The Himalayan Tightrope: A High-resolution Study of the Economy-Environmental Trade-off of Infrastructure Development

Findings

1 Recap: Data

Our data is as follows:

Night lights (NTL) Sourced from Li et al. (2020). Spatial resolution: 1km. Temporal range: 1992 – 2023.

Built-up area (GHSL) Sourced from Global Human Settlements Layer, Copernicus (ESA). Spatial resolution: 0.1km. Temporal range: 1975 – 2030 in 5-year gaps.

Land surface temperature (LST) Sourced from MOD11A2, Terra MODIS (NASA). Spatial resolution 1km. Temporal range: 2000 – 2020.

Forest cover (VCF) or Vegetation Continuous Fields. Sourced from DevDataLab cf. MOD44B, Terra MODIS (NASA). Spatial resolution: 0.25km. Temporal range: 2001 – 2020.

PM2.5 air pollution (PM2.5) Sourced from DevDataLab cf. Van Donkelaar et al. (2021). Spatial resolution: 1km. Temporal range: 2000 – 2020.

2 Indexation

We have two methods of indexation: *uniform weighting* (UW) or *principal component analysis* (PCA). Principal component analysis is helpful in the presence of high multi-collinearity. To check for high correlation between variables, i.e., to understand if PCA is required, we undertake two tests: the *Kaiser–Meyer–Olkin (KMO) test* and *Bartlett's test of sphericity*.

Variables	KMO MSA	Bartlett p-value
VCF, PM2.5	0.5, 0.5	0
VCF, PM2.5, LST	0.68, 0.57, 0.56	0

3 Models

$$\text{NTL} = \underbrace{\text{VCF} + \text{PM2.5} + \text{LST}}_{\text{PCA required}} \tag{1}$$

Our first model has the highest T (i.e., 20), and is thus more efficient for estimating our epochs. We can consider showing only this for epoch analysis, considering that GHSL is correlated with NTL anyway. I am obliged to estimate this both with and without the LST term, owing to the justification we have provided for Model 2.

$$\underbrace{\text{NTL} + \text{GHSL}}_{\text{Both PCA and UW}} = \underbrace{\text{VCF} + \text{PM2.5}}_{\text{Both PCA and UW}} \tag{2}$$

Our second model is motivated by LST having more long-run effects than the other environmental indicators. It is limited by having $T = 5$, which necessarily makes the estimation of one epoch an FD instead of FE model.

$$\underbrace{\text{NTL} + \text{GHSL}}_{\text{Both PCA and UW}} = \underbrace{\text{VCF} + \text{PM2.5} + \text{LST}}_{\text{PCA required}} \tag{3}$$

This is the unrestricted model with all indicators. Also has $T = 5$.

4 Some Worries: Assumptions during Analysis

The following are assumptions made and possible weaknesses in our analysis identified after compiling the full dataset. They can be read as pain points, which can be fixed in future versions, or as questions that may arise during presentations, and thus require explanations to be prepared.

Mean vs Median We currently use the median cell value while building our zonal statistics for LST, and mean for all others. This needs to be qualified by theory and on some understanding of the data distribution.

Sample selection bias We are currently deleting *shrids* where data is missing for some particular indicator: ref line 63-65 and 132 of `regression.R`, which leads to losing upto 1% of the population.

Cluster standard errors Should we cluster standard errors by group, or by time, or simply report heteroskedasticity-robust standard errors?

Standardization We are currently standardizing our variables (i.e., $\mu = 0$ and $\sigma^2 = 1$) before estimating coefficients. Does this take away from our analysis? Is this commonplace?

5 Results

Table 1: Descriptive statistics.

Statistic	N	Mean	St. Dev.	Min	Max
NTL	652,620	6.471	6.525	0.000	63.000
PM2.5	652,620	30.542	6.748	12.386	77.000
VCF	652,500	75.112	10.964	20.381	100.000
LST (°F)	651,840	292.511	3.050	264.125	300.335
GHSL	130,524	118.525	204.381	0.000	3,917.117

Table 2: Regression results from Model 1. UW implies uniform weighting, Env implies PCA index. Results 1-3 are for Epoch 1, and 4-6 are for Epoch 2.

	Dependent variable:					
	Nighttime lights (std.)					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Envindex}_{\text{full}}$	0.075^{***}			-0.069^{***}		
	(0.000)			(0.000)		
$\text{Envindex}^2_{\text{full}}$	0.071^{***}			-0.013^{***}		
	(0.000)			(-0.000)		
$\text{Env}_{\text{VCF,PM2.5}}$		0.024^{***}			-0.158^{***}	
		(0.000)			(0.000)	
$\text{Env}^2_{\text{VCF,PM2.5}}$		0.073^{***}			-0.012^{***}	
		(0.000)			(-0.000)	
VCF+PM2.5_{UW}			0.034^{***}			-0.223^{***}
			(0.000)			(0.000)
VCF+PM2.5^2_{UW}			0.147^{***}			-0.025^{***}
			(0.000)			(-0.000)
Observations	325,860	325,860	325,860	325,860	326,250	326,250
R^2	0.060	0.051	0.051	0.006	0.029	0.029
Note:						*p<0.1; **p<0.05; ***p<0.01