Who Supplies the Nutrients? The puzzle of crop zinc heterogeneity and low-zinc market crops in Uganda

By Leah EM Bevis*

December 28, 2015

Micronutrient deficiencies impair the health and productivity of over 2 billion people worldwide, yet reliable estimates of these "hidden" deficiencies are scarce. Rather than measuring micronutrient status directly, we often rely on food recall data and food composition tables (FCTs) to calculate vitamin or mineral intake for a given population. FCTs report sample-specific mean nutrient content by food, ignoring that nutrient content is a distribution and shifts over space and time. For this reason micronutrient intake estimated via FCTs will tend to under-estimate deficiency prevalence, and will fail to detect key vulnerable populations. In rural Uganda, crops sampled at market are far lower in nutrient status than newly harvested crops sampled from homes; children reliant on those market-purchased crops are differentially vulnerable to zinc deficiency. This paper models crop zinc content for farms across Uganda using a unique dataset containing crop nutrient content and a larger, nationally representative panel dataset. It explores the spatial, household-specific, and time-varying factors that drive selection into the staple market, and therefore influence crop zinc content at market in any given season. The low zinc content of market crops is explained primarily by regional selection into market, but zinc content at market also shifts over time. Temperature variation, in particular, drives selection to market and shapes the nutrient content of market crops.

*Charles H. Dyson School of Applied Economics, Cornell University. *Email*: leb99@cornell.edu *Phone*: 614.288.4008 *Acknowledgements*: Special thanks to the UNC survey PIs Clark Gray, Ephraim Nkonya, Darrell Shultze, Chris Barrett and Leah VanWay. Thanks to HarvestPlus, the National Science Foundation and Cornell CIIFAD for funding, and to the Kampala IFPRI and HarvestPlus offices for support. Thanks to all the UNC surveyors, particularly Agaba Choice and Sentumbwe George. Thanks to Mike Rutzke, Ross Welch and Raymond Glahn for guidance, to Tembi Williams, Maia Call and Tonny Bukeera for research assistance, and to Chris Barrett, David Just, Shanjun Li and various seminar participants for feedback on earlier drafts. All remaining errors are solely my responsibility. *Keywords:* micronutrients, minerals, zinc, deficiency, crops, soil, market participation *JEL Codes:* I14, I15, Q12, Q18, O13

1 Introduction

The most recent FAO estimates indicate that 840 million people worldwide lack energy from their diets — they are hungry. Yet the global toll of "hidden hunger," those suffering from micronutrient deficiencies, is far higher, probably exceeding two billion people (Kennedy, Nantel and Shetty, 2003; Horton, Alderman and Rivera, 2009). Most of these individuals suffer from insufficient intake of vitamins or minerals despite eating an adequate number of calories. In some cases visible indicators, such as goiter or blindness, mark their deficiencies. But for the majority of those suffering, micronutrient deficiencies are impossible to detect clinically, hence the term "hidden hunger."

Yet the ramifications of these invisible deficiencies are far-reaching, and grave. Micronutrient malnutrition leads to a sustained loss of productivity, depressed immune functioning, blindness, and an increased risk of maternal and infant mortality. Micronutrient malnutrition in utero or during early childhood is particularly dangerous, irreversibly degrading both physical and cognitive capacity (Black, 2003).

Despite the prevalence and risks associated with micronutrient malnutrition, precise deficiency rates are difficult to estimate. Clinical diagnosis of micronutrient deficiencies is often impossible, and testing for micronutrient status is expensive and in some cases infeasible. For this reason, regional or national prevalence rates are generally estimated via dietary micronutrient intake (Joy et al., 2014). This calculation relies on food composition tables (FCTs), which define the "average" nutrient intake associated with consumption of any given food (USDA, 2013).

Yet the micronutrient content of food is not fixed; it is highly heterogeneous, a conditional distribution that shifts over space and time.[1] Because FCTs do not capture this heterogeneity, intake estimates that rely on FCTs will not capture the true heterogeneity in micronutrient intake, nor in micronutrient deficiency rates. Populations dependent on lower-than-average nutrient staples are at higher risk for micronutrient malnutrition, but this will be missed in FCT-based estimates.

Additionally, consumers lack the information that might propel them to demand higher-nutrient crops, as US consumers currently demand iodized salt or vitamin D enriched milk. While producers and traders may possess fuzzy information on crop nutrient content (through knowledge about crop variety, management techniques, etc.), consumers do not. Yet Ugandan consumers do value micronutrient-dense foods once information on nutrient content is provided (Chowdhury et al., 2011), and consumers with lower dietary diversity may place differentially greater value on enriched foods (Stevens and Winter-Nelson, 2008). Improved information regarding the heterogeneity of crop micronutrient content could therefore catalyze a shift in consumer demand for higher-nutrient crops, as well as improving the capacity of governments/non-profits to target vulnerable populations.

[1] Crop nutrients vary according to soil nutrients and soil pH, rainfall and temperature, agricultural management techniques and trade patterns. In Malawi, low-zinc soils and maize puts semi-subsistence farming families at risk of zinc deficiency (Chilimba et al., 2011, 2012). In Bangladesh, rice farmers with low-zinc soil and low-zinc rice have lower zinc status themselves (Mayer et al., 2007). English and Belgian populations dropped to dangerously low selenium status when European Union trade restrictions barred the import of high-selenium wheat from the United States.

This paper begins to fill this knowledge gap, documenting and explaining the heterogeneity of crop zinc content in the context of rural Uganda. I provide evidence that this heterogeneity, combined with spatial trends in production and marketing, causes certain types of families to be particularly vulnerable to zinc deficiency in a way that standard estimates do not capture, and that the families themselves are unlikely to recognize.

In particular, I investigate a curious empirical finding: that staple crops sampled from rural markets across Uganda appear to be lower and less variable in zinc content than staple crops sampled from rural household farms. The difference is so great that apparent risk of zinc deficiency doubles within a population of children under 5 if the zinc content of food is defined according to the mean zinc content of market-purchased crops rather than farm-sampled crops.[2] Such a difference in estimated deficiency rates suggests that unobserved heterogeneity of crop zinc content leaves certain children — those particularly dependent on market-purchased crops — at much higher risk for zinc deficiency than a standard estimate would account for.

Why is the zinc content of staples at market so low? Part of the market-household differential can be explained by regional selection into marketing — a tiny minority of farmers supply the bulk of staples at market, and these farmers tend to grow crops slightly lower in zinc content than average.

To show this, I estimate a production function for crop zinc using unusually rich data on crop nutrient contents, household characteristics and geospatial factors. I choose the optimal form for this function via cross-validation and then project crop zinc content into a larger, nationally-representative panel dataset in a manner similar to Small Area Estimation (Elbers, Lanjouw and Lanjouw, 2003).

By combining these zinc content predictions with survey data on crop production and crop marketing, I compare the predicted zinc content distribution of produced crops (grown for consumption and/or market) to that of marketed crops. The result mimics what was observed in crop samples: the zinc content of crops sold to market is significantly lower, on average, than the zinc content of crops produced on farms across the country. For certain crops (e.g. maize, cassava) the predicted differential explains a significant portion of the observed household-market differential. For other crops (e.g., legumes), the predicted differential is smaller.

Additionally, the household-market differential varies over time, driven by shifts in the spatial pattern of supply. Although the differential is large and statistically significant in most seasons, in a few seasons the distribution of crop zinc content at market nearly mirrors the underlying distribution of crop zinc content in all produced crops. To examine time-varying factors that might shift the market distribution, I estimate a triple hurdle model of crop supply to market, using three rounds of data between 2009 and 2011. Temperature and rainfall shocks are incorporated into the model, which includes three stages of farmer decisions: the production decision, the selling decision, and the sale quantity conditional on production and selling.

[2] A child is at risk of zinc deficiency if their intake falls below the Recommended Daily Allowance (RDA) for zinc — i.e., the intake necessary to meet the needs of 95 percent of the relevant population. The RDA for zinc ranges from 2 mg/day (children under 6 months) to 5 mg/day (children 4-8 years).

In this model, temperature shocks significantly influence supply, allowing for the possibility that, if distributed non-randomly with respect to spatial patterns in crop zinc content, they may change the mean zinc content of crops at market. Temperature variation during the period directly prior to crop sampling is examined for low-zinc, medium-zinc, and high-zinc farmers. Positive temperature shocks were experienced by low-zinc farmers during this period, shocks about twice the magnitude of high zinc farmers. This additional stimulus may have further decreased the zinc content of crops at market during the period when samples were gathered.

While the precise spatial patterns examined here are specific to Uganda, the underlying conditions are common to almost any developing country. Crop nutrient content varies across space in almost any context (Mayer et al., 2007; Chilimba et al., 2011; Sillanpää, 1972), and a very small, non-random selection of farmers supply the market for staples in most developing countries and certainly in sub-Saharan Africa (Barrett, 2008). This suggests that families consuming low-nutrient staples, whether via auto-consumption or purchases, may fall through the cracks of traditional deficiency estimates. This also suggests that the nutrient content of crops at market is unlikely to be representative of a broader, country-wide distribution. The disparity between these distributions will vary by context, however, and across time.

The paper proceeds as follows. Section 2 provides evidence on heterogeneity of crop zinc in rural Uganda, provides agronomic background on the inputs to crop zinc production, and outlines three hypotheses regarding the observed household-market differential in crop zinc content. Section 3 provides an overview of the datasets used in this paper. Section 4 estimates and then projects estimates of crop zinc content, for each of 6 staple crops, in order to compare the distributions of zinc content in produced and marketed crops. Section 5 estimates staple supply to market and the average partial effect of random weather realizations, in order to examine how weather shocks temporarily shift the distribution of zinc content in marketed crops. Section 6 concludes.

2 Motivation and Hypotheses

Zinc deficiency is one of the most common micronutrient deficiencies, with over 2 billion people estimated to be at risk worldwide (Hotz and Brown, 2004).[3] It is also one of the most dangerous. Zinc deficiency in utero or during childhood inhibits skeletal growth, retards cognitive capacity, depresses motor development, is associated with diarrhea and acute lower respiratory infections, and delays sexual maturity in adolescence (Prasad, 2003; Hotz and Brown, 2004). Zinc and vitamin A deficiency alone are estimated to account for 9 percent of disability-adjusted life-years (DALYs) for children under five (Black et al., 2008), and zinc deficiency is a primary cause of child stunting (Brown et al., 2002; Imdad and Bhutta, 2011; Brown et al., 2013).[4] The association is so strong, in fact, that the International Zinc Nutrition and Consultative Group (IZiNCG)

[3] Its prevalence surpasses that of iodine deficiency, suffered by slightly under 2 billion people (WHO, 2004), may be on part with that of vitamin A deficiency, impacting a third of preschool age children and fifteen percent of pregnant women (WHO, 2009), and approaches that of iron deficiency, estimated to impact between 4 and 5 billion people, with 2 billion severely deficient, or anemic (WHO, 2008).

[4] Stunting is defined as child height-for-age between -2 and -3 z-scores below the median World Health Organization child growth standards. Severe stunting occurs below -3 z-scores.

considers widespread stunting to be a key indicator for likely zinc deficiency (Hotz and Brown, 2004).

The effects of zinc deficiency are also severe for adults, who lose muscle mass under zinc deficiency in order to release zinc for maintenance of vital organs. Adult zinc deficiency is associated with a number of diseases/conditions including chronic liver disease, diabetes, and macular degeneration (Prasad, 2003). Zinc deficiency is also associated with a loss of appetite, therefore contributing to deficiency in other nutrients (Hotz and Brown, 2004). Thus, zinc deficiency in children is capable of irreversibly degrading future productivity, while zinc deficiency in adults decreases current productivity.

Moreover, mild or moderate zinc deficiency is almost impossible to diagnose clinically, as it presents with few observable symptoms besides stunting. Zinc is so closely regulated by the body that it is difficult to diagnose even with blood plasma tests, which basically indicate dietary zinc intake the day before testing, but not true zinc status (Fischer Walker and Black, 2007). (Ideally bone tissue is analyzed). For this reason, and because of the diverse functions in the human body that depend on zinc and the diverse symptoms of zinc deficiency, Graham et al. (2007) call zinc deficiency the "ultimate hidden hunger." This is also why we generally estimate risk of zinc deficiency by measuring dietary zinc intake.

Calculating dietary zinc intake relies, of course, on food composition tables (FCTs). For instance, work by HarvestPlus regarding dietary nutrient intake in Uganda defines food nutrient content according to an FCT that was particularly tailored for Ugandan dishes, but also draws food nutrient contents primarily from the US Department of Agriculture (USDA) FCT (Hotz et al., 2012*b*,*a*; USDA, 2008). If the mean zinc content of Ugandan food items differs from the mean zinc content provided by the USDA FCT, zinc intake and deficiency rates may be systematically under- or over-estimated.

Crop samples from Uganda illustrate the heterogeneity of crop zinc content, as well as the fact that FCT values do not necessary represent true mean or median crop zinc values. Figures 1 and 2 illustrate the zinc content distribution of maize and cassava, which were sampled at both smallholder farms and markets in communities across Uganda. Sampling occurred during harvest period, and in the weeks directly following harvest — the crops sampled from smallholder farmers were therefore grown both for auto-consumption and for sale to market. Sales to market had not yet occurred, as the post-harvest drying and processing period had not yet taken place. Crops sampled from markets can therefore be thought of loosely as a sub-set of crops sampled from households, though from one or two seasons prior. They are the crops that were produced on smallholder farms, but eventually sold to market.

In both Figures 1 and 2, the distribution of zinc content for household farms is wide, with a long tail to the right. The USDA FCT values for maize zinc content under-estimate both mean and median zinc content for household producers; the FCT value for cassava zinc content captures median zinc content, but under-estimates mean zinc content. Appendix 1 provides additional figures illustrating crop zinc heterogeneity — the FCT crop zinc content values under-estimate median crop zinc for farmer-grown sorghum, over-estimate median crop zinc for farmer-grown sweet potato, and almost perfectly estimate median crop zinc for farmer-grown groundnuts.

Figures 1 and 2 illustrate a second pattern: the zinc content of crops found at market appears to be lower and less variable than the overall zinc content of produced crops that were sampled on farms. Median cassava zinc content is 40 percent lower at market, and median maize content is 83 percent lower. While these were the only two crops sampled at both homes and markets, suggestive evidence points to similar patterns for other crops. More details can be found in Appendix 1.

In a society where, for instance, the majority of zinc intake comes through the consumption of animal-sourced foods, mis-specifying the zinc content of crops might not significantly change calculated zinc intake or estimated zinc deficiency rates. In societies where minerals are drawn largely from plant-based foods, as in much of sub-Saharan Africa, the consequences of such mis-specification may be drastic.

This is the case in Uganda. Figure 3 illustrates three kernel density distributions for child zinc intake (mg/day), both calculated according to identical food recall data collected from 299 Ugandan children under 5 years of age. One of the distributions values the zinc contribution of crops according to the mean zinc content of crops sampled at farms. The other values this contribution according to the mean zinc content of crops sampled at markets. The third values the zinc content of crops according to the HarvestPlus FCT. Zinc intake appears highest, on average, when calculated via the mean zinc content of household samples — median zinc intake appears to be 3.21 mg/day when market samples define content, but appears to be 5.91 when household samples define content. When calculated via the FCT, median zinc intake appears to be 4.04 mg/day.

These differences have serious implications for gauging zinc deficiency in children. Figure 4 illustrates three kernel density distributions for child zinc adequacy, measured as calculated intake over age-specific adequate intake. When household samples define content, it appears that 28 percent of children are at risk of zinc deficiency. When market samples define content, this figure rises to 55 percent. When the FCT defines content, 49 percent of children appear to be at risk of zinc deficiency. Appendix 2 holds more information on both of these figures.

Why does the zinc content of crops sold to market differ from the zinc content of crops sampled on farms, directly after harvest? While some hypotheses can be easily discarded, others seem more plausible and can be tested using existing data. Still others may be plausible, but are untestable in the available data.

First, it might be possible that crop zinc content is a credence good; producers may observe crop zinc while consumers do not. If this were the case, producers might selectively sell low-zinc crops to market and keep high-zinc crops at home for consumption. Hoffmann et al. (2013) isolates such a crop-level selection process in Kenya, where smallholder farmers differentially sell crops that are more likely to be aflatoxin-contaminated, and keep less contaminated crops for home consumption. Yet they do this based on visual indicators of aflatoxin contamination. Crop zinc content is not correlated with visual indicators, except in examples of extreme deficiency not observed in these data. It seems unlikely, therefore, that farmers sort their produced crop according to zinc content.

Second, it is well known that certain micronutrients degrade over time. Staples purchased from market were likely to be older than staples sampled at homes; while samples from household farms had just reached maturity and been harvested during the summer of 2013, market samples might have been drawn from the first harvest of 2013 or even the second harvest of 2012. For a nutrient that degrades over time, this household-market time lag might explain the differential between freshly-harvested crops found on farms and potentially older crops found at market. Minerals such as zinc, however, do not degrade over time. Such degradation occurs in vitamins, as well as in certain macronutrients such as proteins, but not in trace metals/minerals such as zinc (Favell, 1998).

Third, crop mineral contents change significantly with variety (Velu et al., 2007, 2011). In particular, it is long established that, in many cases, variety yield is inversely associated with variety mineral and vitamin content (Grzebisz, 2011; Davis, 2009; White and Broadley, 2005; Mayer, 1997; Farnham et al., 2004). Known at the "dilution effect," studies of this inverse ratio generally observe that higher-yielding varieties accumulate relatively more dry matter than nutrients (Farnham et al., 2004). For instance, wheat grain in the UK has been declining in zinc content since the 1960s, a trend that coincided with the introduction of semi-dwarf, high-yielding cultivars (Fan et al., 2008). In such high-yielding cereal varieties, increased carbohydrate production lowers the mineral:carbohydrate ratio. If larger, commercially-focused farmers are more likely to grow hybrid or high-yielding crops, this might explain the differential for cereals, at least. Unfortunately, the variety data necessary to test this hypothesis are not available.

Fourth, crop mineral content is shaped by soil characteristics and soil management techniques (Hotz and Brown, 2004; Shivay, Kumar and Prasad, 2008; Sillanpää and Vlek, 1985; Bevis, 2015). In some cases, crop minerals vary primarily on a regional scale. In China, for instance, human selenium status ranges from <10 mcg/day in areas with selenium deficient soils, to 5000 micg/day in areas with extremely high levels of soil selenium (Thomson, 2004). In other cases crop minerals vary on a household-to-household basis. In rural Bangladesh, for instance, the zinc content of rice grown by smallholder families, as well as the zinc status of the smallholder family members, tracks the soil zinc concentration in rice plots (Mayer et al., 2007).

The established spatial heterogeneity in crop zinc suggests that selection into marketing, if correlated with crop zinc content, could lead to either particularly low (high) zinc content at market if the main suppliers come from areas of particularly low (high) zinc soils. Such a phenomenon seems particularly plausible given that a small minority of farmers supply the market in most sub-Saharan African countries. For instance, less than a quarter of Ethiopian families were found by Levinsohn and McMillan (2007) to be sellers of any common grain, a third of Tanzanian farmers were found to be net sellers of food (Sarris, Savastano and Christaensen, 2006), a quarter of Zambian farmers were found to be net sellers of maize (Jayne, Zulu and Nijhoff, 2006), and a quarter of Malagasy farmers were found to be net sellers of rice (Minten and Barrett, 2008).

This selection into the marketing process is non-random. Propensity to sell staples, net sales and gross sales are all correlated with geographic location and transaction costs (Barrett, 2008; Heltberg and Tarp, 2002; Renkow, Hallstrom and Karanja, 2004; Dorosh et al., 2010), making it plausible that regional selection into selling could influence the nutrient content of crops at market. Farm and farmer characteristics also influence crop sales — for instance, net sales are strongly, positively associated with wealth, land-holdings, and transportation-related assets (Barrett and Dorosh, 1996; Jayne et al., 2001; Levinsohn and McMillan, 2007; Boughton et al., 2007). These factors are often found to be correlated with land management practices and so potentially with soil zinc concentration and crop zinc content.

These two established phenomenon — spatial heterogeneity in crop zinc and non-random selection into crop marketing — suggest two, related mechanisms through which crops at market might end up lower in zinc content than the national average.

Hypothesis 1: The primary suppliers in Uganda grow lower-zinc crops than the average farmer, due to household or regional selection into marketing.

A number of mechanisms might lead to household-level heterogeneity in crop zinc within any given region. Soil pH and soil management practices vary greatly among farms and even among plots within a single farm, and both are known to affect the availability of soil zinc to crops (Sillanpää, 1972). Phosphate fertilizers have also been shown to decrease zinc content in a range of crops (Peck et al., 1980; Moraghan, 1994),[5] and fallowing practices may affect soil zinc concentration, potentially impacting the zinc content of crops (Allaway, 1986). If the primary market suppliers in any given region have particularly basic or particularly alkaline soils, or engage in agricultural practices that impede zinc uptake for other reasons, this might lead to their growing lower-zinc crops.

Crop zinc is equally likely to vary across regions, rather than within them. This could occur due to regional variation in soil or rock type, precipitation levels, soil zinc content or soil pH (Sillanpää, 1972). In this case regional selection into the market, rather than household-level selection, might cause the observed disparity.

Hypothesis 1 implicitly assumes that the primary market suppliers are fixed over time. And many predictors of crop sales are time-invariant, at least in the short run — soil type, land-holdings and access to roads, for instance. But some predictors vary across time. Rainfall and temperature shocks, in particular, might change the distribution of sellers for a single season only. If weather variation substantially alters the geography of staple production or marketing for a season, then the zinc content of market staples might adjust in a corresponding manner. This leads to the second hypothesis.

Hypothesis 2: Low-zinc sellers were more likely to supply the market only in late 2012 or early 2013, due to weather shocks during that period, creating a temporary dip in the zinc content of crops at market.

Unfortunately, the panel dataset into which crop zinc content is projected does not include survey rounds for 2012 or 2013. It is not possible, therefore, to examine whether

[5] The application or even natural presence of minerals such as phosphorus, cadmium, or calcium may impact soil zinc availability to the plant in two ways. Primarily, these minerals compete with zinc for uptake by the roots. Additionally, certain minerals bind with zinc to form compounds less available to plants. Both types of interaction are intermediated by soil pH. For instance, within pH range 6-7, zinc and calcium tends to form insoluble calcium zincates, which makes soil zinc less available to plants (Sillanpää, 1972).

the low-zinc farmers truly did sell more than usual to market in this period. It is possible, however, to observe weather shocks during this time period, using highly detailed geo-spatial data on precipitation and temperature. Estimating the causal association between weather shocks and supply in the panel dataset allows us to examine whether 2012 and/or 2013 weather shocks might have temporarily shifted the geographic distribution of crop suppliers in order to further decrease the zinc content of crops at market during the sampling period.

3 Data

This paper utilizes three rounds of the Ugandan National Panel Survey (UNPS), from 2009, 2010, and 2011.[6] Families within the UNPS sample complete a socio-economic survey as well as a detailed, plot-level agricultural survey. Because Uganda has two agricultural seasons, each family in the UNPS is visited twice during each year, and the agricultural survey (including extensive measurement of agricultural production and sales) is conducted during each visit. The household survey is conducted only once per year — during the first visit for half of the families in each enumeration area, and during the second visit for the other half. For the purposes of seasonal agricultural production and supply, each household is therefore observed up to six times. Including all households from 2009, 2010, and 2011 provides a total of 17,080 household-season observations.

This paper also utilizes soil and crop data from an NSF-funded survey run by the University of North Carolina (UNC) in 2013. The UNC survey collected household-level socio-economic data and plot-level agricultural data for both seasons of 2012, from a total of 839 households. It also gathered plot-level soil samples for macronutrient analysis.

In half of the survey districts, additional soil samples as well as crop samples were collected for micronutrient analysis. Six staple crops were sampled at the household directly during and after harvest: maize, sorghum, sweet potato, cassava, beans and ground nuts. Because these six crops were sampled from households at harvest, they might potentially have gone on to be sold or to be consumed at home. The distribution of crop zinc content drawn from household samples, therefore, is best viewed as a representation of the distribution of zinc content for all produced crops. Additionally, samples were taken from rural, sub-county markets, as well as households. Maize flour and cassava flour were sampled at the market level, along with a number of other crops and food items.

Because both the UNC and the UNPS data are geo-referenced, geo-spatial data on soil and climate conditions can be extracted for each household in both datasets. Additionally, because they contain similar agricultural and household survey modules, a number of survey variables exist for both datasets. Using these jointly available covariates, crop zinc content can be modeled in the UNC dataset and predicted into the

[6]The UNPS is one of a collection of nationally representative panel surveys called the Living Standard Measurement Survey — Integrated Surveys in Agriculture (LSMS-ISA). It is implemented by the national statistics office with technical support from the World Bank Development Economics Research Group.

UNPS dataset, in a manner similar to Small Area Estimation (Elbers, Lanjouw and Lanjouw, 2003).

Table 1 reports summary statistics for all such covariates. The top four variables are sourced from geo-spatial data, and the bottom variable (farm productivity) is taken from UNC and UNPS survey data. Because the UNC data do not include households in the central region of Uganda, the last two columns of Table 1 provide summary statistics for the non-central UNPS households.[7]

Additional UNPS data is used to estimate a triple hurdle model of farmer supply to market. Table 2 displays summary statistics for the additional UNPS variables used in this model.

Geo-spatial raster data on soil characteristics were procured from the Africa Soil Information Service (AfSIS) / World Soil Information project (Hengl et al., 2015). Rasters of season-average rainfall means, season-average rainfall standard deviation, and season-specific rainfall realizations were created using Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data, compiled by month and gridded at 0.05 decimal degrees.[8] Rasters of season-average temperature means, season-average temperature standard deviation, and season-specific temperature realizations were created using Era-interim re-analysis data from the European Center for Medium-Range Weather Forecasts (ECMRWF). These data are predicted daily within a 0.75 degree grid, and then interpolated down to a 0.25 degree resolution.

Growing degree days (DD) are calculated for each day, and then averaged within seasons to indicate seasonal temperature. They are calculated in this manner, where T is the temperature in degrees Celsius:

$$DD(T) = \begin{cases} 0 & \text{if T} \leq 8^\circ \text{ C} \\ T - 8 & \text{if } 8^\circ \text{ C} \leq \text{T} \leq 32^\circ \text{ C} \\ 24 & \text{if T} \geq 32^\circ \text{ C} \end{cases}$$

This formula captures "useful" heat to crops, as most plants cannot absorb heat below 8 degrees Celcius or above 32 degrees Celsius, but absorb heat within that range in a linear manner (Dell, Jones and Olken, 2013).

Figures 3 and 4 are generated using food recall data from a Food Frequency Questionnaire (FFQ) administered to 237 children, each under 5 years of age and living in UNC households for which soil and crop samples were gathered. During the FFQ child care-takers stated how many times their child had consumed each of 53 selected, zinc-rich foods over the course of the last seven days, and the average portion size for each food consumed. In 99 cases a second food recall was conducted, extending the food intake dataset to 336 observations.[9] More information on the FFQ, and the construction of zinc intake from these food recall data, can be found in Appendix 2.

[7]Uganda is broken up into four regions: Central, Western, Eastern and Northern. Each regions is broken into districts, which are broken into counties, then sub-counties, then parishes, then villages.

[8]A decimal degree express latitude and longitude geographic coordinates as decimal fractions. Values around bounded, as with latitude and longitude, by $\pm°90$ and $\pm°180$, respectively.

[9]These 99 children were revisited partway through the survey, as their heights had been improperly recorded by the survey instrument.

4 Producing Crop Zinc

4.1 Estimation Strategy

To test Hypothesis 1, I model observed crop zinc content in the small UNC dataset and then predicted crop zinc content for all farmers within the larger, nationally representative UNPS dataset. This allows comparison between the zinc content of crops grown by the average farmer, by farmers who sell some crops to market, and by major crop sellers. Such comparisons must be done in the UNPS dataset because the UNC dataset is not nationally representative.

This technique, estimating crop zinc in a small, rich dataset in order to predict into a larger, nationally-representative dataset, borrows from Small Area Estimation techniques (Elbers, Lanjouw and Lanjouw, 2003). Two assumptions are crucial for valid predictions. First, the covariates used to estimate zinc in the UNC data and the covariates used to predict zinc into the UNPS data must be measured identically across the two datasets. Geo-spatial variables created in R or ArcGIS, such as interpolated climate or soil qualities or distance to roads, are guaranteed to be measured identically as they are drawn from the same external dataset. Household survey variables such as household head education or farm size must be obtained via the same questions and techniques. Because the UNC questionnaire was modeled on an earlier version of the UNPS surveys, many of the survey modules are nearly identical.[10]

Second, predictive power, measured by Mean Squared Forecast Error (MSFE), must be maximized. The goal is not to obtain unbiased covariate estimates, as would be true if the purpose of estimation was inference about parameters, nor even to maximize R^2. Rather, it is to model crop zinc in such a manner as to minimize error when zinc is predicted into the UNPS dataset. Over-fitting the model with too many covariates, while increasing R^2, will simultaneously increase MSFE. For this reason that Small Area Estimation models are generally parsimonious, including a few variables at most.

To minimize MSFE, one might use cross-validation to test all possible Ordinary Least Square (OLS) models — every permutation of covariates, including all models with one covariate, all models with two covariates, all models with three covariates, etc. It is equivalent and more efficient, however, to minimize MSFE within a rotated space, where the OLS covariate matrix is decomposed via single value decomposition (SVD) into a new matrix of orthogonal eigenvectors and matching single values (eigenvalues). In this rotated matrix, candidate solutions to the OLS minimization are chosen by allowing the number of single values and matching eigenvectors, k, to vary from 1 to k_{max}, where k_{max} is the total number of non-zero singular values (Lawson and Hanson, 1974).

Candidate models with a low k are parsimonious. Candidate models with a larger k are not, and while they provide a better fit to the original data, the solutions are less stable, and after a point have lower forecast power. In this rotated space, an optimal k^* may be chosen in order to minimize MSFE. Essentially, the SVD aids in the end goal of

[10] The UNPS surveys were largely modeled on a 2002/2003 survey run by UBOS. The UNC survey is the second wave of a panel, where the first wave of data was partially drawn from that 2002/2003 survey. This means that many UNPS modules are identical to that of the 2003/2003 UBOS survey, and therefore later UNPS surveys.

choosing the "correct" model by first generating the most parsimonious model for every possible size of model. The size of the model (the number of candidate vectors, k) is then chosen by using cross-validation to minimize MSFE. Appendix 3 outlines the SVD approach in greater detail.

While the precise data generating process for crop zinc is unknown, potential covariates surely include data on soil conditions, crop variety, temperature and precipitation, and all interactions between these variables. Ideally, the model would include data on household-specific soil zinc and soil pH; because the UNPS surveys do not collect soil samples, this is impossible. Therefore, I use geo-spatial data on area-level soil conditions, as they can be extrapolated to both datasets. The Africa Soil Information Network (AfSIS) soil quality rasters include data on soil pH, but not soil zinc. However, soil calcium is highly correlated with soil zinc — in the UNC measured soil data, soil zinc concentration and soil calcium concentration have a bivariate correlation coefficient of 0.507[11] — and soil calcium is included in the AfSIS rasters. It is therefore used to proxy for soil zinc concentration.

Covariates in the crop zinc model therefore include geo-spatial soil calcium and geo-spatial soil pH, average growing degrees per year, average precipitation, and interactions among all of these variables.

Additionally, a number of household-level covariates are also initially included: dummy variables for bicycle ownership, brick walls, a male head of household, the use of organic amendments on soil, the use of inorganic amendments on soil, and the existence of erosion-controlling structures on the farm, as wells as household head age (quadratically) and a measure of farm productivity: crop value per hectare, calculated at the plot level and then averaged across plots. Because AfSIS soil characteristics can only capture landscape-level variation, these household-level variables, known to be associated with soil management and nutrient loss, proxy for household-specific deviations from the area mean. If these household-specific factors are significantly associated with impact soil zinc, then farmer-level selection into the market may play a role in explaining the low zinc content of crops at market, as predicted by Hypothesis 1.

Modeling and prediction is done for three sets of crop samples: for cereals (maize and sorghum) for tubers (cassava and sweet potato) and for legumes (beans and ground nuts).[12] Each crop is demeaned individually, however.

Hypotheses 1 and 2 are examined by comparing an unweighted distribution of predicted zinc content for producers, and a sale-weighted distribution of predicted zinc content for sellers. The first distribution reproduces the zinc content one would observe by randomly sampling maize from any farmer who had produced a positive quantity of maize, as done in the 2013 UNC survey. The second distribution weights by selling quantity for all expected crop sellers, thereby reproducing the zinc content one might observe at market if all sales were pooled across the entire country.

[11] In fact, within the UNC sample, crop zinc content can be explained better by interpolated soil calcium and soil pH than by measured soil zinc and soil pH.

[12] Samples are pooled within crop types due to limited sample size. The underlying assumption is that functional form and error distribution are the same within crop type.

4.2 Results

Initially, household-specific covariates are included in the zinc production model. With one exception, however, these covariate are given zero weight by the SVD. (Appendix 4 addresses this further.) It seems, therefore, that area-level factors have a greater influence on crop zinc production than do household-specific factors. One household-specific covariate does predict crop zinc, however — farm productivity, measured in log value produced per hectare. This variable is therefore included in the final models, while the other variable are dropped. The final model set of covariates available for single variable decomposition is therefore soil pH, soil calcium, average precipitation, average temperature, and log crop value per hectare, as well as all interactions amongst these variables, including squared terms.

This matrix of covariates is decomposed into orthogonal vectors, and the optimal vector k is chosen to minimize mean squared forecast error (MSFE). Figure 5 illustrates mean squared forecast error (MSFE), matrix condition number and R^2 for the cereal model, each over k. Appendix 4 holds the same figure for tubers and legumes. Choosing $k = 5$ minimizes MSFE, making this the optimal model for predicting crop zinc content out of sample. In the case of legumes and tubers, the optimal k is 3, and the models for predicting legume and tuber zinc content are chosen accordingly.

Because the in-sample explanatory power of a model, after a certain point, is inversely correlated with out-of-sample predictive power, R^2 is fairly low for each of the three models. After demeaning log crop content by individual crop (individually for maize and for sorghum, for instance, within the cereal sample), the R^2 for explaining remaining variation is 0.042, 0.081, and 0.023 for cereals, tubers and legumes, respectively. (Total R^2 is higher, as crop means explain X, Y, and Z percent of cereal, tuber and legume zinc on their own.) This is similarly the case with Small Area Estimation, where minimizing forecast error is also the primary goal and predictive models are therefore chosen to be parsimonious. For instance, Chandra (2013) chooses two explanatory variables that predict 26 percent of outcome variation.

With a low R^2, the predicted crop zinc content for any individual farmer is, of course, wrong. Yet the goal is merely to create and compare averages across groups of farmers — crop sellers and non-sellers, high-sellers and low-sellers. If error around predicted crop zinc is orthogonal to these categories, then comparison of predictions across categories is valid.

Figure 6 illustrates the non-parametric distribution of forecast residuals for farmers who sell cereals, farmers who do not sell cereals, and for major sellers — farmers in the top 25 percent of sales quantity. The distributions are noisy, and appear to be indistinguishable from one another.[13] Averages of predicted crop zinc may therefore be compared across groups of farmers without concern that omitted variables bias these distributions.

Having predicted zinc content into the UNPS dataset, Hypotheses 1 is examined by comparing an unweighted distribution of predicted zinc content for all producers

[13] These histograms show population within each bin. Therefore, column length differs according to category population size — because few people are "major sellers," these columns are lower.

("produced crops"), and a sale-weighted distribution of predicted zinc content for sellers only ("marketed crops"). The unweighted distribution mimics the distribution that would occur if one sampled crops randomly from any household that was a producer. The sales-weighted distribution mimics the distribution that expected at market. If selection on market participation is correlated with predicted crop zinc content, then these distributions will differ.

Table 3 compares the central moments of these distributions for each season of the UNPS panel, for maize. For all but one season, maize produced on farms is higher in zinc, on average and by median, than the maize expected at market, and in each case this difference is significant at the 0.001 percent level. Mean crop zinc content at market is 16 percent lower than mean crop zinc content of all farm-produced maize in the first season of 2009, and 13 percent lower in the first seasons of 2010 and 2011. In the second seasons the difference shrinks — crop zinc content at market is 3 and 9 percent lower, on average, than farm-produced zinc content in the second seasons of 2009 and 2011, and is 6 percent higher, on average, than farm-produced zinc content in the second season of 2010.

Table 4 displays the difference in zinc content between crops found at market and crops grown at home, for all crops and for each individual season as well as by seasonal average. For every crop, there is a notable difference between the expected zinc content at market, when compared to unweighted household production: marketed crop zinc content is significantly lower for almost every crop-season combination. The bottom two lines of this table display statistics from observations pooled across years, within season. For all but beans, the difference is larger in season one than season two. However, the difference is particularly notable for maize (for which the difference in season 1 is twice that of season 2) and cassava and groundnuts (for which the difference in season 1 is triple that of season 2).[14]

Additionally, the season-to-season variation in the market-household differential is striking for certain crops. For sweet potato, it is as low as 3 percent and as high as 29 percent. For cassava, it is as low as 9 percent and as high as 33 percent — not far from the 40 percent differential observed in the analyzed cassava samples.

Tables 3 and 4 support Hypothesis 1: data on staple sellers and staple sales quantities, examined in conjunction with predicted crop zinc content for all producers, indicates that marketed staples will be lower in zinc content, on average, than the staples produced by an average Ugandan farmer. The magnitude of this differential varies across crops and time. While farm productivity does predict crop zinc content, all other predictors are regional — household-level covariates appear less important than regional soil and weather patterns (Appendix 4).

Because the predicted distribution of zinc content in marketed crops shifts over seasons, Hypothesis 2 is plausible.[15] The next section therefore estimates the impact of

[14]Excluding the central region from the producer's distribution, as was done when sampling crops in 2013, leads to as slight increase in the household-market differential for most crops, and particularly in season 2. Results available upon request.

[15]Theoretically, it is possible that zinc content in crops might change over time, particularly with relation to weather shocks. Because only cross-sectional data on crop zinc content are available, however, it is

time-varying weather shocks on sales, in order to examine the possibility that rainfall or temperature shocks drove particularly low-zinc farmers to supply the market in the period directly prior to crop sampling.

5 Marketing Crop Zinc

5.1 Estimation Strategy

Because farmers make production and selling decisions sequentially, a triple hurdle model is appropriate for modeling market supply of staples. I model three stages of choice: the binary decision to produce the crop in question, the binary supply decision, and the continuous supply quantity decision. Modeling all three stages of choices allows the external factors such as rainfall or agricultural shocks to influence both production and selling decisions, rather than holding the early stages constant. The model here follows closely the triple-hurdle model used by Burke, Myers and Jayne (2015) to describe farmer production and market participation in Kenya's dairy sector, which itself builds on the ordered tobit model of Bellemare and Barrett (2006). Additionally, it is equivalent to a three-stage version of the "Exponential Type II Tobit Model" discussed by Wooldridge (2010), and I adapt his notation in this paper.

The model differs from that of Burke, Myers and Jayne (2015) or Bellemare and Barrett (2006) in its focus on gross sales, rather than net sales. In the first and second stages farmers are separated into producers, non-producers, sellers and non-sellers (rather than net producers and net sellers), because the nutrient distributions of interest are for all home-produced crops and all crops sold to market.

The participation decisions can be stated as in Equations 1 and 2, where $s_1 = 1$ indicates a decision to produce a given crop, and $s_2 = 1$ indicates a decision to sell that crop. (If $s_1 = 0$, necessarily $s_2 = 0$.) The matrix x_1 includes data on elevation, soil characteristics (soil clay content, soil pH), weather expectations (mean and standard deviation for rain and growing degree days over the decade prior to the panel's start), household endowments (household labor and farm size), other household characteristics (household size, household head age and education, non-farm income status, wealth index, urban/rural status), and transaction costs (distance to major town, distance to major road, distance to Kampala).[16] The matrix x_2 includes all of the same variables, but including realized weather occurrences (coefficients of variation for rainfall and growing degree days, around the means already controlled for).

$$s_1 = 1[\mathbf{x_1}'\gamma + v_1 > 0] \tag{1}$$

$$s_2 = 1[\mathbf{x_2}'\delta + v_2 > 0] * s_1 \tag{2}$$

Following Wooldridge (2010) closely, the latent intensity equation can be written as in Equation 3, and observed outcomes (in this case crop sales) is given by Equation 4. The intensity variable to be estimated is $log(y)$, as in Equation 6, where y denotes crop

not possible to estimate a production function for zinc that accounts for time-varying factors. Variation in crop zinc content over time is therefore likely to be underestimated.

[16]Major towns are those towns with...

sales. The matrix x_3 is identical to x_2 except that it includes transaction costs, relevant for the binary sale decision but not for the sale quantity decision.

$$w^\star = exp(\mathbf{x_3}\,\beta + u) \tag{3}$$

$$y = s_2 * w^\star \tag{4}$$

$$m^\star = log(w^\star) \tag{5}$$

$$log(y) = log(w^\star) \text{ when } y > 0 \tag{6}$$

I allow the error terms to be correlated as in Equation 7, where $\mathcal{N}_3$ signifies a trivariate normal distribution, ρ_{12} give both the covariance and correlation between the stage 1 and stage 2 error terms, ρ_{23} is the correlation between stage 2 and stage 3 error terms, and $\sigma\rho_{23}$ is the covariance between stage 2 and stage 3 error terms.[17] Conditional on v_2, v_1 and u are assumed independent. This structure is a methodological generalization beyond the assumptions of Burke, Myers and Jayne (2015) and Bellemare and Barrett (2006).

$$(v_1, v_2, u) \sim \mathcal{N}_3(\mathbf{0}, \mathbf{\Sigma}_3) \text{ with } \mathbf{\Sigma}_3 = \begin{bmatrix} 1 & \rho_{12} & 0 \\ \rho_{12} & 1 & \sigma\rho_{23} \\ 0 & \sigma\rho_{23} & \sigma \end{bmatrix} \tag{7}$$

Given the covariance matrix of Equation 7, the log likelihood is given by Equation 8. Here, Φ_2 signifies the CDF of a standard bivariate normal (specified in Equation 5) and ϕ_3 signifies the probability distribution function (PDF) of the specified trivariate normal distribution.

$$\begin{aligned} f(s_1, s_2, y|\gamma, \delta, \beta, \rho_{12}, \rho_{12}, \sigma) = {} & \Big[1 - \Phi(\mathbf{x_1}\,\gamma)\Big]^{1[w_1=0]} * \Big[\Phi_2(\mathbf{x_1}\,\gamma, -\mathbf{x_2}\,\delta, \rho_{12})\Big]^{1[w_1=1, w_2=0]} \\ & * \int_{-\mathbf{x_1}'\gamma}^{\infty} \int_{-\mathbf{x_2}'\delta}^{\infty} u_3 * \phi_3(u_1, u_2, u_3, \rho_{12}, \rho_{23}, \sigma) du_2 du_1^{1[w_1=w_2=1]} \end{aligned} \tag{8}$$

Consistent and efficient estimates of all parameters could be obtained through maximizing the likelihood specified in Equation 8. Consistent but inefficient estimates may also be obtained via a two-step process. In the first step, the joint likelihood for the first two stages is maximized, as specified in Equation 9. This likelihood function is identical to that of the bivariate probit under the special case — true in this model — that s_2 is never found to equal 1 when s_1 is 0. Thus, a bivariate probit consistently estimates γ, δ, and ρ_{12}.

$$\begin{aligned} f(s_1, s_2|\gamma, \delta, \rho_{12}) = {} & \Big[1 - \Phi(\mathbf{x_1}\,\gamma)\Big]^{1[s_1=0]} * \Big[\Phi_2(\mathbf{x_1}\,\gamma, -\mathbf{x_2}\,\delta, \rho_{12})\Big]^{1[s_1=1, s_2=0]} \\ & * \Big[\Phi_2(\mathbf{x_1}\,\gamma, \mathbf{x_2}\,\delta, \rho_{12})\Big]^{1[s_1=1, s_2=1]} \end{aligned} \tag{9}$$

In the second step, β, σ and ρ_{23} must be estimated. For those with production and sales greater than zero, expected log sales can be given as in Equation 10. However,

[17] The parameter ρ_{12} gives both the covariance and correlation because stage 1 and stage 2 errors are standard normal. The term $\sigma\rho_{23}$ gives covariance between stage 1 and stage 2 errors for the same reason.

because v_1 is independent from u, conditional on v_2, the expectation within Equation 10 may be re-written as $E[u|v_2 > -\mathbf{x_2}\,\delta]$. Additionally, $E[u|v_2] = \sigma\rho_{12}v_2$, and $E[v_2|v_2 > -\mathbf{x_2}\,\delta] = \lambda(\mathbf{x_2}\,\delta)$, where λ is the Inverse Mills Ratio. Thus, Equation 10 may be re-written as in Equation 11, and consistent estimates of β and of $\sigma\rho_{23}$ can be obtained by regressing $log(y)$ on $\mathbf{x_3}$ and on $\lambda(\mathbf{x_2}\,\delta)$ for all observations with $log(y) > 0$.[18]

$$E[y|s_1 = 1, s_2 = 1] = \mathbf{x_3}\,\beta + E[u|v_1 > -\mathbf{x_1}\,\gamma, v_2 > -\mathbf{x_2}\,\delta] \tag{10}$$

$$E[y|s_1 = 1, s_2 = 1] = \mathbf{x_3}\,\beta + \sigma\rho_{23}\lambda(\mathbf{x_2}\,\delta) \tag{11}$$

The density of y, conditioned only on x_1, x_2, and x_3, is obtained by multiplying the conditional density $f(y|x_1, x_2, y > 0)$ by $P(y > 0|x_1, x_2) = P(s_2 = 1]$, as in Equation 12 (Wooldridge, 2010) .

$$E[y|x_1, x_2, x_3] = \left(\mathbf{x_3}\,\beta + \sigma\rho_{23}\lambda(\mathbf{x_2}\,\delta) \right) * \Phi_2(\mathbf{x_1}\,\gamma, \mathbf{x_2}\,\delta, \rho_{12}) \tag{12}$$

The derivative of this quantity with respect to x_j represents the total increase in (log) sales that occurs, due to a change in x_j, through shifts in production patterns, shifts in selling patterns, or shifts in the distribution of quantity sold. This derivative is given by Equation 13. Examining the derivative with respect to rainfall or temperature, for instance, tells us how much we can expect to see log sales change (or perhaps shift geographically) in response to random weather realizations or evolving weather expectations.

$$\frac{dE[y|x_1, x_2, x_3]}{dx_j} = \frac{d\left(\mathbf{x_3}\,\beta + \sigma\rho_{23}\lambda(\mathbf{x_2}\,\delta) \right)}{dx_j} * \Phi_2(\mathbf{x_1}\,\gamma, \mathbf{x_2}\,\delta, \rho_{12}) + \left(\mathbf{x_3}\,\beta + \sigma\rho_{23}\lambda(\mathbf{x_2}\,\delta) \right) * \frac{d\Phi_2(\mathbf{x_1}\,\gamma, \mathbf{x_2}\,\delta, \rho_{12})}{dx_j} \tag{13}$$

5.2 Results

Table 5 reports regression results for each stage of the triple hurdle model for maize supply. (Full results for the other crops are available upon request.) Both soil factors and weather factors are significantly associated with decisions at each stage. Temperature expectations, captured by average, season-specific total growing degree days (DD) during the decade previous to UNPS data collection, are positively associated with the probabilities of both producing and selling a crop, though with diminishing returns. The standard deviation of growing degree days, capturing temperature risk, is negatively associated with both decisions. Rainfall expectations appear far less important than temperature expectations, especially in the first two stages.

Temperature realizations positivity, significantly drive both the likelihood of selling (Column 2 of Table 5 and sales quantity (Column 3). Because this variable is randomly distributed, and orthogonal to temperature means, the association estimated is causal.

[18]Standard errors are bootstrapped, in order to be valid for inference (Heckman, 1979). I plan to do this, but have not yet gotten around to figuring out how to bootstrap with survey weights. It's next up on the agenda, however.

It is also logical; farmers who experience an anomalously high number of growing degree days during the production period are likely to capitalize on their good luck by selling crops, or selling more crops, to market. In both stages, however, the association is smaller for those farmers who are further from major towns, and likely face greater transaction costs in the marketing process. Rainfall realizations follow a similar pattern in both Columns 2 and 3, but the effects are only significant in the third stage.

Table 6 reports the coefficient estimates on rainfall realizations for each of the additional 5 hurdle models. These models were specified in precisely the same manner as the model presented in Table 5. (Full results are available upon request.) Rainfall realizations seem to have little impact on crop sales decisions, except in the case of maize and ground nuts. Temperature realizations positively impact sorghum sales decisions, as with maize, but negatively impact sweet potato sales decisions, and negatively impact the binary sales decision regarding cassava. In general, weather realizations have greater impact on decisions regarding cereal sales than they do on decisions regarding tuber and legume sales — logical given that grasses (cereals) are particularly sensitive to weather conditions.

To shift supply, rainfall and temperature realizations must have a significant impact over both the binary and continuous selling decision. The parameter estimate of greatest interest is therefore the average partial effect (APE) of temperature and rainfall, which changes according to distance to town, given the interaction between climate realizations and distance to town.

Figures 7 and 8 display the APEs for rainfall and temperature for cereal crops, the only crops for which weather realizations significantly shift crop supply. As Table 6 suggests, these APEs are greater for farmers closer to market. The difference is quite small though — Figures 9 and 10 display the same APE estimates with bootstrapped 95 percent confidence intervals, and the gradient is almost unnoticeable.

At median distance to market, a positive growing degree realization of one coefficient of variation (CV) results in a 3.5 percent increase in maize supply, and a 1.8 percent increase in sorghum supply. A negative growing degree realization of one CV results in a decrease in supply of the same magnitude. At the same distance quintile, a positive rainfall realization of one CV results in 1.4 percent increase in maize supply, and a 0.7 percent decrease in sorghum supply. These rainfall effects are insignificant, however, and especially so for sorghum. (This is logical given that sorghum is a drought-resistant cereal.)

Figures 11 and 12 illustrate rainfall and temperature variation over time, measured in CVs, and broken out across cereal zinc content quintiles. (Because the zinc content of maize and sorghum are modeled together, their quintiles are identical, while their mean zinc content differs.) Rainfall realizations in 2012 and 2013 do not appear to vary with cereal zinc content.

However, 2012 and 2013 variation in degree days does differ markedly by crop zinc content. The farmers with lower crop zinc content are mostly likely to have received high degree day realizations, resulting in positive marketed supply concentrated among these farmers. The farmers with the lowest crop zinc content are least likely to have received high degree day realizations. This differential is particularly large in the first season of 2012 and the first season of 2013, where the highest zinc farmers would have supplied 0.35 percent more maize than usual to the market.

These figures suggest that random weather variation in 2012 and 2013 may have further shifted down the zinc content distribution of cereals at market. Whether it can fully explain the remaining gap observed between the zinc content of maize sampled from markets and the zinc content of maize sampled from farms across the country is difficult to say. It is notable, however, that random shocks in rainfall and temperature have the ability to shift the zinc content of crops going to market, by shifting the geography of suppliers for a particular season.

6 Conclusion

This paper documents the heterogeneity of crop zinc content in Uganda, and explores how regional selection into staple marketing drives a downward shift in the nutrient content of crops at market. This shift varies by crop and by season, and is particularly notable for maize and for tubers. For instance, maize sold to market during the first agricultural season is 13.5 percent lower in zinc content, on average, than maize grown across the country. Sweet potatoes and cassava sold to market market during that same season are 19.5 percent and 16.7 percent lower in zinc content, respectively, than their counterparts grown across the country. Given Ugandans' heavy reliance on these three crops, such differentials are concerning.

Additionally, time-varying factors such as rain or temperature shocks have the ability to increase/decrease supply from either low-zinc farmers or high-zinc farmers, shifting the distribution of crop zinc content at market for only one or two seasons. Farmers growing low-zinc cereals experienced particularly high numbers of growing degree days in 2010, 2012 and 2013, while farmers growing high-zinc cereals experienced the lowest numbers of growing degree days. Since these shocks are causally associated with sorghum and maize supply, they likely shifted the zinc content of cereal at market to be even lower than normal.

The precise patterns found here cannot be expected in every context. However, these findings as well as decades of literature on crop nutrient heterogeneity and market participation suggest that we should never expect, a priori, that the nutrient content of crops at market should perfectly mirror the nutrient content of crops produced country-wide. It is well established that crop nutrient content varies over space, and particularly with soil nutrient concentration and soil pH, with crop management practices and with crop variety. It is equally well established that a select group of farmers supply the staple market in almost any developing context, and that these farmers are better off than average and own more land than average. They may manage their land differently than the average farmer, or grow different varieties than the norm.

It is logical, therefore, that if factors driving regional selection into marketing (e.g. distance to capital city, soil nutrients, temperature) or factors driving household-level selection into marketing (e.g. land quantity, crop variety) are correlated with crop nutrient content, the nutrient density of crops found at market will differ from the nutrient density of crops found on farms across the country. Moreover, crops produced for auto-consumption will differ across regions and households in terms of their nutrient density. There is rarely reason to assume that the nutrient content of food is identical across two samples.

This type of heterogeneity has a number of policy implications. To begin with, if crop nutrient content varies primarily by region (rather than by household), and if families rely largely on home-produced crops (rather than market-purchased crops), then traditional estimates may vastly under- or over-estimate micronutrient deficiencies for particular regions. Conversely, gathering even low-resolution data on the nutrient density of crops in various regions might greatly increase effective targeting of the most vulnerable populations.

Additionally, in Uganda specifically, children and families highly reliant on market-based crops will be more vulnerable to zinc deficiency than traditional assessments would estimate. This may also be true in other contexts. While access to markets often leads to dietary diversification and a greater reliance on animal sourced foods, low-nutrient staple crops at market may mitigate the positive effects that market access has on micronutrient intake and health. Families reliant on markets for food consumption, but with low dietary diversification (e.g. the urban poor), may be far more vulnerable to micronutrient deficiencies than traditional estimates would portray. This finding provides a new dimension to prior research on the consumption and health implications of market participation.

This paper also highlights the fact that both production and marketing patterns drive the nutrient content of crops at market, making each of these processes a potential area for micronutrient-focused interventions. Biofortified crops, such as those promoted by HarvestPlus, or fortified foods such as iodized salt, enriched oils, or milk with added vitamin D are all examples of production-level solutions to micronutrient deficiencies.

Additionally, trade provides a means for moving nutrients across space, a fact that has been noted in many developed contexts. For instance, many European citizens would be at risk for selenium deficiency if high-selenium wheat was not imported from the United States. In fact, in the early 1990s European Union import restrictions decreased reliance on imported wheat from the US, and selenium status in England, Belgium, and Scotland dropped dangerously (La Daniels and Simmer, 2000; Rayman, 1997).

For any of these production or marketing solutions to be feasible, however, governments and consumers must have access to information regarding the nutrient content of crops. While the geography of crop production and supply influences human micronutrient status and health worldwide, and particularly in developing contexts, policy-makers notice only in the most extreme cases. Consumers, particularly in poor countries, have almost no information on the nutrient content of purchased crops, and are therefore unable to demand changes to the food system that dictates much of their nutrient intake. To better understand how alterations in crop production techniques, market supply patterns and trade patterns might mitigate micronutrient deficiencies, a body of knowledge must first be built around these processes.

References

Allaway, W H. 1986. "Soil-plant-animal and human intrrelationships in trace element nutrition." 465–488. New York, NY:Academic Press.

Barrett, Christopher B. 2008. "Smallholder market participation: concepts and evidence from eastern and southern Africa." *Food Policy*, 33(4): 299–317.

Barrett, Christopher B, and Paul A Dorosh. 1996. "Farmers' welfare and changing food prices: Nonparametric evidence from rice in Madagascar." *American Journal of Agricultural Economics*, 78(3): 656–669.

Bellemare, Marc F, and Christopher B Barrett. 2006. "An ordered Tobit model of market participation: Evidence from Kenya and Ethiopia." *American Journal of Agricultural Economics*, 88(2): 324–337.

Bevis, Leah EM. 2015. "Soil-to-Human Mineral Transmission with an Emphasis on Zinc, Selenium, and Iodine." *Springer Science Reviews*, 3(1): 77–96.

Black, Maureen M. 2003. "Micronutrient Deficiencies and Cognitive Functioning." *Journal of Nutrition*, 3927S–3931S.

Black, Robert E, Lindsay H Allen, Zulfiqar A Bhutta, Laura E Caulfield, Mercedes De Onis, Majid Ezzati, Colin Mathers, and Juan Rivera. 2008. "Maternal and child undernutrition: global and regional exposures and health consequences." *The Lancet*, 371(9608): 243–260.

Boughton, Duncan, David Mather, Christopher B Barrett, Rui Benfica, Danilo Abdula, David Tschirley, and Benedito Cunguara. 2007. "Market participation by rural households in a low-income country: An asset-based approach applied to Mozambique." *Faith and Economics*, 50(1): 64–101.

Brown, Kenneth H, Janet M Peerson, Juan Rivera, and Lindsay H Allen. 2002. "Effect of supplemental zinc on the growth and serum zinc concentrations of prepubertal children: a meta-analysis of randomized controlled trials." *The American journal of clinical nutrition*, 75(6): 1062–1071.

Brown, Kenneth H, Sonja Y Hess, Stephen A Vosti, and Shawn K Baker. 2013. "Comparison of the estimated cost-effectiveness of preventive and therapeutic zinc supplementation strategies for reducing child morbidity and mortality in sub-Saharan Africa." *Food & Nutrition Bulletin*, 34(2): 199–214.

Burke, William J, Robert J Myers, and Thomas S Jayne. 2015. "A Triple-Hurdle Model of Production and Market Participation in Kenya's Dairy Market." *American Journal of Agricultural Economics*, aav009.

Chandra, Hukum. 2013. "Exploring spatial dependence in area-level random effect model for disaggregate-level crop yield estimation." *Journal of Applied Statistics*, 40(4): 823–842.

Chilimba, Allan DC, Scott D Young, Colin R Black, Katie B Rogerson, E Louise Ander, Michael J Watts, Joachim Lammel, and Martin R Broadley. 2011. "Maize grain and soil surveys reveal suboptimal dietary selenium intake is widespread in Malawi." *Scientific reports*, 1.

Chilimba, Allan DC, Scott D Young, Colin R Black, Mark C Meacham, Joachim Lammel, and Martin R Broadley. 2012. "Agronomic biofortification of maize with selenium (Se) in Malawi." *Field Crops Research*, 125: 118–128.

Chowdhury, Shyamal, JV Meenakshi, Keith I Tomlins, and Constance Owori. 2011. "Are consumers in developing countries willing to pay more for micronutrient-dense biofortified foods? Evidence from a field experiment in Uganda." *American Journal of Agricultural Economics*, aaq121.

Davis, Donald R. 2009. "Declining fruit and vegetable nutrient composition: What is the evidence?" *HortScience*, 44(1): 15–19.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken. 2013. "What do we learn from the weather? The new climate-economy literature." National Bureau of Economic Research.

Dorosh, Paul, Hyoung Gun Wang, Liang You, and Emily Schmidt. 2010. "rop production and road connectivity in Sub-Saharan Africa: A Spatial Analysis." World Bank Policy Research Working Paper 5385.

Elbers, Chris, Jean O Lanjouw, and Peter Lanjouw. 2003. "Micro–Level Estimation of Poverty and Inequality." *Econometrica*, 71(1): 355–364.

Fan, Ming-Sheng, Fang-Jie Zhao, Susan J Fairweather-Tait, Paul R Poulton, Sarah J Dunham, and Steve P McGrath. 2008. "Evidence of decreasing mineral density in wheat grain over the last 160 years." *Journal of Trace Elements in Medicine and Biology*, 22(4): 315–324.

Farnham, MW, PE Wilson, KK Stephenson, and JW Fahey. 2004. "Genetic and environmental effects on glucosinolate content and chemoprotective potency of broccoli." *Plant Breeding*, 123(1): 60–65.

Favell, D. J. 1998. "A comparison of the vitamin C content of fresh and frozen vegetables." *Food Chemistry*, 62(1): 59–64.

Fischer Walker, Christa L, and Robert E Black. 2007. "Functional indicators for assessing zinc deficiency." *Food & Nutrition Bulletin*, 28(Supplement 3): 454S–479S.

Graham, Robin D, Ross M Welch, David A Saunders, Ivan Ortiz-Monasterio, Howarth E Bouis, Merideth Bonierbale, Stef De Haan, Gabriella Burgos, Graham Thiele, Reyna Liria, et al. 2007. "Nutritious subsistence food systems." *Advances in Agronomy*, 92: 1–74.

Greffeuille, Valérie, AP Polycarpe Kayodé, Christèle Icard-Vernière, Muriel Gnimadi, Isabelle Rochette, and Claire Mouquet-Rivier. 2011. "Changes in iron, zinc and chelating agents during traditional African processing of maize: effect of iron contamination on bioaccessibility." *Food chemistry*, 126(4): 1800–1807.

Grzebisz, Witold. 2011. "Magnesium–food and human health." *Journal of Elementology*, 16(2).

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica: Journal of the econometric society*, 153–161.

Heltberg, Rasmus, and Finn Tarp. 2002. "Agricultural supply response and poverty in Mozambique." *Food Policy*, 27(2): 103–124.

Hengl, Tomislav, Gerard BM Heuvelink, Bas Kempen, Johan GB Leenaars, Markus G Walsh, Keith D Shepherd, Andrew Sila, Robert A MacMillan, Jorge Mendes de Jesus, Lulseged Tamene, et al. 2015. "Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions." *PloS one*, 10(6): e0125814.

Hoffmann, Vivian, Samuel Mutiga, Jagger Harvey, Michael Milgroom, and Rebecca Nelson. 2013. "A Market for lemons: Maize in Kenya." *College Park, MD. University of Maryland.*

Horton, Sue, Harold Alderman, and Juan A Rivera. 2009. "Hunger and malnutrition." *Global Crises, Global Solutions: Costs and Benefits.*

Hotz, Christine, and Kenneth H Brown. 2004. *Assessment of the risk of zinc deficiency in populations and options for its control.* International nutrition foundation: for UNU.

Hotz, Christine, Cornelia Loechl, Abdelrahman Lubowa, James K Tumwine, Grace Ndeezi, Agnes Nandutu Masawi, Rhona Baingana, Alicia Carriquiry, Alan de Brauw, Jonnalagadda V Meenakshi, et al. 2012*a*. "Introduction of β-carotene–rich orange sweet potato in rural Uganda resulted in increased vitamin A intakes among children and women and improved vitamin A status among children." *The Journal of nutrition*, 142(10): 1871–1880.

Hotz, Christine, Lubowa Abdelrahman, Cristina Sison, Mourad Moursi, and Cornelia Loechl. 2012*b*. "A food composition table for Central and Eastern Uganda." *Washington, DC: International Food Policy Research Institute and International Center for Tropical Agriculture.*

Imdad, Aamer, and Zulfiqar A Bhutta. 2011. "Effect of preventive zinc supplementation on linear growth in children under 5 years of age in developing countries: a meta-analysis of studies for input to the lives saved tool." *BMC public health*, 11(Suppl 3): S22.

Jayne, Thomas S, Takashi Yamano, J Nyoro, Tom Awuor, et al. 2001. "Do farmers really benefit from high food prices? Balancing Rural Interests in Kenya's Maize Pricing and Marketing Policy." *Tegemeo Institute for Agricultural Policy and Development Working Paper B*, 2.

Jayne, Thorn S, Ballard Zulu, and JJ Nijhoff. 2006. "Stabilizing food markets in eastern and southern Africa." *Food Policy*, 31(4): 328–341.

Joy, JM, EL Ander, SD Young, CR Black, MJ Watts, ADC Chilimba, B Chilimba, EWP Siyame, AA Kalimbira, R Hurst, SJ Fairweather-Tait, AJ Stein, RS Gibson, PJ White, and MR Broadley. 2014. "Dietary mineral supplies in Africa." *Physiologia Plantarum.*

Kennedy, Gina, Guy Nantel, and Prakash Shetty. 2003. "The scourge of "hidden hunger": global dimentions of micronutrient deficiencies." *FNA ANA*, 32: 8–16.

La Daniels, R, and Gibson K Simmer. 2000. "Indicators of selenium status in Australian infants." *Journal of Paediatrics and Child Health*, 36: 370–374.

Lawson, Charles L, and Richard J Hanson. 1974. *Solving least squares problems.* Vol. 161, SIAM.

Levinsohn, James, and Margaret McMillan. 2007. "Does food aid harm the poor? Household evidence from Ethiopia." In *Globalization and Poverty.* 561–598. University of Chicago Press.

Mayer, Anne-Marie. 1997. "Historical changes in the mineral content of fruits and vegetables." *British Food Journal*, 99(6): 207–211.

Mayer, Anne-Marie B, Michael C Latham, John M Duxbury, Nazmul Hassan, and Edward A Frongillo. 2007. "A food-based approach to improving zinc nutrition through increasing the zinc content of rice in Bangladesh." *Journal of Hunger & Environmental Nutrition*, 2(1): 19–39.

Minten, Bart, and Christopher B Barrett. 2008. "Agricultural technology, productivity, poverty and food security in Madagascar." *World Development*, 36(5): 797–822.

Moraghan, JT. 1994. "Accumulation of zinc, phosphorus, and magnesium by navy bean seed." *Journal of plant nutrition*, 17(7): 1111–1125.

Peck, N. H., D. K. Grunes, R. M. Welch, and G. E. MacDonald. 1980. "Nutritional Quality of Vegetables Crops as Affected by Phosphorus and Zinc Fertilizers." *Agronomy Journal*, 72: 528–534.

Prasad, Ananda S. 2003. "Zinc deficiency." *Bmj*, 326(7386): 409–410.

Rayman, MP. 1997. "Dietary selenium: time to act." *British Medical Journal*, 314(387-388).

Renkow, Mitch, Daniel G Hallstrom, and Daniel D Karanja. 2004. "Rural infrastructure, transactions costs and market participation in Kenya." *Journal of Development Economics*, 73(1): 349–367.

Sarris, Alexander, Sara Savastano, and Luc Christaensen. 2006. *Agriculture and Poverty in Commodity-dependent African Countries: A Rural Household Perspective from the United Republic of Tanzania.* Food & Agriculture Org.

Shivay, Yashbir Singh, Dinesh Kumar, and Rajendra Prasad. 2008. "Effect of zinc-enriched urea on productivity, zinc uptake and efficiency of an aromatic rice-wheat cropping system." *Nutrient Cycling in Agroecosystems*, 81(3): 229–243.

Sillanpää, M, and PLG Vlek. 1985. "Micronutrients and the agroecology of tropical and Mediterranean regions." In *Micronutrients in tropical food crop production.* 151–167. Springer.

Sillanpää, Mikko. 1972. *Trace elements in soils and agriculture.* Food & Agriculture Org.

Stevens, Robyn, and Alex Winter-Nelson. 2008. "Consumer acceptance of provitamin A-biofortified maize in Maputo, Mozambique." *Food Policy*, 33(4): 341–351.

Thomson, Christine D. 2004. "Selenium and iodine intakes and status in New Zealand and Australia." *British Journal of Nutrition*, 91: 66–672.

USDA. 2008. "USDA National Nutrient Database for Standard Reference, release 21." Beltsville (MD), Agricultural Research Service, USDA Nutrient Data Laboratory.

USDA. 2013. "USDA National Nutrient Database for Standard Reference, Release 26." U.S. Department of Agriculture, Agricultural Research Service.

Velu, G, KN Rai, V Muralidharan, T Longvah, and J Crossa. 2011. "Gene effects and heterosis for grain iron and zinc density in pearl millet (Pennisetum glaucum (L.) R. Br)." *Euphytica*, 180(2): 251–259.

Velu, G, KN Rai, V Muralidharan, VN Kulkarni, T Longvah, and TS Raveendran. 2007. "Prospects of breeding biofortified pearl millet with high grain iron and zinc content." *Plant Breeding*, 126(2): 182–185.

White, Philip J, and MR Broadley. 2005. "Historical variation in the mineral composition of edible horticultural products." *Journal of Horticultural Science and Biotechnology*, 80(6): 660–667.

WHO. 2004. "Iodine status worldwide: WHO global database on iodine deficiency."

WHO. 2008. "Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia."

WHO. 2009. "Global prevalence of vitamin A deficiency in populations at risk 1995-2005: WHO global database on vitamin A deficiency."

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data.* MIT press.

Figures

Figure 1: Maize Zn Distribution: Household vs. Market Samples

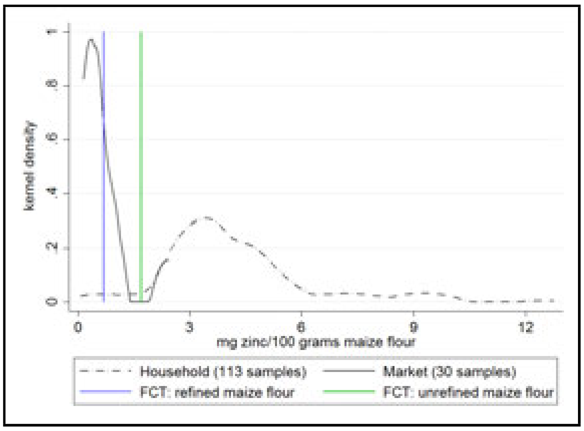


Figure 2: Cassava Zn Distribution: Household vs. Market Samples

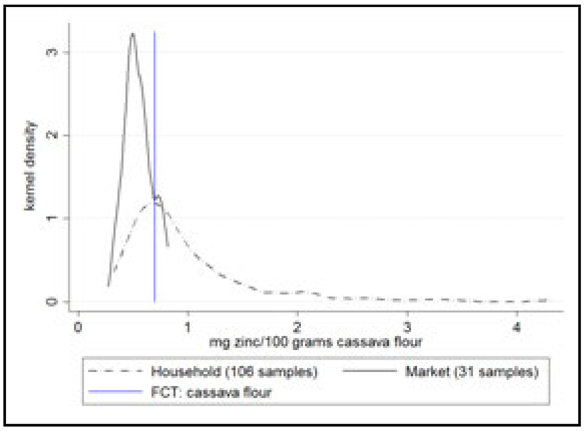


Figure 3: Child Zinc Intake: Household vs. Market Values

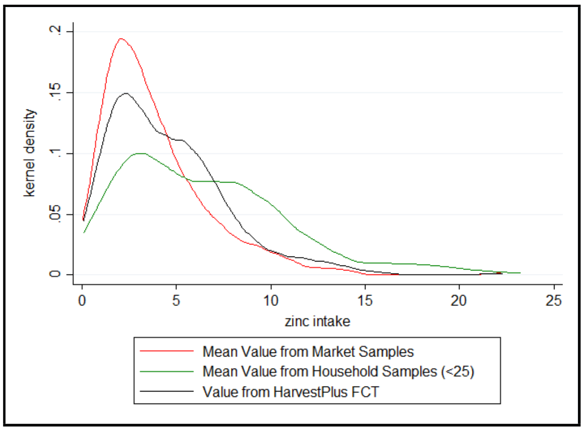


Figure 4: Child Zinc Adequacy: Household vs. Market Values

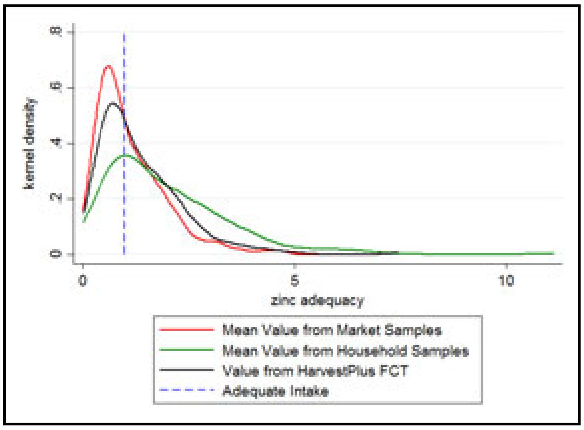


Figure 5: Cereals: Choosing k by RMFE

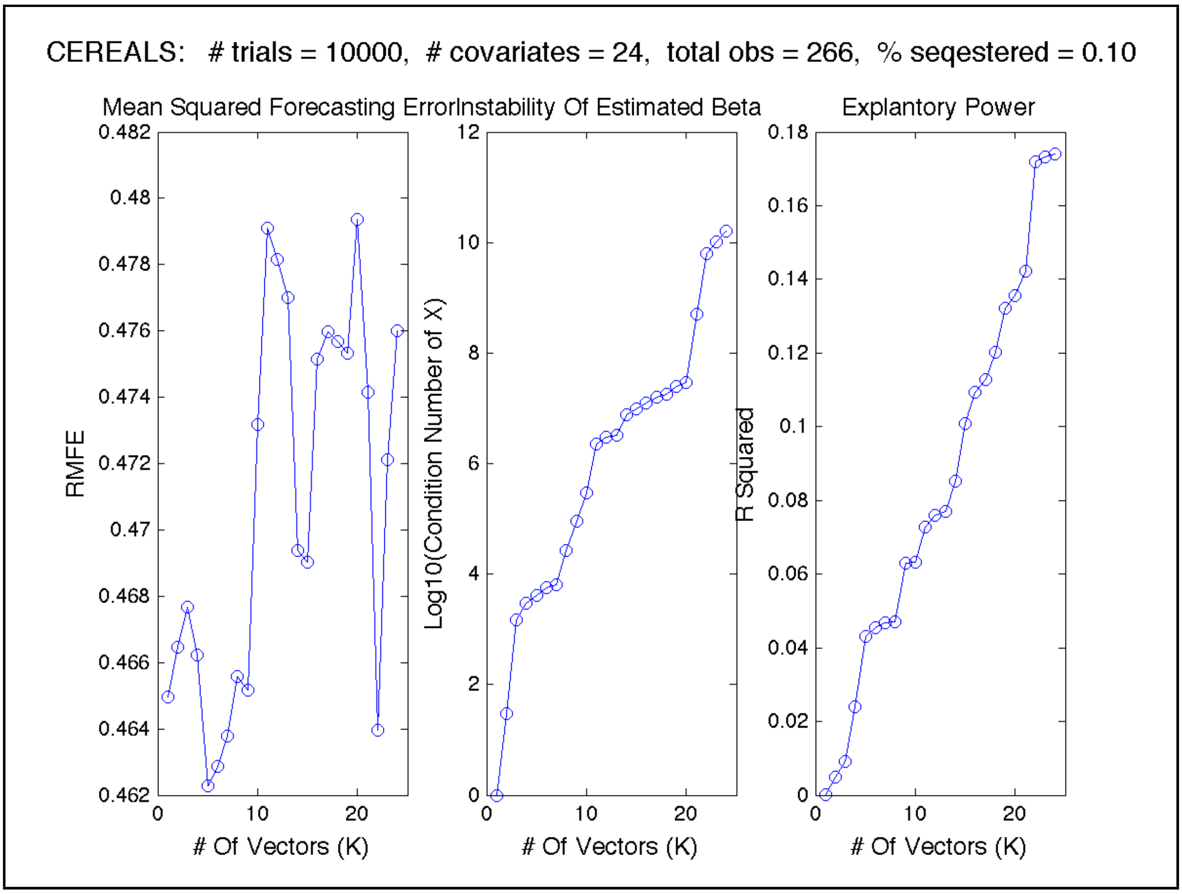


Figure 6: Cereal: Residuals (k=5) for Producers, Sellers, and Major Sellers

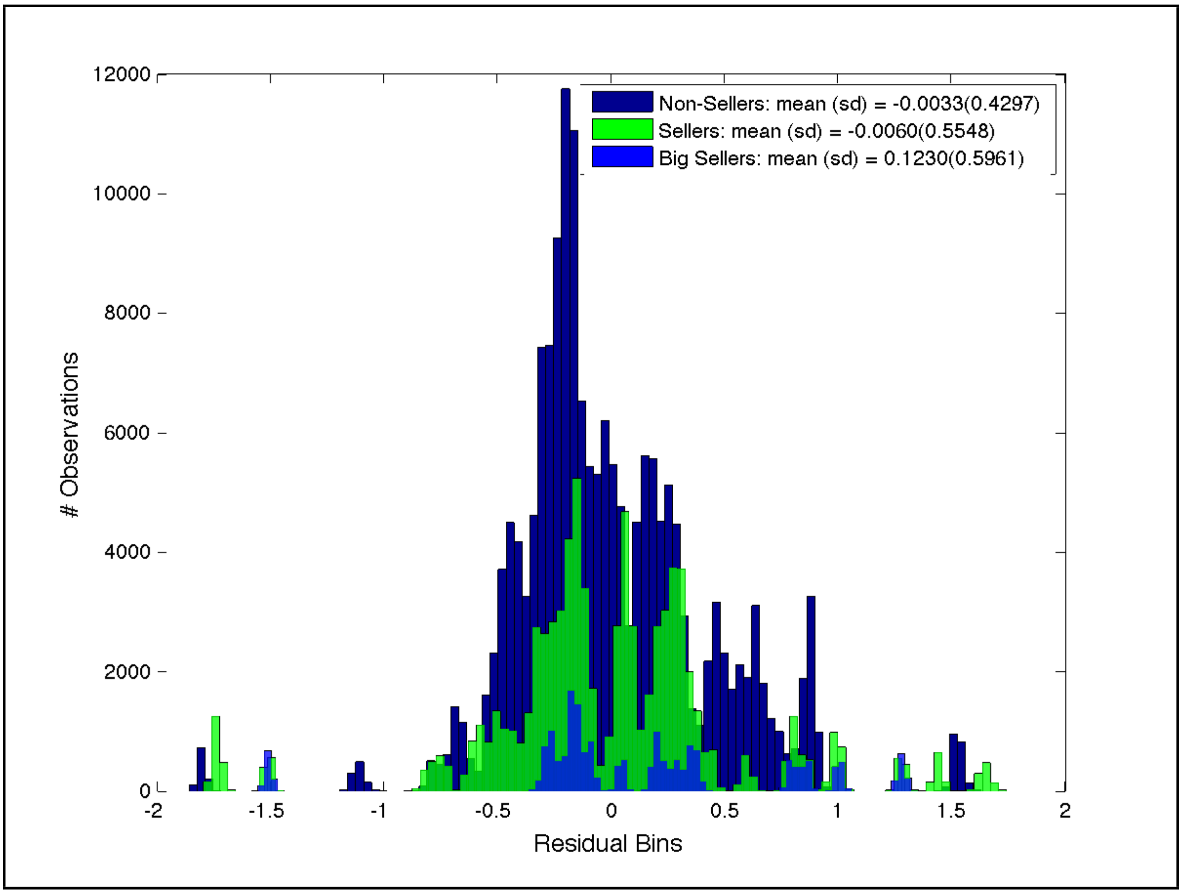


Figure 7: APE of Rainfall CV for Cereals, over km to Major Town

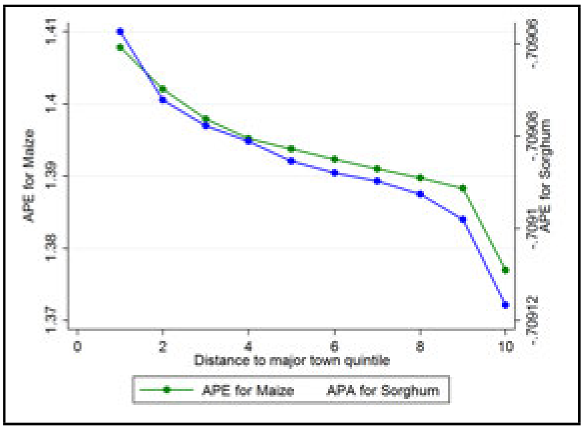


Figure 8: APE of Degree Day CV for, Cereals, over km to Major Town

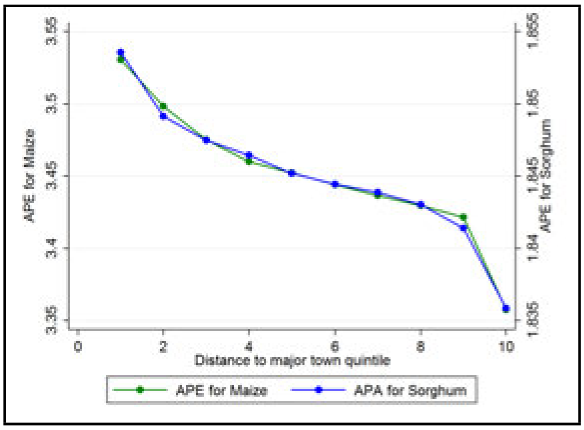


Figure 9: APE of Rainfall CV for Cereals, over km to Major Town

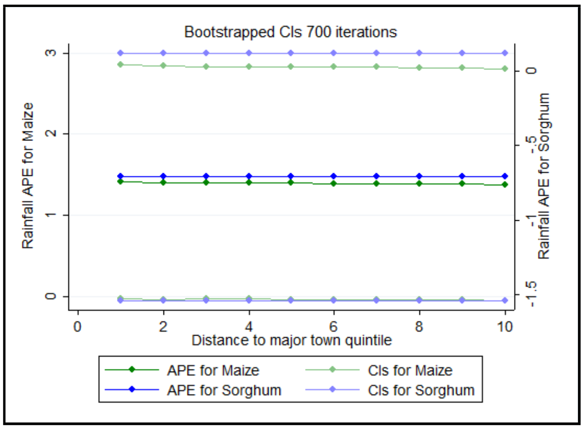


Figure 10: APE of Degree Day CV for, Cereals, over km to Major Town

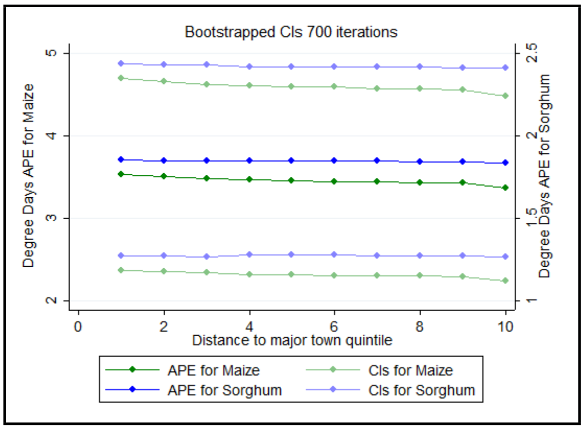


Figure 11: Mean Rainfall CV 2009-13, by Cereal Zinc Content Quintile

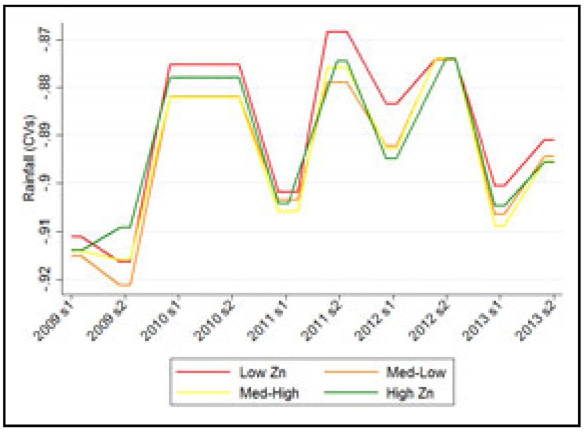
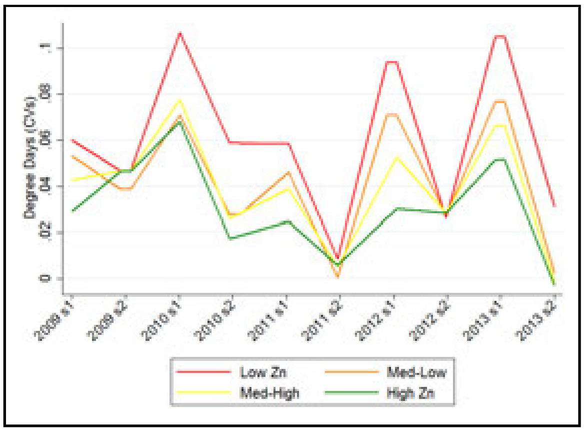


Figure 12: Mean Degree Days CV 2009-13, by Cereal Zinc Content Quintile



Tables

Table 1: UNC and UNPS Covariates for Predicting Crop Zinc (Summary Stats)

	UNC Data		UNPS Data		UNPS Non-Central	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
Soil Ca (%)	7.078	3.048	7.778	2.751	7.762	3.054
Soil pH (pH 10)	5.680	0.314	5.815	0.349	5.841	0.399
Annual mean temp (DD)	6283	847	12506	2053	12566	2258
Annual mean precip (mm)	22.019	4.339	22.019	4.339	22.914	2.726
Farm productivity (log Ugx/ha)	13.216	1.419	11.623	3.0445	12.283	2.313

All statistics displayed un-weighted

Table 2: UNPS Non-Seasonal Covariates for Predicting Crop Sales (Summary Stats)

	Median	Mean	Stdev
Elevation n	1177.00	1226.86	235.88
Soil Clay (%)	40.16	39.95	5.49
Non-farm income (binary)	0.00	0.32	0.46
Wealth (pca index)	-0.34	-0.00	1.75
Urban (binary)	0.00	0.23	0.42
HH labor (# able-bodied)	2.00	2.59	1.61
Log farm area (log acres)	0.00	-1.27	3.04
No farm land (binary)	0.00	0.34	0.47
Distance to Kampala (km)	188.26	181.45	107.69
Distance to major town (km)	30.24	34.17	31.18
Distance to road (km)	2.89	5.30	6.68
Rainfall expectation (avg total mm)	6296.76	6255.72	1243.08
Rainfall risk (sd)	51.90	53.04	11.13
Degree Days expectation (avg daily DD)	11.11	11.34	1.22
Degree Days risk (sd)	0.34	0.34	0.08
Rainfall Realization (CV)	-0.89	-0.89	0.02
Degree Days Realization (CV)	0.03	0.04	0.04

Statistics are un-weighted
Risk and expectation variables are season-specific, and calculated over 1999-2008
Rain and expectation realizations are given by CV around the decade means

Table 3: Predicted Log Zn Content of Home-Produced vs. Market-Sold Maize

Year	Season	Produced Crops			Marketed Crops			Difference	
		Median	Mean	Stdev	Median	Mean	Stdev	% Diff	t-stat
9	1	6.628	8.282	6.242	5.993	6.98	4.607	-0.157	343
9	2	6.417	6.994	2.984	6.588	6.768	2.066	-0.032	28
10	1	6.663	8.060	5.422	6.383	7.021	3.023	-0.129	265
10	2	6.410	7.098	3.356	6.769	7.531	2.716	0.061	243
11	1	6.833	8.148	5.422	6.497	7.100	3.975	-0.129	295
11	2	6.651	7.519	3.950	6.649	6.843	2.754	-0.090	194

Household statistics: weighted by survey weights and production binary
Market statistics: weighted by survey weights and sale quantity (kg)
T-statistic according to the distribution of log crop zinc
% Diff = (Market Mean - Household mean)/(Household mean)

Table 4: Predicted Log Zn Content of Home-Produced vs. Market-Sold Crops

Year	Season	Maize		Sorghum		Sw Potato		Cassava		Beans		Gnuts	
		% Diff	t-stat	% Diff	t-stat	% Diff	t-stat	% Diff	t-stat	% Diff	t-stat	% Diff	t-stat
9	1	-0.157	343	-0.020	30	-0.152	118	-0.142	70	-0.067	320	-0.024	81
9	2	-0.032	28	-0.050	80	-0.036	19	0.114	178	-0.017	79	-0.029	76
10	1	-0.129	265	-0.029	88	-0.289	221	-0.325	330	-0.025	100	-0.025	108
10	2	0.061	243	-0.022	7	-0.215	175	-0.092	61	-0.121	786	-0.025	74
11	1	-0.129	295	-0.109	170	-0.182	114	-0.128	82	-0.026	115	-0.046	206
11	2	-0.090	194	-0.020	0	-0.121	91	-0.083	27	-0.017	70	0.033	103
Avg	1	-0.135	701	-0.074	223	-0.195	315	-0.167	282	-0.032	337	-0.03	278
Avg	2	-0.073	301	-0.065	172	-0.136	225	-0.058	-40	-0.051	591	-0.011	75

Household statistics: weighted by survey weights and production binary
Market statistics: weighted by survey weights and sale quantity (kg)
T-statistic according to the distribution of log crop zinc
% Diff = (Market Mean - Household mean)/(Household mean)

Table 5: Triple Hurdle Model Estimated for Maize (Page 1)

	(1)	(2)	(3)
Dependent variable:	w_1=1	w_2=1	Log Sales
Soil and Environment			
Elevation	-0.000525***	-0.000649***	-0.000850
	(9.64e-05)	(0.000114)	(0.000640)
Soil clay content (%)	0.0392***	0.0344***	0.0764**
	(0.00328)	(0.00408)	(0.0319)
pH	4.828***	5.088***	15.91***
	(0.951)	(1.415)	(5.122)
$(\text{pH})^2$	-0.404***	-0.412***	-1.253***
	(0.0796)	(0.120)	(0.417)
Climate Expectations			
Rain seasonal mean (mm)	-0.000106	1.04e-05	-0.000597**
	(0.000105)	(0.000148)	(0.000296)
$(\text{Rain seasonal mean})^2$	-1.69e-09	-8.45e-09	3.15e-08
	(7.65e-09)	(1.15e-08)	(2.42e-08)
Rain seasonal sd (mm)	0.00816**	0.00448	0.0165**
	(0.00383)	(0.00411)	(0.00823)
Temp seasonal mean (DD)	0.876***	1.305***	1.607
	(0.242)	(0.334)	(1.354)
$(\text{Temp seasonal mean})^2$	-0.0326***	-0.0544***	-0.0702
	(0.0109)	(0.0148)	(0.0570)
Temp seasonal sd (DD)	-2.380***	-1.016***	-1.334
	(0.268)	(0.332)	(1.086)
Climate Realizations			
Rain realization (CV)		0.455	6.770**
		(1.464)	(3.357)
(Rain CV)*(km to major town)		-0.0117	-0.0971***
		(0.0113)	(0.0251)
Temp realization (CV)		2.973***	7.984**
		(1.144)	(3.685)
(Temp CV)*(km to major town)		-0.0658***	-0.179**
		(0.0230)	(0.0791)
Input Endowments			
Household labor (# laborers)	-0.0212	-0.0269*	-0.000488
	(0.0131)	(0.0159)	(0.0355)
Farm size (log hectares)	0.0387***	0.123***	0.351***
	(0.0145)	(0.0170)	(0.113)
No land owned (dummy)	-0.916***	-0.168	0.818***
	(0.0975)	(0.116)	(0.283)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Triple Hurdle Model Estimated for Maize (Page 2)

Dependent variable:	(1) w_1=1	(2) w_2=1	(3) Log Sales
Transaction costs			
Distance to major town (km)	0.208***	0.0984	-0.0779
	(0.0452)	(0.0681)	(0.189)
(Distance to major town)2	7.08e-05***	-4.41e-05	-0.000255***
	(1.34e-05)	(3.98e-05)	(6.83e-05)
Distance to major road (km)	0.0101*	0.0222***	0.0597**
	(0.00565)	(0.00640)	(0.0235)
(Distance to major road)2	-0.000168	-0.000519**	-0.00112
	(0.000186)	(0.000208)	(0.000730)
Distance to Kampala (km)	-0.000928**	-0.00185***	-0.00941***
	(0.000433)	(0.000528)	(0.00194)
Household			
Household size (people)	0.0231	0.0180	0.00487
	(0.0153)	(0.0163)	(0.0306)
Head age (years)	0.00404	-0.00131	-0.0152
	(0.00415)	(0.00499)	(0.0112)
Head education (years)	0.0181*	0.00883	-0.0345
	(0.00941)	(0.0111)	(0.0222)
No Non-farm income (log Ush)	0.0629	0.0499	0.261***
	(0.0392)	(0.0467)	(0.0954)
Wealth (pca index)	0.0255*	0.0256	0.0560
	(0.0152)	(0.0195)	(0.0604)
Urban (dummy)	-0.269***	-0.224***	-0.354
	(0.0536)	(0.0645)	(0.286)
Year FE	Yes	Yes	Yes
Season FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
CRS	Yes	Yes	Yes
ρ_{12}		2.691***	
		(0.237)	
IMR coefficient			1.877
$(\sigma * \rho_{32})$			(1.197)
Observations	17,026	17,026	2,590
R-squared			0.270

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Climate Realizations on Sale Propensity and Sale Quantity

	Maize		Sorghum		Sw Potato	
	$w_2 = 1$	Sales	$w_2 = 1$	Sales	$w_2 = 1$	Sales
Rain	0.455	6.770**	-2.806	-8.768	0.209	3.984
	(1.464)	(3.357)	(2.071)	(10.59)	(2.253)	(4.907)
Rain*Km	-0.0117	-0.0971***	-0.000231	-0.0948*	0.00182	-0.0439
	(0.0113)	(0.0251)	(0.0103)	(0.0484)	(0.0207)	(0.0447)
DD	2.973***	7.984**	4.750***	27.13	-4.888***	-16.52
	(1.144)	(3.685)	(1.426)	(17.38)	(1.751)	(40.97)
DD*Km	-0.0658***	-0.179**	-0.0691**	-0.498*	0.0312	-0.0463
	(0.0230)	(0.0791)	(0.0335)	(0.263)	(0.0353)	(0.276)
	Cassava		Beans		Gnuts	
	$w_2 = 1$	Sales	$w_2 = 1$	Sales	$w_2 = 1$	Sales
Rain	0.555	3.639	1.308	-2.006	-1.756	-19.59**
	(1.915)	(6.319)	(1.568)	(3.161)	(1.652)	(8.210)
Rain*Km	0.00327	-0.00551	-0.0171	0.0198	-0.00440	-0.0479*
	(0.0102)	(0.0446)	(0.0114)	(0.0303)	(0.00710)	(0.0283)
DD	-3.504**	30.88	0.434	0.240	0.0710	-5.452
	(1.443)	(20.33)	(1.149)	(2.463)	(1.495)	(3.667)
DD*Km	0.0243	-0.283*	-0.0134	0.00147	0.00753	0.121
	(0.0283)	(0.167)	(0.0229)	(0.0499)	(0.0278)	(0.0841)

These results were estimated via a crop-specific triple hurdle model of sales, specified exactly as in Table 7

Appendix 1

Figures 1 and 2 illustrate the zinc density distribution of maize and cassava, which were sampled at both homes and markets in communities across Uganda. Median cassava zinc content drops by 40 percent at market, and median maize content drops by 83 percent.

While these were the only two crops sampled at both homes and markets, suggestive evidence points to similar patterns for other crops. Figure A1 illustrates the zinc content of millet purchased at market and sorghum sampled at households, two cereals generally considered to have almost identical zinc content. (The HarvestPlus Food Composition Table lists millet and sorghum as having 1.7 and 1.6 mg of zinc per 100 grams of crop, respectively.) The median zinc content of millet purchased from market is only 56 percent of the median zinc content of sorghum sampled from homes.

Figure A2 illustrates the zinc content distribution of cowpeas purchased at market and beans sampled at households. While not strictly comparable, these crops are again considered to have similar zinc content, and the observed pattern is the same: market-purchased cowpeas are lower and less variable in zinc than home-produced beans. While market samples were not gathered for sweet potatoes or groundnuts, Figures A3 and A4 illustrate heterogeneous zinc content with long upper tails, as with other sampled crops.

It is unlikely that processing accounts for the difference between household and market zinc content, as there was no difference between the market vs. home processing of cassava, millet and sorghum, or beans and cowpeas. Beans and cowpeas were purchased whole from the market or sampled whole from the household. Millet and sorghum are purchased as grain from the market, and sampled as grain from the household. While cassava was purchased in the form of flour from the market, and purchased as tubers from the household, the household tubers were then processed in exactly the same way that tubers are processed at market; the tubers were peeled and sliced into pieces, the pieces were dried, and the dried cassava was ground into flour.

There was a slight processing difference for maize; household-sampled maize was ground into unrefined maize flour, while maize flour purchased at market had been ground at local mills, and often refined or partially refined. But while refinement might lead to 10-50 percent decrease in maize zinc content, it is unlikely to account for an 83 percent decrease. HarvestPlus FCT values indicate that maize processing decreases zinc content by 60 percent, substantially less than the 83 percent differential observed in our data. And even this value may be higher than appropriate, given that the HarvestPlus value for refined flour is derived from the USDA FCT value for refined, de-germed cornmeal. (This is NDB No 20022 in the most recent USDA table, Nutrient Database Release 26.) Locally-owned Ugandan mills de-husk, but do not de-germ, cereals, therefore preserving a greater proportion of zinc and other nutrients. In a study of such locally-owned mills in Benin, Greffeuille et al. (2011) finds that processing dry maize grain (as is done in Uganda) decreases zinc content by only 11 percent, and processing wet, washed maize grain decreases zinc content by 54 percent.

Crops were analyzed for trace metal content via a Vulcan 84 Digestion, which gives the mineral quantity in parts per million. Degradation of nutrients over time is also unlikely to account for the differential, given that metals do not degrade over time. (Vitamins do.)

Figure A1: Millet and Sorghum Zinc Distributions: Household vs. Market

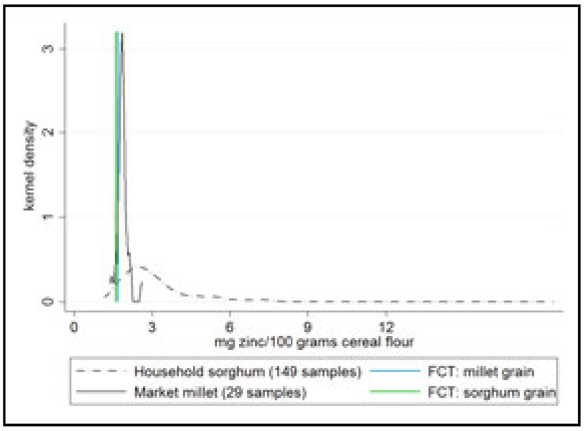


Figure A2: Bean and Cowpea Zinc Distributions: Household vs. Market

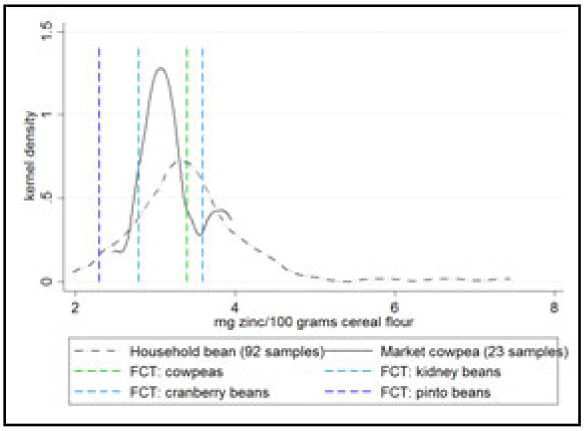


Figure A3: Sweet Potato Zinc Distributions: Household Only

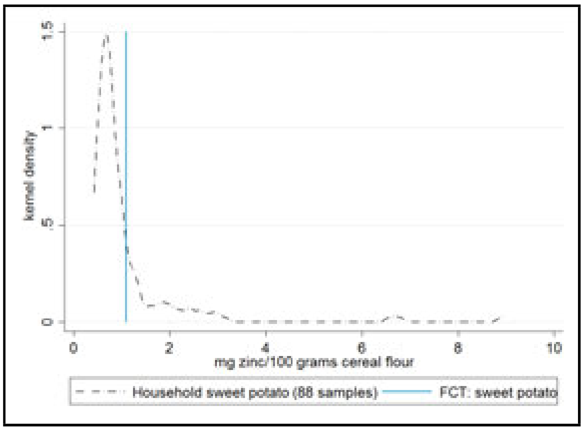
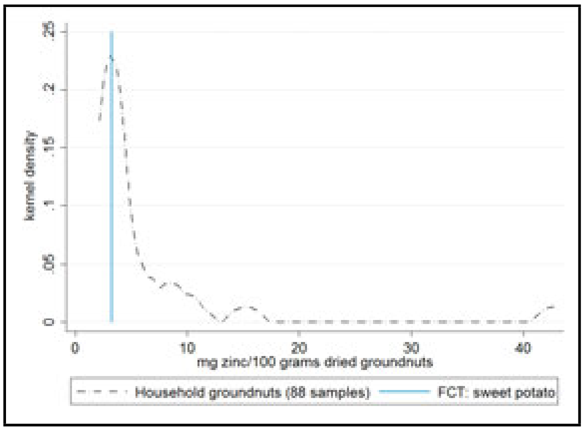


Figure A4: Groundnut Zinc Distributions: Household Only



Appendix 2

Food Frequency Questionnaires (FFQs) are a common tool for procuring the dietary intake of a select group of vitamins or minerals, rather than attempting a to gather comprehensive data on all dietary intake. In an FFQ, an individual (or caretaker) gives the number of times that each item on a list of foods was eaten during a particular recall period — in our case one week. In some cases, and in our survey, an average portion size is also chosen for each food item. We illustrated portion sizes (small, medium and large) with a book of pictures, as shown in Figures A5 and A6. Caretakers were also shown the physical plate on which the pictures were taken.

Foods listen on an FFQ are chosen to represent the major dietary sources of the nutrients of interest, according to prior and more comprehensive food recall data. Our nutrients of interest were zinc and selenium, and the foods listened on our FFQ capture 98 percent of all zinc consumption in Ugandan children under five, according to detailed food recall data gathered in central and eastern Uganda by the Reaching End Users (REU) project, a HarvestPlus initiative introducing orange-fleshed sweet potato to farmers in rural Uganda (Hotz et al. 2012).[19]

The FFQs were conducted for one child per household, covering all zinc-relevant food intake in the preceding week. Qualifying children were under 5 years of age, present at the time of the survey, and not exclusively breastfeeding. If a household had multiple children present who fit these criteria (e.g. a son, a niece, and a cousin), surveyors chose the biological child of the household head. If the household head had multiple biological children present, all fitting the necessary criteria, the surveyors chose the oldest child.

Three zinc intake variables are calculated according to the food recall data provided by the FFQ — the difference between them is in the assumed zinc content of particular crops. For the first zinc intake variable, the mean zinc content of crops sampled on smallholder farms is used to define the contribution of staple crops to dietary zinc intake. For the second, the mean zinc content of crops sampled from rural markets is used. For the third, crop zinc content is defined according to the HarvestPlus food frequency table (FCT) for Uganda.

For crops that were sampled at homes but not at markets (sweet potato, ground nuts and cow peas), the nutrient content from the HarvestPlus FCT is used in this second, market-based variable. Sorghum and millet are assumed to have identical zinc content, and so sorghum samples from households define the contribution of both crops to dietary zinc intake in the first variable, and millet samples from markets define the contribution of both crops to dietary zinc intake in the second variable.

For all three zinc intake variables, the zinc content of foods not sampled at homes or markets (generally foods low in zinc content such as leafy greens, or foods for which no U.S. import permit may be obtained such as goat meat) is defined according to the HarvestPlus food composition table value.

[19] In using the REU food recall data to construct a list of foods for the FFQ, I followed protocols laid out by a technical document written by Christine Hotz, the primary architect of the REU Food Frequency Recall. That food recall was focused on vitamin A consumption, and so the recall itself was not useful for our purposes, but the methodology — and data to conduct it — was highly useful.

Figure A5: Portion Size Picture of Mukene (Small Fish)

Figure A6: Portion Size Picture of Katogo (Cassava & Beans)

Appendix 3

Lawson and Hanson (1974) explain how Single Value Decomposition (SVD) may be used to sold the Ordinary Least Squares (OLS) problem. Their technique may be explained in the following way.

In the typical approach to OLS we minimize $||\mathbf{X}\beta - \mathbf{y}||^2$, where $\mathbf{X}$ is an $m \times n$ matrix, β is an $m \times 1$ matrix, and $\mathbf{y}$ is an $m \times 1$ matrix.

Using SVD, we may decompose $\mathbf{X}$ follows

$$\mathbf{X} = \mathbf{USV^T}$$

where $\mathbf{U}$ is an orthogonal $m \times n$ matrix, $\mathbf{S}$ is a diagonal $n \times n$ matrix with successive, positive and non-decreasing entries, and $\mathbf{V^T}$ is an orthogonal $n \times n$ matrix.

Because $\mathbf{U}$ is orthogonal, $\mathbf{U}^T$ is also orthogonal, and both are therefore distance-preserving under multiplication. Thus,

$$\begin{aligned} ||\mathbf{X}\beta - \mathbf{y}||^2 &= ||\mathbf{U}^T(\mathbf{X}\beta - \mathbf{y})||^2 \\ &= ||\mathbf{U}^T(\mathbf{USV^T}\beta - \mathbf{y})||^2 \\ &= ||\mathbf{SV^T}\beta - \mathbf{U}^T\mathbf{y})||^2 \\ &= ||\mathbf{S}\gamma - \mathbf{g}||^2 \end{aligned}$$

where the third line follows from the fact that $\mathbf{U}$ is an orthogonal matrix, and we define $\gamma = \mathbf{V^T}\beta$ and $\mathbf{g} = \mathbf{U}^T\mathbf{y}$, both $n \times 1$ matrices.

To minimize $||\mathbf{X}\beta - \mathbf{y}||^2$, we can therefore choose a $\hat{\gamma}$ to minimize $||\mathbf{S}\gamma - \mathbf{g}||^2$. The original parameter vector $\hat{\beta}$ is calculated as $\mathbf{V}\hat{\gamma}$.

Predicted outcome $\hat{\mathbf{y}}$ may be equivalently calculated as either $\mathbf{X}\hat{\beta}$ or $\mathbf{USg}$, since $\mathbf{X}\beta \simeq \mathbf{y}$ and $\mathbf{S}\hat{\gamma} \simeq \mathbf{g}= \mathbf{U}^T\mathbf{y}$. The residual $\hat{\mathbf{r}}$ may be equivalently calculated as either $\mathbf{y} - \mathbf{X}\hat{\beta}$ or $\mathbf{U}(\mathbf{g} - \mathbf{S}\hat{\gamma})$ since $\mathbf{y} - \mathbf{X}\beta = \mathbf{Ug} - \mathbf{USV^T}\beta = \mathbf{U}(\mathbf{g} - \mathbf{S}\gamma)$.

Because $\mathbf{S}$ holds successively non-increasing diagonal values, the elements of $\hat{\gamma}$ become increasingly insignificant to $\hat{\beta} = \mathbf{V}\hat{\gamma}$. The candidate solution $\gamma^{(k)}$ may therefore be considered, where each element of $\gamma^{(k)}$ is identical to that of the full solution $\hat{\gamma}$ up until the k'th element, and all subsequent elements are zero, as below.

$$\gamma^{(k)} = \begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In many cases the candidate solution $\gamma^{(k)}$, for some particular $k < n$, minimizes mean squared forecasting error (MSFE) better than the full solution γ. This is because the rows of matrix $\mathbf{V}$ hold "averages" for the columns of matrix $\mathbf{X}$, but with each row explaining less and less of the variation in $\mathbf{X}$. One might imagine that, past some particular column k, the rows of matrix $\mathbf{V}$ hold only sample-specific variation in $\mathbf{X}$, rather than variation that can be predicted out of sample. In other words, a solution vector $\hat{\beta}$ that captures such variation is over-fitting the model, leading to an increase in R^2 within sample, but also an increase in MSFE out of sample. A solution vector $\hat{\beta}$ that captures only the variation within the first k vectors of $\mathbf{V}$ will better minimize MSFE.

Appendix 4

An initial model includes a number of household-specific covariates in the model for crop zinc production. For each of the three crop categories (cereals, tubers and legumes), these household-specific covariate are given zero weight, with the exception of the log productivity per hectare variable. Figure A7 illustrates this fact for the cereal model — the vertical axis gauges the weight of each covariate listed below, and the first 14 covariates (bicycle owned, motorcycle owned brick walls, cash crop grown, legumes grown, education of household head, head age, head sex, household size, trees planted on plots, inorganic fertilizer used, organic fertilizer used, erosion-controlling structures) used are given zero weight. Interacting these covariates, and/or interacting these covariates with the area-level covariates, yields the same result. This is true for the cereal model and for the tuber and legume models. (Results available upon request.)

The only household-specific covariate to receive weight in the model is log productivity per hectare. Productivity per hectare has implications for household-specific soil nutrient content — in this way, the variable may proxy for household-specific deviation from the area-level values of soil pH and soil calcium that are included in the model. This variable, and all of it's interaction, is included in the final model. In Figure A7, the variable is called "lugxperhec" and it's interactions are the last 6 variables (luph, luph2, luca, lupr, luau, lulu).

The final model used includes all interactions between soil calcium, soil pH, average precipitation, average temperature, and log productivity per hectare. Average precipitation is a raster-grid average of monthly precipitation data from 1999-2008 (the decade before the UNPS survey began), and average precipitation is a raster-grid average of monthly temperature data from the same period.

For cereals, $k = 5$ minimizes RMFE, at which point R^2 is 0.43. (See Figure 5.) For both tubers and legumes, $k = 3$ minimizes RMFE, at which point R^2 is 0.81 and 0.23, respectively. Figures A8 and A9 display these results.

Figure 6 illustrates the cereal model residual distribution for producers, sellers and major sellers. Figures A10 and A11 do the same for the tuber model and the legume model. In neither case do residual distributions appear to differ by selling status, indicating that comparisons of predicted zinc should be valid across these categories.

Figure A7: Zero-Weighted Household-Specific Covariates (Cereal Model)

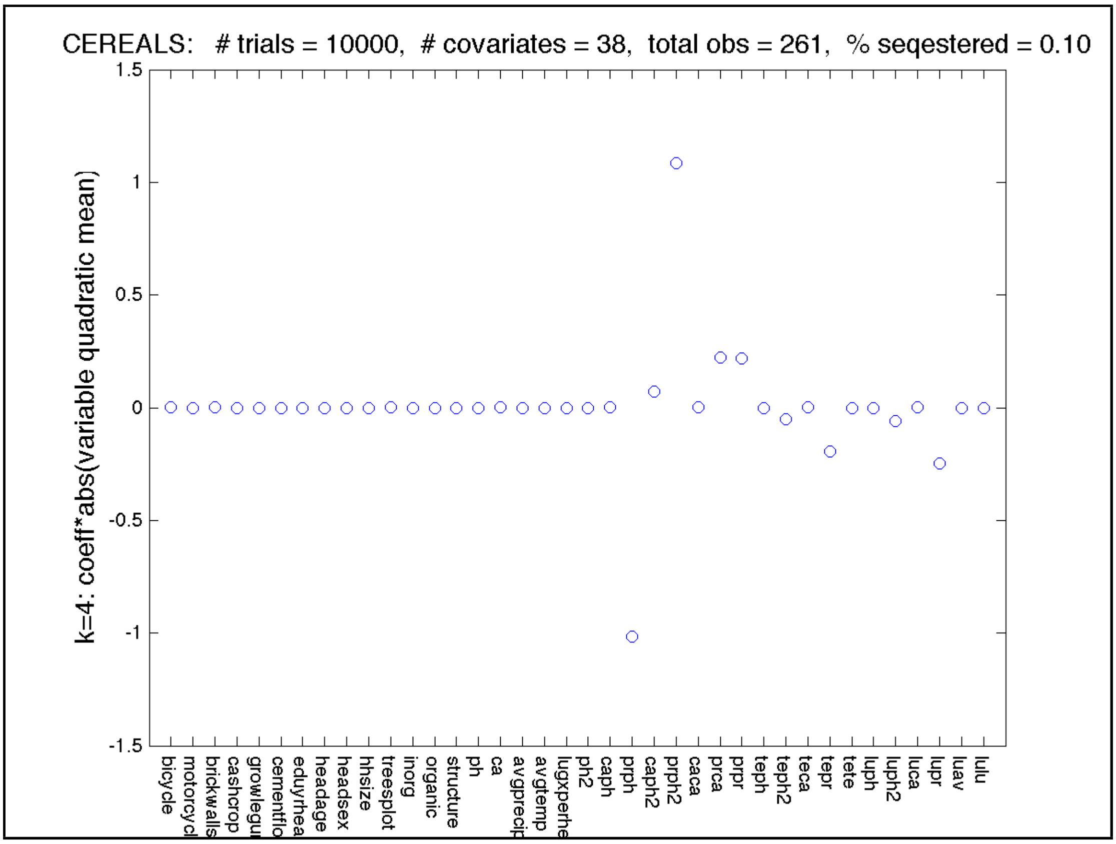


Figure A8: Tubers: Choosing k by RMFE

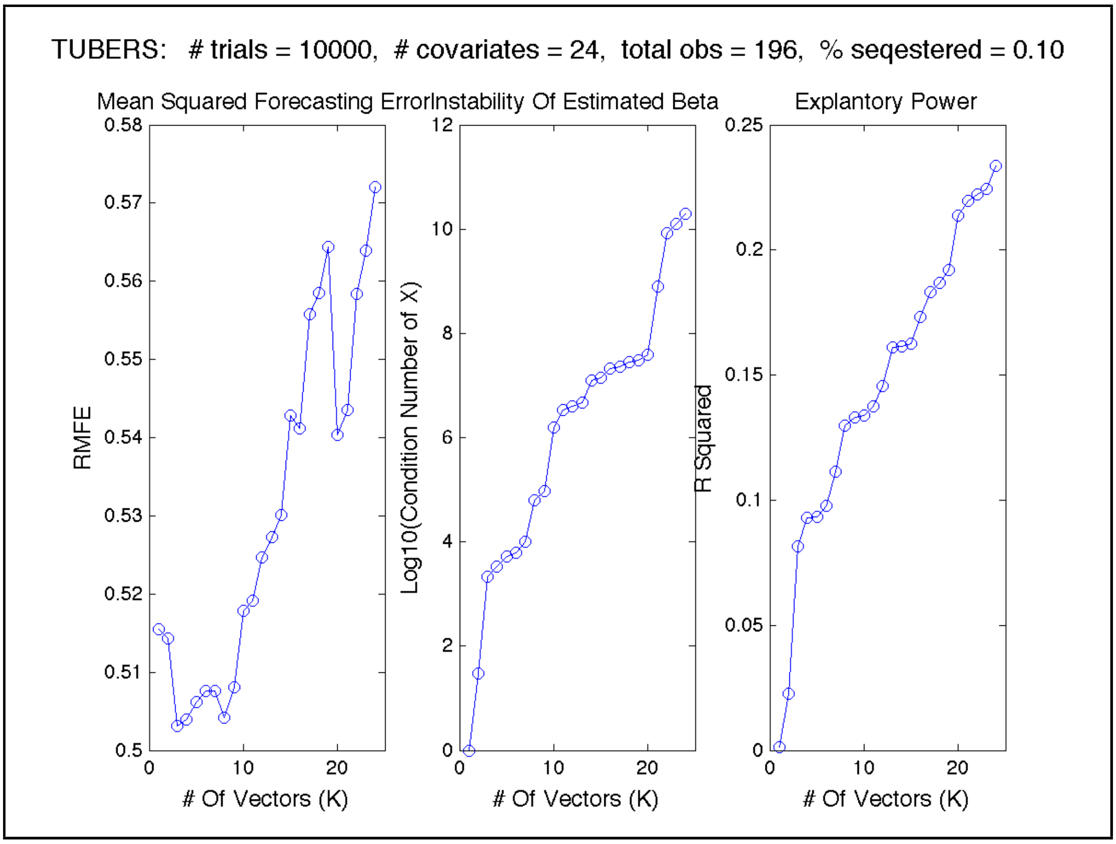


Figure A9: Legumes: Choosing k by RMFE

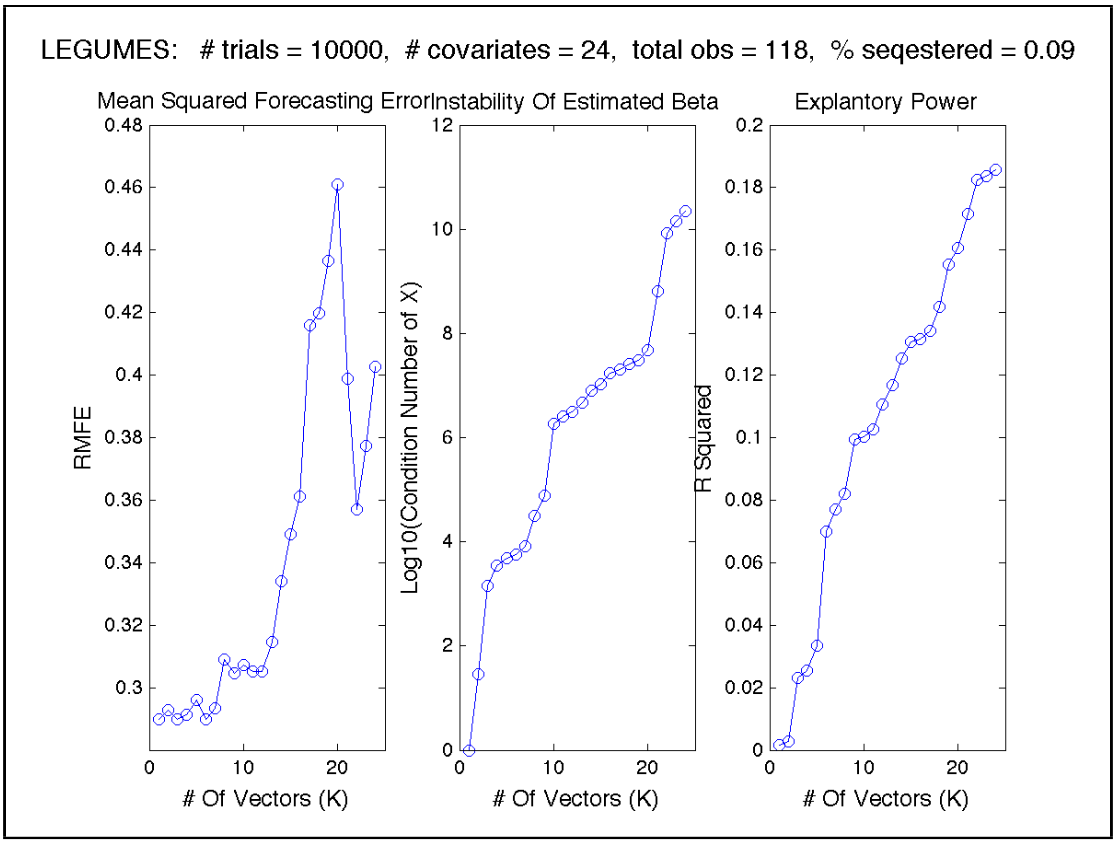


Figure A10: Tubers: Residuals (k=5) for Producers, Sellers, and Major Sellers

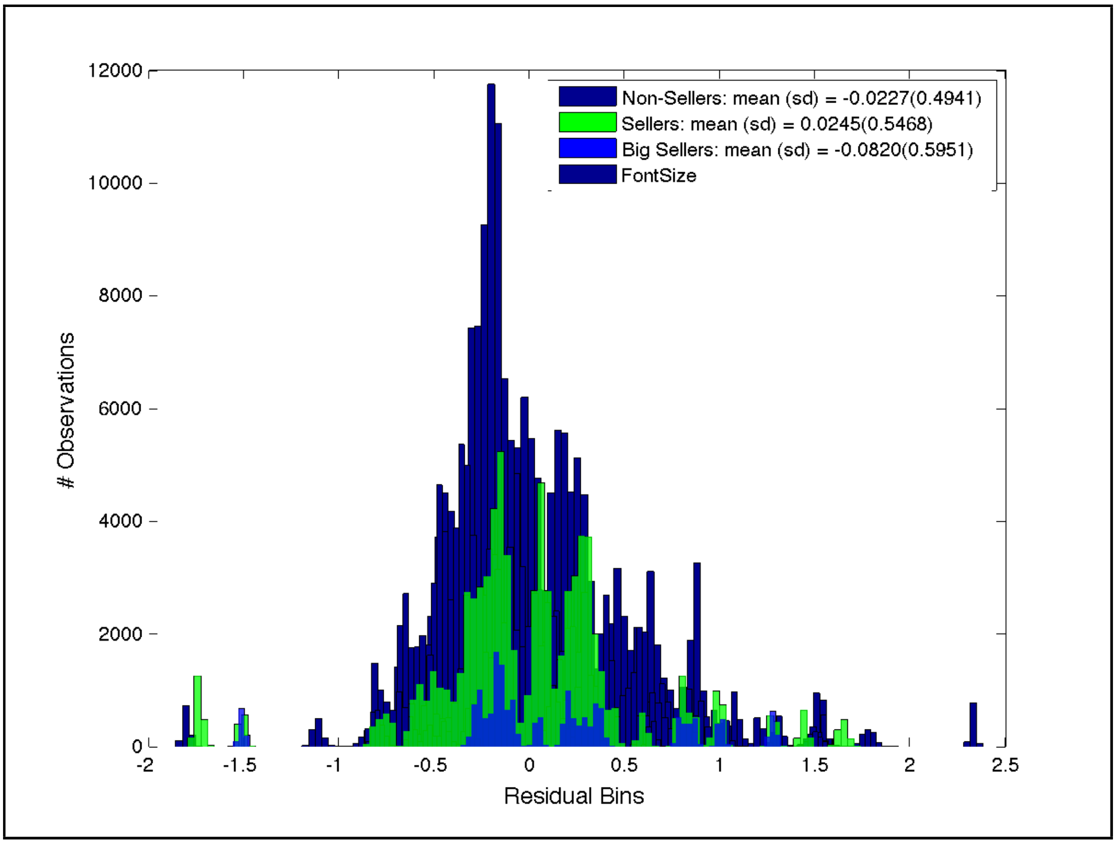


Figure A11: Legumes: Residuals (k=5) for Producers, Sellers, and Major Sellers