

This article examines the impact of the 2008 Beijing Olympics on pollution concentrations and pollution emissions in and around Beijing. The article focuses on changes in the outcome variables, particularly air quality in Beijing and surrounding cities, using a diff-in-diff approach. The authors state that, with the exception of SO₂, pollution levels in Beijing decreased temporarily as a result of the Olympics.

Major Comments

Contribution of the Paper:

As written, the contribution of the paper seems limited. If the results are to be believed, air quality, as measured by pollution levels of capital cities in and around Beijing improved “for a short period of time” as a result of the attempts to improve air quality for the Olympics. This is not terribly surprising given the drastic steps that the Chinese government took to ensure that the city was relatively clean during this short time frame. Furthermore, other research cited in the paper has already established this (Wang et al 2010, Schleicher et al 2012). It might be useful to frame your paper differently and describe your findings as being long (or at least medium) term effects. The paper tries to describe these findings as short term effects which is problematic given that the data is annual.

Of more potential importance is the impact of the Olympics on long run outcomes. The authors dance around this issue but the idea is never fleshed out. Perhaps strong short-run pushes to improve environmental outcomes can have lasting long-run impacts. Consider the case of car owners who, during the Olympics, were only allowed to drive every other day. Maybe some of these commuters were forced to take public transit and in doing so learned that public transit was easy to use. This might have encouraged a non-trivial amount of people to switch to using public transit. A similar story could be told for pollution emitting factories that were forced (for a brief period) to adopt environmentally friendly practices. Maybe some of these practices persist in the long run because firms ended up liking some of the new techniques they were forced to use and discovered that their implementation was easier than expected. Maybe this did not happen, but if your results gave a conclusive answer to this question, it would strike me as an important finding.

Accuracy of Analysis:

1) Diff-in-Diff Estimation

- 1) Interpretation of results: The major finding, as stated by the authors is that, apart from SO₂, air quality in Beijing improved for “a short time period” as a result of the Olympics.

Given the lack of statistical significance in the majority of their specifications, it is hard to justify this statement using their results. They report results for three pollutants (SO₂, PM₁₀ and NO_x) using two different treatment groups (Groups B and C) and two different start date times (2007 and 2008). There is a statistically significant finding for PM₁₀ and NO_x levels at the 10% level in only one of their four specifications. Also, the authors should pick a preferred specification and interpret the DD coefficient for the reader. Currently the reader does not know whether these are big or small effects.

Also, what is meant by “a short time period”? The DD specification as described in equation 3 tests whether air quality is different in the treated region after the treatment went into effect. This is a period of three years. Is there evidence that air quality briefly improved and then worsened? If so, this should be made clear.

- 2) Standard Errors: There is no discussion of how the standard errors are estimated. Failing to cluster SE's in DD models is known to significantly underestimate the size of the SE's and make coefficients falsely appear to be statistically significant. Bertrand et al (2004) is one of many papers that discuss this issue. SE's should be robust and clustered at the level at which the error terms are correlated. I suspect this will be at the city level.
- 3) Trends: DD specifications should always test for differing trends in the treated and control groups. A plot of the data would be very useful here. If the treated group is trending differently than the control group prior to the treatment then the coefficient cannot be interpreted as the treatment effect. I suspect that the growth in economic activity within China has not been homogenous.
- 4) Possible spillovers: You may be overstating your results if the Olympics forced firms to shift their production activities from Beijing to different locations in China. In this case the treatment group would be impacted by the Olympics because “dirty” production is shifted away from Beijing and to other parts of the country. A discussion of this possibility is important.
- 5) Start Date and Control Group: You state in the Methodology and Data section that 2008 will be the first “Post” period in your analysis. Then the first results you present use 2007 as the first “Post” period. There is no reason given why 2007 is used, other than the fact that the results are statistically significant when 2007 is used and not when 2008 is used. This immediately makes the reader suspicious of your results. If there is a logical reason to use 2007 then this should be stated. Additionally, there should be more discussion of your selection of treatment and control groups. The introduction provides a brief description but more logic behind the choice of each specification is needed. State in detail why this particular set of cities might have experienced improved environmental outcomes. The policies were designed to improve Beijing's air quality and not the air quality of surrounding cities. The most obvious treatment group to me would be Beijing alone but these results are not reported.

- 6) With DD estimators it is always good to see the raw data in a plot and in summary statistics. You cite Hotchkiss et al (2003), who provide a nice figure as well as some a good write up of their DD results.
 - 7) Use of fixed effects: The authors could easily include city fixed effects to remove any time invariant differences in pollution levels between cities. Year fixed effects could also be included to account non parametrically for changes in air pollution levels that are common to all cities in a given year.
- 2) Data: The air pollution data they use comes from and is China's Statistical Yearbook which contains annual pollution levels for 29 "main" cities (these are later described as capital cities) between 2003 and 2010. The primary concern regarding the data is its frequency. Annual data is unlikely to capture the short run impact of the Olympics. It could however capture long run impacts. There is very little discussion of the data. It is said to be the average level that year. Is average obtained from daily highs or monthly highs or daily averages? This is very important since air quality was by far the best during the month of the Olympics. Was air quality improved the entire year, only one month or for all months following the Olympics? This is very important to the interpretation and the annual data limits the ability to answer these questions.

Also, is data in China's Statistical Yearbook accurate? Any evidence you could provide here would be good.

Minor Comments

- 1) There should be more discussion of the five year plans for readers with limited knowledge of Chinese environmental policy. What are they and why are these important to environmental outcomes?
- 2) The paper did not flow well at times. For example, the last sentence of the third paragraph on page seven struck me as particularly clunky and difficult to follow.
- 3) My personal preference is to include stars in the results indicating the significance. It was difficult to immediately see which results were significant and which were not.
- 4) Consistent use of subscripts when referring to NO_x and SO_2 .
- 5) The last sentence in the second paragraph of the Introduction contains a typo: "8s5%"