Grantham Research Institute on
Climate Change and
the Environment

Statistical corruption in Beijing's air quality data has likely ended in 2012

Thomas Stoerk

May 2015

Grantham Research Institute on Climate Change and the Environment

Working Paper No. 194

The Grantham Foundation for the Protection of the Environment
Global Green Growth Institute
The London School of Economics and Political Science

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has nine research programmes:

1. Adaptation and development
2. Carbon trading and finance
3. Ecosystems, resources and the natural environment
4. Energy, technology and trade
5. Future generations and social justice
6. Growth and the economy
7. International environmental negotiations
8. Modelling and decision making
9. Private sector adaptation, risk and insurance

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: http://www.lse.ac.uk/grantham.

This working paper is intended to stimulate discussion within the research community and among users of research, and its content may have been submitted for publication in academic journals. It has been reviewed by at least one internal referee before publication. The views expressed in this paper represent those of the author(s) and do not necessarily represent those of the host institutions or funders.

Title: Statistical corruption in Beijing's air quality data has likely ended in 2012

Author: Thomas Stoerk[1,2]*

First Version: 4 June 2013

This Version: 20 May 2015

Affiliations:

[1]Department of Economics, Universitat Pompeu Fabra, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain.

[2]Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

*Correspondence to: thomas.stoerk@upf.edu.

Abstract:

This research documents changes in likely misreporting in official air quality data from Beijing for the years 2008 to 2013. It is shown that, consistent with prior research, the official Chinese data report suspiciously few observations that exceed the politically important *Blue Sky Day* threshold and an excess of observations just below that threshold. Similar data, measured by the US Embassy in Beijing, do not show this irregularity. To document likely misreporting, this analysis compares the air quality measurements to Benford's Law, a statistical regularity known to fit air pollution data. I find that the Chinese data fit Benford's Law poorly until a change in air quality measurements at the end of 2012. From 2013 onwards, the Chinese data fit Benford's Law closely. The US Embassy data, by contrast, exhibit no variation over time in the fit with Benford's Law, implying that the underlying pollution processes remain unchanged. These findings suggest that misreporting of air quality data for Beijing has likely ended in 2012.

Introduction:

Ambient air pollution is the cause of a major health crisis in current-day China, with pollution from ambient particulate matter alone being the fourth most important health burden in terms of age-standardized disability-adjusted life year rates [1]. At the same time, political stakes are high: to fix air pollution or to "bring his head" was the unequivocal message that Beijing's mayor received from the central government earlier this year [2]. To address public pressure, the Beijing Municipal Environmental Protection Bureau (BMEPB) regularly publishes statistics on the city's air quality. These data, however, have not gained public trust as they often stand in stark contrast to measurements reported by the US Embassy in Beijing [3-5].

And the public might have a point: the 'leaders make numbers' phenomenon has a tradition in China [6]. Air quality data in China have been found to be suspicious in several cities until 2010 based on both a discontinuity test [7-8], and on a detailed analysis of histograms and location of measurement stations for Beijing until 2007 [9]. These studies suggest that misreporting is prevalent around the politically important *Blue Sky Day* threshold, possibly due to the strict enforcement of promotion criteria [10-11]. This label designates days with an air quality index (AQI) of 100 or less, and the number of *Blue Sky Days* enters the performance assessment of local officials.

The present research adds to this literature by employing a systematic approach to identifying misreporting that goes beyond the analysis of a discontinuity. Moreover, the analysis uses more recent data. This improves on the literature in two aspects: Firstly, this research is the first to track the anomaly over time, and uncover new regularities. Secondly, my method allows to compare the similar, but slightly different US Embassy data to the BMEPB data and thus distinguish misreporting from temporary measures such as factory closures on borderline *Blue Sky Days*.

Methods (Data):

This research uses a new database of Beijing ambient air quality data that was collected and merged from two different sources (see *SI 1* for the data sources). The first source is the BMEPB. The BMEPB has been reporting daily air quality measurements from 31st December 2007 until 18th March 2013. The second source comes from the measurements that the Embassy of the United States in Beijing provides via their Twitter account. These data report hourly and run from 8th April 2008 3pm to 31st December 2013 11pm, with a missing period from 6th November 2008 to 7th February 2009.

While following the same calculation procedure, the AQIs of the BMEPB and the US Embassy aggregate different pollutants and use slightly different break points to convert pollutant concentrations into AQI values (see *SI 2* for the full definitions). The following paragraphs explain the exact composition of both datasets and provide evidence to show that both data sources are comparable because they reflect the same underlying pollution processes.

The BMPEB measures three pollutants until 30th December 2012 (PM_{10}, NO2, SO2) and, following a change in its AQI, six pollutants from 31st December 2012 onwards ($PM_{2.5}$, PM_{10}, NO2, SO2, O3, CO). The US Embassy, by contrast, measures only $PM_{2.5}$ throughout. While the differing definitions might suggest that the datasets might not be sufficiently comparable due to the possibility of the BMEPB measurements being based on pollutants other than particulate matter, in practice this concern is muted by the clear prevalence of particulate matter-driven observations in the BMEPB data. The BMEPB data report the main pollutant for all observations with an AQI exceeding 50, and, as shown in *SI 2*, the AQI of a given observation is only determined by the concentration of the main pollutant on that day, that is by the pollutant concentration defined as most harmful to human health compared to the other pollutant concentrations measured on the same day.

Using the information on the main pollutant, I extract the share of particulate matter-driven observations as 96.93% for the full BMEPB sample, as 98.03% until 30th December 2012 and as 78.72% thereafter (see *SI3*: Fig. 7). Note that the true share of particulate matter-driven observations after 30th December 2012 is likely higher than 78.72% as 10.85% of the observations fail to specify a main pollutant and it is reasonable to suppose that the main pollutant would have been $PM_{2.5}$ or PM_{10} for part of the observations. The clear majority of the BMEPB observations are thus based on AQIs set by particulate matter[a].

A remaining concern is that the BMEPB measures only PM_{10} until 30th December 2012, whereas the US Embassy measures $PM_{2.5}$. The analysis until 30th December 2012 therefore presupposes a high correlation between PM_{10} and $PM_{2.5}$ in Beijing. A sharp discontinuity in the pollution processes, for instance, would appear in both the BMEPB and the US Embassy data only if this correlation were high.

Several pieces of evidence exist to show that this correlation is, in fact, very high. Firstly, I use the information on the main pollutant for observations with an AQI beyond 50 to convert the BMEPB data into concentrations of PM_{10} and $PM_{2.5}$. Then, I correlate these concentrations to the US Embassy's $PM_{2.5}$ concentrations for those US Embassy observations that would have yielded an AQI beyond 50 using the BMEPB's AQI definition to ensure comparability. This exercise shows a high correlation between both the inferred BMEPB PM_{10} concentrations and the

[a] As a robustness check, *SI3* shows that the findings drawn from the BMEPB data are unaffected when dropping all observations that report a main pollutant other than $PM_{2.5}$ or PM_{10}.

US Embassy $PM_{2.5}$ concentrations (correlation coefficient: 0.71) and the inferred BMEPB $PM_{2.5}$ concentrations and the US Embassy $PM_{2.5}$ concentrations (correlation coefficient: 0.91).

Secondly, recent research [12] on ambient air quality in Beijing finds similarly high correlations between 0.69 and 0.85 between the BMEPB PM_{10} concentrations and the US Embassy $PM_{2.5}$ concentrations for a 423 day period between 10 May 2010 and 6 December 2011, which is part of my earlier subsample. Such a high correlation is not specific to Beijing and borne out by research from other parts of the world (for Mumbai, India, see [13]; for Milan, Italy, see [14]).

Overall, this evidence suggests that both data sources move very closely together, and thus reflect the same underlying pollution processes. In terms of data quality, the US Embassy data is less likely to be influenced by the *Blue Sky Day* threshold than the BMEPB data because the number of *Blue Sky Days* in Beijing does not enter the evaluation of US officials.

Methods (Analysis):

An anomaly in the official air pollution data. First, I investigate whether the raw data contain suggestive evidence for manipulation of the official air quality data due to incentives. The histograms of the Beijing air quality data from the BMEPB show a striking anomaly: there are disproportionally many observations just at and below the politically important *Blue Sky Day* threshold of 100, and surprisingly few values just above 100 (Fig. 1, a). Next, the sample period is split on 31st December 2012, the date on which the BMEPB started to include measurements of $PM_{2.5}$. Strikingly, the missing values above 100 come entirely from the earlier subsample (Fig. 1, b). The later subsample does not show this pattern (Fig. 1, c).

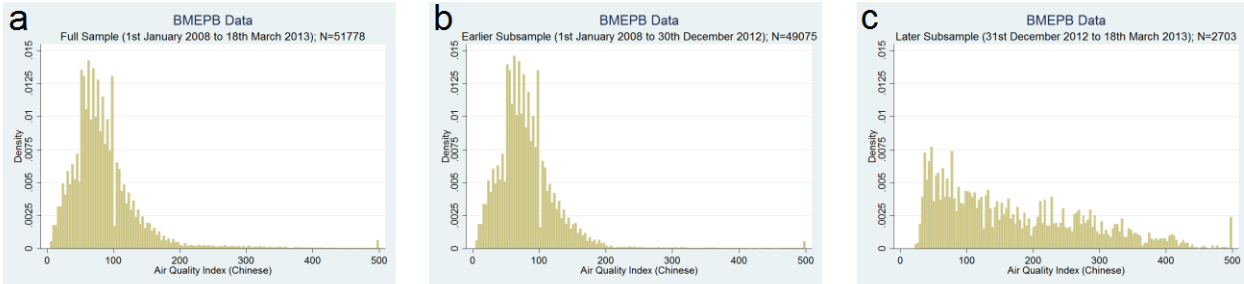


Fig. 1: Histogram of air pollution levels (BMEPB data). *Histograms of the BMEPB data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample. AQI values of 100 and less constitute* Blue Sky Days.

Evidence for likely misreporting. While misreporting seems to be a likely explanation, and was so interpreted by [9] who found the same anomaly until 2007, it is not the only one. Instead, the anomaly might be due to emergency measures such as temporary driving bans that Beijing's local authorities might have implemented on borderline *Blue Sky Days*. If emergency measures rather than misreporting explained the missing values in the BMEBP data, the US Embassy data should exhibit the same anomaly. The histograms for the US Embassy data show that this not the case, neither for the full sample (Fig. 2, a) nor for the different subsamples (Fig 2, b and c). This evidence suggests that misreporting by the BMEPB was prevalent until 2012 and then stopped.

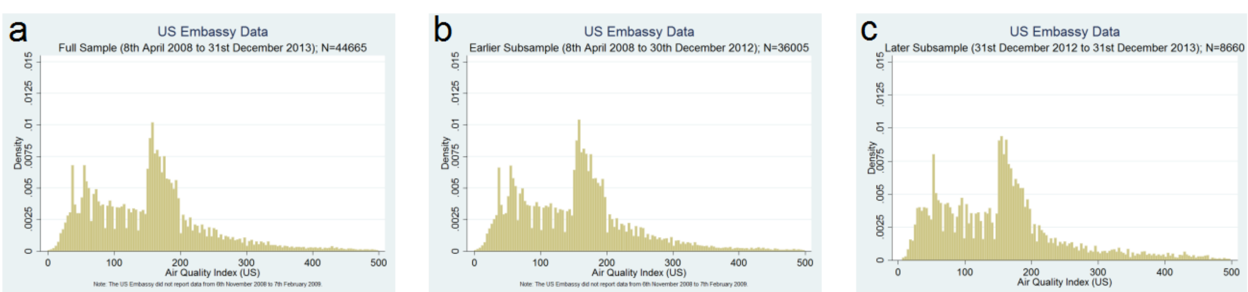


Fig. 2: Histogram of air pollution levels (US Embassy data). *Histograms of the US Embassy data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

Yet, an important caveat is that the two data sources are not comparable in levels due to the late inclusion of $PM_{2.5}$ measurements in the BMEBP data and to the use of slightly different breakpoints (see *SI2*)[b]. To make both data sources comparable and find a direct indication for misreporting, this research compares the goodness-of-fit of both the BMEPB and the US Embassy data to statistical regularity called Benford's Law.

Benford's Law is a distribution that closely characterizes the distribution of the first significant digit in many naturally occurring large data sets [15]. According to Benford's Law, the frequency for the first non-zero digit is approximately governed by the following distribution:

$$Frequency(i) = \log_{10}\left(1 + \frac{1}{\mathrm{i}}\right) \quad where\ i = 1,2, \dots 9.$$

The goodness-of-fit with Benford's Law as an indication for fraudulent data has been used in economics in general [16-18] and for ambient air pollution data in particular [19-21]. This analysis uses Benford's Law in two ways. First, graphical evidence on the fit between Benford's Law and relevant subsamples of the datasets is shown to give an indication of whether misreporting occurs in the BMEPB's data and whether it has changed over time. Second, the analysis computes a goodness-of-fit measure between the predicted frequency and the empirical frequency for both data sources and plots this measure over time.

The goodness-of-fit measure is the χ^2 statistic, which is an appropriate statistical measure for comparing data with discrete categories to a predicted distribution[c] [22]. In the case of

[b] The different breakpoints in the AQI used by the US Embassy also explain the hump-shaped distribution of the US Embassy's histogram (Fig. 2, a-c). *SI4* provides further illustration of this point and shows the close comparability of Fig. 1c and Fig. 2c when using the same breakpoints.

[c] Other measures commonly used to test goodness-of-fit with Benford's Law include the Kolmogorov-Smirnov statistic, the Kuiper statistic, the d (distance) statistic and the statistic m (max) [23]. The conclusions drawn from Figures 5 and 6 are robust to using any of these goodness-of-fit measure because of their high correlation with the χ^2 statistic in both samples: Kolmogorov-Smirnov: 0.94 (BMEPB data), 0.90 (US Embassy data); Kuiper: 0.96

Benford's Law, there are 9 discrete categories corresponding to the digits 1 to 9. The χ^2 statistic is defined as:

$$\chi^2 = \sum_{i=1}^{9} \frac{(\text{Frequency Observed}_i - \text{Frequency Benford}_i)^2}{\text{Frequency Benford}_i}$$

Results:

As can be seen, the BMEPB data match Benford's Law poorly for the full sample (Fig. 3: a), suggesting misreporting. This misreporting seems to happen exclusively in the earlier subsample (Fig. 3: b). In the later subsample (Fig. 3: c), by contrast, the BMEPB data match Benford's Law closely. Misreporting is therefore the likely margin of action. To verify that this improvement in goodness-of-fit is not due to a change in the underlying processes that generate the air pollution, Figure 4 shows that the fit of the US Embassy data is good and unchanged throughout all periods.

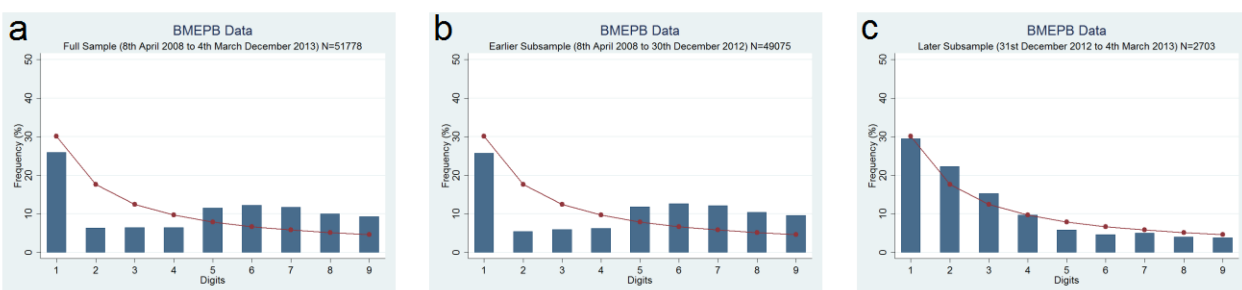


Fig. 3: Observed frequencies and Benford's Law (BMEPB data). *Observed frequencies for the first significant digits (blue bars) against the frequencies predicted by Benford's Law (red connected dots) for the BMEPB data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

(BMEPB), 0.90 (US Embassy); d (distance): 0.93 (BMEPB), 0.90 (US Embassy); m (max): 0.73 (BMEPB), 0.89 (US Embassy).

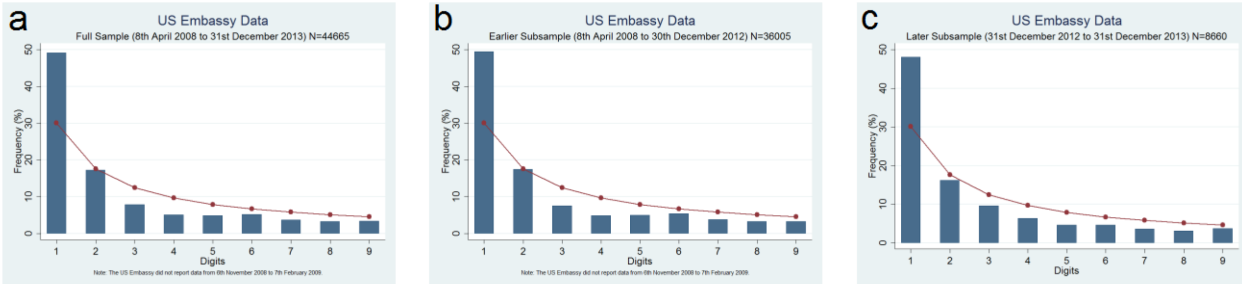


Fig. 4: Observed frequencies and Benford's Law (US Embassy data). *Observed frequencies for the first significant digits (blue bars) against the frequencies predicted by Benford's Law (red connected dots) for the US Embassy data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

Likely misreporting over time. To ensure that the improvements in the goodness-of-fit with Benford's Law for the BMEPB data reflect a general trend rather than being an artifact of the date at which I split the sample, the analysis computes the goodness-of-fit over time. Figure 5 shows that the goodness-of-fit between the BMEPB data and Benford's Law markedly improved after the BMEPB started measuring $PM_{2.5}$ on 31st December 2012. To check that the underlying pollution processes remain the same, Figure 6 shows that the goodness-of-fit for the US Embassy data has not changed. The reason for the improved goodness-of-fit for the BMEPB data must therefore lie in the Chinese measurements.

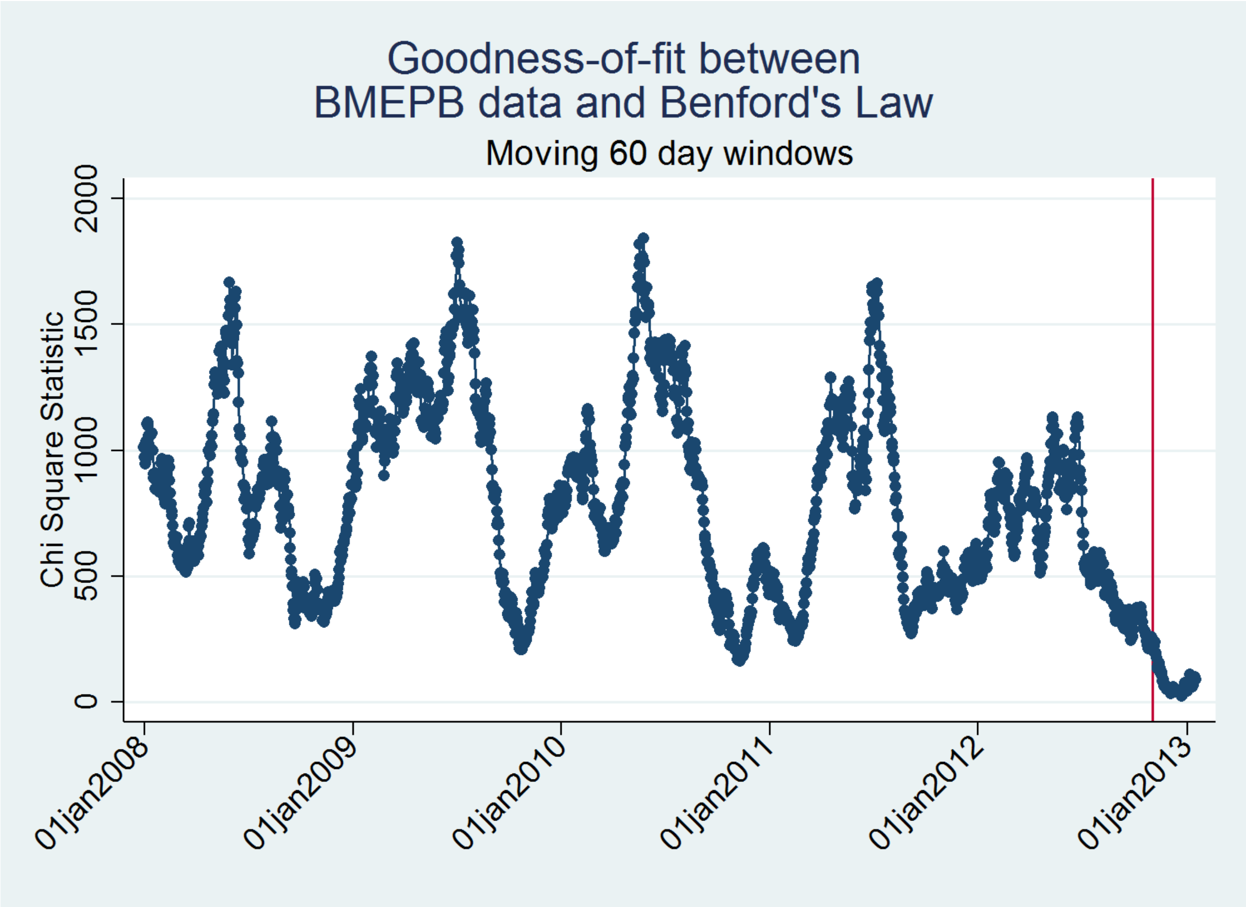


Fig. 5: Goodness-of-fit with Benford's Law over time (BMEPB data). *χ^2 statistic comparing the BMEPB data to Benford's Law. Computed over moving 60 day windows. The red, solid line marks earliest time window during which the BMEPB started measuring $PM_{2.5}$ (from 31st December 2012 onwards).*

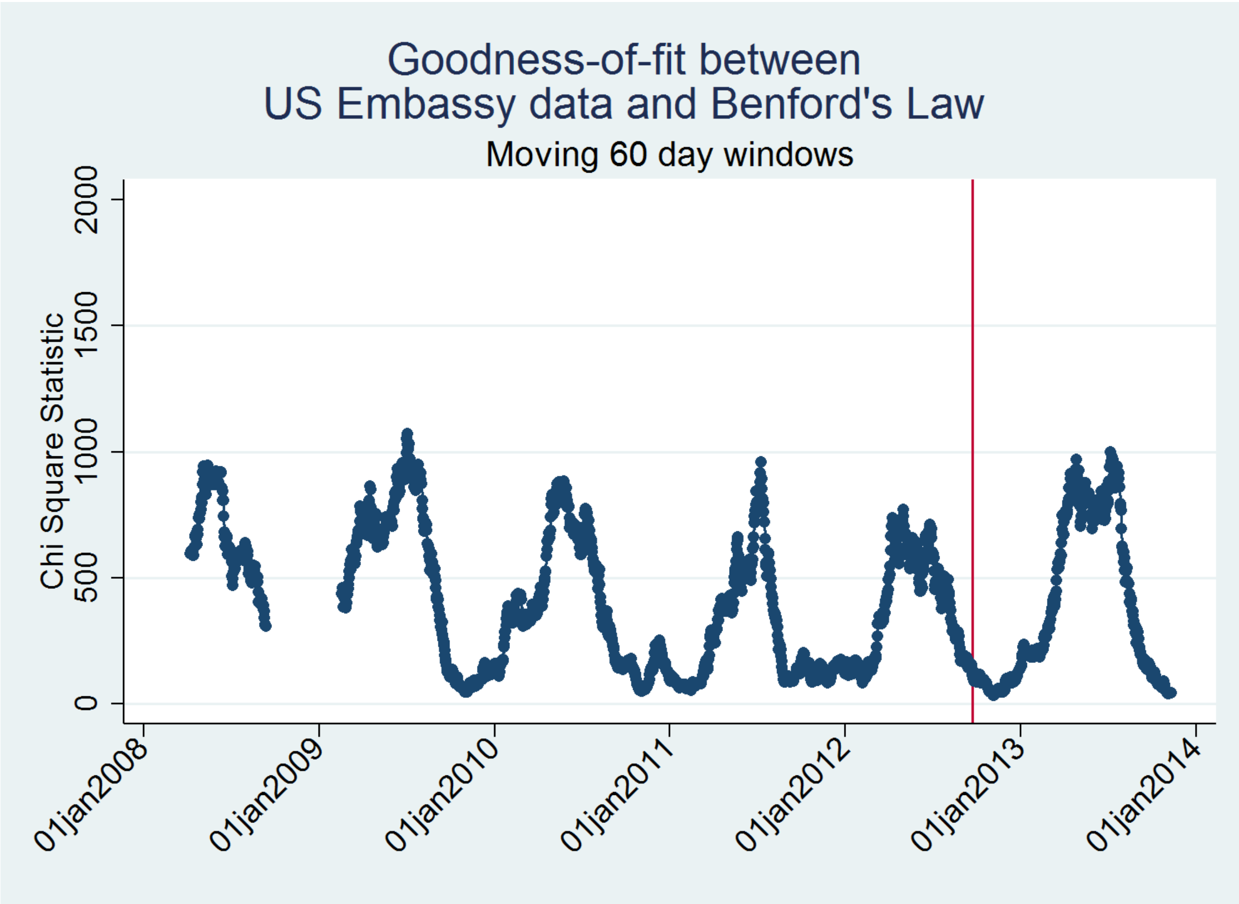


Fig. 6: Goodness-of-fit with Benford's Law over time (US Embassy data). *χ^2 statistic comparing the US Embassy data to Benford's Law. Computed over moving 60 day windows. The red, solid line marks earliest time window during which the BMEPB started measuring $PM_{2.5}$ (from 31st December 2012 onwards). The US Embassy did not report data between 6th November 2008 and 7th February 2009.*

Discussion and Concluding Remarks:

One important caveat regarding these results is in order. The analysis relies on two pieces of evidence: a sharp discontinuity in the histograms of the raw data, and the goodness-of-fit with

Benford's Law. Both pieces of evidence can detect a relabeling of AQI values from just beyond the *Blue Sky Day* cutoff to just below it. More sophisticated methods of misreporting, however, such as shifts in the entire distribution of AQI values, would go undetected. While this limitation is important in theory, in practice this is less so: previous studies that detected misreporting of air quality data in Beijing until 2007 [9] and other cities in China until 2010 [7-8] found misreporting to consist of the relabeling practice that the analysis can track. The methodology used in this study is thus accurate in closely following known misreporting over time. To establish that all possible kinds of misreporting have stopped, however, would require further analysis.

Political pressure to fix air pollution can result in "statistical corruption": government authorities respond by misreporting the desired data. Using air quality data from Beijing, this research shows that the official data exhibit such misreporting, thus confirming that the misreporting analyzed in [9] extended well beyond 2007. Furthermore, this research makes use of the unique setting in Beijing, where air quality is independently measured by both the Chinese authorities and the US Embassy. Employing a statistical misreporting measure that makes both data sources comparable, this analysis suggests that the authorities in Beijing likely manipulated air quality from 2008 to 2012. From 2013 onwards, this has changed: despite ongoing suspicion regarding the quality of Chinese air pollution data [8], the 'leaders make numbers' phenomenon seems to have been overcome.

Environmental governance in China is currently at a crossroads. Proposals have called for policies to decouple economic activity from pollution [24], to improve health outcomes by reducing pollution [25], and to realize the co-benefits of reducing air pollution for climate change mitigation [26]. At the same time, research has shown the significant impact of public policy on ambient air pollution in China [27], thus highlighting the role for policy. Whichever strategies China may decide to implement, reliable air quality data are needed for successful implementation and evaluation. The findings of the present research are thus a reason for optimism: unlike only a few years ago, statistics on atmospheric pollutants now seem to allow for evidence-based environmental policy making in China.

References:

[1.] G. Yang, Y. Wang *et al.*, Rapid health transition in China, 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet*. **381**, 1987-2015 (2013).

[2.] People's Daily, "北京市长向国务院立'生死状': 治不好空气污染'提头来见'" (18 January 2014, http://js.people.com.cn/html/2014/01/18/283276.html)

[3.] D. Grammaticas, "China's pollution, a toxic issue" (BBC, 7 December 2011; http://www.bbc.co.uk/news/world-asia-16068926).

[4.] E. Wong, "On Scale of 0 to 500, Beijing's Air Quality Tops 'Crazy Bad' at 755" (New York Times, 12 January 2013; www.nytimes.com/2013/01/13/science/earth/beijing-air-pollution-off-the-charts.html?_r=0).

[5.] Le Monde, "Chine: la pollution de l'air pousse Pékin 'à accepter plus de transparence'" (28 January 2013; http://www.lemonde.fr/asie-pacifique/article/2013/01/28/chine-la-pollution-de-l-air-pousse-pekin-a-accepter-plus-de-transparence_1823481_3216.html).

[6.] J. Liu, H. Yang. *Science*. **325**(5941), 675-676 (2009).

[7.] Y. Chen, G. Z. Jin, N. Kumar, G. Shi, Gaming in Air Pollution Data? Lessons from China. *B.E. J. Econ. Anal. Poli.* **12**(3) 1-41 (2012).

[8.] D. Ghanem, J. Zhang, 'Effortless Perfection': Do Chinese cities manipulate air pollution data? *J. Environ. Econ. Manag.* **68**(2014), 203-225 (2014).

[9.] S. Q. Andrews, Inconsistencies in air quality metrics: 'Blue Sky' days and PM_{10} concentrations in Beijing. *Environ. Res. Lett.* **3**(034009), 1-14 (2008).

[10.] S. Zheng, M. E. Kahn, Understanding China's Urban Pollution Dynamics. *J. Econ. Lit.* **51**(3), 731-772 (2013).

[11.] A. L. Wang, The Search for Sustainable Legitimacy: Environmental Law and Bureaucracy in China, *Harv. Envtl. L. Rev*. **37**(2), 365-440 (2013).

[12.] J.-F. Wang *et al.*, Estimation of Citywide Air Pollution in Beijing, *PLoS One*, **8**(1), 1-6 (2013).

[13.] R. Kumar, A.E. Joseph, Air pollution concentrations of PM2.5, PM10 and NO2 at ambient and kerbsite and their correlation in Metro City - Mumbai. *Environ. Monit. Assess.*, **119**(1-3), 191-199 (2006).

[14.] G.M. Marcazzan, S. Vaccaro, G. Valli, R. Vecchi, Characterisation of PM10 and PM2.5 particulate matter in the ambient air of Milan (Italy). *Atmos. Env.*, **35**(27), 4639-4950 (2001).

[15.] F. Benford, The law of anomalous numbers. *P. Am. Philos. Soc.* **78**(4), 551-572 (1938).

[16.] H. Varian, Benford's law. *Am. Stat.* **23**, 65-66 (1972).

[17.] R. M. Abrantes-Metz, S. B. Villas-Boas, G. Judge, Tracking the Libor rate. *Appl. Econ. Lett.* **18**(10), 893-899 (2011).

[18.] T. Michalski, G. Stoltz, Do countries falsify economic data strategically? Some evidence that they might. *Rev. Econ. Stat.* **95**(2), 591-616 (2013).

[19.] C.F. Dumas, J. H. Devine, Detecting Evidence of Non-Compliance In Self-Reported Pollution Emissions Data: An Application of Benford's Law, paper presented at the Annual Meeting of the American Agricultural Economics Association, Tampa, FL, 30 July - 2 August 2000.

[20.] R. J. C. Brown, Benford's Law and the screening of analytical data: the case of pollutant concentrations in ambient air. *Analyst* **130**, 1280-1285 (2005).

[21.] S. De Marchi, J. T. Hamilton, Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory. *J. Risk Uncertainty* **32**, 57-76 (2006).

[22.] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York, 1956), p. 45.

[23.] J. Morrow, Benford's Law, Families of Distributions and a Test Basis (CEP Discussion Paper No 1291, Centre for Economic Performance, 2014).

[24.] H.-W. Xiao, A.S. Mujumdar, L. Che, Pollution: uncouple from economy boom. *Nature* **517**, 145 (2015).

[25.] V. Hughes, Where there's smoke. *Nature* **489**, S18-S20 (2012).

[26.] J. Schmale, D. Shindell, E. von Schneidemesser, I. Chabay, M. Lawrence, Clean up our skies. *Nature* **515**, 335-337 (2014).

[27.] D. Almond, Y. Chen, M. Greenstone, H. Li, Winter Heating or Clean Air? Unintended Impacts of China's Huai River Policy. *Am. Econ. Rev. P&P* **99**(2), 184-90 (2009).

Acknowledgments: While all errors are my own, I have greatly benefitted from comments by A. Dechezleprêtre, P. Dimakopoulos, B. Kolb, G. Leon, H. Llavador, M. Martinez Carrasco, J. Morrow, R. Nagel, O.Papaspiliopoulos, G. Ponzetto, J. Quoidbach, D. Tang, Y. Zou, Y. Zylberberg and seminar audiences at Universitat Pompeu Fabra and the London School of Economics and Political Science. A special thank you to V. Fouka for her generous help in programming. Financial support from the Spanish Ministry of Economy and Competitiveness (ECO2011-25295) is gratefully acknowledged.

Financial interests statement: The author has no competing financial interests.

Supplementary Information for

Title: Statistical corruption in Beijing's air quality data has likely ended in 2012

Author: Thomas Stoerk[1,2]*

Affiliations:

[1]Department of Economics, Universitat Pompeu Fabra, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain.

[2]Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

*Correspondence to: thomas.stoerk@upf.edu.

This file includes:

SI 1: Data Sources

SI 2: Definition of the Air Quality Indices (AQIs) used by the Beijing Municipal Environmental Protection Bureau and the US Embassy

SI3: Robustness check: Analysis after dropping BMEPB observations from non-particulate matter AQIs

SI4: Robustness check: Figure 2 using BMEPB AQI breakpoints

SI 1: Data sources

The original, public data sources are:

BMEPB web interface: http://www.bjepb.gov.cn/air2008/Air1.aspx. The historical dataset could not be downloaded directly; instead, a webscraping algorithm was used. This web interface stopped reporting in March 2013.

US Embassy web interface: http://www.stateair.net/web/historical/1/1.html. In conformity with the US Department of State's data use statement, I note that the US Embassy data used in this study are not fully verified or validated and are released with the sole purpose of providing health information to US citizens who are travelling abroad. I furthermore give attribution to the US Department of State for providing the data.

SI 2: Definition of the Air Quality Indices (AQIs) used by the Beijing Municipal Environmental Protection Bureau and the US Embassy

The AQI is an index that is used to inform the population about the health effects of different air pollutants. The AQIs used by the BMEPB and the US Embassy are calculated according to the same two-step procedure:

In the first step, the concentration of each pollutant is converted into an individual air quality index value (IAQI) via the following formula:

$$IAQI_p = \frac{IAQI_{High} - IAQI_{Low}}{BP_{High} - BP_{Low}} (C_P - BP_{Low}) + IAQI_{Low}$$

where C_p is the measured concentration of pollutant p, BP_{High} is the breakpoint that is higher than or equal to C_p while BP_{Low} is the breakpoint that is lower than or equal to C_p. $IAQI_{High}$ and $IAQI_{Low}$ are the AQI scores that correspond to the BP_{High} and BP_{Low} according to the following tables:

Breakpoints for Different Pollutants Before 31st December 2012				
	BMEPB[i]			US Embassy[ii]
AQI[iii]	***SO_2***	***NO_2***	***PM_{10}***	***$PM_{2.5}$***
0	0	0	0	0
50	50	80	50	15.5
100	150	120	150	40.5
150	[iv]	[iv]	[iv]	65.5
200	800	280	350	150.5
300	1600	565	420	250.5
400	2100	750	500	350.5
500	2620	940	600	500

All breakpoints are for concentrations over a 24 hour period (in µg/m^3).

Notes:

[i]GB 3095-1996, SEPA Announcement [2000] No. 1 (Amendment to GB 3095-1996).

[ii]EPA-454/B-12-001.

[iii]The official name for the air quality index used by the BMEPB was "Air Pollution Index" until the 31st of December 2012. To avoid confusion, the name "Air Quality Index" is used throughout this article.

[iv]The BMEPB does not employ a separate breakpoint for an AQI of 150 before 31st December 2012.

Breakpoints for Different Pollutants from 31st December 2012 Onwards							
	BMEPB[i]						US Embassy[ii]
AQI	***SO_2***	***NO_2***	***CO***	***O3***	***PM_{10}***	***$PM_{2.5}$***	***$PM_{2.5}$***
0	0	0	0	0	0	0	0
50	50	40	2	100	50	35	15.5
100	150	80	4	160	150	75	40.5
150	475	180	14	215	250	115	65.5
200	800	280	24	265	350	150	150.5
300	1600	565	36	800	420	250	250.5
400	2100	750	48	[iii]	500	350	350.5
500	2620	940	60	[iii]	600	500	500

All breakpoints are for concentrations over a 24 hour period (in μg/m^3); excepting CO (measured in mg/m^3) and O_3 (measured over an 8 hour period).

Notes:
[i]HJ 633-2012.
[ii]EPA-454/B-12-001.
[iii]O_3 concentrations beyond 800 μg/m3 per 8 hour period do not have a corresponding air quality index.

In the second step, the individual air quality index scores are aggregated by picking the highest amongst the IAQI values. The AQI therefore reflects the pollutant with the highest IAQI.

$$AQI = \max_{p\, \in \{SO2, NO2, CO, O3, PM10, PM2.5\}} \{IAQI_p\}$$

An example to illustrate the above definition. Assume hypothetically that the US Embassy measured a concentration of PM2.5 of 34μg/m^3 over a 24 hour period. Then, the corresponding AQI would be

$$IAQI_{PM2.5} = \frac{100-50}{40.5-15.5}(34-15.5)+50 = 87$$

SI3: Robustness check by dropping BMEPB observations from non-particulate matter AQIs

To guard against concerns that the reported findings might be influenced by the minority of BMEPB observations from days for which particulate matter was not the main pollutant, Figures 1 and 3 are redrawn after excluding observations that are identified as based on neither PM_{10} nor $PM_{2.5}$ as the main pollutant on a given day.

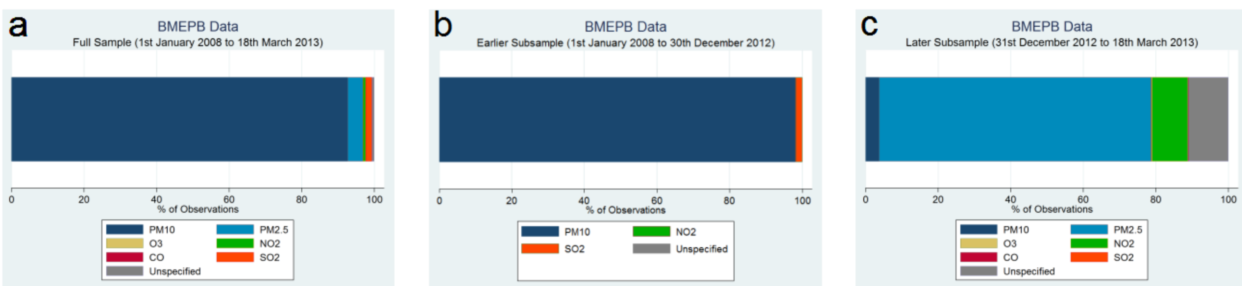


Fig. 7: Pollutant composition of the BMEPB dataset. *Share of main pollutants for all observations with an AQI beyond 50.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

Fig. 7 shows the share of each pollutant in the BMEPB data based on the main pollutant reported for observations with an AQI exceeding 50 for the full sample and the two subsamples. PM10 and PM2.5 drive 96.96% of observations for the full sample, 98.03% of the observations for the earlier subsample, and 78.72% of the observations in the later subsample.

Excluding these observations is a conservative approach because it is likely that some of the AQI observations that fail to specify a pollutant are in fact based on $PM_{2.5}$ or PM_{10}. Furthermore, dropping the observations identified as based on neither PM10 nor PM2.5 rather than replacing them with their nearest neighbouring observation based on $PM_{2.5}$ or PM_{10} is likely to decrease the chance of finding a close goodness-of-fit with Benford's Law as the greater part of these observations come from the lower end of the AQI distribution rather than the distribution as a whole.

As can be seen from Figures 1' and 3', the conclusions from analysis are not driven by the minority of observations identified as not from particulate matter. As in the main analysis, the anomaly of missing values at the *Blue Sky Day* threshold is present in the full sample (Fig. 1', a), comes entirely from the earlier subsample (Fig. 1', b) and vanishes in the later subsample (Fig. 1', c).

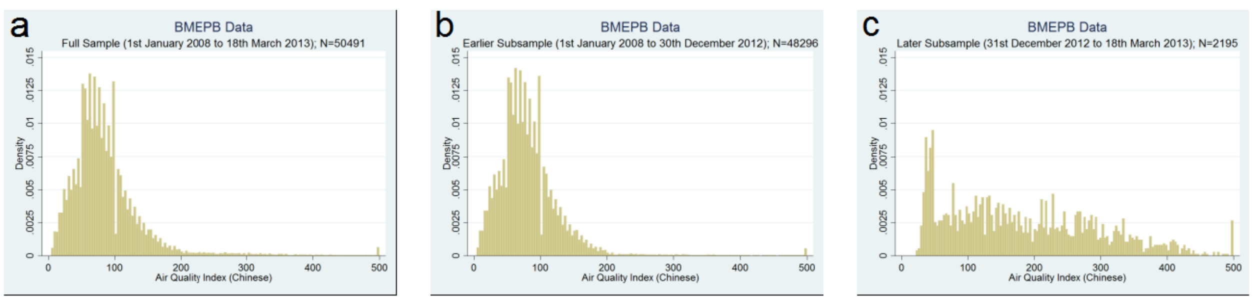


Fig. 1': Histogram of air pollution levels (Restricted BMEPB data). *Histograms of the BMEPB data after dropping observations identified as neither from PM_{10} nor $PM_{2.5}$.*

***(a)** full sample, **(b)** earlier subsample, **(c)** later subsample. AQI values of 100 and less constitute* Blue Sky Days.

The same pattern emerges from the analysis based on Benford's Law: the data fit Benford's Law poorly for the full sample and the earlier subsample (Fig. 3', a and b) and the fit markedly improves for the later subsample (Fig. 3', c).

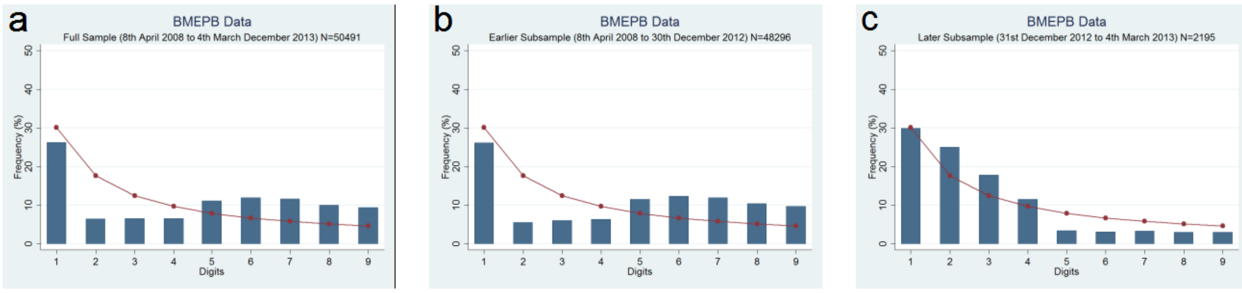


Fig. 3': Observed frequencies and Benford's Law (Restricted BMEPB data). *Observed frequencies for the first significant digits (blue bars) against the frequencies predicted by Benford's Law (red connected dots) for the BMEPB data after dropping observations identified as neither from* PM_{10} *nor from* $PM_{2.5}$. ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

SI4: Robustness check: Figure 2c using BMEPB AQI breakpoints

This section explains why the histograms of the AQIs measurements by the BMEPB and the US Embassy appear different when comparing Fig. 1c and Fig. 2c and shows their close comparability through converting the US Embassy data from the US AQI to the BMEPB AQI.

From *SI2*, recall that the US Embassy and the BMEPB use different breakpoints when converting PM2.5 concentrations into AQI values. As illustrated in Fig. 8 below, the US Embassy maps PM2.5 concentrations from 65.5 to 150.5 µg/m^3 to an AQI of 150 to 200, whereas the BMEPB only maps PM2.5 concentrations from 115 to 150 µg/m^3 to an AQI ranged 150 to 200. This explains why the US Embassy AQI observations look bunched between AQI values of 150 and 200 compared to the BMEPB data.

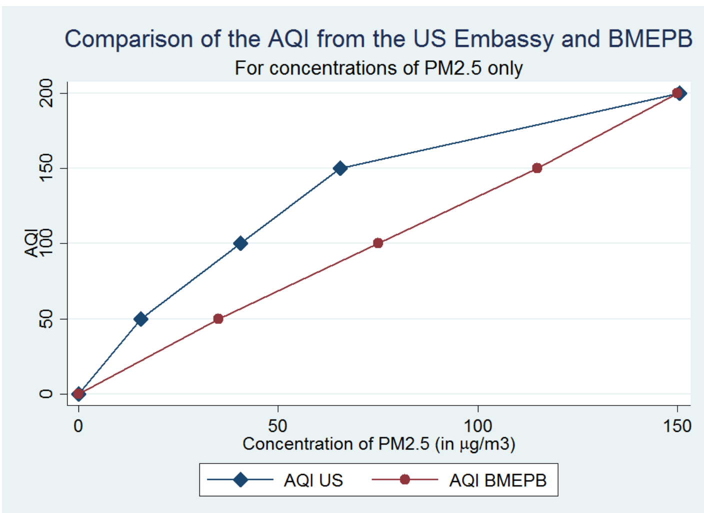
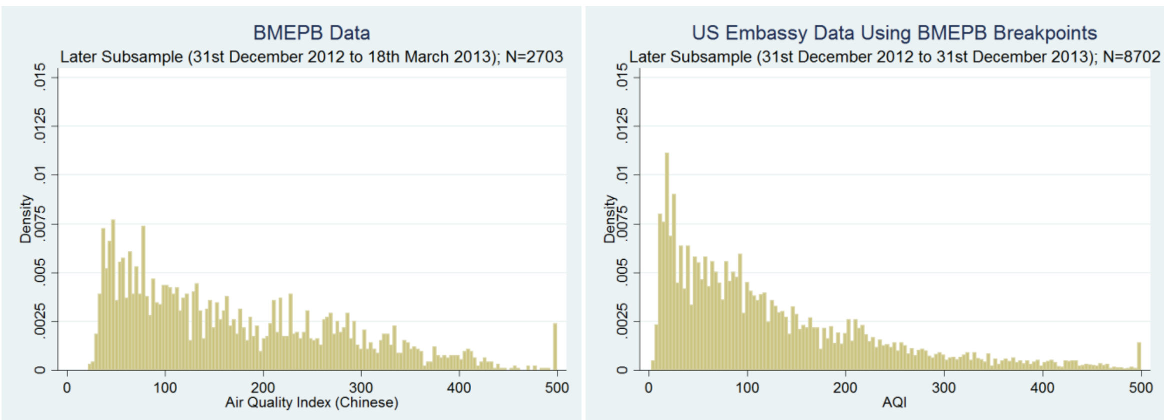


Fig. 8: Illustration of the different PM2.5 breakpoints. *This graph illustrates the different breakpoints used by the US Embassy and the BMEPB in the later subsample for AQI values up to 200. The figure is based on the AQI definitions reported in* SI2.

To illustrate this point, the left panel in the composite figure below reproduces the original BMEPB histogram for the later subsample (Fig.1, c) and compares it to the US Embassy data mapped into the Chinese AQI in the right panel (Fig. 2', c). The similarity between both histograms provides further evidence that the underlying raw data of both the BMEPB and the US Embassy reflect the same pollution processes and thus offer a good degree of comparability.



Figs. 1c and 2c': Histograms of air pollution levels (BMEPB data and US Embassy data).

Histograms of the BMEPB data ***(left panel)*** *and the US Embassy data converted into the AQI used by the BMEPB* ***(right panel)****.*