Long-term Effects of Early Life Maize Yield on Maize Productivity and Efficiency in Rural Malawi

Mussa, Richard

Department of Economics, Chancellor College, University of Malawi

4 January 2017

Long-term Effects of Early Life Maize Yield on Maize Productivity and Efficiency in Rural Malawi

Richard Mussa*

January 4, 2017

Abstract

The paper assesses the effects of maize yields just prior to birth (*in utero*), in the first and the second years of life on adult life productivity and efficiency of maize farmers born between 1984 and 1995 in rural Malawi. To ensure that early life maize yields are not confounded by omitted local chacteristics, they are transformed into relative maize yields by using a cumulative gamma distribution. I find that maize yield just prior to birth significantly increases maize output in a farmer's adult life. However, relative maize yields in the first and second years of life have no long-term effects on maize production. Furthermore, there is no long-term impact of early life maize yields on the technical efficiency of maize production. These findings survive a number of robustness checks including alternative definitions of early life maize yields, controlling for migration and allowing for serial correlation. Furthermore, the results are not driven by sample selection originating from survival induced by maize yields in early life. Thus, low maize productivity in early-life begets low maize productivity in adult life. The paper finds that the impact of inputs under the farm input subsidy programme (FISP) on maize productivity is almost of the same order of magnitude as the long-term impact of maize yield *in utero*.

Keywords: Productivity; *In utero*; Malawi

1 Introduction

There is a growing body of literature which looks at the long-term impacts of shocks in the fetal period, infancy, and early childhood. on various economic outcomes (see survey papers by Strauss and Thomas (1998), and Currie and Vogl (2013)). These early-life shocks take different forms including: weather (Maccini and Yang, 2009; Gørgens et al., 2012; Dercon and Porter, 2014), disease (Bleakley, 2010; Kelly, 2011; Lin and Liu, 2014), and war (Mansour and Rees, 2012; Grimard and Laszlo, 2014). The various shocks have been found to have lasting effects on adult outcomes such as health, disability, income, and cognitive and non-cognitive skills.

Due to a number of reasons, the long-term effects of early-life shocks are likely to be more pronounced in developing countries. First, the shocks occur more frequent in many

*Department of Economics, Chancellor College, University of Malawi, Box 280, Zomba, Malawi, rimussa@yahoo.co.uk.

developing countries than in the industrialized world (Currie and Vogl, 2013). Second, the limited access to formal savings and insurance in developing countries leads to a limited availability and effectiveness of shock mitigation strategies, and this in turn suggests that poor households are forced to trade-off between short-run consumption and longer-run earnings and human capital accumulation (Ferreira and Schady, 2009; Fiszbein et. al.,2009; Maccini and Yang, 2009; Currie and Vogl, 2013).

Unlike the other shocks which are mostly extreme negative events with inherent problems of generalizability (Maccini and Yang, 2009; Hoynes et. al., 2016), weather shocks taking the form of either precipitation shocks or temperature shocks can be characterised as more typical. This paper focuses on these more typical weather shocks; however, a key point of departure for this paper is that instead of indirectly measuring the long-run impacts of early-life weather shocks by using either rainfall or temperature anomalies a direct approach is adopted with anomalies in maize production.

A key attraction of using maize yields is that it allows the study to look at an output instead of an input such as rainfall or temperature. Furthermore, using maize yields is useful as it enables the study to avoid the possible challenges that are inherent in an indicator like fluctuations in precipitation or temperature. These fluctuations may influence other environmental conditions correlated with economic activity and public health, including forest fires, floods and landslides, the availability of potable water, and agricultural pest control (Maccini and Yang, 2009).

Similar to many African countries, maize is a primary staple crop in Malawi, and is the best direct indicator of incomes especially rural incomes (Burke et al., 2014). Maize accounts for more than two-thirds of caloric availability (Ecker and Qaim, 2011). Compared to neighbouring countries, food consumption is less diversified in Malawi. For instance, Malawi's per capita maize consumption of 133.1 kg/per person per year is 2.5 times that of Mozambique, and 2.3 times that of Tanzania. Only Zimbabwe (110.4 kg/per person) and Zambia (110.2 kg/per person) are the closest to Malawi (Mussa, 2015). As a result of this low food diversification, national food security continues to be defined in terms of access to maize.

It is not just food consumption which is skewed towards maize, crop production by smallholder agriculture is dominated by maize production. For instance, NSO (2012) found that 85% of households in Malawi cultivated maize (69% in urban areas, and 88% in rural areas). According to Smale (1995) given its importance "maize is life" in Malawi. As a result of this, maize availability takes a special place in political, social, and economic discourse.

Despite its significance not just in Malawi but across the continent, there is a dearth of literature on whether or not maize production shocks or anomalies in a farmer's early life affect his/her productivity in producing the crop later in adulthood. The paper closes this gap in knowledge by answering the following question: Does low maize productivity in early-life beget low maize productivity in adult life? More specifically, the paper sets out to examine whether early-life maize yield affects adulthood maize production directly as an input in the production function, and/or indirectly as a factor narrowing the technology gap in the inefficiency effect function.

The rest of the paper is structured as follows. In Section 2 the methodology is presented, and the variables and data used are discussed. This is followed by the empirical results in Section 3. Section 4 concludes.

2 Empirical Methods

2.1 Data and Variables

In this paper a farmer is defined as a household member who makes decisions concerning crops to be planted, input use and the timing of cropping activities on a field. The focus on the farmer rather than the household head is motivated by Udry (1996) who finds that in a situation where many plots are controlled by different household members the assumption that resource allocation within the household is pareto efficient does not hold. Thus, the unitary household model is inappropriate as households members compete as well as cooperate.

The data used in the paper come from two sources. The early life district level-birth year maize yield data is compiled from crop production data from the Ministry of Agriculture and Food Security. The data is collected annually and in every district through the Agriculture Production Estimates Survey (APES) in which extension workers act as data collectors. The APES collects data on area cultivated, yield, and production of crops. It also collects data on livestock and fisheries. Of interest in this paper is the maize yield which is measured in metric tonnes per hectare. For each district and year, the maize yield is calculated as a total of local maize, hybrid maize, and composite maize.

Using year of birth and district of birth, this data is then linked to adult life production and farmer characteristics data taken from the Third Integrated Household Survey (IHS3). The IHS3 is statistically designed to be representative at national, district, urban and rural levels. The survey was conducted by the National Statistical Office from March 2010 to March 2011. The survey collected information from a sample of 12271 households; 2233 (representing 18.2%) are urban households, and 10038 (representing 81.8%) are rural households. The survey collected socio-economic data at the household level and on individuals within the households. It also collected data on farming activities including crop output, land, labour and other inputs.

This paper focuses on rural areas as this where maize production is more likely to happen. The APES data starts from 1984, and the youngest farmer in the IHS3 is 15 years old. Consequently, the matched sample includes farmers born between 1984 and

1995. Although some maize in Malawi is produced by irrigation, the most dominant form of maize production is rainfed. I thus use rainfed early life and adult life maize yield data. The harvest period for maize in Malawi is March-May of every year. For farmers born between January and June, the maize yield just prior to birth (*in utero*) assigned to them is from the previous maize growing season while for farmers born between July and December, their maize yield *in utero* is from their year of birth. Yields in the first and second years of life are then generated as one period and second period leads of the yield *in utero* respectively.

Currently, Malawi has 28 districts, however, the government of Malawi has since 1994 been splitting some of the districts to form news ones. The new districts are: Balaka formerly part of Machinga, Phalombe formerly part of Thyolo, Neno formerly part of Mwanza, and Likoma formerly part of Nkhatabay. Since early life maize yields between 1984-1995 are used in this paper, the new districts are merged back into the old ones to end up with 24 districts. Total maize yields from the merged districts are then used. After data cleaning, I end up with data on 1275 rural maize farmers who cultivate a total of 1626 maize fields.

The evolution of maize yields (in metric tonnes per hectare) by birth year for the period 1984-1995 is shown Figure 1. It is evident that the maize yields per hectare have been fairly volatile over the study period. Yields per hectare were mostly above 5 metric tonnes per hectare over the period 1984-1991, and with the exception of 1993, yields remained below 5 metric tonnes per hectare between 1992 and 1995, reaching a low of 1.73 metric tonnes per hectare in 1992. The sharp fall in maize yield experienced in 1992 can be explained by the severe drought which Malawi and the rest of Southern Africa experienced.

Maize yield in a farmer's early life may depend on observed and unobserved local characteristics thus making them potentially endogenous. To ensure that the early life maize yields are not confounded by these omitted local characteristics, I follow Burke et al. (2014) and Flatø et al. (2016) and transform the actual maize yields into relative maize yields by using a cumulative gamma distribution. This transformation ensures that in each year, each district receives a value which reflects the probability of having a maize yield at that level or below in that particular district. This in turn means that the level of relative maize yield in a given year is orthogonal to local characteristics.

Figure 2 shows a box plot of the relative maize yield prior to birth, in the first of year of life, and in the second year of life. by a farmer's birth year. The plots capture the interquartile range of relative maize yield values for 24 districts in Malawi, and the line within each box is the median. The highest and lowest values within 1.5 times the interquartile range from the bounds of the boxes are represented by whiskers. Two things stand out: first, there is a wide range of values of relative maize corresponding to each year of birth. Second, across the birth years, there is no discernible monotonic trend in the median values of relative maize yields.

Table 1 reports production characteristics of the maize farmers across quintiles of relative maize yield in early life. The relative maize yield in the first quintile reflects the lowest maize yield (bottom 20%) while the highest maize yield is represented by the fifth quintile (the top 20%). Moving across the quintiles indicates that there is a bivariate positive relationship between relative maize yields in early life and maize yields in adult life. For instance, the adult life yield corresponding to the bottom 20% of relative maize yield *in utero* is 317.07kg, and this almost triples to 825.56kg when one looks at the top 20% of relative maize yield *in utero*. A similar pattern can be noted when quintiles of relative maize yields in the first and second years life are used.

With the exception of seeds, the quantities of land, fertiliser, labour, capital used by farmers vary directly with relative maize yields in early life. For example, farmers in the first quintile of relative maize yield *in utero* on average use about 629.33 Malawi Kwacha (US$4.17) of capital, this however doubles to 1317.71 Malawi Kwacha (US$8.74) for the fifth quintile. Farmers in the first quintile of early life maize yields have more secure land than those in the fifth quintile. In contrast, farmers in the first quintile of early life maize yields access more government or private extension services than their counterparts who are in the fifth quintile. Overall, this bivariate analysis points to suggestive evidence that higher maize yields in a farmer's early life are associated with higher maize productivity in a farmer's adult life. In this paper, I use multivariate methods to more rigorously investigate the existence and nature of this relationship.

2.2 Estimation

I model the effect of maize yield just prior to the farmer's birth (t), maize yield in the farmer's first year of life ($t+1$), and maize yield in the farmer's second year of life ($t+2$) by looking at them as production inputs and as factors which affect a farmer's efficiency. This allows for the simultaneous measurement of the efficiency and production based effects of maize yield in early life on a farmer's maize production later in adult life. I employ the non-neutral production function by Dinar et al. (2007) which in turn is a simplified version of Huang and Liu's (1994) non-neutral frontier model.

Let the production structure for a maize farmer be specified using a single-output, multi-input Cobb-Douglas stochastic production frontier as follows

$$\begin{aligned} \ln q_{ijt+a} &= \beta_0 + \sum_{k=1}^{5} \beta_k \ln x_{kijt+a} + \sum_{k=0}^{3} \phi_{t+k} y_{ijt+k} \\ &\quad + \sum_{k=1}^{2} \delta_k D_{kijt+a} + F_j + T_t + M_t + v_{ijt+a} - u_{ijt+a} \end{aligned} \tag{1}$$

$$v_{ijt+a} \sim N\left(0, \sigma_v^2\right) \tag{2}$$

$$u_{ijt+a} \sim N^{+}\left(\mu_{ijt+a}, \sigma_u^2\right) \tag{3}$$

$$\mu_{ijt+a} = \sum_{k=0}^{3} \gamma_{t+k} y_{ijt+k} + z_{ijt+a}\alpha + F_j + T_t + M_t + \omega_{ijt+a} \tag{4}$$

where; i indexes a farmer, j is a farmer's district of birth, t is a farmer's birth year, and a is the farmer's age. q_{ijt+a} is a rainfed maize output index. I use a maize output index instead of actual maize produced because some of the maize fields owned by farmers are mixed stand with more than one crop planted in a season. Consequently, most inputs (land, fertiliser and labor) are at the field level, and cannot be uniquely assigned to maize production only. The maize output index generated as follows (Liu and Myers, 2009)

$$q_{ijt+a} = \begin{cases} \frac{\sum_m p_m q_{ijm+a}}{p_1} & \text{if intercropped field} \\ q_{ij1} & \text{if monocropped field} \end{cases} \tag{5}$$

where q_{ijt+a} is the maize output index, p_m is the median price in the community at harvest time of crop m, q_{ijtm} is the yield of crop m, and crop 1 is maize. For monocropped fields, maize yield is simply the actual yield. β_0 is an intercept,.β_k ($l = 1...5$) are output elasticities with respect to inputs x_i. Five inputs are used namely; land measured in acres, own and hired labour measured in man days, capital measured as the total monetary value in Malawi Kwacha of farm implements (hoes, slashers, axes, oxcarts, oxploughs) owned by a household, seed measured in kilograms, organic and inorganic fertiliser measured in kilograms. The three principal variables of interest are; y_{ijt} the relative maize yield just prior to the farmer's birth, y_{ijt+1} the relative maize yield in the farmer's first year of life, and y_{ijt+2} the relative maize yield in the farmer's second year of life. ϕ_{t+k} are the corresponding coefficients for the variables. F_j, T_t and M_t are district of birth, year of birth and month of birth fixed effects.

v_{ijt+a} is a two sided random variable which captures random variations in the economic environment facing production units, reflecting luck, weather, measurement errors, and omitted variables from the model. u_{ijt+a} is a technical inefficiency effect which is a non-negative truncation of a normal random variable. It represents deviations from potential output that reflect inefficiency such as farm-specific knowledge, the will and skills of farmers, and other disruptions to production. The notation "+" means that the underlying distribution is truncated from below at zero so that realized values of the random variable u_{ijt+a} are positive. It is assumed that v_{ijt+a} and u_{ijt+a} are independent of each other. Equation (4) is a technical ineffiency model where z_{ijt+a} is a vector of controls which determine inefficiency, α is a coefficient vector, and ω_{ijt+a} is an error term.

The inefficiency and the stochastic frontier production function in equations (1) to (4) are jointly estimated by using maximum likelihood estimation to achieve both efficiency and consistency. Farm-specific estimates of technical efficiency are obtained via the conditional expectation $E[exp(u_{ijt+a}|v_{ijt+a})]$ (Battese and Coelli, 1988). To measure the relationship between maize yield in early life and maize production later in life I use the coefficients ϕ_{t+k} $(k = 0, 1, 2)$ and γ_{t+k}. For example, a positive (negative) sign of ϕ_t implies that maize yield *in utero* increases (decreases) maize production later in life, and positive (negative) of γ_t means that maize yield *in utero* decreases (increases) efficiency i.e. reduces the gap between potential and actual maize output. Effects for maize yield in the first and second years of life are computed analogously.

2.3 Model specification tests

To ensure that the modeling structure as represented by equations (1) to (4) is valid, the paper tests a number of hypotheses sequentially using the Wald test (hypotheses 1-4, 6 and 7), and a third-moment test developed by Coelli (1995) (hypothesis 5).

1. $H_0 : \beta_1 = \cdots = \beta_5 = \phi_t = \phi_{t+1} = \phi_{t+2} = \delta_1 = \delta_2 = 0$, this null hypothesis means that all variables included in the frontier production function are jointly insignificant.

2. $H_0 : \sum_{k=1}^{3} \phi_k + \sum_{k=1}^{5} \beta_k = 1$, the null hypothesis means that there are constant returns to scale.

3. $H_0 : \phi_t = \phi_{t+1} = \phi_{t+2} = 0$, the null hypothesis means that there is no production based effect of early life relative maize yield i.e. treating maize yield as a production input is inappropriate.

4. $H_0 : \mu = 0 = \sigma_u^2 = 0$, the null hypothesis implies that there is no inefficiency component. If the null hypothesis is true, then the truncated-normal model reduces to a linear regression model with normally distributed errors.

5. $H_0 : \gamma_t = \gamma_{t+1} = \gamma_{t+2} = \boldsymbol{\alpha} = 0$, the null hypothesis specifies that the included exogenous determinants of technical inefficiency are jointly insignificant. A rejection of this null implies that the included exogenous factors together influence technical inefficiencies.

6. $H_0 : \gamma_t = \gamma_{t+1} = \gamma_{t+2} = 0$, the null hypothesis means that early life relative maize yield does not influence efficiency i.e. there is efficiency based effect of maize yield on maize production.

In addition to the independent variables already discussed, I control for the age of the farmer measured in years, gender, security of land tenure, average years of schooling in a farmer's household, and whether the farmer received agricultural extension services. I also include a community level economic infrastructure index to measure availability of and access to economic infrastructure in a community. The infrastructure index is constructed by using multiple correspondence analysis (MCA) (see e.g. Asselin (2002) and Blasius and Greenacre (2006) for more details). The economic infrastructure index is based on the presence of the following in a community: a perennial and passable main road, a daily market, a weekly market, a post office, a commercial bank, and a microfinance institution. Table 2 presents summary statistics of the variables used in the study.

3 Results

3.1 Model specification results

Table 3 shows model specification tests results. The Wald test results indicate that all the variables included in the Cobb-Douglas production frontier are jointly statistically significant. The three early life maize yield variables in the production function are jointly significant; this suggests that there are production based effects of early life maize yield. The third-moment test results lead to the rejection of the null hypothesis of no inefficiency component, and this means that there are technical inefficiency effects. Thus, the mean of the inefficiency term can be modeled as a linear function of a set of covariates.

The Wald test results show that the determinants of inefficiency included in the technical inefficiency model are jointly significant. However, the results further reveal that the three variables capturing early life maize yields are jointly insignificant in the inefficiency model. This suggests that there are no efficiency based effects of maize yield *in utero*, in the first and second years of life for maize farmers in rural Malawi. Consequently, including early life maize yield in both the production and inefficiency components would lead to an incorrect model specification. I now turn to a discussion of the results for the production frontier, technical inefficiency and production uncertainty models.

3.2 Econometric results

Maximum likelihood results of the Cobb-Douglas production frontier and the technical inefficiency models are reported in Table 4. They indicate that all the five conventional inputs of maize production have statistically significant effects on output. The output elasticities are fairly sizable, and there is a clear ranking in terms of the sizes of the elasticities. The elasticity of maize output with respect to seeds is the smallest while the elasticity of maize output with respect to fertiliser is the largest. The output elasticity of fertiliser suggests that holding other factors constant a 1% increase in fertiliser is associated with an increase maize output of 0.32% while the corresponding change arising a from a *ceteris paribus* increase in seeds is 0.05%.

The returns to scale is about 0.78, and this calculated as a sum of the output elasticities of seeds, land, fertiliser, labour, and capital. This suggests that maize production in rural Malawi exhibits decreasing returns to scale. This is further supported by the

Wald test result in Table 3 which rejects the existence of constant returns to scale. This evidence of decreasing returns to scale in maize production is similar to what other studies on cereal production find such as Weir and Knight (2007) in Ethiopia and Asadullah and Rahman (2009) in Bangladesh. A previous study by Mussa (2015) which was based on a Translog production also finds that maize production in Malawi is characterised by decreasing returns to scale with a returns to scale coefficient of 0.86.

The results show that a number of factors significantly influence efficiency of maize production in rural Malawi. Gender matters when it comes to efficiency Similar to and consistent with the findings of Liu and Myers (2009), male maize farmers are more technically efficient than female farmers. This gender difference in efficiency could be explained by the fact that female farmers in most agrobased developing countries do not have the same inheritance rights as males, and this may act as a disincentive for hardwork. There is statistically significant negative relationship between a farmer's age and inefficiency. This means that other things being equal, older farmers are likely to be more efficient. This finding agrees with a contention by Coelli and Battese (1996) that older farmers are likely to be more efficient because they have more farming experience.

The results indicate that farmers that have secure land tenure such that the land was inherited or was purchased with a title deed are more efficient. This could possibly by explained by the fact that secure land tenure may lead to more investment such as soil conservation and tree planting (Deininger and Jin, 2006) which may turn lead to increased farm productivity. As argued by Binar et al. (2007), agricultural extension services may speed up the diffusion process and the adoption of new varieties and technologies as well as leading to the efficient utilization of existing technologies by improving farmers' know-how. Consistent with this argument, the paper finds that extension services lead to higher efficiency. As noted by Asadullah and Rahman (2009), underdeveloped infrastructure can have negative effects on efficiency as farmers may not acquire inputs at the right time, or not at all. The results confirm that availability of economic infrastructure in a community improves the efficiency of maize farmers.

I now turn to the key focus of this paper, and look at the existence and nature of the relationship between early-life maize yields and maize productivity in adult life. The Wald test results (see Table 3) discussed earlier have shown that the relationship between early life maize yields and maize production is asymmetric. Specifically, early-life maize yields jointly influence maize production i.e. there is direct effect of early life maize yields on productivity, however, early-life maize yields do not influence technical efficiency i.e. there is no indirect effect of early life maize yields on productivity. Furthermore, results in Table 4 show that in the inefficiency model, the coefficients on maize yield *in utero*, maize yield in the first year of life, and maize yield in the second year of life are all individually statistically insignificant. In light of this asymmetry, the rest of the paper focuses on the production frontier results.

The frontier model results in Table 4 indicate that the coefficient on maize yield in the first year of life is negative while maize yield *in utero* and in the second year of life carry a positive coefficient. However, at all the conventional levels of significance only maize yield just prior to a farmer's birth significantly increases maize output in a farmer's adult life. Thus, maize yields in the first and second years of life have no long-term effects on maize production. The coefficient on maize yield *in utero* is 0.85, this means that holding other factors constant, a 1% decrease in the mean of relative maize yield *in utero* (which amounts to a reduction of 0.005 in the cumulative gamma distribution from 0.49 to 0.48) is associated with a reduction in maize output of 0.42%. Clearly, this is a quantitatively large effect.

Is the long-term of early-life maize yields on maize productivity gendered? The literature on long-term effects of early-life environmental conditions on various economic outcomes find that these effects can vary with gender. For instance, Maccini and Yang (2009) find that higher early-life rainfall has large positive effects on the adult outcomes of women, but not of men in Indonesia. In their case, early-life rainfall refers to rainfall in the year of birth, and not in the year prior to birth i.e. *in utero.* Since this paper finds that it is only maize yield just prior to birth (i.e. *in utero*) that has long-term impacts on the productivity of maize farmers, it follows that the effect is not gendered. For a developing country like Malawi the use of ultrasound is limited or non-existent, this coupled with the fact that there is no known evidence of sex selective abortion, it makes sense to conclude that the gender of a child *in utero* is unknown.

What exactly is the nature of the relationship between maize yield *in utero* and maize productivity in a farmer's adult life? To get a better understanding of the exact pattern of this relationship, I use Figure 3 which depicts a nonparametric local polynomial regression of predicted maize yield and maize yield *in utero.* The regression is conditional on the conventional maize production inputs, birth-month fixed effects, birth-year fixed effects, birth-district fixed effects. The solid line is a nonparametric regression estimate, and the shaded bounds are 95% confidence intervals. The plot reveals that the relationship between a farmer's maize yield later in life and maize yield just prior to birth is fairly flat for low relative yields (below 0.3) and high relative yields (above 0.7), and it is positive and steep in between. The rising portion of the relationship corresponds to actual maize yields *in utero* of between 2.57 and 4.50 metric tonnes per hectare.

It is important to put these results in some context. The government of Malawi has been implementing a farm input subsidy programme (FISP) since the 2005/6 growing season. Every growing season, FISP provides low-cost fertilizer and improved maize seeds to poor smallholders who are mostly rural based (Chirwa and Dorward, 2013). It is a massive undertaking on the part of government; for instance, in the 2012/13 financial year, the programme represented 4.6% of GDP or 11.5% of the total national budget (World Bank, 2013). The frontier results offer some interesting insights into how the impact on maize productivity of the two inputs under FISP compare with the impact of maize yield *in utero*.

The estimated elasticities suggest that the combined effect on maize output in adulthood of a 1% increase in seed and fertilizer is 0.37%. This translates into an increase in the average maize yield of 2.14kg, from 570.31kg to 572.45kg. In contrast, a 1% increase in the average relative maize yield *in utero* (this is equivalent to an increase in actual maize yield *in utero* from 3.38 metric tonnes per hectare to 3.41 metric tonnes per hectare) on adult-life maize output is 0.42%. This converts into an increase in the average maize yield of 2.37kg, from 570.31kg to 572.68kg. This means that the impact of inputs under FISP on maize productivity is almost of the same order of magnitude as the long-term impact of maize yield *in utero*.

3.3 Robustness Checks and Potential Pathways

I examine whether our principle result-that maize yield *in utero* in a farmer's district of birth has positive long term effects on maize productivity- is robust to a number of specification issues namely; migration, age restriction, serial correlation, alternative early-life maize yield definition, and sample selection. I also look at the potential pathways through which maize yield *in utero* affects maize productivity later in a maize farmer's adult life.

A possible concern about the key finding of this paper is that maize yield shocks in early life could have forced parents of some of the farmers to move with their children out of their district of birth to better districts. This migration can potentially bias the results to the extent that migration is correlated with maize yields in early. The IHS3 collects information about migration in terms of whether household members (including those involved in farming) still live in or outside their district of birth at the time of the survey. Using this information, I am able to distinguish those maize farmers whose current district is the same as their district of birth from those who moved from their district of birth. Out of a total of 1275 maize farmers used earlier, I end up with 1123 farmers who indicated that they were born in a village or town in their current district. The results in column of 1 of Table 5 allay this migration concern as maize yield *in utero* still has positive and statistically significant effect on maize productivity in this restricted sample.

One would expect that if maize yield *in utero* has a really lasting effect on maize productivity, this effect would be more pronounced and more evident for older farmers. To check this, I re-estimated the frontier model on a sub-sample of farmers who are aged 20 and above. With this restriction, the overall sample of 1275 farmers is reduced to 1202. The results in column 2 of Table 5 indicate that the coefficient on maize yield *in utero* of 0.949 is not just positive and significant, but is indeed larger than the 0.852 seen earlier for the overall sample.

Another specification concern is that maize yields *in utero* could be serially correlated with maize yields two years before birth or longer. Hence, the effect of maize yield *in utero* on maize productivity could be picking up this lagged effect. To alleviate this concern, I re-estimated the Cobb-Douglas stochastic frontier with relative maize yield two years prior to birth included as an additional covariate. The results for this sensitivity check are reported in column 3 of Table 5. Controlling for the lagged effect of maize yield does not change our earlier conclusion that maize yield *in utero* has a lasting impact on maize productivity in adult life. Besides, though there is a positive effect of maize yield two years prior to birth on productivity, the effect is statistically not different from zero.

The key result of this paper has been based on a transformation of early-life maize yields to get relative maize yields by using the cumulative gamma distribution. I check the robustness of the result by alternatively defining the three maize yield variables in early life as deviations from the local mean in a year (see e.g. Jayachandran (2006), Mancini and Yang (2009), Tiwari et al. (2017)).The three deviations in maize yield are generated as the natural log of maize yield minus the natural log of average annual maize yield in the district of birth. This means that the deviations capture maize yield from the norm in one's birth district. In computing the mean yield, the yield in the individual's district of birth is excluded. In terms of interpretation, a value of 0.05 suggests that maize yield was approximately 5% higher than normal. The frontier results for the re-defined variables are shown in Table 6. Just like before, only the deviation of maize yield *in utero* has a positive and statistically significant effect on maize productivity. The key result of this paper is therefore insensitive to an alternative definition of early-life maize yields.

Finally, selection effects might confound the key finding of this paper. There are two selection concerns. First, there is potential for positive selection parents of farmers born in years with good maize yields. To alleviate this concern, I estimated a linear regression of parental characteristics on early life maize yields. I use years of schooling of a farmer's father and mother separately to capture parental characteristics. The results in Table 7 indicate that there is no statistically significant relationship between maize yields in a farmer's early life and parental characteristics. Second, farmers are included in the data if they were alive in 2010/11. A concern with this that there is a potential problem of sample selection if early-life maize yield influences the likelihood of a farmer surviving through 2010/11.

To check for this possible mortality selection, I estimated two linear regressions of male and female farmers' birth-district and birth-year cohort sizes on early life maize yields. The disaggregation by gender is critical because as found by Waldron (1983) boys are more vulnerable than girls to dying in childhood. As a result of this, one would expect mortality selection to be more evident among males than females The results are reported in Table 8. All the three maize yield variables have no statistically significant relationship with birth-district and birth-year cohort sizes for male and female farmers.

The results suggest that the effects of maize yield shocks may still be felt many years or even decades later. As pointed out by Hoynes et. al.(2016) causal mechanisms through which early-life events have long-run effects are best understood for nutrition. Thus, from the nutrition perspective, there are three potential pathways through which maize yields in early would influence maize yields later in life.

First, high maize yield *in utero* could reflect higher household incomes which in turn could be used to purchase and provide better nutrition (Maccini and Yang, 2009). This represents an indirect effect of maize production on nutrition. Second, and a rather more direct channel, the existence of long-run effects of early-life maize yield might be explained by the "foetal origins hypothesis" or Barker's hypothesis (Barker 1992; Almond and Currie, 2011). The "foetal origins hypothesis" postulates that adult outcomes are strongly influenced by experiences in the womb, in infancy and in early childhood. Hence, low maize yield in early life may reflect food inavailability at a critical period of life which may have long-run negative irreversible effects on maize productivity.

Finally, in Malawi maize is a source of 56-72% of B vitamins (Ecker and Qaim, 2011). B vitamins are critical in health and brain function (Kennedy, 2016), and moreover, Vitamin B12 helps prevent a type of anemia called megaloblastic anemia that makes people tired and weak. Deficiencies in these vitamins may affect their capacity to invest in learning during childhood and may harm their long-run outcomes (Hoynes et. al., 2016), and in the case of this paper these outcomes include productivity.

4 Conclusion and Implications

The paper has looked at the effects of maize yields at different times in early life namely; just prior to birth (*in utero*), in the first and second years of life on adult life productivity and efficiency of maize farmers in rural Malawi. To ensure that early life maize yields are not confounded by omitted local characteristics, they are transformed into relative maize yields by using a cumulative gamma distribution. I find that maize yield just prior to birth significantly increases maize output in a farmer's adult life.

However, maize yields in the first and second years of life have no long-term effects on maize production. Furthermore, there is no long-term impact of early life maize yields on the technical efficiency of maize production. These findings survive a number of robustness checks including alternative definitions of early life maize yields, controlling for migration, and allowing for serial correlation. Furthermore, the results are not driven by sample selection originating from survival induced by maize yields in early life.

The results have useful policy implications. First, as shown by Mussa (2015), maize productivity for Malawi was 2.1 metric tonnes per hectare between 2006 and 2012, but this is significantly lower than the corresponding maize productivity levels of 4.1 and

9.3 for South Africa and the United States of America respectively. The findings of this study thus suggest that contemporaneous productivity enhancing interventions such as FISP alone may have a limited impact on closing this maize productivity gap unless a longer term view of productivity is adopted. This has to involve a realisation that low maize productivity in early life has permanent and irreversible effects on the productivity of maize farmers, and that interventions such as weather insurance or policies that ensure food security especially maize availability to infants is critical.

Second, a number of studies have assessed the impact of FISP on maize productivity and other outcomes (see e.g. Ricker-Gilbert et al. (2011), Ricker-Gilbert et al. (2013) Chirwa and Dorward (2013)). Although these studies cast some doubts over the reported magnitude of the increase in maize production which is attributable to FISP, the findings of this study point to a possibility of the existence of some long-term benefits of FISP which have hitherto not been measured, and which are yet to be realised. The paper has shown that the impact of inputs under FISP on maize productivity is almost of the same order of magnitude as the long-term impact of maize yield *in utero*. Thus, infants who have been exposed to the increased maize productivity arising from FISP may benefit through increased maize productivity when they become maize farmers in adulthood.

Finally, the findings imply that there is partial consumption smoothing among households in rural Malawi. The fact that temporary shocks in early life have permanent effects suggests that households have limited smoothing ability possibly arising from a lack of mitigation strategies such as formal and informal support networks (Dercon and Hoddinot, 2003; Islam and Maitra, 2012)). All this then means that households are forced to trade-off between short-run consumption and longer-run earnings and human capital accumulation (Ferreira and Schady, 2009; Fiszbein et. al.,2009; Maccini and Yang, 2009; Currie and Vogl, 2013). As pointed out by Dercon and Hoddinot (2003) such temporary negative shocks may lead to a poverty trap characterised by a permanently lower equilibrium income stream (through lower maize productivity) in adulthood, making previously feasible outcomes impossible. As shown by Islam and Maitra (2012), microcredit has a significant shock mitigating effect such that microcredit organizations and microcredit have an insurance role to play.

References

Almond D, Currie J. 2011. Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives* 25: 153-172.

Asadullahah, M. N. and Rahman, S. 2009. Farm productivity and efficiency in rural Bangladesh: the role of education revisited, *Applied Economics*, 41, pp. 17-33.

Asselin, L.M. 2002. Multidimensional poverty: Composite indicator of multidimensional poverty. Levis, Quebec: Institut de Mathematique Gauss.

Blasius J, Greenacre M. 2006. Correspondence analysis and related methods in practice. In M.Greenacre, and J. Blasius (Eds.), Multiple correspondence analysis and related methods (pp. 3-40). London: Chapman and Hall.

Barker, DJP. 1992. Fetal and Infant Origins of Adult Disease. *British Medical Journal* 301: 1111.

Battese, G. E. and Coelli, T. J. 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data, *Journal of Econometrics*, 38: 387-399.

Bleakley H. 2010. Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure. *American Economic Journal of Applied Economics* 2:1-45.

Burke, M., Gong, E., and Jones, K. 2014. Income shocks and HIV in africa. *The Economic Journal*, 125, 1157-1189.

Chirwa, E. and Dorward, A. 2013. *Agricultural Input Subsidies: The Recent Malawi Experience*, Oxford: Oxford University Press.

Coelli, T. J. (1995) Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis, *Journal of Productivity Analysis*, 6, pp. 247-268.

Coelli, T. J. and Battese, G. E. 1996. Identification of Factors which Influence the Technical Inefficiency of Indian Farmers, *Australian Journal of Agricultural Economics*, 40, pp. 103-128.

Currie, J. and Vogl, T. 2013. Early-Life Health and Adult Circumstance in Developing Countries, *Annual Review of Economics,* 5:1-36.

Deininger, K. and Jin, S. 2006. Tenure security and land related investment: evidence from Ethiopia, *European Economic Review*, 50, pp. 1245-1277.

Dercon, S. and Hoddinott, J. 2003. Health, shocks and poverty persistence, WIDER Discussion Papers, World Institute for Development Economics (UNUWIDER), No. 2003/08

Dercon, S., and C. Porter. 2014. Live Aid revisited: Long-term impacts of the 1984 Ethiopian famine on children. *Journal of the European Economic Association* 12: 927-948.

Dinar, A., Karagiannis, and Tzouvelekas, G. V. 2007. Evaluating the impact of agricultural extension on farms performance in Crete: a nonneutral stochastic frontier approach, *Agricultural Economics*, 36, pp.133-144.

Ecker O., and Qaim, M. 2011. Analyzing Nutritional Impacts of Policies: An Empirical Study for Malawi, *World Development*, 39, pp. 412-428.

Flatø, M., Muttarak , R., and Pelser, A. 2016. Women, Weather, and Woes: The Triangular Dynamics of Female-Headed Households, Economic Vulnerability, and Climate Variability in South Africa, *World Development*, http://dx.doi.org/10.1016/j.worlddev.2016.08.015.

Ferreira, F. and Schady, N. 2009. Aggregate economic shocks, child schooling, and child health. *World Bank Research Observer*, 24:147-181.

Fiszbein, A, Schady ,N., Ferreira F., Grosh, M., Keleher, N., Olinto, P., and Skoufias, E. 2009. Conditional Cash Transfers : Reducing Present and Future Poverty, World Bank Publications, The World Bank, number 2597.

Gørgens, T. Xin M, Vaithianathan R. 2012. Stunting and selection effects of famine: A case study of the Great Chinese Famine. *Journal of Development Economics* 97:99-111.

Grimard, F. and S. Laszlo. 2014. Long-Term Effects of Civil Conflict on Women's Health Outcomes in Peru. *World Development*, 54: 139-155.

Hoynes, H., DW.,Schanzenbach and Almond D. 2016. Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review*, 106: 903-934.

Huang, C. J. and J. T. Liu. 1994. Estimation of a Non-Neutral Stochastic Frontier Production Function, *Journal of Productivity Analysis*, 5, pp. 171-180.

Islam, A. and Maitra P. 2012. Health shocks and consumption smoothing in rural households: Does microcredit have a role to play? *Journal of Development Economics* 97: 232-243.

Jayachandran, S. 2006. Selling labor low: wage responses to productivity shocks in developing countries. *Journal of Political Economy*, 114: 538-575.

Kelly E. 2011. The Scourge of Asian Flu *In utero* Exposure to Pandemic Influenza and the Development of a Cohort of British Children. *Journal of Human Resources* 46:669-694.

Kennedy, DO. 2016. B Vitamins and the Brain: Mechanisms, Dose and Efficacy-A Review. *Nutrients,* 8, 68. http://doi.org/10.3390/nu8020068.

Lin MJ, Liu E. 2014. Does *In utero* Exposure to Illness Matter? The 1918 Influenza Epidemic in Taiwan as a Natural Experiment. *Journal of Health Economics*, 37: 152-163

Liu, Y, and Myers, R. 2009. Model selection in stochastic frontier analysis with an application to maize production in Kenya, *Journal of Productivity Analysis*, 31, pp. 33-46.

Maccini S, Yang D. 2009. Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *American Economic Review* 99:1006-1026.

Mansour H, Rees DI. 2012. Armed conflict and birthweight: Evidence from the al-Aqsa Intifada. *Journal of Development Economics* 99:190-199.

Mussa, R. 2015. Do the Poor Pay More for Maize in Malawi?, *Journal of International Development*, 27:546-563.

NSO (National Statistics Office). 2012. Integrated Household Survey 2010-2011. Report, National Statistics Office, Zomba, Malawi.

Ricker-Gilbert J, Jayne TS, Chirwa E. 2011. Subsidies and Crowding Out: A Double-Hurdle Model of Fertilizer Demand in Malawi. *American Journal of Agricultural Economics* 93: 26-42.

Ricker-Gilbert J, Mason NM, Jayne TS, Darko F, Tembo S. 2013. What are the Effects of Input Subsidies on Maize Prices? Evidence from Malawi and Zambia. *Agricultural Economics* 44: 671-686.

Smale M. 1995. Maize is life: Malawi's delayed green revolution. *World Development*, 23: 819-831.

Strauss J, Thomas D. 1998. Health, Nutrition, and Economic Development. *Journal of Economic Literature* 36: 766-817.

Tiwari, S., Jacoby, H.G. and Skoufias, E. 2017. Monsoon babies: rainfall shocks and child nutrition in Nepal. *Economic Development and Cultural Change* 65: 167-188.

Udry, C. 1996. Gender, Agricultural Production and the Theory of the Household, *Journal of Political Economy*, 104, pp. 1010-1046.

Waldron, I. 1983.Sex differences in human mortality: the role of genetic factors, *Social Science and Medicine*, 17, 321-333.

Weir, S. and Knight, J. 2007. Production externalities of education: evidence from rural Ethiopia, *Journal of African Economies*, 16, pp. 134-65.

World Bank. 2013. Malawi Public Expenditure Review, Report No. 79865-MW.

Figure 1. Actual maize yield by year of birth for farmers born 1984-1995

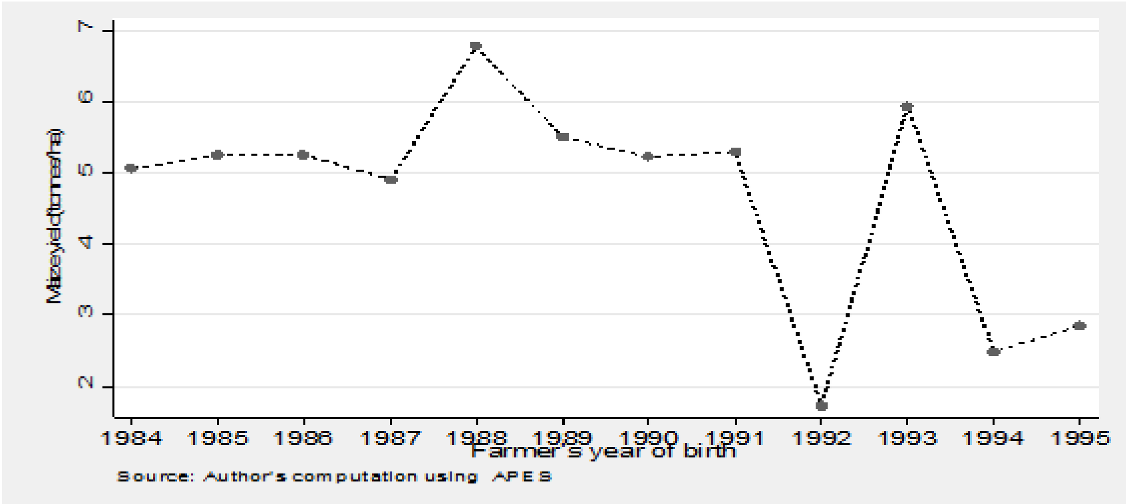

Figure 2. Boxplots of relative maize yield in early life by year of birth for farmers born 1984-1995

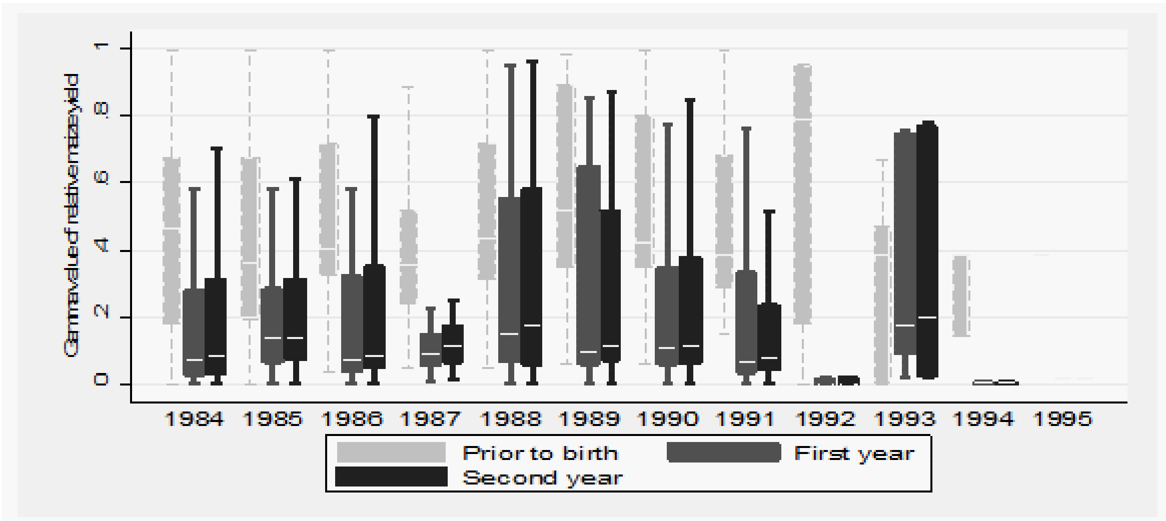

Table 1. Selected production characteristics by quintiles of relative maize yield in early life of farmers born 1984-1995

Variable	Maize yield in utero		Maize yield first year of life		Maize yield second year of life	
	First	Fifth	First	Fifth	First	Fifth
yield	317.07	825.56	354.25	773.78	364.33	776.65
seed	8.69	8.56	12.04	9.06	11.93	9.16
land	0.74	0.98	0.80	0.96	0.76	0.96
fertilizer	54.04	78.94	56.54	80.17	55.13	80.70
labour	24.99	27.59	25.26	27.37	25.02	27.42
capital	629.33	1317.71	616.27	1328.02	648.65	1330.57
sex of farmer	0.75	0.77	0.73	0.78	0.74	0.78
age of farmer	23.84	23.18	23.37	23.22	23.35	23.25
accessed govt. or private extension	0.28	0.17	0.28	0.19	0.29	0.19
has secure land tenure	0.82	0.86	0.76	0.87	0.76	0.87
Observations	335	318	338	321	327	322

Table 2. Summary statistics, farmers born 1984-1995

Variable	Mean	SD
yield	570.314	995.527
seed	9.129	39.734
land	0.833	0.646
fertilizer	64.537	117.527
labour	26.210	11.655
capital	838.876	4490.752
relative maize yield in utero	0.487	0.295
relative maize yield first year of life (t+1)	0.242	0.285
relative maize yield second year of life (t+2)	0.257	0.290
sex of farmer	0.755	0.430
age of farmer	23.396	2.179
average years of schooling in a HH	3.761	2.386
accessed govt. or private extension	0.242	0.428
has secure land tenure	0.831	0.374
index of economic infrastructure	-0.219	0.775
Observations	1626	

Table 3. Model specification tests

No.	Hypothesis	Wald /Z statistic	P-value	Conclusion
1.	$H_0 : \beta_1 = \cdots = \beta_5 = \phi_t = \phi_{t+1} = \phi_{t+2} = \delta_1 = \delta_2 = 0$	2247.05	0.00	Frontier variables jointly significant
2.	$H_0 : \sum_{k=1}^{3} \phi_k + \sum_{k=1}^{5} \beta_k = 1$	12.22	0.00	No constant returns to scale
3.	$H_0 : \phi_t = \phi_{t+1} = \phi_{t+2} = 0$	34.41	0.00	There are production based effects of early life maize yield
4.	$H_0 : \mu = 0 = \sigma_u^2 = 0$	-8.33 [a]	0.00	Inefficiency effects are present
5.	$H_0 : \gamma_t = \gamma_{t+1} = \gamma_{t+2} = \alpha = 0$	1118.54	0.00	Efficiency variables jointly significant
6.	$H_0 : \gamma_t = \gamma_{t+1} = \gamma_{t+2} = 0$	1.17	0.95	No efficiency based effect of early life maize yield

[a] This is based on the standard normal statistic. DF is degrees of freedom.

Table 4. Cobb-Douglas stochastic production frontier and inefficiency effects models

Variable	Frontier Model	Inefficiency Model
log of seed	0.055***	
	(0.012)	
log of land	0.197***	
	(0.061)	
log of fertilizer	0.319***	
	(0.079)	
log of labour	0.096***	
	(0.017)	
log of capital	0.110***	
	(0.036)	
relative maize yield *in utero*	0.852***	2.892
	(0.213)	(4.952)
relative maize yield first year of life (t+1)	-0.155	-1.301
	(0.674)	(7.195)
relative maize yield second year of life (t+2)	0.119	-1.213
	(0.562)	(4.020)
sex of farmer		-3.979**
		(1.762)
age of farmer		-5.740*
		(3.359)
average years of schooling in HH		-0.691***
		(0.252)
accessed govt. or private extension		-1.947***
		(0.099)
has secure land tenure		-2.901***
		(0.305)
index of economic infrastructure		-0.106*
		(0.057)
month of birth fixed effects	Yes	Yes
year of birth fixed effects	Yes	Yes
district of birth fixed effects	Yes	Yes
Observations	1626	1626

Notes: Relative maize yield is the cumulative gamma distribution of maize yield. In parentheses are standard errors clustered at the district level. ***Indicates significant at 1%, **at 5% and *at 10%.

Figure 3. Boxplots of relative maize yield in early life by year of birth

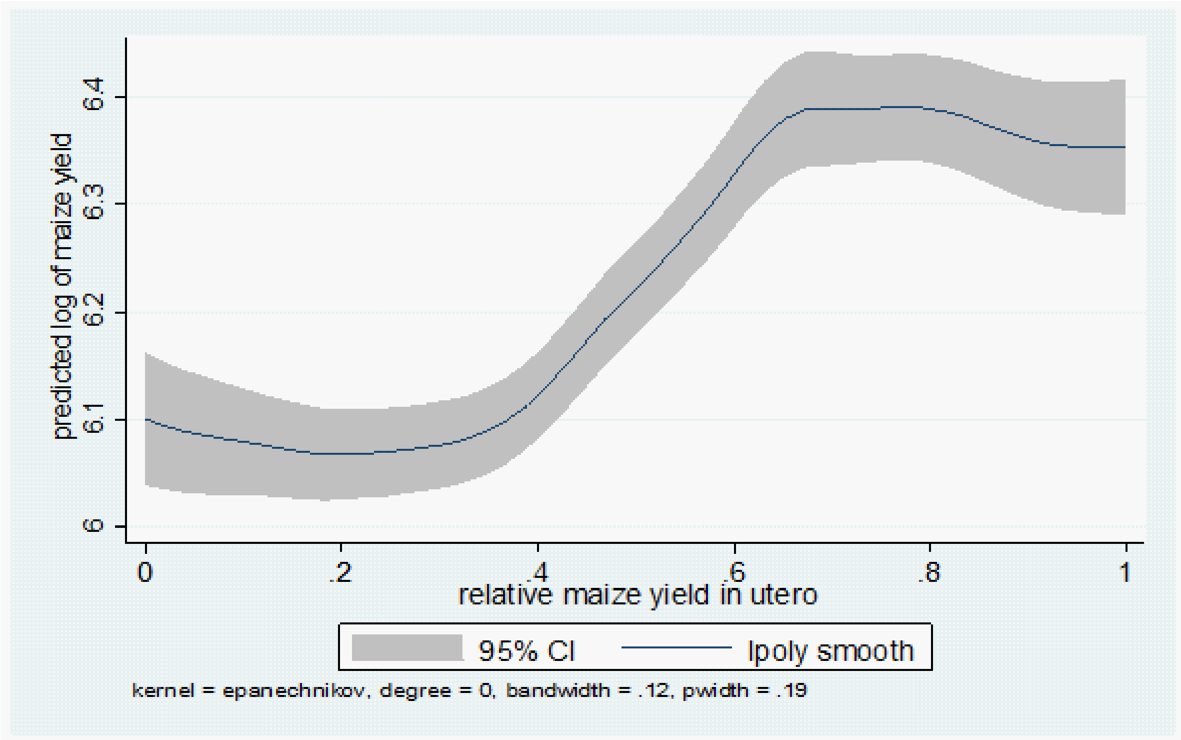

Table 5. Robustness checks for the Cobb-Douglas production frontier

Variable	(1)	(2)	(3)
	Same District	Aged above 20	Lagged Effect
relative maize yield *in utero*	0.874***	0.949***	0.532*
	(0.236)	(0.258)	(0.307)
relative maize yield first year of life (t+1)	-0.286	-0.591	-0.109
	(0.662)	(0.873)	(0.678)
relative maize yield second year of life (t+2)	0.246	0.481	0.113
	(0.577)	(0.602)	(0.558)
relative maize yield two years prior to birth (t)			0.514
			(0.370)
Observations	1427	1540	1626

Notes: Relative maize yield is the cumulative gamma distribution of maize yield. The conventional inputs, month of birth, year of birth, and district of birth fixed effects are included in the estimation. In parentheses are standard errors clustered at the district level. ***Indicates significant at 1%, **at 5% and *at 10%.

Table 6. Cobb-Douglas stochastic production frontier with maize yields as deviations

Variable	Frontier
deviation of maize yield *in utero*	0.263***
	(0.045)
deviation of maize yield in the first year of life (t+1)	0.258
	(0.192)
deviation of maize yield in the second year of life (t+2)	-0.017
	(0.181)
Observations	1626

Notes: The deviations in maize yield are generated as the natural log of maize yield minus the natural log of average annual maize yield in the district of birth. The conventional inputs, month of birth, year of birth, and district of birth fixed effects are included in the estimation. In parentheses are standard errors clustered at the district level. ***Indicates significant at 1%, **at 5% and *at 10%.

Table 7. Linear regression results of parental years of education on early life maize yield

Variable	Father	Mother
relative maize yield *in utero*	-0.715	0.305
	(0.591)	(0.323)
relative maize yield first year of life (t+1)	1.249	0.710
	(1.805)	(0.982)
relative maize yield second year of life (t+2)	-0.195	-0.528
	(1.663)	(0.903)
F-statistic	2.24	2.56
R-squared	0.06	0.07
Observations	1618	1597

Notes: Relative maize yield is the cumulative gamma distribution of maize yield. In parentheses are standard errors. ***Indicates significant at 1%, **at 5% and *at 10%.

Table 8. Linear regression of cohort size in a district and year of birth on average maize yield

Variable	Male Farmers	Female Farmers
mean relative maize yield *in utero*	-0.282	-1.155
	(1.775)	(1.232)
mean relative maize yield first year of life (t+1)	7.151	4.938
	(6.260)	(5.342)
mean relative maize yield second year of life (t+2)	-7.344	-7.175
	(6.177)	(5.212)
F-statistic	10.07	2.64
R-squared	0.70	0.48
Observations	189	138

Notes: The dependent variable is the cohort size in a farmer's district and year of birth. Relative maize yield is the cumulative gamma distribution of maize yield. The mean yield is for the district and year of birth. In parentheses are standard errors. ***Indicates significant at 1%, **at 5% and *at 10%.