The Long-term Impact on Academic Performance of Organochlorine Pollution in Soil and Groundwater

Nguyen Vuong

23rd July 2021

Does organochlorine (OC) pesticide pollution in soil and groundwater affects educational outcome?

Introduction - Motivation

- OC pesticides are persistent organic pollutants. Soil and metabolic half-life is long. DDT is an example
  - Air half-life is 2 days
  - Soil half-life is 2-15 years
  - Half-life in human is 7 years
- OC pesticides are toxic to human health, and to cognitive function.
- OC pesticides were popular in agriculture and malaria prevention.
- But, OC-pesticide pollution was not addressed in the literature.

Introduction - Context

- In 70s and 80s, a number of countries banned DDT and other OC pesticide. It started with Hungary in 1968.
- Stockholm convention on banning POPs (including OC pesticide) is effective from 2004.
- India, China, and North Korea are still producing DDT. India still uses DDT in agriculture.
- Vietnam banned OC pesticides (DDT, Aldrin, Dieldrin, etc.) in 1997. But, DDT was not on the list of permitted pesticides from 1991.
- Until June 2013, there are 1652 OC-polluted hot spots, which were mostly former agricultural warehouses. Many of them now becomes schools or residential areas.

Introduction - Scope

Reframed research question

Do OC-polluted hot spots have any impact on academic performance? Can we observe it with cross-sectional data?

To answer this question, I am going to

- Estimate the impact of OC-polluted hot spots on test scores.
- Use 2017 test scores from 10th-grade entrance exam in Thanh Hoa province.
- Construct a recentered instrument.

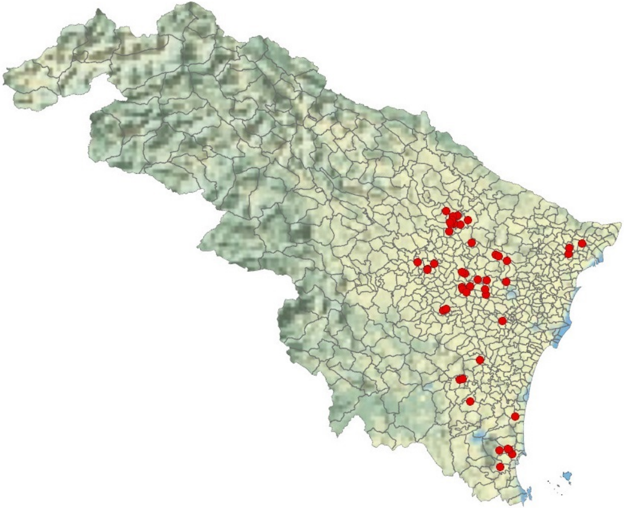
Potential contribution

- Construction of a recentered instrument in the case of distribution system. This method is developed from Borusyak and Hull (2020).
- Contribute to a growing literature on early exposure to pollution.

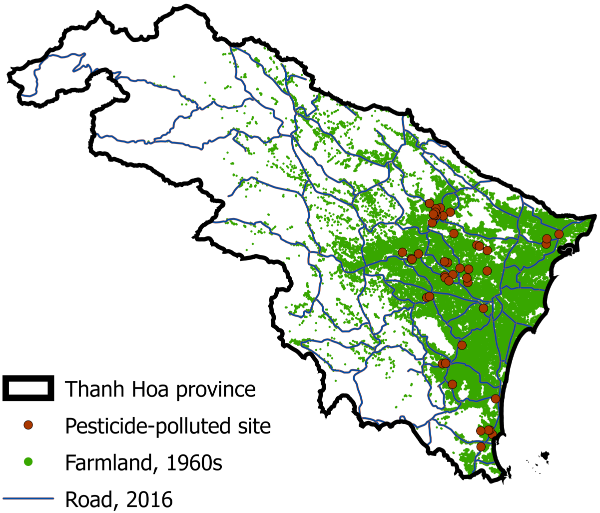
Data

- List of OC-polluted hot spots
  - Decision No. 1946/QD-TTg has 355 hot spots, including 10 in Thanh Hoa.
  - Decision 103/QD-UBND and 1448/QD-UBND issued by Thanh Hoa.
- Test score data from the 10th-grade entrance exam in Thanh Hoa.
  - Double-blind
  - Three subjects: Math, Literature, and English.
  - Administrated by provincial department of education and training.
- Rice farmland: US Army Map Service (1965).
  - Use OpenCV to detect rice growing area
  - Reshape it into diamond grid with 0.0025 degree longitude and latitude diameters.

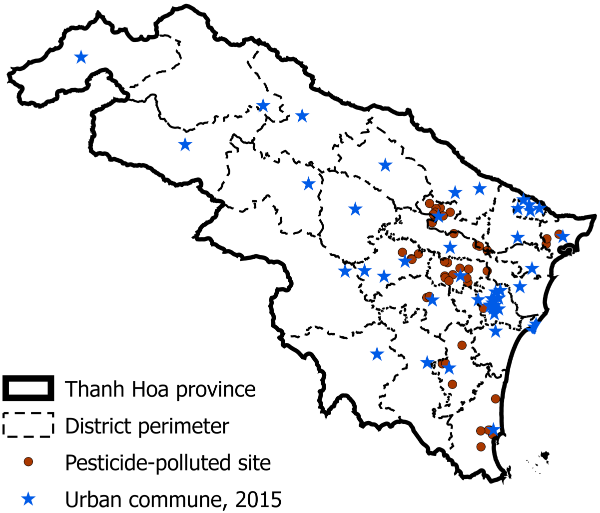
OC-polluted hot spots



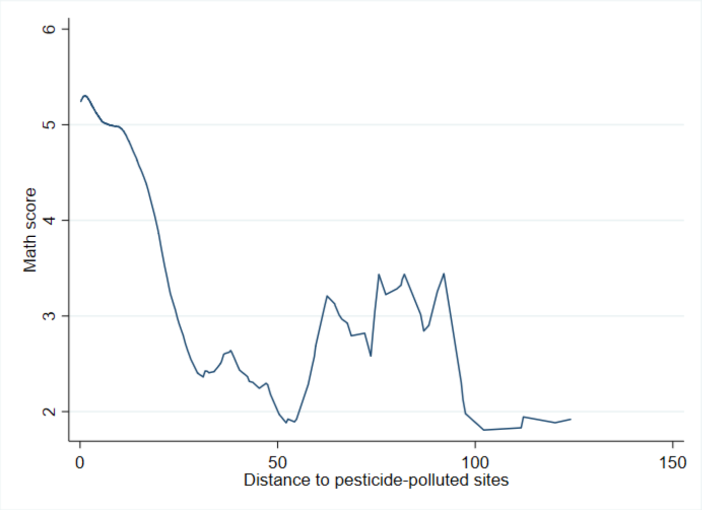
OC-polluted hot spots, rice farmland and road



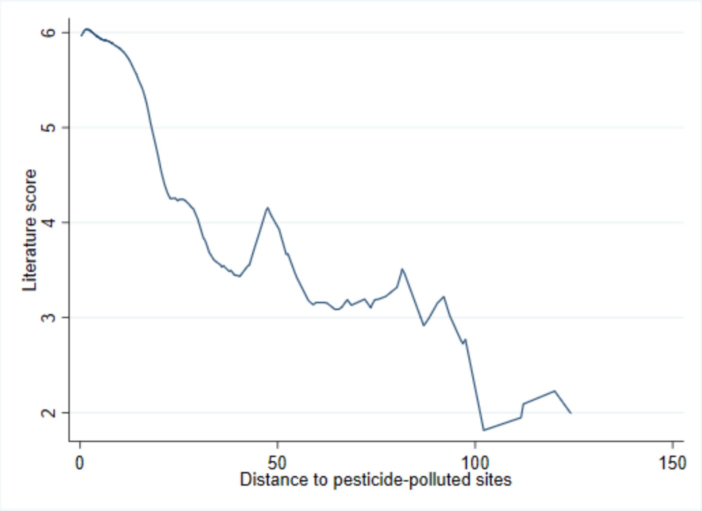
OC-polluted hot spots, towns and urban areas



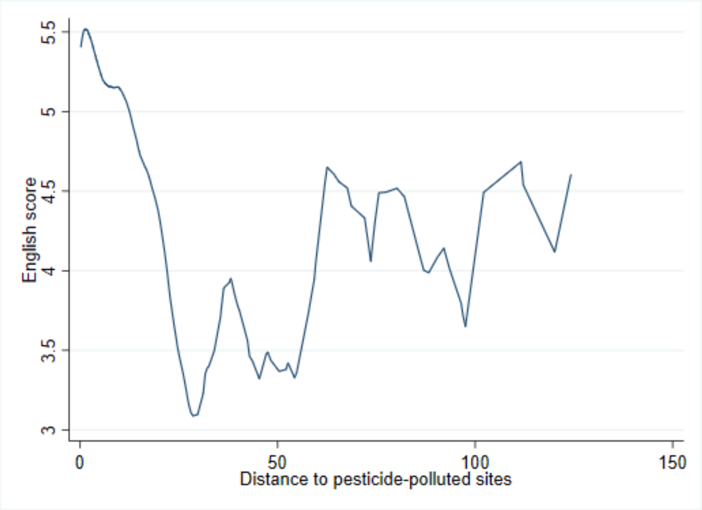
Distance to the nearest hot spot and test scores



Distance to the nearest hot spot and test scores (con't)



Distance to the nearest hot spot and test scores (con't)



Endogeneity issue

- T_i is a dummy for affected communes. The OLS specification is

$$y_i = \beta_0 + \beta_1 . T_i + \varepsilon_i$$

- True effect of T_i is β_1.
- Estimated effect: $\mathbb{E}(\hat{\beta}_1) = \beta_1 + \mathbb{E}(\varepsilon_i | T_i = 1) - \mathbb{E}(\varepsilon_i | T_i = 0)$
- Endogeneity issue arises as $\mathbb{E}(\varepsilon_i | T_i = 1) > 0$ and $\mathbb{E}(\varepsilon_i | T_i = 0) < 0$, or

$$\mathbb{E}(T_i \varepsilon_i) > 0$$

Compare affected and nearby non-affected communes

- The characteristic of neighboring communes should be similar.
- But groundwater and soil pollution does not travel far from the source.
- Limitation: school quality could be different between schools in district capitol and rural communes. We cannot observe the counterfactual.
- The econometric specification is

$$Y_{icd} = \beta_0 + \beta_1 . T_{cd} + \gamma X_{icd} + \eta_d + \varepsilon_{icd}$$

- The sample is restricted to communes that are near the OC-polluted hot spot.

Preliminary result

	10th grade entrance exam in 2017		
	Math score	Literature score	English score
	(1)	(2)	(3)
Panel A: The nearest dumping sites is less than 3 kilometers away			
Affected communes	-0.147**	-0.182***	-0.394***
	(0.065)	(0.043)	(0.067)
Observations	5,991	5,991	5,991
R-squared	0.132	0.204	0.158
Panel B: The nearest dumping sites is less than 5 kilometers away			
Affected communes	-0.104*	-0.123***	-0.204***
	(0.058)	(0.040)	(0.059)
Observations	12,151	12,151	12,151
R-squared	0.135	0.186	0.193

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Recentered instrument

- Given the geographic location of a commune g_i, then $\mathbb{E}(T_i|g_i) = \Pr(T_i = 1)$ and $\varepsilon_i = e(g_i)$.
- By the law of iterated expectations

$$\begin{aligned} \mathbb{E}(T_i\varepsilon_i) &= \mathbb{E}[\mathbb{E}(T_i|\varepsilon_i)\varepsilon_i] \\ &= \mathbb{E}[\mathbb{E}(T_i|g_i)\varepsilon_i] \\ &= \mathbb{E}(\Pr(T_i = 1)\varepsilon_i) \end{aligned}$$

- $\mathbb{E}[(T_i - \Pr(T_i = 1))\varepsilon_i] = 0$, and $z = T_i - \Pr(T_i = 1)$ is a valid instrument.
- Given P_i is the predicted probability and $P_i = \Pr(T_i = 1) + u_i$. Assume that $E(u_i\varepsilon_i) = 0$, then $\hat{z} = T_i - P_i$ is a valid instrument.

Constructing instrument

- OC-polluted hot spots are former warehouses.
- During the command economy, each commune has an agricultural collective and a warehouse. There are 600 communes in Thanh Hoa provinces, but only a relative small number of communes have polluted hot spots.
- These former warehouses could be the nodes in the pesticide distributing system. Two reasons to support this:
  - Distributing through the nodes is cheaper.
  - The amount of pesticide is limited.
- By modeling the distribution system, I could predict the probability of being affected.

Modeling the distributing system

- Denote W as the set of distribution nodes.
- Commune j have a rice production function: $f(l_j, p_j)$. Price of rice is 1.
- Denote c_j as the cost of pesticide plus the transportation costs (factory $\rightarrow$ warehouse $\rightarrow$ commune j).
- Profit from sending pesticide to commune j is

$$\pi_j = f(l_i, p_j^*) - c_j p_j^* - f(l_i, 0)$$

- The optimal amount of pesticide to send to commune j from the nearest node is p_j^*

$$\frac{\partial f}{\partial p_j}(l_i, p_j^*) = c_j$$

Modeling the distributing system (con't)

- Node i would send pesticide to commune j is $p_j^* > 0$ and $\pi_j > 0$. If there are two nodes that could send pesticide to commune j, then they would send from the nearest. Therefore, the coverage set C_i for each node could be constructed.
- There is also a fixed cost c^f for each warehouse. It requires $\sum_{j \in C_i} \pi_j \geq c^f$.
- Without considering the fixed cost, the profit of W would be

$$\Pi_W = \sum_{i \in W} \sum_{j \in C_i} \pi_j . \mathbb{1}\{\pi_j \geq 0\}$$

- Given the fixed cost c^f, the optimal set W would maximize Π_W.

Parameterize the model

- Use the production function from Norwood and Marra (2003) that

$$f_i(l_i, p_i) = A_i l_i [1 - \exp(\gamma_0 - \gamma_1 p_i / l_i)]$$

- The optimal amount of pesticide to commune j and profit are

$$\begin{aligned} p_j^* &= \frac{l_j \gamma_0}{\gamma_1} - \frac{l_j}{\gamma_1} \ln\left(\frac{c_j}{\gamma_1 A_j}\right) \\ \pi_j &= A_j l_j e^{\gamma_0} - \frac{c_j l_j}{\gamma_1} - p_j^* c_j \\ &= l_j \left(A_j e^{\gamma_0} - \frac{c_j}{\gamma_1} - \frac{c_j \gamma_0}{\gamma_1} + \frac{c_j}{\gamma_1} \ln\left(\frac{c_j}{\gamma_1 A_j}\right)\right) \end{aligned}$$

- The condition that $p_j^* \geq 0$ and $\pi_j \geq 0$ is

$$A_j e^{\gamma_0} \gamma_1 \geq c_j$$

Parameterize the model (con't)

- The cost $c_j = c(1 + \theta d_{ij} + \theta' d_i)$.
- The productivity of communes that are near the river is higher. $A_j = A(1 + \delta r_j)$.
- Denote $M = Ae^{\gamma_0}\gamma_1/c$. The profit to supply pesticide from node i to commune j now could be rewritten as

$$\gamma_1 \pi_j = l_j \left[M(1 + \delta r_j) - (1 + \theta d_{ij} + \theta' d_i) \left[1 + \ln \left(\frac{M(1 + \delta r_j)}{1 + \theta d_{ij} + \theta' d_i} \right) \right] \right]$$

- The criteria function is

$$\Pi_W = \sum_{i \in W} \sum_{j \in C_i} l_j \left[M(1 + \delta r_j) - (1 + \theta d_{ij} + \theta' d_i) \left[1 + \ln \left(\frac{M(1 + \delta r_j)}{1 + \theta d_{ij} + \theta' d_i} \right) \right] \right]$$

Solve the problem

- The location of the factory is fixed, find the positions for 30 warehouses.
- Calibrate the values of M, δ, θ, θ' to find a set of locations that is fitted best to the reality.
- Algorithm
  - Start with an arbitrary set of points
  - Fix 29 points, find a point that maximize Π_W given other 29 points.
  - Keep doing that until no improvement is found.
- Given the parameters, I will introduce noise to simulate to find the probability of having a warehouse in a commune.

THANK YOU!

Comments and feedback are welcome!