ECONOMETRICS
FINAL PROJECT

Impact of Agricultural production on Health Outcomes over a period of time

Saanvi Sehwag
Shelja Agarwal
Shivam Yadav
Vishnu Vardhan
Vedant Bothra

Introduction

With 2.4 million babies dying in 2020, the first month of infancy is the most vulnerable phase for child survival.

Children who die during the first 28 days of life suffer from illnesses and diseases linked to a lack of quality treatment during or shortly after delivery, as well as throughout the early days of life.

Introduction

Most neonatal fatalities are caused by preterm birth, intrapartum problems (birth asphyxia or inability to breathe at birth), infections, and birth abnormalities.

Almost all deaths from asphyxia (97.8%) occur within the first week of life, with 70% happening within the first 24 hours (day 0)

Introduction

Most of the cases of Asphyxia are seen in countries with low and middle-income countries with lesser per capita GDP as compared to other developed or developing nations.

It may be because of the fact that these low-income countries don't have proper health conditions due to which mother has some illnesses.

Motivation

Parent illnesses like stress and hypertension cause the infant born to have issues like less birth weight, which is one of the major causes of asphyxia.

Birth weights between 2.5 kg to 4 kg are considered to be normal and below 2.5 kg is very low and possess a major risk of asphyxia.

Motivation

Many acute and chronic health issues, including as cardiovascular and respiratory disease, arthritis, skin cancer, hearing loss, and amputations, are more common in farmers.

Other health effects, such as stress and unfavourable reproductive outcomes, have received minor attention in the agricultural sector.

Motivation

There are various other factors like yield index, the population of a district, Kharif and Rabi crops that might affect the variable v41 in that district.

Different types of crops are grown excessively in different regions of India. For example, wheat is grown excessively in north India which is a Kharif crop, while in the South, the rabi crop rice is grown more.

Approach

This project aims to study the impact of agricultural production on health outcomes over a period of time.

We will analyze the relationship between the states and union territories in India for a period of some consecutive years.

We will investigate for the causal linkage between agricultural growth and health outcomes.

Data Summary

- Removed Outliers.
- Number of States and districts - 29 + Islands, 736 districts
- Time Period - 2011 - 2019
- Seasons - 2 Rabi and Kharif
- Total Data Points - 736*2*9

Variables and their Descriptions

Variable v41 is the dependent variable and the independent variables are defined further.

Variable	Description	Acronym	Type
V41	Percentage of infant deaths due to Asphyxia (to total reported infant deaths)	v41	Dependent
V28	Percentage of newborns having a weight less than 2.5 kg (to newborns weighed at birth)	v28	Dependent

Variable	Description	Acronym	Type
V16	Percentage of safe deliveries (to total reported deliveries) gdp It is the gdp of that state on that particular year	v16	Dependent
GDP	It is the gdp of that state in that particular year.	gdp	Independent
BEDS	It is the state wise number of beds as of 2020	beds	Independent
TAPS	It is district wise tap water access	taps	Independent
INDEX	It is the yield index of districts of different states	index	Independent

Variable Description (Dependent)

Variable v41 = Percentage of infant deaths due to Asphyxia (to total reported infant deaths)

Variable v28 = Percentage of newborns having a weight less than 2.5 kg (to newborns weighed at birth)

Variable v16 = Percentage of safe deliveries (to total reported deliveries)

Variable Description (Independent)

Variable GDP = It is the GDP of that state on that particular year

Variable beds = It is the statewide number of beds as of 2020.

Variable Tap = It is district-wise tap water access.

Variable Index = It is the yield index of districts of different states.

After considering all the specifications we have chosen said models:

$$H_{d,t} = \beta_0 + \beta_1(gdp_{d,t}) + \beta_2(beds_{d,t}) + \beta_3(tap_{d,t}) + \beta_4(yieldindex_{c,d,t}) + \beta_5(v28_{d,t}) + \mu_{d,t}$$

Season : kharif

d : district, t : year

$$H_{d,t} = \beta_0 + \beta_1(gdp_{d,t}) + \beta_2(beds_{d,t}) + \beta_3(tap_{d,t}) + \beta_4(yieldindex_{c,d,t}) + \beta_5(v28_{d,t}) + \mu_{d,t}$$

Season : Rabi

d : district, t : year

$$H_{d,t} =$$

$$\beta_0 + \beta_1(gdppercap_{d,t}) + \beta_2(beds_{d,t}) + \beta_3(tap_{d,t}) + \beta_4(yieldindex_{c,d,t}) + \beta_5(v28_{d,t}) + \mu_{d,t}$$

d : district, t : year

$$H_{d,t} = v41$$

NULL Hypothesis:

H_0:

- $\beta_{1,kharif} < 0, \beta_{2,kharif} < 0, \beta_{3,kharif} < 0, \beta_{4,kharif} < 0, \beta_{5,kharif} > 0$
- $\beta_{1,rabi} < 0, \beta_{2,rabi} < 0, \beta_{3,rabi} < 0, \beta_{4,rabi} < 0, \beta_{5,rabi} > 0$

Here, we have taken what we think about how the dependent variable v41 would depend on the independent variables from our perception or point of view before running the actual model.

We will then compare this with the results obtained from the model defined.

Alternative Hypothesis:

H_a:

- $\beta_{1,kharif} \geq 0, \beta_{2,kharif} \geq 0, \beta_{3,kharif} \geq 0, \beta_{4,kharif} \geq 0, \beta_{5,kharif} \leq 0$
- $\beta_{1,rabi} \geq 0, \beta_{2,rabi} \geq 0, \beta_{3,rabi} \geq 0, \beta_{4,rabi} \geq 0, \beta_{5,rabi} \leq 0$

Hypothesis Explanation:

Variable v41 vs GDP

v41 is the percentage of infant deaths due to Asphyxia. Higher GDP of a district equates to greater human progress, because it means more valuable goods and services have been created. .A higher GDP of a state or country implies that v41 there should be low . So we hypothesize a greater v41 if the GDP of the district is low, v41 and district should negatively correlate.

Variable v41 vs Beds

Here beds independent variable denotes the district-wise number of beds available in a particular year. A higher number of beds in a state should imply that v41 there should be low as it has better health facilities which are indicated by more number of beds. We may say v41 and Beds are negatively correlated and we hypothesize when Beds are higher in number, then v41 should be less.

Variable v41 vs Taps

Variable 'taps' represent the district-wise tap water access. Districts with a higher value of variable 'taps' should basically show a declining nature in the curve of variable v41.

It is because access to tap water is a very basic amenity and should be available to everyone.

If these basic facilities are not available to everyone, then it is a health hazard and definitely variable v41 would increase. We may say v41 and tap are negatively correlated.

Variable v41 vs Index

Variable 'index' basically represents the yield index of districts of different states. The yield index is the total agricultural yield per land area of the district. An increase in yield index over a period of time should reflect a declining movement in variable v41 as if yield index increases then there must be a good yield of crops and enough food supply is available, as a result, the people of the district must be getting adequate quantities of food. So their health condition should be stable or improved. Therefore, low risk of getting asphyxia or other diseases.We may say v41 and yield index are negatively correlated and we hypothesize when yield index is higher in number, then v41 should be less.

Variable v41 vs v28

v41 is the percentage of infant deaths due to asphyxia while v28 is the percentage of newborns with low body weight.

An increase in v28 would simply possess a greater risk on the newborns to be affected with asphyxia and then unfortunate deaths. So, a higher number of newborns with low body weight implies that v41 would increase.

Reference:
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6053756/)

Regression Results:

Table 1:

$$H_{d,t} = \beta_0 + \beta_1(gdp_{d,t}) + \beta_2(beds_{d,t}) + \beta_3(tap_{d,t}) + \beta_4(yieldindex_{c,d,t}) + \beta_5(v28_{d,t}) + \mu_{d,t}$$

Season : kharif

d : district, t : year

Table 2:

$$H_{d,t} = \beta_0 + \beta_1(gdp_{d,t}) + \beta_2(beds_{d,t}) + \beta_3(tap_{d,t}) + \beta_4(yieldindex_{c,d,t}) + \beta_5(v28_{d,t}) + \mu_{d,t}$$

Season : Rabi

d : district, t : year

```
Call:
lm(formula = v41 ~ index + beds + gdp + tap + v28, data = main_kharif_df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.116  -8.275  -2.720   4.623  92.909

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.240e+00  2.107e-01  39.111  < 2e-16 ***
index       -3.928e-02  1.162e-02  -3.382 0.000722 ***
beds         1.021e-06  1.744e-06   0.586 0.558045
gdp          4.702e-08  4.130e-09  11.385  < 2e-16 ***
tap         -1.306e-02  3.890e-03  -3.356 0.000792 ***
v28          4.823e-02  9.720e-03   4.962 7.03e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.6 on 16845 degrees of freedom
  (11978 observations deleted due to missingness)
Multiple R-squared:  0.02788,   Adjusted R-squared:  0.02759
F-statistic: 96.62 on 5 and 16845 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = v41 ~ index + gdp + beds + tap + v28, data = main_rabi_df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.975  -8.043  -2.563   4.618  92.227

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.621e+00  2.186e-01  34.865  < 2e-16 ***
index        2.068e-02  1.114e-02   1.857  0.06339 .
gdp          5.661e-08  4.447e-09  12.731  < 2e-16 ***
beds        -1.998e-06  1.791e-06  -1.116  0.26460
tap         -1.232e-02  4.059e-03  -3.035  0.00241 **
v28          6.927e-02  1.060e-02   6.536 6.54e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.48 on 15171 degrees of freedom
  (10561 observations deleted due to missingness)
Multiple R-squared:  0.03255,   Adjusted R-squared:  0.03223
F-statistic: 102.1 on 5 and 15171 DF,  p-value: < 2.2e-16
```

Results:

For Rabi Season:

$\beta_{1,rabi}$ and $\beta_{4,rabi}$ are found out to be positive, opposite to what was expected.
Also $\beta_{1,rabi}$ is significant. So we reject our null Hypothesis.

$\beta_{2,rabi}$, $\beta_{3,rabi}$, , $\beta_{5,rabi}$ all have similar signs to the Null Hypothesis.

For Kharif Season:

$\beta_{1,kharif}$ and $\beta_{2,kharif}$ are found out to be positive, opposite to what was expected.
Also $\beta_{1,kharif}$ is significant.Though $\beta_{2,kharif}$ is positive, it was found to be insignificant. So we reject our null Hypothesis.

$\beta_{3,kharif}$, $\beta_{4,kharif}$, , $\beta_{5,kharif}$ all have similar signs to the Null Hypothesis.

Conclusion:

In both Kharif and in Rabi the GDP has a positive correlation with the variable v41. So higher the GDP in a district the higher the v41 variable is.This may be due to population, the states with the highest population, top the charts in GDP ranking.

It can be due to the large population size as well. large population somewhat correlates to higher gdp, state like Uttar Pradesh is at the third position when we compare GDP of all UT's and states of India, but it is placed at the very bottom by Niti Aayog health index.

So a better way to measure this would be GDP per capita . So, we modeled with the same variables, replacing gdp with gdppercapita. AAs expected gdppercapita negatively correlates with v41.

gdp per capita

```
Call:
lm(formula = v41 ~ index + beds + gdp_cap + tap + v28, data = main)

Residuals:
    Min      1Q  Median      3Q     Max
-16.761  -8.278  -2.538   4.794  91.751

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.117e+00  1.356e-01  59.859  < 2e-16 ***
index        1.845e-02  3.685e-03   5.006 5.58e-07 ***
beds         3.044e-06  1.040e-06   2.926  0.00344 **
gdp_cap     -4.295e-08  2.591e-09 -16.575  < 2e-16 ***
tap         -1.119e-02  2.472e-03  -4.524 6.08e-06 ***
v28          5.765e-02  6.337e-03   9.097  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Variable v41 vs Taps

In both Kharif and in Rabi the tap variable has a negative correlation with the Variable v41. That is, if access to tap water increases in a district, then v41 decreases there.

This is because the availability of tap water affects the health of individuals in the district so there is a lower risk of health-related problems. Hence, v41 which is infant deaths due to asphyxia can be prevented by safe tap water access.

Variable v41 vs v28

In both Kharif and Rabi, variable v28 has a positive correlation with the Variable v41. That is, while v28 increases, v41 also increases.
This is because it is very natural when higher number of newborns have weight fewer than 2.5 kgs, are at a greater risk of being diagnosed with asphyxia. So there will be a higher number of deaths of newborns with asphyxia.

Variable v41 vs yield index

During Kharif season, v41 and yield index show a negative correlation as hypothesized. That is, if the yield index is high in a district, then the variable v41 is low. While during Rabi, when the index decreases, variable v41 increases.

On calculating the average of yield index's for both the season for a particular year, We get to know that they are very close, so what could be the reason that during Kharif season yield negatively correlates, but during rabi season it positively correlates.One of the reasons could be stubble burning.

Variable v41 vs index

During winter season mostly north India is blanketed under smog which is mostly contributed by farm stubble burning, causing pollution, low visibility, breathlessness to major three states & UT -Delhi, Punjab and Haryana.[1] This matches the duration of Rabi season, Rabi cropping season is from October-March (winter).

.The Punjab Preservation of Subsoil Water Act (2009) made it mandatory for farmers to transplant paddy late during the Kharif season to prevent loss of water[1]. So the smog produced due to stubble burning, continues to cover the atmosphere which leads to various respiratory diseases. One of them is Asphyxia.

References:

1) Ihttps://media.neliti.com/media/publications/344386-stubble-burning-a-rationale-for-asphyxia-c6e552b4.pdf
2)https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6053756/
3)https://en.wikipedia.org/wiki/List_of_Indian_states_and_union_territories_by_GDP
4)https://www.hindustantimes.com/india-news/niti-aayog-health-index-2021-see-ranking-of-states-and-uts-based-on-performance-101640601217921.html
5)https://www.weforum.org/agenda/2015/04/how-agriculture-can-improve-health-and-nutrition/

Team Members

Econometrics Project

Saanvi Sehwag - 2020462
Shelja Agarwal - 2020470
Shivam Yadav - 2020471
Vishnu Vardhan - 2020480
Vedant Bothra - 2020260