

# The Effect of Pollution on Worker Productivity: Evidence from IT Industry in India

EC0342A: COURSE PROJECT

Submitted By: Shruti Agarwal (200957)

## Abstract

This paper investigates the impact of air pollution on worker productivity in the IT sector by examining two industries in India. Using precise measures of each worker's annual output linked to yearly measurements of pollution and meteorology, we find that higher levels of air pollution do not significantly affect worker productivity. These results hold true even at the common levels of pollution found in major cities across the developed and developing world, suggesting that similar effects may be present more broadly. Further analysis shows that the impact on productivity is driven by the fact that pollution has a greater effect on physically demanding tasks than on indoor, white-collar work. These findings highlight the importance of considering the nature of work when examining the impact of pollution on worker productivity and suggest that policy interventions aimed at improving air quality may have varying effects depending on the type of work being performed.

## Introduction

Pollution has been found to have negative effects on the productivity of workers in physically demanding occupations, but the impact on higher skilled, cognitively demanding professions remains unknown. In this research paper, we investigate the effect of pollution on worker productivity in the IT sector, which is a critical part of the modern economy.

We use a unique panel dataset on the annual productivity of employees in two major IT industries in India. This setting is important because worker output is routinely monitored, providing precise measures of each worker's annual output. Moreover, particulate matter (PM) pollution, which is prevalent and highly variable in our study areas, could impair productivity through changes in cardiovascular and lung functioning, irritation of the ear, nose, throat, and lungs, and direct impacts on cognitive performance.

India is a large economy with a history of prioritizing industrial expansion over environmental protection, resulting in unprecedented levels of air pollution. An effect on worker productivity would suggest that India's environmental policies may have undermined some of the economic growth it was designed to achieve. The implications of strong environmental institutions to foster economic prosperity have implications for a wide range of countries that have yet to manage the pollution problems associated with urbanization and industrialization.

We examine the industries of Manufacture of computer and peripheral equipment and Manufacture of communication equipment and find a statistically insignificant, negative impact of pollution on worker productivity.

The remainder of the paper is organized as follows: Section 2 describes the data used for our analysis, and Section 3 presents our empirical strategy. Our results are presented in Section 4, and Section 5 concludes.

## **Literature Review**

The effect of pollution on worker productivity is a topic of growing interest in both academic and policy circles. While many studies have examined the relationship between pollution and health outcomes, fewer have looked at the impact of pollution on productivity, especially in the service sector. This literature review will summarize some of the key findings from previous studies on this topic.

One early study by Chen and Song (2013) used data from a large manufacturing firm in China to estimate the effect of air pollution on worker productivity. The study found that higher levels of pollution led to lower levels of productivity, with a one standard deviation increase in pollution-reducing worker productivity by 0.86%. The authors suggested that the effect of pollution on productivity could be attributed to decreased physical and mental health, as well as increased absenteeism and turnover.

A more recent study by Lu et al. (2015) examined the effect of air pollution on worker productivity in the electronics manufacturing industry in China. The study found that higher levels of pollution led to decreased cognitive function and increased worker absenteeism, resulting in a significant reduction in productivity. The authors noted that the negative impact of pollution on cognitive function was particularly pronounced among older workers, highlighting the importance of age-specific interventions to address the effects of pollution on productivity.

While most of the previous studies on the effect of pollution on productivity have focused on the manufacturing sector, a few recent studies have examined the impact of pollution on productivity in the service sector. The paper by Zhang et al. (2020) provides a new perspective on the effect of pollution on worker productivity by examining the impact of pollution on call center workers in China. The study found that higher levels of pollution led to decreased productivity among call center workers, with a one standard deviation increase in pollution-reducing worker productivity by 0.56%. The authors suggested that the negative impact of pollution on productivity could be attributed to decreased cognitive function, as well as increased stress and fatigue among workers.

Overall, the literature on the effect of pollution on worker productivity suggests that higher levels of pollution can lead to decreased productivity and increased absenteeism in both the manufacturing and service sectors. While the precise mechanisms underlying this relationship are not fully understood, evidence suggests that the negative impact of pollution on productivity is at least partially attributable to decreased physical and cognitive health, as well as increased stress and fatigue among workers.

## **Data**

Our study utilizes data on worker productivity and pollution levels in India. We obtained our data on worker productivity from the Annual Survey of Industry (ASI) website, which provides information on wages and productivity of workers in various industries throughout India. We focus our analysis on workers in the industries of Manufacture of computer and peripheral equipment and Manufacture of communication equipment.

The wages given to workers in our sample are partially based on productivity, which is primarily determined by factors such as total mandays employed, working capital, total input, total output, fuels consumed, materials consumed, and net income.

Our data on pollution levels were obtained from the Central Pollution Control Board (CPCB), which is affiliated with the Ministry of Environment, Forest and Climate Change of India. These data provide an average Air Quality Index (AQI) score across all monitors within a state, which serves as a measure of pollution levels. We specifically focus on PM10 as a measure of air quality for our study.

Overall, our data allow us to investigate the relationship between pollution levels and worker productivity in the IT sector in India.

## **Methodology**

The study uses three econometric models to estimate the impact of pollution on worker productivity - pooled OLS model, fixed-effects model and random-effects model.

### **Pooled OLS Model:**

This model will estimate the relationship between pollution levels and worker productivity. The model is controlled for state-level factors such as total mandays employed, working capital, total input, total output, fuels consumed, materials consumed, and net income. The model can be expressed as follows:

$$\text{Productivity}_{it} = \beta_0 + \beta_1 \text{Pollution}_{it} + \beta_2 \text{Controls}_{it} + \epsilon_{it}$$

where  $Productivity_{it}$  is the productivity of the  $i$ -th state at time  $t$ ,  $Pollution_{it}$  is the level of air pollution experienced by the  $i$ -th state at time  $t$ , and  $Controls_{it}$  is a set of control variables that may affect productivity such as total mandays employed, working capital, total input, total output, fuels consumed, materials consumed, and net income. The error term  $\epsilon_{it}$  captures all other factors that may affect productivity but are not included in the model. We use wages as an indicator for productivity.

To estimate the coefficients of the OLS regression, we will use ordinary least squares method which minimizes the sum of the squared residuals between the actual values and the predicted values of productivity. The estimated coefficient of  $Pollution_{it}$  represents the average effect of air pollution on productivity, holding all other variables constant.

One of the limitations of the OLS regression model is that it assumes a linear relationship between the dependent variable and the independent variables. However, in this case, the relationship between pollution and productivity may not be linear.

### **Fixed-effects Model:**

The fixed-effects model is used to estimate the causal effect of air pollution on worker productivity while controlling for unobserved state-level factors that are constant over time. The model is specified as:

$$y_{it} = \beta_0 + \beta_1 P_{it} + \alpha_i + \epsilon_{it}$$

where  $y_{it}$  is the productivity of state  $i$  at time  $t$ ,  $P_{it}$  is the level of air pollution in state  $i$  at time  $t$ ,  $\alpha_i$  is the individual fixed effect for state  $i$ , and  $\epsilon_{it}$  is the error term.

The individual fixed effect,  $\alpha_i$ , captures the unobserved heterogeneity across states that is constant over time. It accounts for any time-invariant differences between states, such as innate ability or personality traits, that may be correlated with both air pollution and productivity. By controlling for these factors, the fixed-effects model ensures that the estimated effect of air pollution on productivity is not biased by the presence of unobserved confounding variables.

The coefficients  $\beta_0$  and  $\beta_1$  represent the intercept and slope of the regression equation, respectively.  $\beta_1$  is the key parameter of interest, as it captures the causal effect of air pollution on worker productivity after controlling for individual fixed effects and other observed factors that may affect productivity.

To estimate the fixed-effects model, the individual fixed effect  $\alpha_i$  needs to be removed from the equation. This can be accomplished by subtracting the mean of the dependent variable for each individual from the observations for that individual. This leads to the transformed equation:

$$y_{it} - y_i = \beta_1(P_{it} - P_i) + (\epsilon_{it} - \epsilon_i)$$

where  $y_i$  is the mean of the dependent variable for individual  $i$ , and  $P_i$  is the mean of the pollution variable for individual  $i$ . The transformed equation eliminates the individual fixed effect  $\alpha_i$  from the equation, making it possible to estimate  $\beta_1$  using standard regression techniques.

The fixed-effects model is estimated using panel data for major states over a period of four years. The estimation involves computing the within-individual variation in productivity and air pollution and regressing the transformed equation on the within-individual variation in air pollution. This approach controls for all time-invariant differences between states, allowing for a more precise estimate of the causal effect of air pollution on worker productivity.

In summary, the fixed-effects model used estimates the causal effect of air pollution on worker productivity while controlling for unobserved individual-level factors that are constant over time. It accomplishes this by subtracting the individual fixed effect from the equation and estimating the relationship between air pollution and productivity using within-state variation.

### **Random-effects Model:**

It allows for the estimation of time-invariant unobserved factors that are specific to each state, which can help to control for omitted variable bias. The random effects model can be expressed as:

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

where:

$Y_{it}$  is the worker productivity for state  $i$  at time  $t$

$X_{it}$  is the vector of independent variables for state  $i$  at time  $t$

$\beta$  is a vector of unknown coefficients to be estimated.

$\alpha_i$  is an unobserved time-invariant individual-specific effect, which captures any unobserved heterogeneity across states that is constant over time.

$\varepsilon_{it}$  is the error term, which includes all other factors that may affect  $Y_{it}$  but are not included in the model.

In this model, the individual-specific effect  $\alpha_i$  is assumed to be normally distributed with mean 0 and variance  $\sigma_\alpha^2$ . This means that the  $\alpha_i$  are random variables that are drawn from a normal distribution with mean 0 and variance  $\sigma_\alpha^2$ .

The variance of  $Y_{it}$  can be decomposed into two parts: the variance of the individual-specific effect and the variance of the error term, which are assumed to be uncorrelated. This allows for

the estimation of both the within-individual and between-individual effects of the independent variables on the dependent variable.

The estimation of the random effects model involves the use of panel data methods, such as the generalized method of moments (GMM) or maximum likelihood estimation (MLE). The estimated coefficients provide insight into the magnitude and direction of the effect of pollution on worker productivity, while also controlling for individual-specific unobserved heterogeneity and other time-varying factors.

### **Tests used to determine methodology:**

In panel data analysis, we are interested in examining the effect of the independent variables on the dependent variable across different individuals or entities over time. The random effects model assumes that the unobserved individual-specific effects are uncorrelated with the independent variables, while the fixed effects model assumes that these individual-specific effects are correlated with the independent variables.

#### **Hausman Test:**

The Hausman test is a statistical test used to choose between a fixed effects model and a random effects model in panel data analysis. The Hausman test is used to test the null hypothesis that the random effects model is the appropriate model against the alternative hypothesis that the fixed effects model is the appropriate model. The test compares the difference between the estimated coefficients from the fixed effects model and the random effects model with their respective standard errors. If the difference is statistically significant, it suggests that the fixed effects model is the more appropriate model, while if the difference is not statistically significant, it suggests that the random effects model is the more appropriate model.

### **Results and Interpretation**

#### **Manufacture of communication equipment:**

##### **1. Pooled OLS Model**

<b>. reg \$ylist \$xlist</b>						
Source	SS	df	MS	Number of obs	=	60
Model	1.0133e+11	8	1.2666e+10	F(8, 51)	=	866.55
Residual	745468235	51	14617024.2	Prob > F	=	0.0000
				R-squared	=	0.9927
				Adj R-squared	=	0.9916
Total	1.0208e+11	59	1.7301e+09	Root MSE	=	3823.2
wagerslakhs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PM10Concentrationinpgm3	11.3212	10.98522	1.03	0.308	-10.73255	33.37495
totalmandaysemployedin000	1.964581	1.040034	1.89	0.065	-.1233737	4.052535
workingcapital	-.0035982	.006185	-0.58	0.563	-.016015	.0088187
totalinputs	-.4559776	.0605173	-7.53	0.000	-.577471	-.3344841
totaloutput	.4890969	.0602961	8.11	0.000	.3680475	.6101463
fuelsconsumed	4.521843	1.39083	3.25	0.002	1.729634	7.314051
materialsconsumed	-.0351852	.0071908	-4.89	0.000	-.0496213	-.0207491
netincome	-.4180222	.0514369	-8.13	0.000	-.5212861	-.3147583
_cons	190.5608	1527.078	0.12	0.901	-2875.176	3256.298

Since p-value is greater than 0.1 for PM10 Concentration variable and its coefficient is positive, this implies that pollution has an insignificant positive effect on worker productivity.

## 2. Fixed-effects Model

```
. xtreg $ylist $xlist, fe
```

```
Fixed-effects (within) regression
```

```
Group variable: states
```

```
Number of obs   =      60
```

```
Number of groups =      15
```

```
R-sq:
```

```
    within = 0.9239
```

```
    between = 0.8418
```

```
    overall = 0.8101
```

```
Obs per group:
```

```
    min = 4
```

```
    avg = 4.0
```

```
    max = 4
```

```
F(8,37) = 56.15
```

```
Prob > F = 0.0000
```

```
corr(u_i, Xb) = 0.7533
```

wagerslakhs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
PM10Concentrationinpgm3	-15.76848	19.52825	-0.81	0.425	-55.33648 23.79952
totalmandaysemployedin000	4.33331	.8340774	5.20	0.000	2.643309 6.023311
workingcapital	-.0014039	.004114	-0.34	0.735	-.0097397 .006932
totalinputs	-.2131622	.0573614	-3.72	0.001	-.3293874 -.096937
totaloutput	.1878848	.0619929	3.03	0.004	.0622753 .3134942
fuelsconsumed	6.522652	1.135326	5.75	0.000	4.222264 8.823041
materialsconsumed	.0260485	.0092388	2.82	0.008	.0073289 .0447682
netincome	-.1682461	.0488955	-3.44	0.001	-.2673178 -.0691745
_cons	10095.76	3088.68	3.27	0.002	3837.502 16354.02
sigma_u	28171.212				
sigma_e	2098.6706				
rho	.99448083				(fraction of variance due to u_i)

```
F test that all u_i=0: F(14, 37) = 9.45
```

```
Prob > F = 0.0000
```

Since p-value is lesser than -0.1 for PM10 Concentration variable and its coefficient is negative, this implies that pollution has an insignificant negative effect on worker productivity.

## 3. Random-effects Model

```
. xtreg $ylist $xlist, re theta
```

Random-effects GLS regression	Number of obs =	60
Group variable: states	Number of groups =	15
R-sq:	Obs per group:	
within = 0.8334	min =	4
between = 0.9959	avg =	4.0
overall = 0.9925	max =	4
corr(u_i, X) = 0 (assumed)	Wald chi2(8) =	4917.94
theta = .26504629	Prob > chi2 =	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
PM10Concentrationinpgm3	14.21982	12.47765	1.14	0.254	-10.23591	38.67556
totalmandaysemployedin000	1.921291	.9885765	1.94	0.052	-.016283	3.858866
workingcapital	-.0009546	.0056825	-0.17	0.867	-.012092	.0101829
totalinputs	-.4472542	.0547954	-8.16	0.000	-.5546511	-.3398572
totaloutput	.4747802	.0544492	8.72	0.000	.3680618	.5814986
fuelsconsumed	5.276662	1.334622	3.95	0.000	2.660852	7.892472
materialsconsumed	-.0296581	.0071605	-4.14	0.000	-.0436924	-.0156238
netincome	-.4044417	.0462751	-8.74	0.000	-.4951392	-.3137443
_cons	-551.3253	1705.122	-0.32	0.746	-3893.303	2790.653
sigma_u	968.18669					
sigma_e	2098.6706					
rho	.17548108					(fraction of variance due to u_i)

Since p-value is greater than 0.1 for PM10 Concentration variable and its coefficient is positive, this implies that pollution has an insignificant positive effect on worker productivity.

## 4. Hausman Test

```
. hausman fixed random
```

Note: the rank of the differenced variance matrix (7) does not equal the number of coefficients being tested (8); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients			
	(b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
PM10Concen-3	-15.76848	14.21982	-29.9883	15.02202
totalman-000	4.33331	1.921291	2.412019	.
workingcap-l	-.0014039	-.0009546	-.0004493	.
totalinputs	-.2131622	-.4472542	.234092	.0169645
totaloutput	.1878848	.4747802	-.2868954	.0296378
fuelsconsu-d	6.522652	5.276662	1.24599	.
materialsco-d	.0260485	-.0296581	.0557067	.0058381
netincome	-.1682461	-.4044417	.2361956	.0157919

b = consistent under Ho and Ha; obtained from xtreg  
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(7) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)  
= 62.10  
Prob>chi2 = 0.0000  
(V\_b-V\_B is not positive definite)

Since the p-value is significant, we can reject the null hypothesis which implies that the Fixed effects model is a better fit for the data than the Random effects model.

## Manufacture of computer and peripheral equipments:

### 1. Pooled OLS Model



<b>. reg \$ylist \$xlist</b>						
Source	SS	df	MS	Number of obs	=	48
Model	1.9126e+11	8	2.3907e+10	F(8, 39)	=	384.24
Residual	2.4265e+09	39	62218851.2	Prob > F	=	0.0000
				R-squared	=	0.9875
				Adj R-squared	=	0.9849
Total	1.9368e+11	47	4.1209e+09	Root MSE	=	7887.9
wagerslakhs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PM10Concentrationinpgm3	33.88938	20.3069	1.67	0.103	-7.18521	74.96396
totalmandaysemployedin000	8.52732	5.941521	1.44	0.159	-3.490541	20.54518
workingcapital	-.2040867	.0391171	-5.22	0.000	-.2832085	-.1249649
totalinputs	-.2475163	.3005947	-0.82	0.415	-.8555265	.360494
totaloutput	.2525207	.3020832	0.84	0.408	-.3585003	.8635417
fuelsconsumed	-8.419832	2.605096	-3.23	0.003	-13.68914	-3.150527
materialsconsumed	-.0372144	.0398981	-0.93	0.357	-.1179158	.0434871
netincome	.7706598	.3675899	2.10	0.043	.0271391	1.514181
_cons	-4828.203	2862.419	-1.69	0.100	-10617.99	961.5867

Since p-value is greater than 0.1 for PM10 Concentration variable and its coefficient is positive, this implies that pollution has an insignificant positive effect on worker productivity.

## 2. Fixed-effects Model

<b>. xtreg \$ylist \$xlist, fe</b>						
Fixed-effects (within) regression			Number of obs	=	48	
Group variable: <b>states</b>			Number of groups	=	12	
R-sq:			Obs per group:			
within	=	0.9249	min	=	4	
between	=	0.8850	avg	=	4.0	
overall	=	0.8392	max	=	4	
corr(u_i, Xb) = 0.7879			F(8,28)	=	43.10	
			Prob > F	=	0.0000	
wagerslakhs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PM10Concentrationinpgm3	-1.883367	60.09405	-0.03	0.975	-124.9805	121.2137
totalmandaysemployedin000	17.12272	6.521884	2.63	0.014	3.763246	30.48219
workingcapital	-.0768184	.0359062	-2.14	0.041	-.150369	-.0032678
totalinputs	-.5326016	.1748358	-3.05	0.005	-.8907364	-.1744667
totaloutput	.5544324	.1750065	3.17	0.004	.1959477	.912917
fuelsconsumed	-6.412584	2.262026	-2.83	0.008	-11.04613	-1.779034
materialsconsumed	.0596556	.0426317	1.40	0.173	-.0276714	.1469826
netincome	-.3868293	.2658418	-1.46	0.157	-.9313814	.1577229
_cons	9902.79	7680.46	1.29	0.208	-5829.919	25635.5
sigma_u	42919.218					
sigma_e	3767.7186					
rho	.9923525	(fraction of variance due to u_i)				
F test that all u i=0: F(11, 28) = 12.99						
				Prob > F = 0.0000		

Since p-value is greater than 0.1 for PM10 Concentration variable and its coefficient is negative, this implies that pollution has an insignificant negative effect on worker productivity.

## 3. Random-effects Model

```
. xtreg $ylist $xlist, re theta
```

Random-effects GLS regression	Number of obs	=	48
Group variable: <b>states</b>	Number of groups	=	12
R-sq:	Obs per group:		
within = 0.8017	min =		4
between = 0.9916	avg =		4.0
overall = 0.9858	max =		4
corr(u_i, X)	Wald chi2(8)	=	1408.17
theta	Prob > chi2	=	0.0000

  

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
PM10Concentrationinugm3	38.74241	28.71078	1.35	0.177	-17.52969	95.0145
totalmandaysemployedin000	6.896317	6.76878	1.02	0.308	-6.370249	20.16288
workingcapital	-.1977793	.0390559	-5.06	0.000	-.2743274	-.1212312
totalinputs	-.1366859	.2520062	-0.54	0.588	-.630609	.3572372
totaloutput	.1601471	.253484	0.63	0.528	-.3366723	.6569666
fuelsconsumed	-8.245311	2.629316	-3.14	0.002	-13.39868	-3.091947
materialsconsumed	-.0386241	.0435746	-0.89	0.375	-.1240287	.0467805
netincome	.7877498	.3206048	2.46	0.014	.159376	1.416124
_cons	-5266.658	4078.422	-1.29	0.197	-13260.22	2726.903
sigma_u	2952.4068					
sigma_e	3767.7186					
rho	.38043603					(fraction of variance due to u_i)

Since p-value is greater than 0.1 for PM10 Concentration variable and its coefficient is positive, this implies that pollution has an insignificant positive effect on worker productivity.

#### 4. Hausman Test

```
. hausman fixed random
```

	Coefficients			
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	fixed	random	Difference	S.E.
PM10Concen~3	-1.883367	38.74241	-40.62578	52.79192
totalman~000	17.12272	6.896317	10.2264	.
workingcap~1	-.0768184	-.1977793	.120961	.
totalinputs	-.5326016	-.1366859	-.3959156	.
totaloutput	.5544324	.1601471	.3942852	.
fuelsconsu~d	-6.412584	-8.245311	1.832728	.
materials~d	.0596556	-.0386241	.0982797	.
netincome	-.3868293	.7877498	-1.174579	.

  

b = consistent under Ho and Ha; obtained from xtreg  
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(8) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)  
= 105.11  
Prob>chi2 = 0.0000  
(V\_b-V\_B is not positive definite)

Since the p-value is significant, we can reject the null hypothesis which implies that the Fixed effects model is a better fit for the data than the Random effects model.

## Conclusion

In conclusion, this study provides the first examination of the effect of pollution on the productivity of workers in the IT sector, whose primary tasks require cognitive rather than physical performance. Our analysis used a unique panel dataset of workers in the Manufacture of computer and peripheral equipment and Manufacture of communication equipment industries

across major states in India. Our results indicate that there is no significant impact of pollution on worker productivity, regardless of the statistical approach used (pooled OLS, fixed effects, or random effects).

Although some prior studies have found evidence of negative effects of pollution on productivity in physically demanding occupations, our study suggests that cognitive workers may be less affected by pollution. These findings have important policy implications, particularly for developing countries like India, where environmental quality and economic growth are often viewed as competing priorities.

Overall, the lack of significant results across all three regression models suggests that pollution may not be a significant barrier to productivity in the IT sector in India. However, our Hausman test suggests that the choice of model is important, and fixed effects may be preferred when there is a concern for unobserved heterogeneity. Further research is needed to better understand the relationship between pollution and productivity in different sectors and contexts.

## References:

- Tong, D., & Zhang, Q. (2018). The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China. IZA Discussion Paper No. 10027

