Economics RA Guide Predoc Coding Task 2025

1 Assignment Question

The complete assignment can be found at `https://raguide.github.io/new_email`. The PI finds an article that describes a recent increase in children with asthma in our town. The journalist speculated that the recent opening of a factory producing Pokeballs could be responsible. Pokeballs, while super cute, are made using a highly toxic process that can cause pollution in the city that a factory is in.
The Professor would like to explore a research project on whether the opening of Pokeball factories across the country has had an effect on childhood asthma in a town.
What was the effect of the opening of Pokeball factories on a town's incidence of childhood asthma?
The PI was reminded of a law passed in 2007 that subsidized the opening of Pokeball factories in a number of states. We suspect that the passage of the law was a random and unexpected event.

The deliverable should be a set of evidence that can be used to answer the above questions and a very short write-up. You should focus on stating clearly your assumptions, your findings, and any limitations.
Since you're new to the country, don't worry too much about the institutional and local details of the data and just use what you were given. Nonetheless, you are welcome to suggest that the PI takes a deeper look into patterns that you notice and speculate on mechanisms that exist in the economic literature in other countries (e.g. the US). You should also highlight the limitations of the data in answering these questions and suggest other data sources that we can look into (if they exist in other countries).

Time for the task: 6 hours

2 Solution

2.1 Data Preparation and Merging

2.1.1 Hospital Data

The hospital dataset includes 177 observations, each representing a hospital over the period 2000–2015. It initially appeared in a wide format, with 16 asthma case variables (cases0 to cases15).Steps taken:

1. Verified the absence of missing values or duplicates
2. Created a city variable by extracting the first two words of each hospital name
3. Reshaped the data from wide to long, using:

2.1.2 City Data

The city dataset includes 50 observations and 35 variables from 2000-2015. Imported separately from a CSV file and transformed similarly:

1. Extracted the city name
2. Reshaped to long format with population by city and year

2.1.3 Merging

Two datasets- hospital and city data- were merged using city and year as keys. There were 2,736 merged observations in the final dataset.
Post-merge, the dependent variable **asthma_rate** was created as

```
gen asthma_rate = (cases / population) * 10000
```

This was done to ensure that the measurement of asthma cases could be relative to populations of the respective cities, calculated as the number of reported asthma cases per 10,000 population in a city-year.

3 Emperical Analysis

To assess the causal impact of a 2007 legal change on asthma incidence, this study employs a **panel difference-in-differences (DiD)** design using annual city-level data from 2001 to 2015. The empirical strategy compares changes in asthma rates over time between cities that implemented the law ("treated" cities) and those that did not ("control" cities). This method relies on the key identifying assumption that, in the absence of the policy, treated and control cities would have followed similar trends in asthma rates over time- **the parallel trends assumption**.
I tested the parallel trends assumption — the idea that, in the absence of the treatment, asthma rates in treated and control cities would have evolved similarly over time — using both covariate balance tests and an event-study approach with leads and lags.

3.1 Covariate Balance Tests

To begin, I restricted the dataset to the pre-treatment period (2001–2006) and compared the average number of factories, production of pokeballs and population across treated and untreated cities. A two-sample t-test was performed, and the only statistically significant difference was that in the population variable- the population of control groups cities was higher than treatment group cities at the 5% significance level.
To formally test for differences in pre-trends, I implemented an event study regression by generating relative year dummies for each period surrounding the treatment year (2007). This involved creating binary indicators for each year from six years before the treatment (rel_m6) to eight years after (rel8), excluding the treatment year itself (rel_year == 0) as the reference category. The specification took the form:

```
xtreg asthma_rate rel_m6 rel_m5 rel_m4 rel_m3 rel_m2 rel_m1 rel1 rel2 rel3 rel4 rel5
rel6 rel7 rel8 i.year, fe cluster(city_id)
```

The results, given in Table 1, indicate that pre-treament the parallel trends assumption held fairly well. Most of the pre-treatment coefficients were small and statistically insignificant — except for rel_m2 (corresponding to the year 2005), which showed a statistically significant positive effect. This violation of the pre-trend assumption is non-trivial and suggests that treated cities may have already been experiencing increases in asthma rates

Table 1: Event Study Regression of Asthma Rates on Relative Year Dummies

Variable	**Coefficient**	**Std. Error**	**p-value**
Pre-treatment (Leads)			
rel_m6	0.465	1.196	0.699
rel_m5	-0.283	1.088	0.796
rel_m4	-0.248	1.445	0.865
rel_m3	-1.544	1.089	0.163
rel_m2	**1.945**	**0.793**	**0.018**
rel_m1	-0.579	0.922	0.533
Post-treatment (Lags)			
rel1	2.086	0.977	0.038
rel2	1.491	0.945	0.121
rel3	2.611	1.071	0.019
rel4	2.383	1.324	0.078
rel5	1.406	1.159	0.231
rel6	1.668	1.166	0.159
rel7	3.597	1.466	0.018
rel8	6.049	1.702	0.001
Model Statistics			
Within R^2		0.239	
Number of observations		768	
Number of cities		48	

Notes: Fixed-effects (within) regression with city and year fixed effects. Standard errors are clustered at the city level. Year 2007 (rel_year = 0) omitted as the base period.

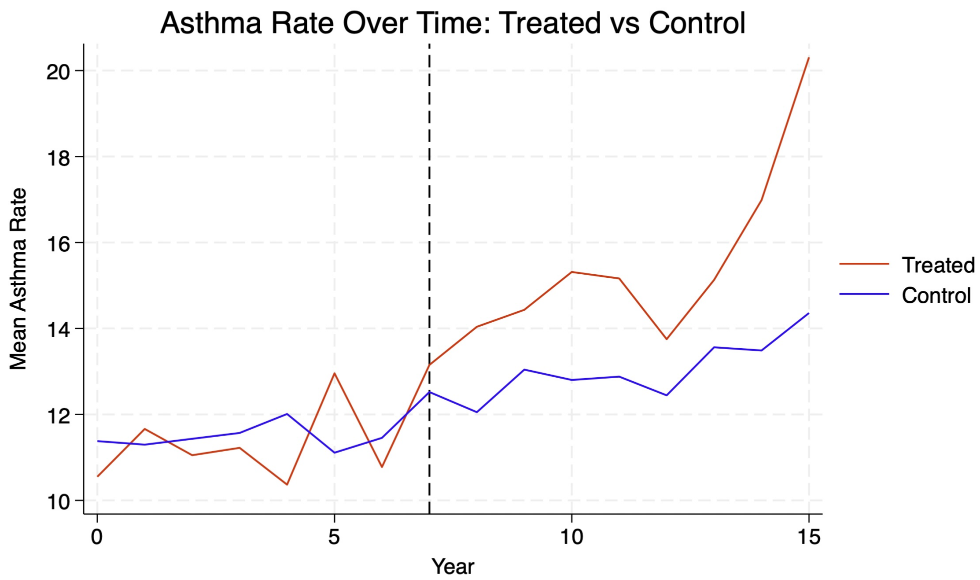

Figure 1: Asthma Rates Over Time in Treated and Control Cities

before the policy was enacted, which could bias the estimated treatment effect. A plot of the coefficients and confidence intervals from the event study is given in Figure 1

Figure 1 displays the evolution of mean asthma rates over time for both treated and control cities, with the year 2007 (marked by a vertical dashed line) representing the timing of the policy intervention. Prior to the policy change, the two groups appear to follow broadly similar trends, lending visual support to the parallel trends assumption required for a difference-in-differences (DiD) analysis. This similarity is especially visible in the relatively flat and aligned movement of asthma rates up to year 6. However, in the post-treatment period, a clear divergence emerges: asthma rates in treated cities rise sharply, while control cities see only a modest increase. By the end of the observed time period, the treated group shows a mean asthma rate significantly higher than that of the control group. This widening gap suggests a potential positive effect of the policy on asthma rates in treated cities, consistent with the findings from the formal DiD and event study regressions. While this graphical evidence does not confirm causality on its own, it provides an important first diagnostic for treatment effect estimation.

3.2 Estimating the Causal Effect

Despite some concerns about pre-trends, I proceeded to estimate the average treatment effect using the standard two-way fixed effects DiD specification. The results are presented in Table 2.

Table 2: Difference-in-Differences Estimation of Asthma Rates

Variable	**Coefficient**	**Std. Error**	**p-value**
post_treated	2.586***	0.859	0.004
Year 1	0.216	0.585	0.713
Year 2	0.166	0.597	0.782
Year 3	0.311	0.524	0.557
Year 4	0.428	0.529	0.423
Year 5	0.399	0.506	0.434
Year 6	0.113	0.540	0.835
Year 7	0.860	0.488	0.084
Year 8	0.731	0.600	0.229
Year 9	1.572***	0.547	0.006
Year 10	1.612***	0.562	0.006
Year 11	1.632***	0.594	0.008
Year 12	0.952*	0.528	0.078
Year 13	2.136***	0.593	0.001
Year 14	2.544***	0.699	0.001
Year 15	4.029***	0.760	0.000
Constant	11.172***	0.383	0.000

Note: Standard errors are clustered at the city level.
* p<0.1, ** p<0.05, *** p<0.01

The key coefficient of interest, *post_treated*, is positive and statistically significant at the 1% level ($p = 0.004$). After 2007, treated cities- those that received the Pokéball factory subsidy- experienced an average increase in asthma rates of 2.59 per 1,000 children relative to control cities. This estimate controls for both year and city fixed effects and is statistically significant at the 1% level ($p = 0.004$). The coefficient on *lawchange_num* was omitted due to collinearity with the fixed effects. Several post-treatment years (e.g., Years 9–15) also show large and significant coefficients, suggesting a persistent increase in asthma incidence following the intervention. These findings point toward a potential unintended health consequence of the policy, and further investigation into possible mechanisms- such as factory emissions or exposure channels- would be warranted.

3.3 Placebo Tests

To further probe the credibility of the estimated treatment effect, I conducted two placebo tests by artificially shifting the treatment year to a period when the policy had not yet been implemented. The rationale is that if the DiD framework is correctly specified, moving the treatment year to a falsified date should produce no significant effect.

In the first placebo test, I assigned the treatment year as 2005, generating a placebo_did variable equal to 1 for treated cities in years equal to or after 2005 and 0 otherwise. The specification remained the same:

```
gen placebo_post = year >= 5
gen placebo_did = placebo_post * lawchange_num

xtreg asthma_rate placebo_did lawchange_num i.year, fe cluster(city_id)
```

A similar exercise was carried out with 2010 as another placebo test. I found that the placebo treatment effect was large, positive, and statistically significant at the 1% level in both cases. This suggests that post-2007 differences in trends between treated and control cities persisted even in the absence of the 2007 policy being redefined- potentially reflecting broader structural changes or differential post-2007 trends not caused by the pokeball factory subsidy law.

4 Results and Discussion

In sum, while the main DiD regression produces a statistically significant effect of the 2007 legal change on asthma rates, the pre-treatment diagnostics and placebo tests complicate a straightforward causal interpretation. The evidence points to underlying differences in trends that may pre-date or outlast the policy window, which cautions against over-reliance on the DiD estimator without further robustness checks or alternative identification strategies.

4.1 Assumptions

We assume that the policy was exogenous and not targeted based on health trends. Our DiD framework relies on the parallel trends assumption- that, in the absence of the subsidy, asthma rates would have evolved similarly across treated and untreated cities. We test this using both covariate balance and an event study (leads and lags) specification.

4.2 Findings

Our findings suggest a statistically significant increase in asthma rates in treated cities following the policy. The main DiD coefficient implies an increase of 2.59 asthma cases per 1,000 children post- 2007 ($p = 0.004$), controlling for city and year fixed effects. The event study shows no major violations of pre- trends overall, though one year (2005) shows a significant anticipatory effect. The divergence in asthma rates becomes more pronounced over time in the post-treatment period.

4.3 Limitations

I found the following limitations:

1. Omitted variables: No controls for pollution levels, healthcare access, or industrial output are included.

2. Significant placebo effects: These suggest that DiD estimates may be upward biased due to non-random post-2007 shocks.

5 Conclusion

This study aimed to evaluate the unintended health consequences of a 2007 policy subsidizing Pokéball factories, using a difference-in-differences framework applied to city-level asthma rates. While the main estimates suggest a statistically significant increase in asthma among children in treated cities, robustness checks—including placebo tests—indicate that caution is warranted in attributing a purely causal interpretation. There is some evidence of trend divergence even before the intervention and potential exposure to unobserved post-treatment shocks.

Nonetheless, this preliminary evidence raises important questions about the environmental and public health consequences of industrial subsidies. Future work should investigate these channels more thoroughly using richer data, including emissions, health records, and air quality measures. Additional institutional insights into how the policy was implemented and where factories were sited would further enhance causal inference. Overall, the results highlight the need for policies that carefully weigh developmental objectives against their potential health externalities.