 2013), $\tilde{Z}_{it} = Z_{it} - \bar{Z}_i$, in which $\bar{Z}_i$ is the average values of $Z_{it}$ for household $i$ across all time periods. Note that, within-transformed IVs are uncorrelated with $c_i$, unlike the means of these time-variant variables or other time-invariant variables which are later used as determinants of $c_i$.

We then estimate the associations between estimated values of $c_i$, $\hat{c}_i$ and other factors, including the de-similarity index $D_i$. Similar approaches of estimating unobserved fixed effects and regressing them on potentially associated factors have been used in past studies.[6]

$c_i$ is often interpreted as an indicator of agricultural productivity in the literature. However, $c_i$ masks potential time-variant efficiency and, depending on the distribution of such efficiency, may be biased. We therefore also estimate another indicator that captures time-variant efficiency. Following Jondrow et al. (1982) and Greene (2005) and its extension to IV estimators (Amsler et al. 2016), we estimate

$$y_{it} = \alpha^{y} + c_i + X_{it}\beta + W_{it}\gamma + \varepsilon_{it} = \alpha^{y} + c_i + X_{it}\beta + W_{it}\gamma + v_{it} - u_{it} \qquad (4)$$

where $u_{it}$ is a half-normal technical efficiency term, and $v_{it}$ is idiosyncratic error. From this, we obtain $c_i^{*} = c_i + u_{it}$.

**Non-parametric model - Data Envelopment Analysis**

Nonparametric models, such as Data Envelopment Analysis (DEA) is sometimes a better method to estimate farm efficiency, particularly where the markets for inputs and outputs are imperfect (Charnes et al. 2013; Lovell 1993), as in much of the agricultural sector in Nigeria.

Specifically, modifying the notations of Cooper et al. (2011) to our case, DEA is estimated as a constrained optimization problem,

$$\min \theta - \epsilon\left(\textstyle\sum_{m=1}^{M} s_k^{-} + s_y^{+}\right), \qquad (5)$$

subject to

---

[5] Strictly speaking, this captures also variations that are attributable to household characteristics, such as management ability. They cannot be strictly separated from pure variation in technological potential. We, however, try to separate as much variation as possible that is due to household characteristics by controlling for observable household characteristics.

[6] Generally, one of the three ways to estimate the effects of time-invariant variables in panel is to estimate the fixed effects and then regress them on time-invariant variables (Dercon 2004). In the agricultural productivity literature, some studies treat time-varying fixed effects as productivity (Foster & Rosenzweig 1996), or regress the estimated fixed effects along a time dimension (Block 2014; Craig et al. 1997; Fan & Pardey 1997; Tack et al. 2015).

$$\sum_{j=1}^{n} x_{mj}\,\lambda_j + s_m^- = \theta x_{mo} \quad m = 1,2,\dots,M;$$
$$\sum_{j=1}^{n} y_j\,\lambda_j - s_y^+ = y_o \ ; \tag{6}$$
$$\lambda_j \geq 0, \ \ j = 1,2,\dots,J$$

in which $j$ indicates the decision-making units (DMUs), $m$ indicates type of inputs and respectively, $J$ and $M$ are the whole sets for DMUs and inputs, respectively. $\theta$ and $\lambda_j$ are the estimated parameters. $x_{mj}$ is the input value, $y_j$ is the output, $s_m^-$ is the slack for $m$, $s_y^+$ is the slack for the output, $\epsilon$ is an element that is smaller than any positive real number, included to facilitate the estimations. Lastly, $j = o$ indicates DMUs for which the efficiency is estimated. The parameter $\theta$, which is estimated for each $j$, constitutes the efficiency score, and takes a value between 0 and 1, with $\theta = 1$ indicating the efficient DMUs that is at the production frontier. One of the well-known drawback of DEA, as opposed to the parametric models presented in (3) and (4), is that the results are sensitive to outliers. Therefore, we apply DEA to the outputs and inputs measured as median values at the level of Enumeration Area (EA) of LSMS-ISA, with the EA as the DMU.

The DEA equations (5) and (6) are estimated using the command "dea" in STATA (Ji & Lee 2010). Since our data are panel data, we follow Charnes et al. (2013) by treating DMUs in each period as different DMUs. From this, we obtain the technical efficiency score $c_j$.

**Estimating the effects on agroecological similarity on productivity indices**

After productivity indices $c_i$ are obtained, we estimate

$$c_i = \alpha^C + D_i\delta + A_i^k\zeta + \bar{Z}_i\eta + \varepsilon_{it}^c \tag{7}$$

or

$$c_i^* = \alpha^{C^*} + D_i\delta^* + A_i^k\zeta^* + \bar{Z}_i\eta^* + \epsilon_i^*$$

in which asterisk (*) indicates parameters correspond to $c_i^*$, while $\varepsilon_{it}^c$ is an idiosyncratic error term. The coefficient $\delta$ is our primary interest as it measures how the agroecological similarity index $D_i$ is associated with the productivity of $c_i$, given the other characteristics of household $i$. The household characteristics are controlled for by $\bar{Z}_i$.

Agroecological variables are included as determinants of productivity and efficiency, as was illustrated in **Error! Reference source not found.** and **Error! Reference source not found.** earlier, while other household characteristics are included to control for the effects of other factors affecting the productivity or efficiency. We show that $D_i$ is still a significant determent of $c_i$, after controlling for all the other conventional correlates of efficiency.

Note that $c_i$ is measured with error (deviates from true productivity). $c_i$ is the dependent variable and thus its measurement error is generally not a concern as long as the errors in $c_i$ are not systematically correlated with the regressors in (7). The standard errors in (7) are asymptotically larger than when $c_i$ is not measured with error. Therefore, the statistical significance of the estimate is a lower bound (Wooldridge 2002 p.72).

**Potential endogeneity of $D_i$ in (7)**

In the second stage, the similarity index $D_i$ is potentially correlated with $A_i^k$ and $\bar{Z}_i$. First, $D_i$ can be correlated with $A_i^k$ as $D_i$ is a function of $A_i^k$, and the locations of research stations may be affected by agroecological factors in the country, although it is not always the case. Second, $D_i$ can be correlated with $\bar{Z}_i$ if households decide to migrate to different locations given their characteristics and $D_i$.

We address this through a Generalized Propensity Score (GPS) model, which has been used in the literature to estimate the impact of potentially endogenous continuous treatment variable like $D_i$ in our case (Imbens 2000; Flores & Mitnik, 2013; Takeshima et al. 2017). GPS differs from IV-methods like GMM, in that it does not require IVs, though requiring stronger assumption, particularly the unconfoundedness assumption, described below. This is because, unlike the FE-GMM model for the production function in which there are reasonable IVs available to instrument inputs decisions (such as input price variables), it is more difficult to find suitable IVs for $D_i$.

Specifically, the model proceeds as follows. First, we estimate the GPS associated with $D_i$, $G(D_i)$. Second, we obtain appropriate weight $\omega_i$

$$\omega_i = \frac{\Gamma(D_i)}{\hat{G}(D_i)} \tag{8}$$

in which $\Gamma(D_i)$ is the density function of a gamma distribution (explained below). Third, we estimate (7) using $\omega_i$ as the GPS model version of the inverse probability weight (IPW).

The idea behind GPS-IPW is that, conditional on $G(D_i)$, or relatedly, conditional on $\omega_i$, the outcome $c_i$ and similarity index $D_i$ are independent and no other common factors affect them both (unconfoundedness assumption), so that the endogeneity of $D_i$ is minimized. This is an extension to the case of a continuous treatment variable of similar assumptions made under the standard propensity score methods with binary treatment variables.

A conventional GPS method uses a normal distribution for $\Gamma(D_i)$, whose mean and standard deviation are estimated as those of $G(D_i)$. Alternatively, $D_i$, when it is a sum of absolute values of random variables, as used here, can be approximated by the class of gamma distribution. As is shown later, the gamma distribution is found to better characterize $D_i$ than the normal distribution. The estimation of GPS using a gamma distribution, is conducted using STATA's gpscore2 command (Guardabascio & Ventura 2014).

The approach described here also combines the GPS model with *regression adjustment*; $c_i$ is regressed on both $D_i$ with weights $\omega_i$, but also other variables $A_i^k$ and $\bar{Z}_i$. If weighting by $\omega_i$ satisfies balancing properties across different values of $D_i$, then other variables can be omitted from the regression. However, in our case, $D_i$ is by design strongly correlated with $A_i^k$, and the distributions of $A_i^k$ and $\bar{Z}_i$ differ significantly across $D_i$ even after being weighted by $\omega_i$. In such circumstances, combining regression adjustment to further improve the balancing properties is a practical way to account for residual differences between subjects with different treatment status (Rubin & Thomas 2000; Imbens 2004; Austin 2011 p.405). The GPS model with regression adjustment is "doubly-robust", and as long as either the $G(D_i)$ or (7) is correctly specified, the model is consistent.

## 5. DATA AND DESCRIPTIVE STATISTICS

Our analyses are conducted using household level data, complemented by spatial agroecological data. The primary farm household data were obtained from the LSMS-ISA, collected over three waves (2010/11, 2012/13, 2015/16) by the World Bank and the National Bureau of Statistics of Nigeria. Each wave of the LSMS-ISA consists of post-planting and post-harvest surveys. The post-planting survey collects data on inputs used from the beginning of the year, which typically overlaps with the beginning of the dry season in northern Nigeria, through planting in the rainy season. The post-planting survey also collects information on dry season output. The post-harvest survey reports outputs from the rainy season production. Each wave of LSMS-ISA is administered to a sample of approximately 5,000 households that were nationally representative in the first wave, and tracked in the second and the third waves. Our analyses focus on the 1,953 farm households in the northern part of Nigeria (North Central, North East and North West geopolitical zones), totaling approximately 5,100 observations for three waves combined.[7]

The LSMS-ISA contains an agricultural module which asks questions that allow us to estimate production functions, including the value of crops produced and the use of inputs like land, labor, non-labor expenses, agricultural capital, and irrigation. In our analyses, all monetary values are converted into real values, deflated by the local average market prices of rice, maize, and sorghum, the three major staple crops in northern Nigeria (Takeshima & Liverpool-Tasie 2015).

### Variable selections

#### Production function variable

Output variable $y$ is measured as the total real value of all crops produced by the household. Use of production values is appropriate in Nigeria where a typical farm household grows many crops, and, because of low share of certified seeds and infrequent seed replacement, the quality of varieties can vary considerably, even when the same crop variety is grown. The variable $y$ consists of the sum of the value of total crop sales, the imputed value of non-sales uses of crops, such as those consumed in the household, and the imputed value of crops in storage.

Production inputs $X_{it}$ consist of land, labor, animal traction, agricultural capital, non-labor expenses, and a dummy variable indicating the use of irrigation. Non-labor expenses include the values of all non-labor inputs, services used, including chemical fertilizer, agrochemicals, or mechanization services. All variables are treated as endogenous as they are likely to be affected by idiosyncratic shocks, except for agricultural capital, which is likely to be fixed in the short term.

The labor variable is constructed by using information on self-employment in the agricultural sector reported over 12 months. Since the LSMS-ISA data ask such information for 12 months prior to post-planting and between the post-planting and the post-harvest surveys, a period of approximately six months, we converted the information in post-planting to 6-month equivalents by simply halving it, and then took the average of the post-planting and post-harvest surveys. Construction of the labor variable also involves applying certain conversion factors for the elderly and the children in the household in order to calculate adult-equivalents. Specifically, we multiplied 0.75 and 0.5 for elderly members and for children, respectively, following Djurfeldt

[7] Three records were not obtained for all 1,953 sample households, as some did not report farming activities in all waves of the survey.

(2013). We also tested slightly different conversion factors and found that the results are generally robust against different factors. This family labor was combined with information on hired-in labor for planting, weeding, and harvesting to generate an overall labor variable. Treating the labor variable as one of the endogenous variables also mitigates measurement errors often associated with farm labor use measurements.

In addition to the use of these inputs variables for $W_{it}$, we also include the amount of rainfall that fell in the survey year. Greater rainfall can increase the harvest, but can also reduce the harvest if it leads to leaching of soil nutrients and applied fertilizers or other agrochemicals or causes flooding that submerges or washes away the plants. In addition, $W_{it}$ includes interaction variables between the survey wave and the geopolitical zone to account for any region-specific time shocks on agricultural production.

**Instrumental variables for production function; control variables for productivity and efficiency**

In production function (3) and (4), within-transformed time-variant exogenous variables ($\tilde{Z}_{it}$) are used as IVs to instrument production inputs. Land use is instrumented by the farm area obtained either through outright purchase or distributed by the village chief. Labor is instrumented by household sizes of different age groups and gender (adult or children, male or female) and the daily wage for male labor land preparation. Labor is also instrumented by the share of non-educated working-age household members, which may account for both the opportunity cost of family labor and factors affecting farm labor productivity. The use of animal traction is instrumented by the real rental rate of draft animals, the value of draft animals (heifer, steer, cow, bull, ox, donkey, horse, camel) owned by the household, average use of animal traction at EA level, which can proxy local traditions of animal traction use, and the size of pasture per livestock head, which can proxy the cost of feeding draft animals. Local pasture area is obtained from Ramankutty et al. (2008). The irrigation dummy is instrumented by the EA share of farm households using irrigation, which proxies local irrigation water access.

Non-labor expenses are instrumented by the distance to the nearest market or administrative center, both of which can affect the general prices of inputs and services, the real price of chemical fertilizer, and asset values (excluding agricultural capital) that affects the liquidity of the household. Importantly, these distance variables are time-variant, as conditions for some households or EAs change over time. Typically, Nigeria's agricultural extension arm, the Agricultural Development Projects, are located in the administrative center. The distance to the nearest administrative center thus serves as a proxy for access to extension services. Lastly, the age and gender of the household head, which are also found to vary over time due to migration, death, or other factors, are also included to instrument general input use. As is shown in the results section, all these IVs are found to satisfy the orthogonality conditions.

While $\tilde{Z}_{it}$ is used as an IV in the production function as above, $\bar{Z}_i$ are used in equations (7) as control variables potentially affecting the productivity or efficiency.

**Agroecological data**

Historical rainfall data and slope of the land are provided in the LSMS-ISA data set. Solar radiation is obtained from NASA (2017). The data of wind speed at 10 meters above ground is obtained from the Climatic Research Unit of the University of East Anglia (Climatic Research Unit 2017). Terrain ruggedness is calculated using elevation data from GTOPO30 (U.S. Geological Survey 1996) applied to a formula by Riley et al. (1999). Soil related data, including bulk density, organic carbon content, cation exchange capacity, and sand and silt composition (%) are taken from 1km resolution soils mapping data (International Soil Reference and Information Centre (ISRIC), 2013; Hengl et al. 2014).

**Descriptive statistics**

**Error! Reference source not found.** presents the descriptive statistics of variables used in the analyses. Some variables are log-transformed when used as IVs because doing so is found to minimize weak-identification problems while maintaining the orthogonality conditions. For these variables, both the non-transformed and log-transformed statistics are presented. Descriptive statistics suggest that our sample consists of heterogenous farm households that provide sufficient variation to allow us to obtain the effects of agroecological similarity on agricultural productivity and efficiency.

**Table 2. Descriptive Statistics of Variables**

| Variables[a] | Raw Mean | Raw Standard deviations | Natural log Mean | Natural log Standard deviations |
|---|---|---|---|---|
| **Production variables** | | | | |
| Total value of crops produced [a] | 63,663.4 | 308,581.6 | 9.1 | 2.3 |
| Land (square meters) | 16,748.2 | 32,330.0 | 9.0 | 1.6 |
| Labor (person-days, adult equivalent) | 482.4 | 461.7 | 5.7 | 1.1 |
| Animal traction (days) | 6.1 | 10.8 | 0.9 | 1.2 |
| Non-labor expenditure [a] | 288.1 | 579.1 | 3.8 | 2.5 |
| Agricultural capital [a] | 83.6 | 3,499.8 | 2.2 | 1.1 |
| Irrigation (% using) | 4.7 | 21.1 | | |
| Rainfall in survey year (mm, 12 months) | 791.1 | 293.6 | | |
| **Time-variant variables** | | | | |
| Age of household head (year) | 48.8 | 14.0 | | |
| Female household head (%) | 4.7 | 21.3 | | |
| Household size, adult male (number) | 1.4 | 0.9 | | |
| Household size, adult female (number) | 1.6 | 0.9 | | |
| Household size, children (number) | 2.8 | 2.1 | | |
| Share of non-educated working-age household member (%) | 45.1 | 41.5 | | |
| Household assets, excluding agricultural capital [a] | 7.1 | 26.1 | 0.9 | 1.5 |
| Value of draft animals [a] | 1,303.8 | 6,234.8 | | |
| Farm area obtained through outright purchase ($m^2$) | 978.5 | 6,415.0 | | |
| Farm area distributed by the village chiefs ($m^2$) | 7,532.1 | 23,367.4 | | |
| Distance to the nearest market center (km) | 75.0 | 40.7 | 4.1 | 0.8 |
| Distance to the nearest administrative center (km) | 92.6 | 53.0 | 4.3 | 0.8 |
| Real fertilizer price (price ratio with staple crops) | 1.8 | 7.0 | 0.0 | 0.8 |
| Daily male labor wage for land preparation [a] | 5.4 | 1.0 | 1.7 | 0.2 |
| Pasture areas per head of livestock ($km^2$ per 1,000 head) | 21.8 | 63.7 | | |
| EA average uses per household of animal tractions (days) | 5.0 | 6.6 | | |
| Rental rates of draft animal [a] | 33.3 | 39.9 | 3.2 | 0.8 |
| EA share of farm households using irrigation (%) | 4.7 | 14.6 | | |
| **Agroecological variables (time-invariant)** | | | | |
| Historical average of annual rainfall (mm) | 987.3 | 308.2 | | |
| Wind speed 10-meter height (m/second, annual average) | 2.8 | 0.5 | | |
| Daily solar radiation (kwh/$m^2$, annual average) | 5.7 | 0.3 | | |
| Slope (%) | 2.6 | 2.8 | | |
| Terrain ruggedness (index) | 33.0 | 46.1 | | |
| Top soil cation exchange (index) | 8.3 | 3.5 | | |
| Top soil acidity (pH) | 6.2 | 0.5 | | |
| Top soil composition – sand (%) | 64.5 | 9.4 | | |
| Top soil composition – silt (%) | 19.6 | 6.6 | | |
| Top soil organic carbon content (g/kg of soil) | 8.0 | 3.2 | | |
| Top soil bulk density (mt/$m^3$) | 1.3 | 0.1 | | |
| Agroecological similarity, average similarity with all major breeding institutes (index) | 0.86 | 0.29 | | |
| Average Euclidean distances to the major breeding institutes (geographical coordinates) [b] | 4.4 | 0.8 | | |

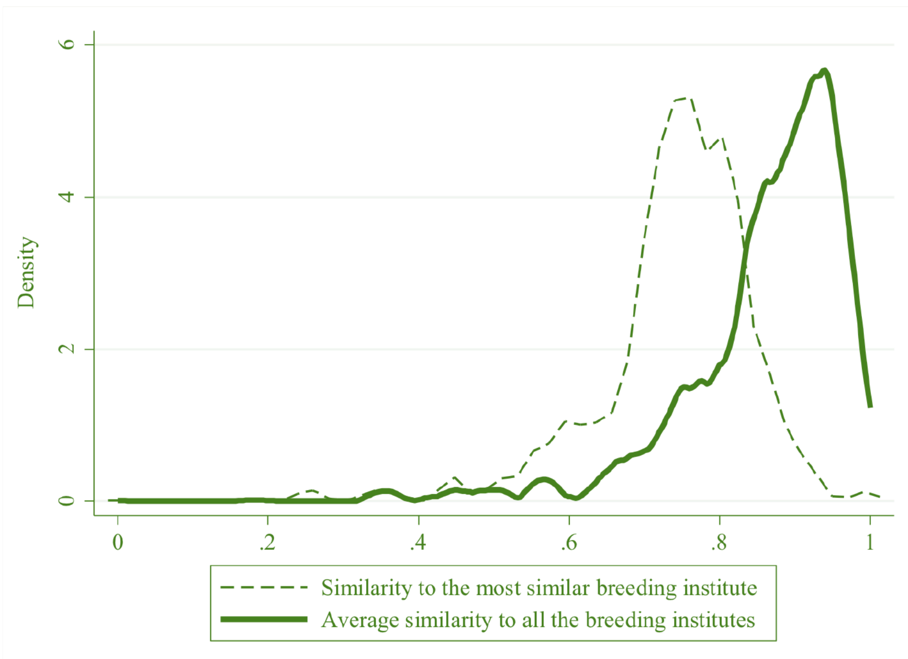
Source: Authors.

[a] For variables measured in values, the units are the values equivalent to kg of staple crops evaluated at local prices. If other units

are used, they are indicated in parentheses.
[b] For example, 4.4 is equivalent to North-South distance of 4.4 latitudinal (longitudinal) degrees, on the same longitude (latitude).

**Error! Reference source not found.** illustrates the distributions of agroecological similarity indices. These are based on the average or the maximum similarity with the six major crop breeding headquarter locations in **Error! Reference source not found.**. Note that the index for the similarity to the most similar breeding institutes is generally below the index for the average similarity to all breeding institutes because each is standardized separately. Similar to Bazzi et al. (2016), these indices have skewed distributions. As was described above, these distributions appear to exhibit gamma distributions. In addition, the distributions somewhat differ between different indices, which motivates robustness checks across different indices. These are presented in the results section.

**Figure 5. Distributions of agroecological similarity indices based on kernel density estimations, 1 = most similar; 0 = least similar**




Source: Authors.

## 6. RESULTS

Our primary interest is the relationship between the agroecological similarity indices and indicators of agricultural productivity and efficiency. We first briefly summarize the production functions from which productivity and efficiency measures are estimated. We then present and discuss the results of the effects of agroecological similarity and our robustness checks.

**Error! Reference source not found.** presents the coefficients of the Cobb-Douglas production functions, estimated through FE and FE-GMM methods. Typically, land, non-labor expenses, and agricultural capital are significant inputs. Labor and animal traction are also significant in FE, although less so in FE-GMM.

Irrigation is also a significant contributor to production, typically leading to two to three times greater output. Results of the specification tests suggest that the models are consistent. The input variables are likely to be endogenous, which are addressed in the FE-GMM model. However, the standard errors indicate that the estimates of the FE-GMM model are less precise. For robustness check purposes, we use the indicators of productivity and efficiency estimated from both models.

**Table 3. Estimated production functions**

| **Dependent variable = ln (total value of crops produced)** | **Estimation model** | | | |
|---|---|---|---|---|
| | **FE** | | **FE-GMM**[b] | |
| **Variables** | **Coefficient** | **Standard error** | **Coefficient** | **Standard error** |
| Land | .179*** | (.031) | .184† | (.115) |
| Labor | .055† | (.035) | .151 | (.331) |
| Animal traction | .107** | (.043) | .096 | (.130) |
| Non-labor expenses | .109*** | (.018) | .293† | (.204) |
| Irrigation | .900*** | (.235) | 2.327*** | (.423) |
| Capital | .089** | (.038) | .073* | (.039) |
| Rainfall | -.179 | (.656) | .299 | (.822) |
| Rainfall squared | -.295 | (.288) | -.495 | (.362) |
| Time dummies | Included | | Included | |
| Time dummies * region dummies | Included | | Included | |
| Intercept | Included | | Included | |
| Number of observations | 5,059 | | 5,041 | |
| Number of farm households in the panel | 1,953 | | 1,953 | |
| p-values | | | | |
| $H_0$: all inputs variables are exogenous | | | .000 | |
| $H_0$: model is not overidentified | | | .121 | |
| $H_0$: model is underidentified | | | .068 | |
| $H_0$: variables are jointly insignificant | .000 | | .000 | |

Source: Authors.
Note: Asterisks indicate the statistical significance: *** 1%; ** 5%; * 10%; † 15%. FE: fixed effects; FE-GMM: fixed effects-generalized method of moments. Standard errors are robust against unknown forms of heteroskedasticity.

**Error! Reference source not found.** presents the distributions of Enumeration Area-median technical efficiency estimated by Data Envelopment Analysis (DEA) (5) and (6). The DEA technical efficiency varies from the lowest median value of 0.238 to the highest median value of 1.000, with a mean of 0.655. The results indicate that variations in technical efficiency across EAs are considerable, further motivating our analyses on the role of agroecological similarity on such efficiency.

**Table 4. Distributions of median Enumeration Area technical efficiencies estimated by Data Envelopment Analysis**

| Percentile distributions | Technical efficiency scores under of variable returns-to-scale (0 = not efficient, 1 = efficient) |
|---|---|
| 0 | .238 |
| 10 | .436 |
| 20 | .485 |
| 30 | .528 |
| 40 | .573 |
| 50 | .622 |
| 60 | .693 |
| 70 | .753 |
| 80 | .827 |
| 90 | .962 |
| 100 | 1.000 |
| mean | .655 |
| sample size | 627 |

Source: Authors.

**Error! Reference source not found.** summarizes the estimated effects of agroecological similarity on agricultural productivity or efficiency, expressed as the elasticity calculated based on equation (7) using various productivity or efficiency indicators obtained from various models. (The full estimation results for the primary specifications are presented in Table 7 and Table 8 in Appendix A). For example, a value of 1.195 suggests that increasing the agroecological similarity by 1 percent leads to a 1.195 percent increase in the productivity or efficiency indicator.

**Table 5. Effects of average agroecological similarity on estimated productivity / efficiency (elasticity)**

| Agroecological similarity indicator ($D_i$) used | Models estimating productivity / efficiency | Productivity: Raw sample | Productivity: GPS-IPW sample | Efficiency: Raw sample | Efficiency: GPS-IPW sample |
|---|---|---|---|---|---|
| Primary specifications | FE | 1.195** | 1.245** | 1.236** | 1.212** |
| | | (.574) | (.588) | (.512) | (.603) |
| | FE-GMM | 1.257** | 1.255** | 1.247*** | 1.190* |
| | | (.543) | (.581) | (.478) | (.615) |
| | DEA for EA median | | | .245*** | .239** |
| | | | | (.092) | (.095) |
| *Robustness check* (a): using equal weights for each group (climate, water, soil) | FE | 1.341** | 1.072* | 1.201*** | 1.052* |
| | | (.594) | (.622) | (.464) | (.635) |
| | FE-GMM | 1.142** | 1.072* | .944** | 1.008† |
| | | (.580) | (.618) | (.450) | (.646) |
| | DEA for EA median | | | .230** | .279*** |
| | | | | (.091) | (.094) |
| *Robustness check* (b) Using the maximum similarity among all breeding institutes | FE | .592** | .696** | .623** | .715** |
| | | (.289) | (.329) | (.270) | (.339) |
| | FE-GMM | .586* | .705** | .560** | .721** |
| | | (.306) | (.335) | (.280) | (.357) |
| | DEA for EA median | | | .125*** | .143*** |
| | | | | (.043) | (.049) |
| *Robustness check* (c) Using the average similarity weighted by the number of developed varieties released | FE | 1.073** | 1.221** | 1.095*** | 1.200** |
| | | (.459) | (.481) | (.417) | (.486) |
| | FE-GMM | 1.084** | 1.231** | 1.041*** | 1.167** |
| | | (.443) | (.485) | (.396) | (.506) |
| | DEA for EA median | | | .219*** | .217*** |
| | | | | (.072) | (.076) |
| *Robustness check* (d) Using the natural log of agroecological variables as control | FE | 1.153*** | .936† | .874** | 1.032* |
| | | (.440) | (.570) | (.369) | (.571) |
| | FE-GMM | 1.319*** | 1.084* | 1.023*** | 1.133* |
| | | (.433) | (.577) | (.367) | (.592) |
| | DEA for EA median | | | .255*** | .339*** |
| | | | | (.074) | (.091) |
| *Robustness check* (e) Using all outstations of breeding institutes | FE | 1.211** | 1.359** | 1.211** | 1.490** |
| | | (.614) | (.663) | (.541) | (.659) |
| | FE-GMM | 1.382** | 1.541** | 1.311** | 1.620** |
| | | (.587) | (.636) | (.508) | (.647) |
| | DEA for EA median | | | .216** | .282*** |
| | | | | (.097) | (.104) |
| *Robustness check* (f) (b) + (e) | FE | .206† | .270† | .167 | .314* |
| | | (.136) | (.173) | (.124) | (.178) |
| | FE-GMM | .325** | .412** | .276** | .444** |
| | | (.132) | (.175) | (.121) | (.183) |
| | DEA for EA median | | | .020 | .044 |
| | | | | (.028) | (.032) |

Source: Authors.
Note: Asterisks indicate the statistical significance: *** 1%; ** 5%; * 10%; † 15%. FE: fixed effects; FE-GMM: fixed effects-generalized method of moments; GPS-IPW: Generalized Propensity Score-Inverse Probability Weighting; DEA: Data Envelopment Analysis; EA: Enumeration Area. Numbers in parentheses are standard errors. Standard errors are estimated through 200 bootstraps.

Estimates based on GPS-IPW samples account for the potential endogeneity of household locations and, thus, agroecological similarity $D_i$. Estimated determinants of GPS are presented in Appendix A (Table 6). As is indicated in Table 6, the specification test of the deviance residuals shows that GPS is consistently estimated through a generalized linear model with a gamma distribution.

**Error! Reference source not found.** suggests that the effects of agroecological similarity are considerably robust across different models used to estimate the productivity or efficiency indicators, and hold for both indicators. The magnitude of the effects for technical efficiency estimated from EA differ from those of the other two models, because these are based on the EA-median values. However, the signs and statistical significance of the effects from the EA-based model are consistent with the other two models.

The results for the primary specification presented in **Error! Reference source not found.** are with respect to $D_i$ based on the average agroecological similarity across all breeding institutes. We further show that these results hold for various other calculations of $D_i$. Specifically, to check the robustness of our results, we check the following:

(a) instead of calculating $D_i$ as the raw average across all parameters, we group them into three types (climate, topography, soil), and apply equal weights to each group, rather than to each parameter;
(b) using the similarity index based on the most similar breeding institute, rather than including all breeding institutes;
(c) using the average similarity weighted by the number of developed varieties released;
(d) adding square terms of agroecological variables as controls;
(e) incorporating not only the breeding institutes, but also all their outstations; and
(f) combining (b) and (e).

The lower rows of **Error! Reference source not found.** summarize the results of these robustness checks. The magnitudes of elasticities vary – in particular, the elasticity of the effect of increasing $D_i$ is much smaller if $D_i$ is based on the maximum similarity among all major breeding institutes (with or without outstations) (results under (b) and (f)). In addition, the elasticity is considerable greater in case (d). Nevertheless, signs and statistical significance are generally consistent. Importantly, in (b) and (f) the estimated effects are statistically insignificant for the raw (unweighted) samples, but are statistically significant with consistent signs for GPS-IPW samples. These results suggest that when we focus on the institutes or substations with the maximum similarity with the households or EA, the bias due to potential endogeneity of $D_i$ becomes substantial. In such cases, using GPS-IPW becomes particularly important. Our findings of the effects of agroecological similarity on the agricultural productivity and efficiency are therefore robust, and the use of GPS-IPW also addresses the endogeneity of $D_i$.

## 7. CONCLUSIONS

The importance of agroecological conditions and their diversity in SSA agriculture is widely acknowledged. The public sector remains a major player in crop breeding and varietal development. Despite the importance of location-specific adaptive breeding research, past reforms of crop breeding systems in SSA countries like Nigeria has focused more on centralizing breeding activities. This has been based partly on the premise that such research systems can still effectively develop a diverse set of varietal technologies suitable for different agroecological conditions through the use of numerous outstations and multilocational trials, regardless of the locations of headquarters or these outstations. However, little empirical evidence exists that support this premise.

Using panel data for agricultural households in northern Nigeria, as well as spatial data on various agroecological factors, this study fills this knowledge gap. Specifically, it empirically shows that the agricultural productivity and technical efficiency at the farm household level is significantly positively affected by the similarity of agroecological conditions between the locations of these households and the locations where major crop breeding institutes are headquartered in Nigeria, namely Maiduguri, Kano, Zaria, Badeggi, Ibadan and Umudike, after controlling for the agroecological conditions and various relevant household characteristics of these households. These results are also robust when we consider similarities in agroecological conditions between households and where outstations of NARIs are located in Nigeria. These findings suggest that where improved varieties are developed and evaluated affects agricultural productivity and technical efficiency at farm level in different locations. The current public crop breeding system in Nigeria does not overcome this diversity in agroecological conditions. Given that the number of breeding institutes and breeders are relatively few in Nigeria and given the size of its arable land compared to other countries outside SSA, these results suggest that the current geographic setup of public crop breeding systems may be partly contributing to continuing low overall agricultural productivity in Nigeria.

Methodologically, the study contributes to the impact evaluation literature by demonstrating how inverse-probability weighting method can be combined with the generalized-propensity-score method and applied to the case where the treatment parameters of interests (or their transformations) may follow a gamma distribution, rather than the normal distribution as is commonly assumed.

The policy implications of the above findings are clear. Locations of research institutes with the mandate for conducting crop breeding and varietal development matters for overall agricultural productivity and efficiency in Nigeria. Thus, reforming the country's crop breeding system, particularly the geographical locations of specific crop breeding institutional headquarters and outstations, is likely to have significant productivity and efficiency enhancing effects. It is, however, impractical to expect substantial agricultural productivity improvement by simply decentralizing the breeding institutes into more locations and reallocating the research funding across the sites. The findings of this study rather should be interpreted as complementary to the results of studies promoting increased overall funding for agricultural research, including crop breeding, in SSA countries including Nigeria (Walker & Alwang 2015; Lynam et al. 2016; Benin 2016). This paper provides useful information for guiding how support for agricultural research, particularly crop breeding, can be increased effectively. One way to do so would be to set up new crop breeding institutes or outstations in manner that results in an improvement in the degree of overall agroecological similarity between the breeding locations for a specific crop and the locations of farm households producing that crop in Nigeria.

Having said this, there are still limitations to this study. We focus primarily on crop varietal development, investigating the role of major public crop breeding institutes that release improved varieties and examining heterogeneity across these institutes in terms of the number of varieties released. Fuller investigations, however, require incorporating other research activities than crop breeding alone, such as research on other production technologies like mechanization, or research on production management, and the intensity of such research activities, measured by research spending or assignment of human capitals and other resources. Similarly, although our analyses account not only for the locations of the headquarters of NARIs but also their outstations, our analyses do not explicitly consider the potential heterogeneity in research intensity across these outstations. Future studies should delve deeper into these issues by gathering detailed information on research intensity, and assess if our findings on the locations of crop breeding are affected by incorporation of this information into the analysis.

## REFERENCES


Alene A, V Manyong, EF Tollens & S Abele. 2009a. Efficiency–equity tradeoffs and the scope for resource reallocation in agricultural research: evidence from Nigeria. *Agricultural Economics 40*: 1-14.

Alene AD, A Menkir, SO Ajala, B Badu-Apraku, AS Olanrewaju, VM Manyong & A Ndiaye. 2009b. The economic and poverty impacts of maize research in West and Central Africa. *Agricultural Economics 40* (5): 535-550.

Alene AD. 2010. Productivity growth and the effects of R&D in African agriculture. *Agricultural Economics 41* (3-4): 223-238.

Alston JM. 2002. Spillovers. *Australian Journal of Agricultural and Resource Economics 46* (3): 315-46.

Alston JM, JS James, MA Andersen & PG Pardey. 2010. *Econometric Estimation and Results*. In *Persistence Pays* (pp. 313-351). Springer New York.

Amsler C, A Prokhorov & P Schmidt. 2016. Endogeneity in stochastic frontier models. *Journal of Econometrics 190* (2): 280-288.

Austin PC. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research 46* (3): 399-424.

Baltagi B. 2013. *Econometric Analysis of Panel Data. 5th ed.* Chichester, UK: Wiley.

Bazzi S, A Gaduh, AD Rothenberg & M Wong. 2016. Skill transferability, migration, and development: Evidence from population resettlement in Indonesia. *American Economic Review 106* (9): 2658-2698.

Beintema NM & GB Ayoola. 2004. *Nigeria.* ASTI Country Brief No. 10. Washington D.C.: International Food Policy Research Institute and International Service for National Agricultural Research.

Benin S. 2016. *Agricultural productivity in Africa: Trends, patterns, and determinants.* Washington DC. Intl Food Policy Res Inst.

Block SA. 2014. The post-independence decline and rise of crop productivity in sub-Saharan Africa: measurement and explanations. *Oxford Economic Papers 66* (2): 373-396.

Brennan JP, IP Singh & LG Lewin. 1997. Identifying international rice research spillovers in New South Wales. *Agricultural Economics 17* (1): 35-44.

Byerlee D & G Traxler. 2001. *The Role of Technology Spillovers and Economies of Size in the Efficient Design of Agricultural Research Systems.* In J Alston, P Pardey & M Taylor, (eds.), *Agricultural Science Policy: Changing Global Agendas*. Baltimore, MD: Johns Hopkins University Press.

Charnes A, WW Cooper, AY Lewin & LM Seiford. 2013. *Data envelopment analysis: Theory, methodology, and applications*. Springer Science & Business Media.
Climatic Research Unit (CRU). 2017. *10m windspeed*. University of East Anglia.

Cooper W, LM Seiford & J Zhu. 2011. *Data envelopment analysis: History, models and interpretations*. Handbook on Data Envelopment Analysis.

Craig B, P Pardey & J Roseboom. 1997. International productivity patterns: accounting for input quality, infrastructure, and research. *American Journal of Agricultural Economics 79* (4): 1064-76.

Dercon S. 2004. Growth and Shocks: Evidence from Rural Ethiopia. *Journal of Development Economics 74* (2): 309-29.

Djurfeldt AA. 2013. African Re-Agrarianization? Accumulation or Pro-Poor Agricultural Growth? *World Development 41*: 217-231.

Evenson R & D Gollin. 2003. *Crop Variety Improvement and Its Effect on Productivity: The Impact of International Agricultural Research*. Wallingford, UK: CABI.

Evenson R & LE Westphal. 1995. *Technological Change and Technology Strategy*. In J Behrman & TN Srinivasan (Eds.), *Handbook of Development Economics*. 2209-2299. Amsterdam: Elsevier.

Fan S & P Pardey. 1997. Research, productivity, and output growth in Chinese agriculture. *Journal of Development Economics 53* (1): 115-137.

Flaherty K, G Ayoola, J Ogbodo & N Beintema. 2010. *Nigeria: Recent Developments in Agricultural Research*. Agricultural Science & Technology Indicators (ASTI).

Flores CA & OA Mitnik. 2013. Comparing treatments across labor markets: an assessment of nonexperimental multiple-treatment strategies. *The Review of Economics and Statistics 95* (5): 1691-1707.

Foster AD & MR Rosenzweig. 1996. Technical change and human-capital returns and investments: Evidence from the green revolution. *American Economic Review 86* (4): 931-953.

Greene W. 2005. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis 23* (1): 7-32.

Griliches, Z. 1991. *The search for R&D spillovers* (No. w3768). National Bureau of Economic Research.

Guardabascio B & M Ventura. 2014. Estimating the dose–response function through a generalized linear model approach. *The STATA Journal 14* (1): 141-158.

Hengl T, de Jesus JM, RA MacMillan, NH Batjes, GB Heuvelink, E Ribeiro, A Samuel-Rosa, B Kempen, JGB Leenaars, MG Walsh & MR Gonzalez. 2014. SoilGrids1km—global soil information based on automated mapping. *PLoS One 9*(8): 1-17.

Iken J & N Amusa. 2004. Maize research and production in Nigeria. *African Journal of Biotechnology 3* (6): 302-307.

Imbens GW. 2000. The role of the propensity score in estimating dose-response functions. *Biometrika 87* (3): 706-710.

Imbens GW. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics 86*: 4-29.

International Soil Reference and Information Centre (ISRIC). 2013. *Soil Property Maps of Africa at 1 km*. Wageningen: ISRIC. http://www.isric.org. Accessed April 1, 2017.

Ji Y & C Lee. 2010. Data envelopment analysis. *Stata Journal 10* (2): 267-280.

Jin S, S Rozelle, J Alston & J Huang. 2005. Economies of scale and scope and the economic efficiency of China's agricultural research system. *International Economic Review 46* (3): 1033-1057.

Johnson M, S Benin, L You, X Diao, P Chilonda & A Kennedy. 2014. *Exploring strategic priorities for regional agricultural research and development investments in southern Africa*. IFPRI Discussion Paper 01318.

Jondrow J, I Materov, K Lovell & P Schmidt. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics 19* (2/3): 233-238.

Lovell CAK. 1993. "Production frontiers and productive efficiency." in: H. Fried, C.A.K. Lovell, S. Schmidt (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press. 3–67.

Lynam J, N Beintema, J Roseboom & O Badiane. 2016. *Agricultural Research in Africa: Investing in Future Harvests*. Washington DC: International Food Policy Research Institute.

Maredia MK & D Byerlee. 1999. *The global wheat improvement system: prospects for enhancing efficiency in the presence of spillovers*. CIMMYT Research Report No. 5. International Centre for Wheat and Maize Improvement, Mexico, D.F.

Maredia MK, R Ward & D Byerlee. 1996. Econometric estimation of a global spillover matrix for wheat varietal technology. *Agricultural Economics. 14*: 159-173.
McCullagh P & JA Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall/CRC.

McMillan DE & MI Meltzer. 1996. Vector-borne disease control in sub-Saharan Africa: A necessary but partial vision of development. *World Development 24*: 569-588.

Morris ML, RP Singh & S Pal. 1998. India's maize seed industry in transition: Changing roles for the public and private sectors. *Food Policy 23*: 55-71.

NACGRAB (National Centre for Genetic Resources and Biotechnology of Nigeria). 2017. *Varietal Release Catalogue.* Available at http://www.nacgrab.gov.ng/images/Varieties_Released_Catalogue.pdf. Accessed April 5, 2017.

National Aeronautics and Space Administration. 2017. *NASA Surface Meteorology and Solar Energy: Global Data Sets.* Accessed April 15, 2017. https://eosweb.larc.nasa.gov/cgi-bin/sse/global.cgi?email=skip@larc.nasa.gov.

Olanya OM, JM Fajemisin & P Oyekan. 1993. Incidence and geographical distribution of downy mildew on maize caused by Peronosclerospora sorghi in Nigeria. *International Journal of Pest Management, 39* (1): 28-34.

Ramankutty N, AT Evan, C Monfreda & JA Foley. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles 22* (1): 1-19.

Riley SJ, SD DeGloria & R Elliot. 1999. A Terrain Ruggedness Index That Quantifies Topographic Heterogeneity. *Intermountain Journal of Sciences 5* (1-4): 23-27.

Roseboom J, N Beintema, PG Pardey & EO Oyedipe. 1994. *Statistical Brief on the National Agricultural Research System of Nigeria.* ISNAR Indicator series project: Phase II Statistical Brief 15. ISNAR, Netherlands.

Rubin DB & N Thomas. 2000. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association 95*: 573-585.

Tack J, A Barkley & L Nalley. 2015. Estimating yield gaps with limited data: An application to United States wheat. *American Journal of Agricultural Economics 97* (5): 1461-1477.

Takeshima H & LS Liverpool-Tasie. 2015. Fertilizer subsidy, political influence and local food prices in sub-Saharan Africa: Evidence from Nigeria. *Food Policy 54*: 11-24.

Takeshima H & O Bakare. 2016. *Chapter 3: Production systems – biophysical and economic environment, and constraints*, in *The Nigerian Rice Economy: Policy Options for Transforming Production, Marketing, and Trade*, K Gyimah-Brempong, M Johnson & H Takeshima (eds.). Philadelphia. University of Pennsylvania Press.

Takeshima H & A Maji. 2016. *Varietal development and the effectiveness of seed sector policies: The case of rice in Nigeria.* IFPRI NSSP Working Paper 34.

Takeshima H, R Adhikari, S Shivakoti, BD Kaphle & A Kumar. 2017. Heterogeneous Returns to Chemical Fertilizer at the Intensive Margins: Insights from Nepal. *Food Policy 69*: 97-109.

Thottappilly G. 1992. Plant virus diseases of importance to African agriculture. *Journal of Phytopathology 134* (4): 265-288.

Tittonell P & KE Giller. 2013. When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. *Field Crops Research 143*: 76-90.

USGS (U.S. Geological Survey). 1996. *GTOPO30*. Sioux Falls, SD, US: USGS Center for Earth Resources Observation and Science.

Walker TS & J Alwang. 2015. *Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa.* CABI.

Wooldridge J. 2002. *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts and London, England: MIT Press.

You L & M Johnson. 2010. Exploring strategic priorities for regional agricultural R&D investments in East and Central Africa. *Agricultural Economics 41* (2): 177-190.

# APPENDIX

## Appendix A: Full estimation results

**Table 6. Determinants of generalized propensity scores of agroecological similarity index estimated through generalized linear model with a variance function specified as a gamma distribution**

| **Dependent variable = agroecological similarity index (average of agroecological similarity with the major breeding institutes)** | | |
|---|---|---|
| **Variables**[a] | **Coefficients** | **Standard Error** |
| Historical average of annual rainfall | -.000*** | (.000) |
| Wind speed 10-meter height | -.290*** | (.021) |
| Daily solar radiation | -.359*** | (.050) |
| Slope | -.119*** | (.003) |
| Terrain ruggedness index | -.009*** | (.000) |
| Top soil cation exchange index | -.038*** | (.003) |
| Top soil acidity | .112*** | (.024) |
| Top soil composition – sand | -.021*** | (.002) |
| Top soil composition – silt | -.010*** | (.002) |
| Top soil organic carbon content | -.035*** | (.004) |
| Top soil bulk density | 1.241*** | (.099) |
| Age of household head | -.001† | (.000) |
| Female household head | -.108*** | (.027) |
| Household size (adult male) | -.006 | (.007) |
| Household size (adult female) | .016** | (.008) |
| Household size (children) | .000 | (.000) |
| Share of non-educated working-age household member | .004 | (.016) |
| Household assets | -.007 | (.005) |
| Value of draft animals | -.000* | (.000) |
| Farm area obtained through outright purchase | -.000 | (.000) |

| | | |
|---|---|---|
| Farm area distributed by the village chiefs | -.000* | (.000) |
| Distance to the nearest market center | .023*** | (.005) |
| Distance to the nearest administrative center | -.069*** | (.005) |
| Real fertilizer price | -.012 | (.011) |
| Daily male labor wage for land preparation | .414*** | (.060) |
| Pasture areas per head of livestock | .159 | (.124) |
| EA average uses per household of animal tractions | .002*** | (.001) |
| Rental rates of draft animal | -.016 | (.012) |
| EA share (%) of farm households using irrigation | .206*** | (.040) |
| Distance to breeding institutes | -.165*** | (.015) |
| Geopolitical zones | Included | |
| Intercept | Included | |
| Number of observations | 1,953 | |
| Log pseudo-likelihood | -3632.8 | |
| p-value ($H_0$: model is correctly specified with gamma distribution)[b] | .223 | |

Source: Authors. Asterisks indicate the statistical significance: *** 1%; ** 5%; * 10%; † 15%.

[a] Averages across all waves are used for time-variant variables.

[b] Based on the skewness-kurtosis test for the normality of the deviance residuals (McCullagh & Nelder 1989).

**Table 7. Elasticity of productivity with respect to each factor**

| Dependent variables = productivity estimated from corresponding models | Productivity estimated from FE | | Productivity estimated from FE-GMM | |
|---|---|---|---|---|
| **Variables** | **Un-weighted** | **GPS-IPW** | **Un-weighted** | **GPS-IPW** |
| Agroecological similarity | 1.195** | 1.245** | 1.257** | 1.255** |
| Historical average of annual rainfall | -1.132† | -.824 | -1.557** | -1.382† |
| Wind speed 10-meter height | -1.099* | -.362 | -.281 | .634 |
| Daily solar radiation | .185 | -.387 | -2.870 | -3.852 |
| Slope | .251*** | .262*** | .252*** | .265*** |
| Terrain ruggedness index | -.009 | .038 | -.049 | -.009 |
| Top soil cation exchange index | -.478* | -.465* | -.439† | -.370 |
| Top soil acidity | -1.276 | -.431 | -1.357 | -.550 |
| Top soil composition – sand | -.850 | -.619 | .212 | .568 |
| Top soil composition – silt | -.266 | -.220 | -.100 | -.025 |
| Top soil organic carbon content | 1.078*** | 1.233*** | 1.175*** | 1.412*** |
| Top soil bulk density | -2.329 | -2.068 | -2.493† | -2.641 |
| Age of household head | -.583*** | -.475*** | -.511*** | -.412** |
| Female household head | -.818*** | -.934*** | -.616** | -.755*** |
| Household size (adult male) | .109 | .112 | .165* | .163* |
| Household size (adult female) | .144† | .069 | .087 | .003 |
| Household size (children) | .000 | .000 | .000 | .000 |
| Share of non-educated working-age household member | .146* | .216*** | .225*** | .304*** |
| Household assets | .077* | .096** | -.028 | .002 |
| Value of draft animals | -.023† | -.014 | -.024† | -.015 |
| Farm area obtained through outright purchase | -.003 | -.002 | -.005 | -.005 |
| Farm area distributed by the village chiefs | -.007 | .008 | -.006 | .007 |
| Distance to the nearest market center | -.087 | -.095 | -.041 | -.045 |
| Distance to the nearest administrative center | -.044 | -.021 | -.002 | .017 |
| Real fertilizer price | .070 | .075 | -.088 | -.057 |
| Daily male labor wage for land preparation | -.987† | -.998† | -1.146* | -1.077* |
| Pasture areas per head of livestock | -.050** | -.051** | -.070*** | -.068*** |
| EA average uses per household of animal tractions | .129† | .116 | .118† | .108 |
| Rental rates of draft animal | .334** | .221 | .307** | .220 |
| EA share (%) of farm households using irrigation | .000 | .001 | -.085*** | -.081*** |
| Distance to breeding institutes | -1.399* | -.932 | -.885 | -.444 |

Source: Authors.

Note: Asterisks indicate the statistical significance: *** 1%; ** 5%; * 10%; † 15%. Standard errors are estimated through 200 bootstraps. FE: fixed effects; FE-GMM: fixed effects-generalized method of moments; GPS-IPW: Generalized Propensity Score-Inverse Probability Weighting.

**Table 8. Elasticity of the efficiency with respect to each factor**

| Dependent variables = efficiency estimated from corresponding models | Efficiency estimated from FE | | Efficiency estimated from FE-GMM | | Efficiency estimated from EA-median DEA | |
|---|---|---|---|---|---|---|
| Variables | Un-weighted | GPS-IPW | Un-weighted | GPS-IPW | Un-weighted | GPS-IPW |
| Agroecological similarity | 1.236** | 1.212* | -1.247*** | 1.190* | .260*** | .254*** |
| Historical average of annual rainfall | -1.078 | -.734 | -1.534* | -1.341† | -.739*** | -.782*** |
| Wind speed 10-meter height | -1.075* | -.223 | -.305 | .876 | .314*** | .304*** |
| Daily solar radiation | .386 | -.702 | -2.545 | -4.171 | -1.088** | -1.199** |
| Slope | .218*** | .251*** | .206*** | .252*** | .011 | .018 |
| Terrain ruggedness index | -.010 | .036 | -.058 | -.017 | -.001 | -.001 |
| Top soil cation exchange index | -.564** | -.479* | -.531** | -.383 | .118** | .117** |
| Top soil acidity | -.427 | -.494 | -.512 | -.609 | -.183 | -.214 |
| Top soil composition – sand | -1.444 | -.908 | -.647 | .166 | .178 | .160 |
| Top soil composition – silt | -.615 | -.343 | -.495 | -.174 | -.039 | -.033 |
| Top soil organic carbon content | 1.100*** | 1.088*** | 1.110*** | 1.231*** | .176*** | .174*** |
| Top soil bulk density | -2.647† | -3.026† | -2.951* | -3.774* | .093 | .069 |
| Age of household head | -.461*** | -.531*** | -.404** | -.468** | -.056 | -.047 |
| Female household head | -.868*** | -.890*** | -.627*** | -.642** | .284** | .296** |
| Household size (adult male) | .024 | .130 | .059 | .159† | .033 | .034 |
| Household size (adult female) | .075 | .065 | .003 | -.034 | -.082*** | -.086*** |
| Household size (children) | .000 | .000 | .000 | .000 | .000 | .000 |
| Share of non-educated working-age household member | .194*** | .243*** | .262*** | .322*** | .061*** | .063*** |
| Household assets | .116*** | .106** | .012 | .009 | .012 | .017 |
| Value of draft animals | -.013 | -.018 | -.014 | -.021 | -.008** | -.009** |
| Farm area obtained through outright purchase | -.001 | .000 | -.004 | -.003 | .000 | .000 |
| Farm area distributed by the village chiefs | .000 | .007 | .000 | .001 | -.005 | -.005 |
| Distance to the nearest market center | -.103† | -.089 | -.038 | -.038 | -.044 | -.046 |
| Distance to the nearest administrative center | -.013 | -.031 | .032 | .009 | -.062 | -.053 |
| Real fertilizer price | .085 | .106 | -.072 | -.027 | -.002** | -.002** |
| Daily male labor wage for land preparation | -1.407** | -1.034† | -1.662*** | -1.064* | -.815*** | -.816*** |
| Pasture areas per head of livestock | -.036 | -.038† | -.060** | -.052** | -.008 | -.009† |
| EA average uses per household of animal traction | .164** | .109 | .156** | .108 | .005 | .004 |
| Rental rates of draft animal | .382** | .164 | .352** | .148 | .108† | .104† |
| EA share (%) of farm households using irrigation | -.001 | .008 | -.089*** | -.077*** | .002 | .002 |
| Distance to breeding institutes | -1.466** | -.566 | -1.101† | -.127 | -.331*** | -.336*** |

Source: Authors.

Note: Asterisks indicate the statistical significance: *** 1%; ** 5%; * 10%; † 15%. Standard errors are estimated through 200 bootstraps.

## Appendix B: List of crop breeding outstations used in the analysis

**Table 9. Locations of Agricultural Research Institutes (HQ) and their outstations (O) in Nigeria**

| Location | State | Int'l Institute of Tropical Agricul-ture | National Cereals Research Institute | National Animal Produc-tion Research Institute | Institute of Agricul-tural Research | Institute of Agricul-tural Research & Training | Lake Chad Research Institute | National Root Crop Research Institute | National Horticul-tural Research Institute | Stored Product Research Institute | National Institute for Fresh-water Fisheries Research | National Veterinary Research Institute | Rubber Research Institute of Nigeria | Nigerian Institute for Oil-Palm Research | Cocoa Research Institute of Nigeria |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ajassor | Cross-River | | | | | | | | | | | | | | **O** |
| Amakama | Abia | | **O** | | | | | | | | | | | | |
| Bacita | Kwara | | **O** | | | | | | | | | | | | |
| Badeggi | Niger | | **HQ** | | | | | | | | | | | | |
| Baga | Borno | | | | | | **O** | | | | **O** | | | | |
| Bagauda | Kano | | | | | | | | **O** | | | | | | |
| Ballah | Kwara | | | | | **O** | | | | | | | | | |
| Benin City | Edo | | | | | | | | | | | | **HQ** | **HQ** | |
| Birnin-Kebbi | Kebbi | | **O** | | | | | | | | | | | | |
| Biu | Borno | | | | | | **O** | | | | | | | | |
| Dadinkowa | Gombe | | | | | | **O** | | **O** | | **O** | | | | |
| Damboa | Borno | | | | | | **O** | | | | | | | | |
| Deba | Gombe | | | | | | **O** | | | | | | | | |
| Gashua | Yobe | | | | | | **O** | | | | | | | | |
| Gembu | Taraba | | | | | | **O** | | | | | | | | |
| Ibadan | Oyo | **HQ** | **O** | | | **HQ** | | | **HQ** | **O** | | | | | **HQ** |
| Ibeku | Abia | | | | | | | | | | | | | | **O** |
| Ibule | Ondo | | | | | | | | | | | | | | **O** |
| Igbariam | Anambra | | | | | | | **O** | | | | | | | |
| Ikenne | Ogun | | | | | **O** | | | | | | | | | |
| Ile-Ife | Osun | | | | | **O** | | | | | | | | | |
| Ilora | Oyo | | | | | **O** | | | | | | | | | |
| Ilorin | Kwara | | | | | | | | | **HQ** | | | | | |
| Iresi | Osun | | | | | | | **O** | | | | | | | |
| Jos | Plateau | | | | | | **O** | **O** | | | | | | | |
| Kabba | Kogi | | | | | | | | | | | | | | **O** |
| Kadawa | Kano | | | | **O** | | | | | | | | | | |
| Kano | Kano | **O** | | | **HQ** | | | | | **O** | | | | | |
| Kishi | Oyo | | | | | **O** | | | | | | | | | |
| Kusuku-Mambilla | Taraba | | | | | | | | | | | | | | **O** |
| Lagos | Lagos | | | | | | | | | **O** | | | | | |
| Malamfatori | Borno | | | | | | **O** | | | | | | | | |
| Maro | Kaduna | | | | | | | **O** | | | | | | | |
| Maiduguri | Borno | | | | | | **HQ** | | | | **O** | | | | |
| Mayo-selbe | Taraba | | | | | | | | | | | | | | **O** |
| Mokwa | Niger | | **O** | | **O** | | | | | | | | | | |
| New Bussa | Niger | | | | | | | | | | **HQ** | | | | |
| New-Marte | Borno | | | | | | **O** | | | | | | | | |
| Ngala | Borno | | | | | | **O** | | | | | | | | |

| Location | State | Int'l Institute of Tropical Agricul-ture | National Cereals Research Institute | National Animal Produc-tion Research Institute | Institute of Agricul-tural Research | Institute of Agricul-tural Research & Training | Lake Chad Research Institute | National Root Crop Research Institute | National Horticul-tural Research Institute | Stored Product Research Institute | National Institute for Fresh-water Fisheries Research | National Veterinary Research Institute | Rubber Research Institute of Nigeria | Nigerian Institute for Oil-Palm Research | Cocoa Research Institute of Nigeria |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Numan | Adamawa | | O | | | | | | | | | | | | |
| Nyanya | Federal Capital Territory | | | | | | | O | | | | | | | |
| Obudu | Cross River | | | | | | O | | | | | | | | |
| Ochaja | Kogi | | | | | | | | | | | | | | O |
| Oguta | Imo | | | | | | | | | | O | | | | |
| Okigwe | Imo | | | | | | | | O | | | | | | |
| Okondi | Cross River | | | | | | | | | | | | | | O |
| Onisere | Ondo | | | | | | | | | | | | | | O |
| Orin-Ekiti | Ekiti | | | | | O | | | | | | | | | |
| Otobi | Benue | | | | | | | O | | | | | | | |
| Owena | Ondo | | | | | | | | | | | | | | O |
| Port Harcourt | Rivers | | | | | | | | | O | | | | | |
| Riyom | Plateau | | O | | | | | | O | | | | | | |
| Samaru | Kaduna | | | HQ | O | | | | | | | | | | |
| Sapele | Delta | | | | | | | | | O | | | | | |
| Talata Mafara | Zamfara | | | | O | | | | | | | | | | |
| Tiga | Kano | | | | | | | | | | O | | | | |
| Uba | Adamawa | | | | | | O | | | | | | | | |
| Ugbenu | Anambra | | | | | | | | | | | | | | O |
| Uhonmora | Edo | | | | | | | | | | | | | | O |
| Umudike | Abia | | | | | | | HQ | | | | | | | |
| Uyo Ubo-ukuku | Akwa-Ibom | | O | | | | | | | | | | | | |
| Vom | Jos | | | | | | | | | | | HQ | | | |
| Warri | Delta | | O | | | | | | | | | | | | |
| Yandev | Benue | | O | | | | | | | | | | | | |
| Yauri | Kebbi | | | | | | | | | | O | | | | |

Source: Author's compilations from the Agricultural Research Council of Nigeria (ARCN) website.