Long-run Fertilizer Production Function and Health Effects

Monika Baloda[1] and Rajveer Jat[2]

[1]Graduate School of Management, University of California, Riverside
[2]Department of Economics, University of California, Riverside

September 16, 2022

Abstract

We estimate long-run production function of crop-yield as a function of fertilizer use per unit land. Using non-parametric functional form tests, we reject the standard functional form used in the literature. We then estimate a highly flexible empirical production function using non-parametric methods. Our analysis suggests that the functional form can be approximated with at-least seven-degree polynomial form. To facilitate easier theoretical analysis, we outline a logistic-style production whose analytical results mimic our empirical outcomes. In second theme of work, we provide evidences of negative impacts of fertilizer use on both human health and soil health/fertility. We then discuss the policy implications of our results. We also provide an alternate theoretical viewpoint on fertilizer use from on bio-sustainability lens.

1 Introduction

In the process of structural transformation, increase in the productivity of agriculture is an essential step. More productive agriculture relaxes food constraints therefore releasing labor to move to more productive modern non-agriculture sector. Fertilizer has been an unequivocal facilitator of increase in agriculture productivity. Developing countries spend huge[1] amount of money in form of subsidies to make fertilizers affordable to farmers. However little is known on long-term impacts of fertilizer use on crop yields. Standard theoretical literature considers fertilizer to be an intermediate input by including it in Cobb-Douglas style production function (Hayami and Rutton (*AER*, 1970)). This theoretical modeling gives us positive and diminishing marginal returns on fertilizer use. Empirical literature employs linear regression model with in a causal identification framework. In this paper, we take one step back and attempt to estimate the long-run production function of crop-yield (y) with fertilizer (x) being the input i.e. $y = f(x)$.

Our main contribution in this paper is to estimate empirical production function using 45 years long time series on 540 present day districts in India. We first check using basic parametric methods such as simple linear regression whether the standard theoretical models work for our data. We find that it explains well and regression results support the standard model. However, one of the major assumption in parametric methods is that the underlying model is linear in parameter. However, there is no obvious way to justify this assumption. We go ahead and test whether underlying relationship between fertilizer intensity and crop-yield can be explained by linear, quadratic, or a complex quadratic plus fixed effect model. Using Li-Wang specification test for functional forms, we reject all these specifications. Thereby concluding that the standard theoretical model is not true. Our large number of observations allows us to use data-intense non-parametric econometric methods. We employ sophisticated technique to assess marginal effects of fertilizer use on wheat yield. Our methodology does not impose any restriction on functional form of marginal effects and production function to begin with. For the given data, we estimate the best possible production function using local linear least square estimation technique. Using bootstrap method, we generate a confidence interval for the estimate of marginal effect at each level of fertilizer use. This facilitate our hypothesis testing problem.

Our finding is that theoretical formulation of "positive and diminishing marginal returns" does not explain the long-run dynamics. Further, empirical literature estimating a constant marginal

[1]Fertilizer subsidies were 0.63% of GDP in the year 2013-14 in India. Source: Indian Agricultural Research Institute, New Delhi

effect (β) using causal identification strategies should consider an alternate functional form beyond linear. Our analysis suggests that if we want to fit the production function in polynomial world, we need to approximate it by a seven-degree polynomial function. Further, we find evidences of negative marginal effects on overuse of fertilizers. Negative effects of fertilizer use is well known and accepted in the toxicology literature but in economics there isn't much work on this. In that light, our work is a new contribution. In order to suggest an alternate simple theoretical modeling, we borrow Weitzman (1974) style model for renewable resources. Our idea is that fertilizers can be seen as an effort to extract nutrients in the soil in form of crop-yield. The logistic-style growth function of nutrients and linear extraction function gives us an long-run equilibrium level of nutrient stock. We analytically show that if we go beyond a certain level of fertilizer use per unit land, the long-run nutrient stock may extinct. These analytical results mimic the empirical findings using non-parametric methods, therefore this model can work as simple but better theoretical framework than standard theoretical model.

Our second contribution is to present an alternate view point from the health economics and environmental toxicology lens. Fertilizer certainly increases the yield (output per unit area) hence income at-least in short run (McArthur, 2017) but its impact on human health is theoretically ambiguous. For a poor farmer, increased use of fertilizer increases her crop yield hence the income and therefore her health budget. Since the farmer now has more money to spend on health, this can positively impact the health, however indirectly. At the same time it possibly may have direct health consequences for the food consumer including farmer herself through increased concentration of hazardous elements in food-produce. Barnwal et al (2017) finds that fertilizer as a crucial input along with water and HYV yields led to around a 3-4 percentage point decrease in infant mortality (from a baseline of 17%), averting around 3-5 million infant deaths per year by 2000. While excess use of fertilizer leads to food-borne diseases (Deb, 2018). Therefore, an informed research on identification of safe limit of fertilizer use is an important question.

The effect of Nitrogen-Phosphorous-Potassium containing fertilizers (hereafter NPK) on Human health is through two main channels: First, direct effect through their entrance into the food chain in the form such as nitrate intake. Second, excess fertilizer use causes plants to increase the uptake of harmful heavy metals like Cadmium (Cd) which then enter into the food chain hence indirect impact (Atafar et al (2010), Ju et al (2007)). Nitrate intake causes a number of health disorders, namely, malformed child births, methaemoglobinemia, gastric cancer, goitre, birth malformations, hypertension, etc. ((Dorsche, 1984), Majumdar and Gupta(2000)). There are reports relating high nitrate intake and increasing incidences of goitre through reduction of iodine assimilation by human body (ECETOC, 1988). M.Crespi and V. Ramazzotti (1991) says that N-nitroso compounds (NOS) are also potent animal carcinogens and their formation, from variety of precursors in the body of animals, has been demonstrated, and this may also occur in humans. Higher phosphate concentration increases the Cd uptake by a plant. In an experiment, Growth chamber studies find that crop Cd concentration was greater when the flax was grown with commercial triple superphosphate or monoammonium phosphate (that contained Cd)than with reagent-grade sources containing only trace amounts of Cd (Jiao et al. 2004). Higher Cd uptake causes mainly Kidney disease (Roberts 2014), in other adverse effects diseases related to pulmonary, cardiovascular, and musculoskeletal systems have been reported. The well known case of Cd toxicity (i.e. itai-itai disease) occurred with subsistence farmers in Japan growing rice on soils contaminated with industrial wastes.

It is intuitively easy to believe that excess fertilizer use impacts human health through soil. Therefore, it becomes a question of vital importance to investigate how fertilizer, in particular excess use of it, impacts soil health. For a variable to be a complete measure of soil health it is essential to take environmental quality and its effects on human health into consideration. In practice, it is very difficult to measure soil health completely. However, soil fertility can be measured in a transparent manner and it can be a credible proxy-measure for soil health. The important point is that the soil fertility is a direct observable and economically meaningful variable, based on which a farmer makes her decision(s). Generally, high organic matter in soil with active soil biology has good soil fertility but farming practices to get better yield using excessive inorganic fertilizers causes nutrient imbalance, lower pest resistance, nutrients are also lost easily and leaching or gas emission further lead to reduced fertilizer efficiency (Chen, 2006). The major effect of fertilizer is observed on soil pH value which was significantly correlated to the differences in bacterial community composition (Lauber et al. 2009) that plays important role in organic matter decomposition, nutrient cycling and other chemical changes (Murphy et al, 2007). The pH of soils vary depending on their initial pH and organic residues (Eghball, 2002; García-Gil et al.; 2004; Butler and Muir, 2006; Meng et al., 2005; Bastida et al., 2008). In soil with a pH below five, there was reduction in microbial biomass whereas fertilization has positive effect on soil with higher pH value (Gesseler, 2014). Higher phosphate concentration increases the Cd uptake by a plant(Jiao et al. 2004). Heavy metals like Cd and Zn uptake is higher in leafy plants as compared to grains, this behaviour is particularly evident in low soil pH (Miller Miller 2000). These examples teach us that fertilizer use impacts the soil health negatively, however fertilizer is necessary to produce enough food to feed the world. Given the two set of constraints, an important research question is to find the 'safe' limit of fertilizer use per unit land.

In sum, we test for the validity of the standard theoretical production function of fertilizer use. We then estimate an empirical production function using a 45-year long time series on 540 districts (540 on present day boundaries, 305 on 1966 boundaries) of India. In second contribution, we present evidences to show negative impacts of fertilizer use on human health. The rest of the paper is organized as follows: section-2 discuss the data. We describe the sources, time-period and cross-sectional units of observations. Due to difficulties in combining all crops into a single index and diverse cropping patterns in India, we construct a sample with sufficient production of a single major crop. We choose wheat as our main crop of interest and construct sample based on share of area allocated to wheat cultivation. Section-3 is the heart of the paper, we start with standard theoretical model of fertilizer production function. We then take it to data. We then formally test whether theoretical functional forms holds in the data. Empirical estimation of production function is done is using non-parametric methods. In the last subsection, we present an alternate theoretical model which mimics the empirical results. Section-4 discusses impact of fertilizer on human health indicators. We start by presenting some descriptive statistics to set stage of our hypothesis. We then employ a linear regression with time and state fixed model to investigate the relationship between the two. We do not claim the results to be causal but argue that it certainly is not a naive correlation by chance result. In section-5, we discuss policy implications of the results obtained in the section-3 and section-4. Our main conclusion is that fertilizer uses per unit land should not be allowed beyond a threshold level. Using a toy model, we show that it is possible for farmers to use too much fertilizer under an open-access or private optimal problem than the social optimal level. Section-6 summarizes everything and concludes the paper.

2 Data

This section describes data sources, sample construction and the definitions of the main variables of interest.

2.1 Data-sources

Our main data-source for the analysis in this paper is ICRISAT district level database of nineteen big states[2] of India covering more than 80% of geography and population. The districts definition in this database relate to 1966 base i.e., data of districts formed after 1966 are given back to their parent districts and removed from the file. Thus, consistent and comparable time series data are available for all districts formed prior to 1966. In total, we have data on 305 districts for 46 years (1966 to 2011, both years including), thereby 14030 number of observations. ICRISAT dataset contains information on large number of variables including fertilizer use per unit land, crop yields, harvest prices, rainfall, irrigation, technology use, soil types etc.

The second main source of data comes from *Global Burden of Disease Study 2019* conducted by Institute for Health Metrics and Evaluation (IHME), 2020, Seattle, United States. Global Burden of Disease data on India is published on the website of the Indian Council of Medical Research (ICMR [3]). This information is available at state level for each years between 1991 and 2011. We use information on Non-communicable disease provided under different definitions such as mortality

[2]Nineteen big states are : Andhra Pradesh, Gujarat, Haryana, Karnataka, Madhya Pradesh, Maharashtra, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, Bihar, West Bengal, Orissa, Assam, Himachal Pradesh, Kerala, Chhattisgarh, Jharkhand, and Uttarakhand.

[3]One can download data from : here.

rate (number of deaths per 100,000 people), YLDs, and DALYs. This data is used for our analysis on fertilizer and human health.

2.2 Sample construction

We are interested in effect of fertilizer use on agricultural output. However, the notion of agriculture output is a bit vague because it is difficult to add-up the productions of different crops. One way to add them up is to convert every crop in value terms using harvest price and quantity produced. However, this introduces another dimension of price which is can make identification hard. Staple crops such as wheat and rice are grown in large number of geographic areas and therefore can credibly reflect the agriculture output. Wheat is a *Rabi* crop grown in winter, farmers mainly inputs such as fertilizers in *Rabi* season. Therefore, we choose *wheat yield* to be the variable representing agriculture output. Although, wheat is a major crop in India but the country is so diverse that not all the districts grow wheat. We therefore choose districts growing wheat sufficiently. For a district i, we define γ_i to be the fraction of total area used for wheat cultivation in a given season. Let γ_0 be the threshold value such that if $\gamma_i \geq \gamma_0$, then we include that observation into our sample. We choose $\gamma_0 = 0.2$, meaning that at-least 20% of the total cultivated area should be allocated to wheat cultivation. We later make sure that our qualitative results are not driven by the choice of γ_0. This gives us a sample of total 5822 observations. We do fertilizer and wheat yield analysis using this sample.

Unfortunately, data on global diseases burden is not available at district level in India, therefore we are restricted at state level. We aggregate the total fertilizer use and total crop-area cultivated in a year at the state level. Then we obtain fertilizer use per unit area cultivated. We have such information for 19 states for 20 years, therefore a total of 380 observations.

2.3 Key Variables

We three main variables of interests are: wheat yield, fertilizer intensity, and health outcomes. We construct normalized measures accurately reflecting the variables of interest. Normalization on per unit land or capita basis makes our variable free from scale effects. Following three bullet points summarizes the same:

1. **Fertilizer** : Urea(NH_3), Phosphate (P_2O_5), and Potash (K_2O) are the three most used commercial fertilizers. These Nitrogen, Phosphorous, and Potassium based fertilizers are represented by the acronym NPK. Our fertilizer measure is the sum of these three fertilizers divided by the total cultivated area at the district level. The unit of this variable used throughout this paper is in kilogram per hectare (kg/hectare). This measure is normalized on per unit land area, therefore is free from scale effects.

2. **Wheat Yield** : Yield basically means the output of a crop produced from a unit area in a given area. Our measure of yield is constructed by dividing total production of wheat in a district by total area used for wheat cultivation. Unit of wheat yield variable is kilogram per hectare ((kg/hectare).

3. **Health Outcome** : Health consequences in the form of non-communicable diseases can be majorly attributed to the broader life-style issues and food-chains. Our main measure of health outcome is *mortality rate* measured as "deaths due to non-communicable diseases per 100,000 people in a given state in a given year". Alternate[4] measures of health outcomes are: DAYLs and YLDs. One DALY represents the loss of the equivalent of one year of full health. YLDs means years of healthy life lost due to disability. Using DALYs, the burden of diseases that cause premature death but little disability (such as drowning or measles) can be compared to that of diseases that do not cause death but do cause disability (such as cataract causing blindness). However, for the sake of easy inference, we use mortality rate as our main measure, but we check whether qualitative results differ if we switch to DALYs.

[4] According to World Health Organization: "Mortality does not give a complete picture of the burden of disease borne by individuals in different populations. The overall burden of disease is assessed using the disability-adjusted life year (DALY), a time-based measure that combines years of life lost due to premature mortality (YLLs) and years of life lost due to time lived in states of less than full health, or years of healthy life lost due to disability (YLDs)."

4. **Other Economic Variables**:ICRISAT complies annual time series data on economic variables such as crop-prices, land, irrigation, land-use, rainfall, technology use etc. We use this information to control for other factors in our main identification strategies.

3 Fertilizer Production Function

3.1 Standard Theoretical Model

3.1.1 Model

Fertilizer enters as an intermediate input in standard Cobb-Douglas style production function of a crop. Mathematically,

$$Y = AK^{\alpha}F^{\beta}L^{1-\alpha-\beta}$$

Y is crop-output per unit land or *yield*, A is technology parameter, K is capital, L is labor input and F is the fertilizer input. The restriction on parameters is standard ones like concave production function i.e. $0 < \alpha, \beta < 1$ and $0 < \alpha + \beta < 1$. This formulation implies that marginal effect of fertilizer use is always positive and is decreasing in intensity of fertilizer use.

3.1.2 Taking this model to data

For a district i in time period j, let's call y_{ij} wheat yield in kg/hectare, x_{ij} be the fertilizer use in kg/hectare. Then in order to take the basic model of last discussion, we estimate the following regression model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \epsilon_{ij} \tag{1}$$

Table 1: Regressing Fertilizer on Wheat Yield : Results

Variable	Model-1	Model-2
β_1 (Fertilizer)	0.007***	.0130***
β_2 (Fertilizer_Sq (x10^{-5})		-1.63***
β_0 (constant)	1.297***	1.015***
R^2	0.45	0.57
Num. Obs.	5822	5822
p-value (F-Stat)	0.000	0.000
***:Significant at 1% level	**: 5% level	*: 10% level

3.1.3 Inferences

The results indicate that there exists evidences to support the standard increasing and concave style theoretical model discussed above. We test this specification using both time and district fixed effects, the qualitative results do not change. Therefore bottom-line conclusion is that standard increasing and concave style production function works.

However the econometric investigation tools used so far are limited to parametric regression family. These class of methods by construction assumes a functional form or specification and then estimate it. However this assumption in the first place should be tested. In the next section, we do the same.

3.2 Parametric Specification Test

In this section we test whether it is good to use a specific functional form for the wheat yield given by $y_{ij} = f(x_{ij}, \beta)$. Here $f(.,.)$ could potentially represent large number of functional form e.g. linear, quadratic, or fixed effect models. We use Li-Wang (1998) test for parametric regression function.

3.2.1 Li-Wang Test

This test is bootstrap based, it gives a more accurate approximation to the null distribution of the test than the asymptotic normal theory result. Suppose a general regression function can be given by the following expression:

$$y_{ij} = m(x_{ij}) + \epsilon_{ij}$$

We are interested to test whether the conditional mean function $E(y_{ij}|x_{ij}) = m(x_{ij})$ takes a specific form i.e. whether $m(x_{ij}) = f(x_{ij}, \beta)$ or not. Formally, we test the following null hypothesis against the alternate hypothesis (determined by the method using data):

$$\begin{array}{lll} \text{Null Hypothesis:} & H_0: & m(x_{ij}) = f(x_{ij}, \beta) \\ \text{Alternate Hypothesis:} & \mathrm{H}_1: & m(x_{ij}) \neq f(x_{ij}, \beta) \end{array}$$

Note that we can test for large number of specification for $m(x_{ij})$ because our $f(x_{ij}, \beta)$ can accommodate both linear and non-linear functional forms in both variable (x) and parameters (β). In particular if $f(x_{ij}, \beta) = x_{ij}\beta$, then we are testing for linear model.

Li-Wang test is a alternate hypothesis based test. Alternate hypothesis is determined by the non-parametric method for our dataset so we have the true alternate hypothesis. The test then employs a *Wald*-test style test statistics to test for the desired specification.

3.2.2 Results

We report the results for three main functional form of interest: linear, quadratic (in x), and quadratic fixed effect model. We test the null hypothesis against the alternate hypothesis determined from the data. We report the LW-test statistics and the associated *p-values*.

Table 2: Specification test for Parametric Regression Functions

Null Hypothesis (functional form of $m(x_{ij})$)	LW-test statistics	p-value
$f(x_{ij}, \beta) = x_{ij}\beta$	206.42	0.0000
$f(x_{ij}, \beta) = x_{ij}\beta_1 + x_{ij}^2\beta_2$	12.81	0.0000
$f(x_{ij}, \beta) = x_{ij}\beta_1 + x_{ij}^2\beta_2 + Controls + FixedEffects$	10.46	0.0000

3.2.3 Inferences

The results shows that associated *p-values* for all three models is less than 0.0001, therefore with more than 99.99 % confidence, we can reject the null hypothesis that conditional mean function $m(x_{ij})$ is equal to the specification $f(x_{ij}, \beta)$. Therefore, we conclude that the standard theoretical model's concave (quadratic) form is not the correct specification for crop-production function. We also reject a possibility of parametric fixed effect model with second order fertilizer term to be the functional form for conditional mean function. Hence, we reject the standard theoretical production function for fertilizer use.

3.3 Non-parametric Estimation

In standard regression problem, we are interested in conditional expectation of our variable of interest Y given X. Denote conditional expectation $E(Y|X) = m(X)$. In parametric form of regression we specify a functional form for $m(X)$ i.e. for example if $E(Y|X) = m(X) = \beta X$ with β being constant then it is our *standard linear regression model.* However, in general, β could vary with X, and further higher order polynomial or non-linear terms of X can be present in the expression of $m(X)$.

Our standard regression problem can be written in a generalized form as follows:

$$scalar-form: \quad y_i = m(x_i) + u_i \qquad vector-form: \quad Y = m(X) + U \tag{2}$$

Let's call $\beta(x)$ be the first derivative of $m(x_i)$ at x i.e. $\frac{\partial m_i(x)}{\partial x_i}|_{x_i=x} = \beta(x)$. We can write down the Taylor series expansion of our $m(x_i)$ in the neighborhood of $x_i = x$ as follows:

$$m(x_i) = m(x) + (x_i - x)\frac{\partial m_i(x)}{\partial x_i}|_{x_i=x} + \frac{1}{2!}(x_i - x)^2\frac{\partial^2 m_i(x)}{\partial x_i^2}|_{x_i=x} + \frac{1}{3!}(x_i - x)^3\frac{\partial^3 m_i(x)}{\partial x_i^3}|_{x_i=x} + ...$$

3.3.1 Local Linear Least Squares

Local linear least squares (LLLS) is an estimation method of $m(X)$ in which we ignore the higher order derivatives starting from the second order derivatives. on the grounds that since x_i is in the close neighbourhood of x, the higher degree terms like $(x_i - x)^2$ becomes close to zero relative to $(x_i - x)$. Also we replace first order derivative by $\beta(x)$ as denoted earlier. With these changes we can re-write the Taylor series expansion as:

$$m(x_i) = m(x) + (x_i - x)\beta(x)$$

Therefore, we can re-write the regression function as:

$$y_i = m(x) + (x_i - x)\beta(x) + u_i$$

Define $z_i(x) = [1 \quad (x_i - x)]$ and $\delta(x) = [m(x) \quad \beta(x)]^T$ where superscript T stands for transpose of a matrix. We can write the regression in the following form:

$$y_i = z_i(x)\delta(x) + u_i$$

Now we define the loss function as:

$$S(x) = \sum_{i=1}^{n} u_i^2 K(\frac{x_i - x}{h}) = \sum_{i=1}^{n} \Big(y_i - m(x) - (x_i - x)\beta(x)\Big)^2 K\left(\frac{x_i - x}{h}\right)$$

The *bandwidth* h is the length of the interval for which we compute kernel weights. Here, $K(.)$ is the kernel function, a continuous weighting device[5].

Following two first order conditions gives us the local linear least square estimators $\hat{m}(x)$ and $\hat{\beta}(x)$.

$$\frac{\partial S(x)}{\partial m(x)} = 0 \qquad \text{and} \qquad \frac{\partial S(x)}{\partial \beta(x)} = 0$$

We can obtain the estimates $\hat{m}(x)$ and $\hat{\beta}(x)$ by jointly solving these two equations. The solution for $\delta(x) = [m(x) \quad \beta(x)]^T$ is given by:

$$\hat{\delta}(x) = [z'(x)K(x)z(x)]^{-1} z'(x)K(x)y$$

This expression is very difficult to compute analytically, therefore, we need to use numerical methods to solve for it. Marginal effect if $\hat{\beta}(x)$ which is the second component of the vector coefficient $\hat{\delta}(x)$. We use bootstrap method to construct confidence intervals.

3.3.2 Results

For y being the wheat yield (output per unit land) in kg/hectare and x being the fertilizer use in same units, we have estimated the relationship between the two using LLLS method described above. Figure-1 reports the marginal effect $\beta(x)$ for each x with confidence interval. On the horizontal axis, we plot the fertilizer use level and corresponding marginal effect $\beta(x)$ on the vertical axis. The red line is represents $\beta(x)$ and dotted lines above and below forms confidence intervals. The horizontal line in yellow color corresponding to $\beta(x) = 0$ is our null hypothesis. Two vertical blue lines are approximately drawn to divide the whole graphs into three regions, it helps us in inferences.

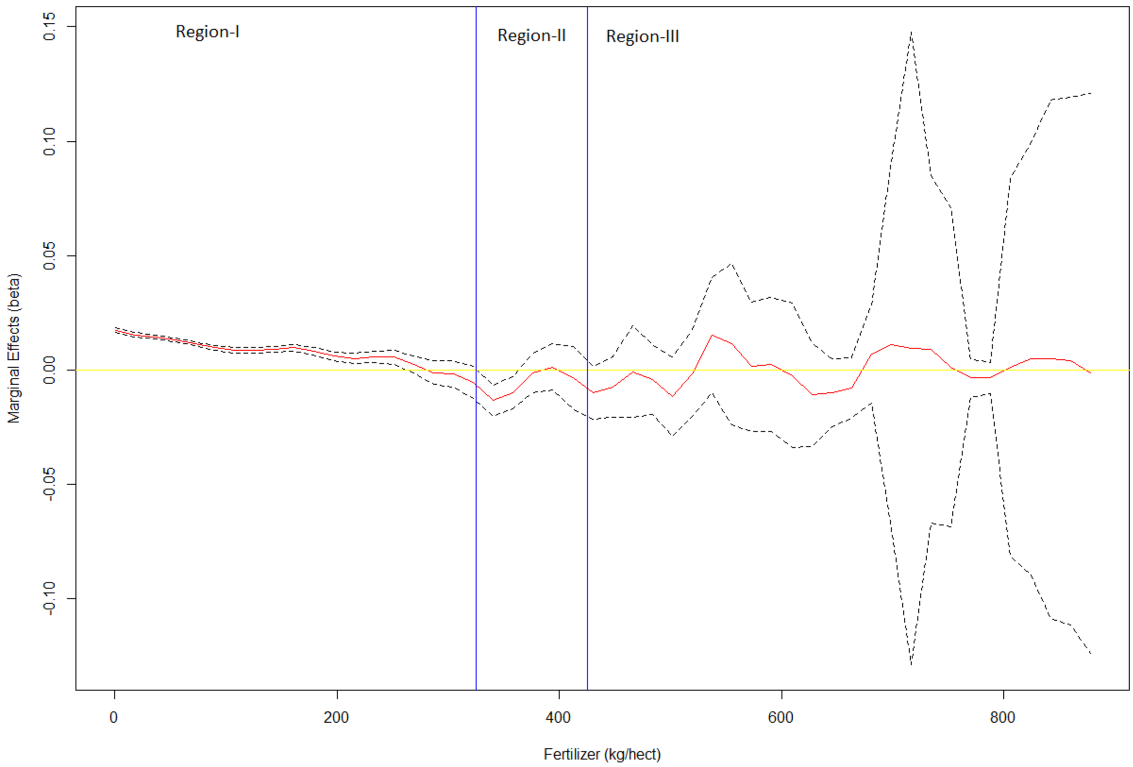
3.3.3 Inferences

The consistency results of the non-parametric estimator of $m(x)$ is given by Liapounov's central limit theorem (Pagan and Ullah (1999)). For n being the number of observations and h being the bandwidth, the rate of convergence of the kernel density estimator is $O(\frac{1}{nh})$ which requires

[5]The basic idea is that we give lesser weight to the observations farther from the point x where we want to estimate. We call this continuous weighing device *Kernel* function. One such kernel function known as "Gaussian" kernel is as follows:

$$K\left(\frac{x_i - x}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x_i - x}{h})^2}$$

Figure 1: Marginal effect of fertilizer use on wheat yield



$nh \to \infty$ as $n \to 0$. This rate is slower than the convergence rate of parametric estimator, in other words, $h \to \infty$ in parametric case which means $O(\frac{1}{nh}) = O(0)$ even if n is finite. In sum, we require lot more data to estimate a parameter consistently using non-parametric methods. In total we have around 6000 observations, so our non-parametric estimator should be consistent. But in the region-III of the figure-1 there aren't many observations, therefore, we do not draw any inference from this region purely because there are too few observation to consistently estimate the parameters. We have good enough number of observations for part-I and II. We draw three main inferences from our results on marginal effects of fertilizer on wheat yield.

First, is that the marginal effect is not constant. It varies with fertilizer use level, if we want to approximate it by polynomial function then it should be a 7-degree polynomial. It is because the marginal effect has 5 peaks (before region-III), which means that $m(x)$ should have 6 peaks and therefore 7 roots and hence 7-degree polynomial. We do not insist on number 7 or 8, but point is that it is a higher-order polynomial certainly not close to quadratic or linear.

Second, is that the effect of fertilizer on wheat yield is positive and significantly different from zero under certain threshold of the fertilizer use level. However, the marginal effect shows steady decrease as fertilizer level increases.

Third is that, it becomes negative after a point (in region-II). We believe that the marginal effect remains negative for any level of fertilizer use after this point. Unfortunately, we are constrained by data to prove it. This is interesting because literature to best of our knowledge does not record negative impact of fertilizer use on yield. Standard concave production says that the marginal effect remains always positive.

3.3.4 Discussion

The main problem that we cannot draw inferences from region-III is that we have lot of parameters to estimate ($\beta(x)$ for every x value) but there are not enough observations. However, if we can somehow reduce number of parameters to be estimated, then we can learn more about region-III. One method to reduce the number of parameters is to use parametric form e.g. $m(x) = x\beta$ or linear regression model. Here we just need to estimate a single parameter, hence can do it efficiently.

We divide the data into four parts based on level of fertilizer use. For the last quantile which is equivalent to the region-III of the figure-1, we find that β of linear regression model is significant and negative. This evidence supports our line of conclusion. Alternatively, it strengthens our claim that fertilizer use beyond a point may result into negative marginal effects on wheat yield or start making land barren.

3.4 Alternate Theoretical Modelling

In this section, we present an alternate theoretical model which can mimic the results obtained from the empirical estimations discussed above. We take help from *renewable resource* literature. We model nutrients in the soil to be a renewable resource. Fertilizer can be seen as a medium to extract these nutrients in form of crop-yield. Higher nutrient presence means higher crop yield. For the simplicity, we assume that we can extract the nutrient only through fertilizer use.

Fertilizer has a complex biological relationship with soil and nutrients. We are mainly interested in the estimation of optimal level of fertilizer for sustainable agriculture. Fertilizer use under a threshold is expected to raise the nutrient level in the soil however overuse may adversely affect the same. We can consider nutrient levels in the soil as renewable and natural resource that can be harvested through the fertilizer input. A logistic concave function of the form $Yield = F(fertilizer)$ can be a good mathematical formulation representing nutrient extraction in the form of crop yield through fertilizer use. The standard literature of renewable and natural resources such as static fishery use this class of models. Broadly, We study two possibilities: first being the optimal level of fertilizer use given that the nutrient level is positive and second case is the possibility of nutrient extinction.

3.4.1 Model

Consider nutrient level of soil as a renewable and natural resource. We employ a Weitzman (1974)-style standard model of renewable resources. In this model, we assume that crop-yield in future is valued as same in the present i.e. no discounting and focus on the long-run equilibrium stock of nutrient. Let X be the stock of some nutrient level in the soil, N be a measure of fertilizer use per unit land i.e. effort of extracting the nutrients, H be the total harvest of nutrients, with:

$$H = qXN \tag{3}$$

The constant q can be interpreted as a technology coefficient which incorporates other factors which makes a difference in nutrient harvest and assumed to be a constant. The nutrient stock grows at the rate given by:

$$\frac{dX}{dt} = \dot{X} = aX(1 - \frac{X}{k}) - H(X, N) \tag{4}$$

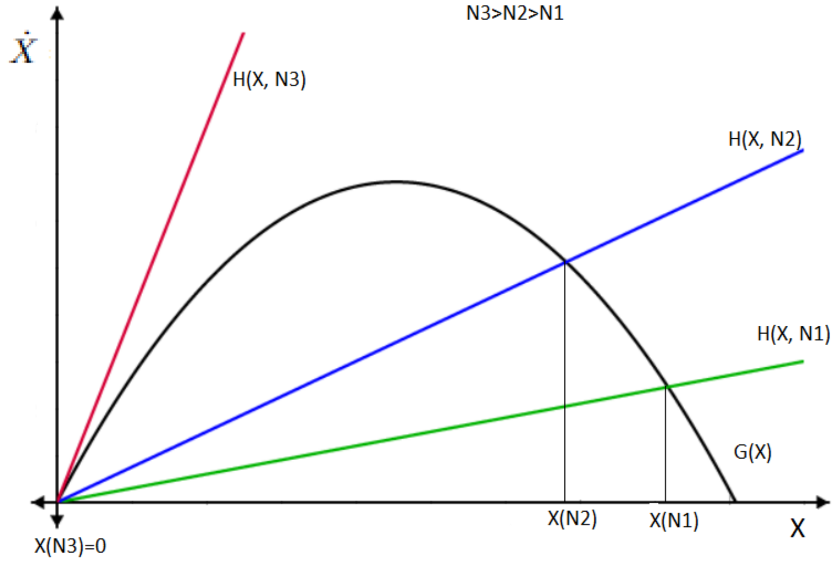
Let's denote $G(X) = aX(1 - \frac{X}{k})$, this first term $(G(X))$ can be interpreted as *recharge function* of nutrients which depends upon level/stock of nutrients, and the second term is the harvest function i.e. how much of nutrient is taken out.

3.4.2 Long Run Nutrient Stock

Rate of change of level of nutrient stock can be written as the difference of recharge function and harvest function $\dot{X} = G(X) - H(X, N)$. Equilibrium level of nutrient stocks can be given by $\dot{X} = 0$. In figure-2, we present a graphical illustration of long-run equilibrium level of nutrient stock growth for different levels of fertilizer use. $N3 > N2 > N1$ are three different level of fertilizer use, equilibrium level of nutrients can be given by $X(Ni)$ for each level of fertilizer use $i = 1, 2, 3$. We observe that for $N1$ and $N2$, the equilibrium level of nutrients X(N) is positive but for a sufficiently higher level $N3$, the equilibrium level of nutrients X(N) is zero. In words, there exist a threshold value of fertilizer use, below which we can have long-run positive level of nutrient in soil. Also, as soon as we cross the the threshold, the nutrient extinction occurs in the long-run. One way to see this is the rate of extraction (slope of H(X,N) in the figure) is higher than the rate of recharge (slope of G(X)) in absolute value at $N3$. We present these observations into a formal mathematical statement and prove the same.

Preposition-1: Suppose that fertilizer use per unit land N is some constant over all time. Then:

Figure 2: Rate of Change of Nutrient Stock ($\dot{X}$) Dynamics



1. For any $N < \frac{a}{q}$, $\quad X(N) > 0$. The equilibrium value of nutrient stock is positive.
2. For $N \geq \frac{a}{q}$, nutrient stock extincts i.e. X(N)=0.

Proof: We have $\dot{X} = G(X) - H$, here $G(X) = aX(1 - \frac{X}{k})$ which is a logistic function representing the growth nutrients, also known as recharge function in natural resource literature. The harvest function $H = qXN$ is linear in X. Given that N is constant over all time, let $N = \bar{N}$ for simplicity.

Let $X(\bar{N})$ represents the equilibrium value of stock for a given $\bar{N}$. The equilibrium value of the stock $X(\bar{N})$ can be given by the expression where the rate of change of X becomes zero ($\dot{X} = 0$) i.e. the stock does not change over time. This gives us the following equation:

$$G(X(\bar{N})) = H(\bar{N}) \tag{5}$$

This implies:

$$aX(\bar{N})(1 - \frac{X(\bar{N})}{k}) = qX(\bar{N})\bar{N}$$

If nutrient stock is not zero i.e. $X(\bar{N}) \neq 0$, this gives:

$$a(1 - \frac{X(\bar{N})}{k}) = q\bar{N}$$

Which implies:

$$X(\bar{N}) = k(1 - \frac{q\bar{N}}{a}) \tag{6}$$

Now if $\bar{N} < \frac{a}{q}$, then $\frac{q\bar{N}}{a} < 1$, this implies that $X(\bar{N}) > 0$. This result is consistent with our one assumption where we use $X(\bar{N}) > 0$ to simplify equations. Hence, the first statement is proved.

Now suppose if $\bar{N} \geq \frac{a}{q}$, then $\frac{q\bar{N}}{a} \geq 1$, which implies that $X(\bar{N}) \leq 0$. $X(\bar{N}) < 0$ does not make any sense because having negative nutrient stock does not hold any meaning. Therefore, $X(\bar{N}) \leq 0$ is equivalent to $X(\bar{N}) = 0$. But during the derivation of this expression we have assumed that $X(\bar{N}) \neq 0$, hence contradiction arises. Therefore, this expression is not valid for $\bar{N} \geq \frac{a}{q}$. The only possible value of $X(\bar{N})$ for $\bar{N} \geq \frac{a}{q}$ is $X(\bar{N}) = 0$

3.4.3 The Optimal Level of Fertilizer Use

Yield as a function of fertilizer use level or yield-fertilizer production function will be equal to harvest:

$$Y(N) = H(X(N), N) \tag{7}$$

Where N is given level of fertilizer use, X(N) is corresponding equilibrium level of nutrient stock. From equation-3 and (6), we can write Y(N) as:

$$Y(N) = qkN(1 - \frac{qN}{a}) \tag{8}$$

Marginal products of fertilizer is defined as the ratio of unit change in yield to unit change in fertilizer use ($\frac{dY}{dN}$). It basically tells us what we get out of what we are putting. In a situation where marginal product is zero, it means that we are getting no increase in yield even if we put an additional unit of fertilizer. Therefore, it is profitable for an economic agent to keep on putting fertilizer until the marginal product becomes zero. Let's calculate the marginal products:

$$\frac{dY(N)}{dN} = qk - \frac{2q^2Nk}{a} \tag{9}$$

If we calculate marginal products at $N = 0$, that is when we put-in first unit of fertilizer, it is equal to $\frac{dY(N)}{dN} = qk$. However more interesting level of fertilizer use is when marginal products becomes zero i.e. it is no longer profitable to put additional fertilizer. This can be obtained by setting $\frac{dY(N)}{dN} = 0$ in equation-9. Because $qk \neq 0$ (by definition), this gives:

$$N = \frac{a}{2q} \tag{10}$$

Hence fertilizer use level $N = \frac{a}{2q}$ is the optimal level of fertilizer use at which the yield can be maximized in the long-run.

4 Fertilizer Use and Human Health

This section investigates how fertilizer use correlates with human health indicators. The main difference in this section is that our analysis is done at state level unlike district level analysis in the previous section. This is because we are limited by data on health outcome which are available at state level for the years 1991-2011. Therefore we use an annual time series sample consists of 19 states for the time period 1991 to 2011, thereby a total of 380 observations. We use a linear regression with fixed effects model to see if the correlation survives after controlling for potential confounding variables.

4.1 Descriptive Analysis

In this subsection we present summary statistics of key variable of interest such as fertilizer used, mortality rate and others. Using bin-scatter (Stepner, 2013), we can combine large number of scatter points into small number of bins in order to present a meaningful time trend.

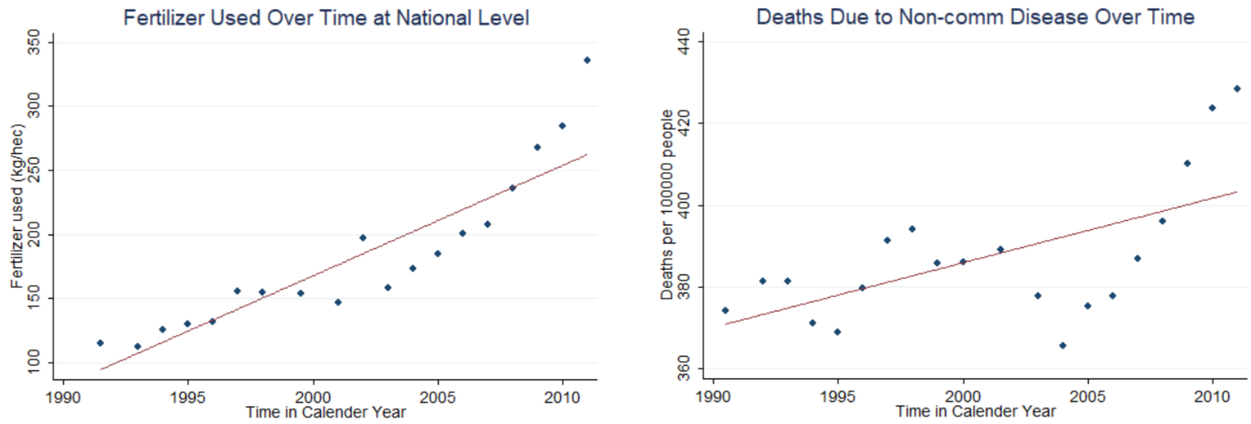
Figure-3 represents trend in fertilizer and mortality rate over time at national level. The vertical axis represents fertilizer (NPK) used in kg per hectare and mortality rate in deaths due to non-communicable diseases per 100,000 people. The horizontal axis represents calendar year. *Bin-scatter* method is used to obtain the following graph.

Figure-3 tells us that both fertilizer use per unit land and mortality rate shows overall increasing trend over time. However there exists a cyclical pattern as well. It shows that the two variables follow a similar time-trend which can be traced back into the cyclical behavior as well. These two similar observations in terms of increasing over time and cyclical behaviour gives us the first hand impression that the two variables are closely related.

4.2 Cluster Analysis

Cluster analysis refers to the problem of partitioning a given data set into homogeneous and separate subsets known as clusters. It looks to find homogeneous subgroups among observations.

Figure 3: Fertilizer Intensity and Mortality Rate at National Level (Source: ICMR, ICRISAT)



This analysis will unfold the underlying patterns in the data for us. We will learn if there exists peculiar subgroups behaving differently, if yes, then it becomes interesting to investigate why do they so.

Our task is to select the variables which can give us maximum information pertaining to the underlying patterns of the data set. Ideally, we select all the variables to get the maximum information but unfortunately, beyond a point adding more variables adds little to our understanding of the data. Further, adding unnecessary variables creates noise hence blur our inferences. Therefore, we look for a subset of variables which have a distribution incorporating large amount of information in the data. In this paper, we mainly study fertilizer use and its consequences on health indicators. So, we choose *fertilizer intensity* (fertilizer use per unit land) and *mortality rate* to perform cluster analysis. These two variables are continuous and has a considerable amount of variation therefore, representing large chunk of information.

How do we quantify the dissimilarity or similarity between two groups? A popular method of representing dissimilarities between groups is to draw a tree like structure called as *dendrogram* which uses euclidean norm or distance. The height in the dendrogram shows dissimilarities among the groups, higher is the height of fusion more dissimilar the groups are. We use *Ward's minimum variance* method to quantify the dissimilarities between the observations.

Table 3: Summary Statistics of Clusters

Group	Fertilizer	Mortality Rate	Yield Value	No. of Obs
Cluster 1	72.17	373.48	7926.94	156
Cluster 2	181.05	385.84	14377.92	142
Cluster 3	304.69	445.77	27149.21	50
Cluster 4	586.63	494.05	13292.59	9
Cluster 5	1060.77	510.78	22341.02	7

After clustering the data based on dissimilarity measures along fertilizer use and mortality dimensions, we report a summary table. In addition to the two aforementioned variables, we also report average yield value[6], a proxy for farm-income level. We find five natural clusters. Here, cluster-1 represents poorer states with least amount of fertilizer use, and least amount of farm-income. Cluster-2 is a group of middle income observations using moderate level of fertilizer and moderate number of mortality rate. Cluster-3 is associated with higher levels of mortality rate and fertilizer intensity, however farm-income is also higher. Cluster 4 and 5 are extreme ones where mortality rate as well as fertilizer used is very high. These interesting observations excite us

[6] Average yield value is the average amount of money value generated from cultivating one unit of area. We sum the price times quantity first and then divide it by total area cultivated for a data-unit i, i.e. $\left(\sum \frac{p_i q_i}{a_i}\right)$

to formally test the relationship between fertilizer and health indicators using rigorous statistical techniques. Therefore, we move towards next section.

4.3 Regression Analysis

There is a large literature in agriculture economics which says that farmers grow staple grain for self-consumption even if it is monetarily less productive (Rivera-Padilla, 2020). Therefore it is not incorrect to say farmers produce majority of their food consumption demand. Further, due to transport cost, population in a state is more likely to eat the food produced in the same state. Therefore, the adverse impacts of excessive use of fertilizers should reflect into the human health indicators of the state. Technically, our model assumes sufficiently positive transport cost between two states such that food-grains are impossible to transport to other state.

4.3.1 Model

We employee a linear regression model with fixed effects to identify the impacts of fertilizer on human health through measures such as deaths due to non communicable disease. For a state i and year j, we estimate the following model:

$$y_{ij} = \beta_0 + \beta_1 f_{ij} + \Gamma X_{ij} + StateFE_i + YearFE_j + \epsilon_{ij} \tag{11}$$

Where y_{ij} is the measure of human health (mortality rate), f_{ij} is the amount of NPK-fertilizer used per unit land in kg per hectare. X_{ij} is the vector of control variables which could potentially be correlated with fertilizer used and influencing human health. $StateFE_i$ and $YearFE_j$ are the state fixed effects and time-trend variables capturing the variations in state and time specific unobservable variables which could confound our identification. ϵ_{ij} is the error term which follows standard normal distribution and is orthogonal to the variables used in the model.

The coefficient β_1 measures the impact of interest i.e. how an additional unit of fertilizer used impacts human health. The employment of sophisticated fixed effect model along with control variables ensures that β_1 measures impact of interest.

4.3.2 Results and Inferences

Table 4: Impact of Fertilizer on Human Health (on state level)

Variable	Model-1	Model-2	Model-3
Fertilizer (kg/ha)	0.1744***	0.1186***	0.0963***
Annual Rainfall (mm)		0.0034***	0.0008
Road-length (km)		0.0034 ***	0.0008
Total Yield Value (Rs/hec)		0.0007***	0.0005**
Time Fixed-Effect			Yes
District Fixed-Effect			Yes
β_0 (constant)	362.74***	308.85***	374.59*
R^2	0.36	0.62	0.93
Num. Obs.	364	275	275
F-Statistics	154.71	106.25	147.05
***:Significant at 1% level	**: 5% level	*: 10% level	

The results show that there is a strong positive association between the fertilizer used and number of people dying due to non-communicable diseases. Quantitatively, an additional unit of fertilizer used is associated with increase of 0.096 deaths per 100,000 people due to non-communicable diseases. We note that the significance of the result is very strong. The p-value is less than 0.001, we use robust standard errors to account for potential clustering. Whether this impact can be said as causal is worth discussing, but is certainly not a naive result. The value of R^2 in model-3 is 0.93, saying that the model explains 93% of the variation in human health indicator. The use of alternate definitions of human health and different specifications of model doesn't change the qualitative results. This suggest that it is fair to say that greater use of fertilizer per unit land deteriorates human health.

Our primary indicator of human health is mortality measured as deaths per 100,000 people due to non-communicable disease. But we check whether the results are sensitive to the use of this measure. We perform the same exercise by using alternate measures such as YLLs and DALYs per 100,000 people. We find that the result remains in the same direction with strong significance, further use and drop of control variables does not play any role in the sign and significance of β_1. The use of different human health indicators and alternate model specification gives us a confidence in saying that the higher degree of per unit land fertilizer use affects human health negatively.

5 Policy Implications

Our results has strong implications for government policy on fertilizer. Indian government subsidize the fertilizer use and the aggregate amount is a massive sum of money, therefore making it an important public policy topic. We discuss the policy suggestions in detail in the following text.

5.1 Soil Health

For economics, fertilizer is always been a very good productivity enhancing technology. However, soil scientists have an alternate view as well. Overuse of anything may disturb the balance of nutrients in soil. We call a soil healthy if it can continue to provide the same yield of output over time without deteriorating the balance of nutrient level. Nutrient level is difficult to observe, however the output or yield is easily observed. If yield from the same land decreases with constant use of fertilizer use over time, then it can be interpreted as an evidence of deteriorating the health of soil. Our results have strong policy implication especially on fertilizer subsidies. The fertilizer production function estimated in the section-3 provides us evidences that overuse of fertilizer may deteriorate the soil fertility over time. Therefore, this calls for a upper limit quota on fertilizer use per unit land.

5.2 Human Health

There are plenty of research in biochemistry literature on how food chains affects the human health. Food-chains are affected by the inputs used in order to produce grains e.g. seeds, fertilizer, pesticides, and water-quality. To the best of our knowledge, there is no research talking about negative health consequences of fertilizer use.

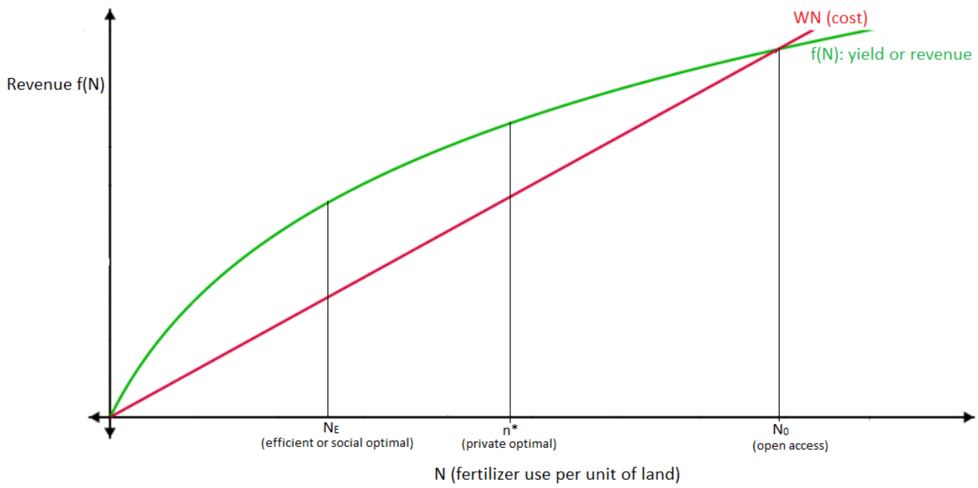
In the section-4, we have shown the unequivocal evidences of strong negative impacts of fertilizer use on human health indicators. One could argue that deaths due to non-communicable diseases may be due to lifestyle issues. However, we control for the time-fixed effect, which should take care of these variation in lifestyle over time. We also include state fixed effects in order to control for geographically regional unobservable factors. We use two variables yield value per unit land and road-length to proxy for the income or development level of the region. We admit that, there could be many other variables that could potentially confound this identification. Nonetheless, we are of the view that these results are not naive, these strong correlations are compelling evidence of negative impacts of fertilizers. Very high R^2-statistics supports this claim.

Interesting point to note is that for soil health fertilizer is helpful upto a certain threshold and then it becomes a bad choice but for human health, fertilizer appears to have negative impact from the beginning. Therefore from human health point of view, chemical (NPK) fertilizer use should be avoided to begin with. A detailed research with district-level dataset could unravel more important dynamics of fertilizer's impact on human health.

5.3 A model of fertilizer use from bio-sustainability point of view

One side-effect of use of chemical fertilizer and pesticides is that they pollute the environment through air, water, and food chains of animal and human. Environment is a common property shared by all the living beings on the earth including humans. The toxic impacts of fertilizers and pesticides are of vital importance from the biodiversity point of view. In this section, we model environment as common property under three different modes of management: private ownership, social optimality, and open-access. We consider chemical fertilizer and pesticides as toxic inputs which causes harm to environment but produces output according to $H = f(N)$, where N is the toxic input per unit land. Let's assume that the stock of nutrients doesn't extinct. Now for simplicity let's assume that yield is a function of toxic input use i.e. $H = f(N)$. For simplicity, we assume that $f(0) = 0$ i.e. no crop-yield is possible without this toxic input, and $f'(N) > 0$ i.e. $f()$ is increasing in N, and $f''(N) < 0$ i.e.$f()$ is a concave function. We present a figure below, which demonstrate the yield and cost of using toxic inputs, it plots N (input per unit land) On horizontal-line and profits on the vertical line.

Figure 4: Optimal Level of Fertilizer Use under Different Ownership



5.3.1 Efficient Level Under Private Optimal:

Private optimal is defined as the solution to the profit maximization problem of an individual producer who takes the aggregate level of N as given i.e. he on his individual capacity cannot influence the aggregate level of toxic inputs N. He chooses his fertilizer level i.e. n_i in order to maximize his own profits. Let $\pi(n_i, N)$ represents the profit, it is a function of aggregate input use and individual input use. Individual's profit depends upon aggregate level of input use because the negative impacts of toxic inputs used by others also impact the yield of an individual. Assuming all agents are identical, a private agent solves:

$$\max_{n_i} \qquad \pi(n_i, N) = \frac{n_i}{N} f(N) - wn_i \tag{12}$$

w is the unit cost of toxic input (N) use. If we define $A(N) = \frac{f(N)}{N}$ which is a function of aggregate input use, then we can rewrite the optimization problem of a private producer:

$$\max_{n_i} \qquad \pi(n_i, N) = n_i(A(N) - w) \tag{13}$$

Since private producer i will take $(N - n_i)$ as given. Now solving the optimization problem, we can write the first order equation:

$$(A(N) - w) + n_i^* A'(N) = 0 \tag{14}$$

5.3.2 Efficient Level Under Social Optimal:

Now consider the optimization problem of social planner which maximizes the aggregate profits. Let's call socially optimal level of input as N_E. Formally,

$$\max_{N} \qquad \pi(n_i, N) = N(A(N) - w) \tag{15}$$

We can write the first order equation:

$$(A(N) - w) + N_E A'(N) = 0 \tag{16}$$

5.3.3 Efficient Level Under Open-Access:

We can also find an important level of toxic input i.e. the level at which aggregate profits becomes zero. Alternatively, a private agent keep on using toxic inputs until the level it directly hurts him back i.e. using an additional unit of input will decrease his profits. We name it as 'open-access' level (N_O) of input use. This can be given by $\pi(n_i, N) = 0$. This gives:

$$\pi(n_i, N) = X(A(N) - w) = 0 \quad \Rightarrow \quad w = A(N)$$

This implies:

$$w = A(N) = \frac{f(N)}{N} \quad \Rightarrow \quad N_O = \frac{f(N_O)}{w} \tag{17}$$

5.3.4 Optimization Result

Preposition-2: $N_E < n^* < N_O$. Fertilizer use per unit land under private ownership is greater than social optimal level. Further, fertilizer use under open access is greater than private optimal level.

Proof:

Step-1: Claim: A'(N)<0

Consider $A(N) = \frac{f(N)}{N}$. We can write:

$$A'(N) = \frac{Nf'(N) - f(N)}{N^2} \tag{18}$$

By assumption our f(N) is a differentiable and concave function on N. Now a function f(x) is concave on X if, for any $x^* \in X$, the tangent line through $(x^*, f(x^*))$ is above the graph of f(x). That is (proof of this property is in appendix):

$$f(x) \leq f(x^*) + f'(x^*)(x - x^*) \tag{19}$$

Now given that f(0)=0 i.e. without any fertilizer, no nutrient can be extracted, and applying the equation-(19) for concave function, for x=0 and $x^* = N$:

$$f(0) \leq f(N) + f'(N)(0 - N)$$

This implies

$$0 \leq f(N) - Nf'(N)$$

Also, our f(N) is strictly concave, therefore, above expression can be written as:

$$Nf'(N) - f(N) < 0 \tag{20}$$

Because N^2 is always positive, with this result we can say that A'(N) as expressed in the equation-(18) is negative.

Step-2: $n^* < N_O$

Now using equation-(14), we say that $A(N) - w = -n^* A'(N)$ is positive because n^* is positive and A'(N) is negative. Since w is the unit cost of fertilizer use and $A(N) = \frac{f(N)}{N}$ is the unit benefit of the fertilizer use, therefore we can interpret this term as profit (or 'rent' in economics jargon).

Suppose a representative agent is using the farm-land under private optimal and open-access settings. Optimization solution of free-access problem gives us the value of profit equals to zero i.e. producer keep on putting the toxic inputs until the profit becomes zero under open access. That is:

$$A(N_O) - w = 0 \qquad \text{and} \qquad A(n^*) - w > 0$$

This implies:

$$A(n^*) > A(N_O)$$

Since A'(N)<0, the above expression implies:

$$n^* < N_O \tag{21}$$

Step-3: $N_E < n^*$

Following the similar reasoning as in step-2, profit under private optimal is $n^*(A(N) - w)$ and under social optimal is $N(A(N) - w)$. Since $N = \sum n$ and all $n > 0$, then we can definitely say that $N > n$ i.e. aggregate N is greater than individual's n. Therefore profit under social optimal is higher than the private optimal. Therefore, we can write:

$$(A(N_E) - w) > (A(n^*) - w) \Rightarrow A(N_E) > A(n^*)$$

Since A'(N)<0, therefore:

$$N_E < n^* \tag{22}$$

5.4 Summary

The soil health and human health implications clearly calls for restricting fertilizer use on an upper limit. One way to enforce this is to restrict the fertilizer subsidies to certain quota per unit land owned by farmer. In analytical model, open access type of ownership may refer to the situation where fertilizer is fully subsidized i.e. free of cost, only effort it takes is to put it in the farm. Model in the previous section predicts that under such a situation farmers may end-up overusing it. In order to ensure social optimal level of fertilizer use, the policy need to increase the cost or fix quotas.

6 Conclusion

We mainly do two things in this paper: First is to estimate long-run empirical production function of fertilizer use with wheat yield as output. Second, we provide convincing evidences on negative impact of fertilizer overuse on soil health and human health. We then discuss the policy implications of the results.

Our main contribution is to empirically estimate the long-run production function of fertilizer use with wheat yield as output. We first start by the standard production function used in the literature and test is using simple parametric regression tools. The standard production function passes the simple tests. However we then employ non-parametric functional form specification test. We then reject the standard production with strong significance. We later estimate our own empirical production function using non-parametric local linear least square methods. Our estimates suggest that if we want to approximate the fertilizer production function using polynomial function then it should be of approximately seven-degree. We, then provide evidences of negative marginal returns on fertilizer use if we cross a certain threshold level. We also discuss a logistic-style theoretical model whose analytical results can mimic the empirical results obtained in our estimation.

Our another contribution is to provide evidences of strong negative health consequences of fertilizer use. We see that our simple linear regression with time and state fixed effects shows that there is strongly significant negative impacts on human health outcomes of fertilizer use. We argue that these results are not naive and should be taken as starting point towards a deeper investigation of how chemical (NPK) fertilizers affect human health through the food-chains. On a similar line, we define soil health to be the ability to produce consistent output with same level of fertilizer use. We find that normal use of fertilizer doesn't harm soil health but overuse does. We then discuss a theoretical framework by modelling food-chain and environment to be a common property resource. Our analytical results show that open access end up using lot more fertilizer per unit land than the socially efficient level. Privately optimal level lies in between. This calls for government intervention on monitoring fertilizer use, the government should design policy in order to discourage fertilizer use per unit land beyond a certain upper-limit.

Our future line of work extends into two directions: First, we want to get deeper understanding on how fertilizer use if affecting human health. For this we need to get the data on district level. We have rich data on the horizontal dimension on diseases (type of diseases etc) but not on granular level (limited to state-level). Second direction of the work is to compute total monetary cost and benefits of the fertilizer use. If we can put an estimate of monetary cost in terms of human health consequences, it can enhance our understanding of the economics of the fertilizer subsidy program.

We end this paper by saying that in the case of a diverse country like India, we found a non-trivial relationship between fertilizer use and health indicators. Fertilizer use below a safe limit is good but beyond that it makes thing worse.

References

[1] Atafar Zahra, Mesdaghinia Alireza , Nouri Jafar, Homaee Mehdi, Yunesian Masoud, Ahmadi-moghaddam Mehdi, and Mahvi Amir Hossein. (2010). "Effect of fertilizer application on soil heavy metal concentration", *Environment Monitoring and Assessment* 160, 83 (2010). weblink

[2] Bogardi Istvan , Kuzelka Robert D., Ennenga Wilma G. (1991). "Nitrate Contamination Exposure, Consequence, and Control", *Springer Berlin, Heidelberg*, Volume 30, *ISBN : 978-3-642-76042-6.*

[3] Chen, J. H. (2006). "The combined use of chemical and organic fertilizers and/or biofertilizer for crop growth and soil fertility", *International workshop on sustained management of the soil-rhizosphere system for efficient crop production and fertilizer use* (Vol. 16, No. 20, pp. 1-11). Land Development Department Bangkok Thailand.

[4] Crespi M., Ramazzotti V. (1991). "Evidence That N-Nitroso Compounds Contribute to the Causation of Certain Human Cancers", *Nitrate Contamination* pp 233-252, weblink.

[5] Deb Papiya. (2018). "Chapter 14 - Environmental Pollution and the Burden of Food-Borne Diseases, Foodborne Diseases", *Handbook of Food Bioengineering*, 2018, Pages 473-500. weblink.

[6] Dorsche M.M., Scragg R.K.R., McMichael A.J., Baghurst P.A. and Dyer, K.F. (1984). "Congenital Malformations and maternal drinking water supplying rural South Australia: A case study", *American Journal of Epidemiology*, 119, 473-486.

[7] ECETOC (European Chemical Industry Ecology and Toxicology Centre). (1988). "Nitrate and drinking water", *Technical Report No. 27, Brussels.*

[8] Grant C. A. and Sheppard S. C. (2008). "Fertilizer impacts on Cadmium availability in agricultural soils and crops", *Human and Ecological Risk Assessment: An International Journal*, 14:2, 210-228.

[9] Hayami, Y., and Ruttan, V. W. (1970). "Agricultural Productivity Differences among Countries", *The American Economic Review*, Vol. 60, No. 5 (Dec., 1970), pp. 895-91

[10] Jiao Y, Grant C A and Bailey L D. (2004). "Effects of phosphorus and zinc fertilizers on cadmium uptake and distribution in flax and durum wheat". *Journal of Science Food Agriculture*, 84: 777–85.

[11] Ju X.T., Kou C.L., Christie P., Dou Z.X., and ,Zhanga F.S. (2007). "Changes in the soil environment from excessive application of fertilizers and manures to two contrasting intensive cropping systems on the North China Plain". *Environmental Pollution*, Volume 145, Issue 2, January 2007, Pages 497-506.

[12] Li Qi, and Wang Suojin. (1998). "A simple consistent bootstrap test for a parametric regression function", *Journal of Econometrics*, 87 (1998) 145—165.

[13] Majumdar Deepanjan, and Gupta Navindu. (2000). "Nitrate pollution of groundwater and associated human health disorders", *Indian Journal of Environmental Health* (2000), 42 (1): 28-39.

[14] McArthur John W., McCord Gordon C. (2017). "Fertilizing growth: Agricultural inputs and their effects in economic development", Journal of Development Economics 127 (2017) 133–152.

[15] Pagan A., and Ullah A. (1999). "Nonparametric econometrics". *Cambridge: Cambridge university press.*

[16] Robertsa Terry L. (2014). "Cadmium and Phosphorous Fertilizers: The Issues and the Science", *Procedia Engineering*, 83(2014), 52–59.

[17] Rivera-Padilla Alberto. (2020). "Crop choice, trade costs, and agricultural productivity", *Journal of Development Economics*, Volume 146, 2020, 102517.

[18] Stepner M. (2013). "Binscatter: Stata module to generate binned scatterplots", *Statistical Software Components* S457709, Department of Economics, Boston College.

[19] Swanson, Timothy M. (1994). "The Economics of Extinction Revisited and Revised: A Generalised Framework for the Analysis of the Problems of Endangered Species and Biodiversity Losses", *Oxford Economic Papers*, New Series, Vol. 46, Special Issue on Environmental Economics (Oct., 1994), pp. 800-821

[20] Weitzman, M.L. (1974). "Free access vs. private ownership as alternative systems for managing common property", *Journal of Economic Theory*, 8(2), pp.225-34.