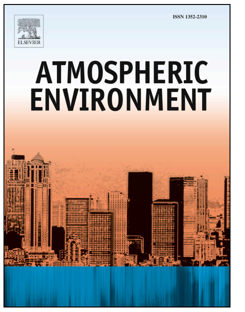
Accepted Manuscript



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

Thomas Stoerk

PII: S1352-2310(15)30633-6
DOI: 10.1016/j.atmosenv.2015.12.055
Reference: AEA 14368

To appear in: *Atmospheric Environment*

Received Date: 20 July 2015
Revised Date: 29 November 2015
Accepted Date: 21 December 2015

Please cite this article as: Stoerk, T., Statistical corruption in Beijing's air quality data has likely ended in 2012, *Atmospheric Environment* (2016), doi: 10.1016/j.atmosenv.2015.12.055.

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.

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, a particular air pollution level used to evaluate local officials, 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 proposes a new way of comparing air quality data via Benford's Law, a statistical regularity known to fit air pollution data. Using this method to compare the official data to the US Embassy data for the first time, 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. Additionally, I use aerosol optical density data to show the general applicability of this method of detecting likely misreporting in air pollution data.

1. Introduction:

Ambient air pollution is the cause of a major health crisis in current-day China. Pollution from ambient particulate matter is the fourth most important health burden in China [1], and coal heating alone has been attributed to causing a loss of 5 life years [2]. At the same time, political stakes are high: "Fix air pollution or 'bring your head'!" was the unequivocal message that Beijing's mayor received from the central government earlier this year [3]. 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 [4-6].

And the public might have a point: the 'leaders make numbers' phenomenon has a tradition in China [7]. Air quality data in China have been found to be suspicious in several cities until 2010 based on both a discontinuity test [8-9], and on a detailed analysis of histograms and the location of measurement stations for Beijing until 2007 [10]. These studies suggest that misreporting is prevalent around the politically important *Blue Sky Day* threshold, possibly due to the strict enforcement of promotion criteria [11-12]. The *Blue Sky Day* 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 central government has, for instance, used the number of *Blue Sky Days* to create a national ranking of air quality in Chinese cities [13], and as one of the indicators for a city to qualify as a national environmental protection model city [8]. The number of *Blue Sky Days* in Beijing has also been widely reported by the media due to its salience. By reporting a higher number of *Blue Sky Days*, local officials can therefore improve their chances for a positive evaluation, while at the same time appeasing the public.

The present research adds to the literature by employing a systematic approach to identifying misreporting that goes beyond the analysis of a discontinuity. This improves on the literature in two aspects: Firstly, my method allows to compare the similar, but slightly different US Embassy data to the BMEPB data for the first time and thus distinguish misreporting from temporary measures such as factory closures on borderline *Blue Sky Days*. Secondly, I can detect changes in misreporting over time.

My contribution additionally lies in a novel application of Benford's Law to compare two similar sets of air quality data to detect misreporting. While Benford's Law is known to fit air pollution in general, it is ex ante difficult to know whether it will fit a particular sample. In my application, I use a control dataset to establish whether Benford's Law fits the particular distribution of air pollution data. Given that the air quality datasets are sufficiently comparable, Benford's Law can then be used to test whether the suspicious dataset conforms to Benford's Law. This application of Benford's Law is particularly useful to detect likely misreporting even when air quality data come from different sources that are not comparable in levels. Furthermore, the method is easily scalable and extends naturally to other settings where two slightly different, yet comparable datasets measure air pollution. The increased availability of aerosol optical density (AOD) data, for example from satellite imagery, offers such settings. To illustrate this use, I show that my analysis holds also when using AOD data rather than the US Embassy data as the control dataset.

2. Data:

This research uses a new database of Beijing ambient air quality data that was collected and merged from three different sources (see *Appendix A* 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 provide 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. The third source is the AERONET measurement station in Beijing, who report aerosol optical density (AOD) measurements from 26th March 2008 to 31st December 2013, with several measurements during different times of each day.

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 *Appendix B* 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 *Appendix B*, 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, and as 98.03% until 30th December 2012 and as 78.72% thereafter (see *Appendix C*: Fig. C.1). 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 $PM_{2.5}$, and then correlating 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 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 [14] on ambient air quality in Beijing finds similarly high

[a] As a robustness check, *Appendix C* further 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}.

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 [15]; for Milan, Italy, see [16]).

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. Additionally, I use aerosol optical density (AOD) measurements by the AERONET measurement station in Beijing as a third data source to ensure that the conclusions regarding the true shape of the distribution of air quality in Beijing are not driven by the possibility of the US Embassy to also misreport, and to demonstrate that my method of analysis extends to AOD data. Due to their high accuracy, a typical use AERONET AOD data is verification of satellite-based AOD data [17-18]. Interpolated AOD at 550nm correlates well with US Embassy data (64.1%), while the correlation with the BMEPB data is reasonable, but significantly weaker (34.8%).

3. Methods:

To make the BMEPB and the US Embassy data 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 a 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 [19]. 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 both in economics [20-22] and, less frequently, for ambient air quality data [23-25]. While Benford's Law is known to fit air quality data in general[b], it is ex ante difficult to know whether it will fit a particular sample. This analysis therefore uses Benford's Law in a new way: 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. The datasets from the US Embassy and AERONET are then used as a control dataset to verify that air quality data for Beijing fit Benford's Law. 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] [27]. In the case of Benford's Law, there are 9 discrete categories corresponding to the digits 1 to 9. The χ^2 statistic is defined as:

[b] See *Appendix F* for an illustration of the fit of Benford's Law with a generic air pollution dataset.
[c] Other measures commonly used to track goodness-of-fit with Benford's Law include the Kolmogorov-Smirnov statistic, the Kuiper statistic, the d (distance) statistic and the statistic m (max) [26]. 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 (BMEPB), 0.90 (US Embassy); d (distance): 0.93 (BMEPB), 0.90 (US Embassy); m (max): 0.73 (BMEPB), 0.89 (US Embassy).

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

4. Results:

4.1 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 as a result of 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).

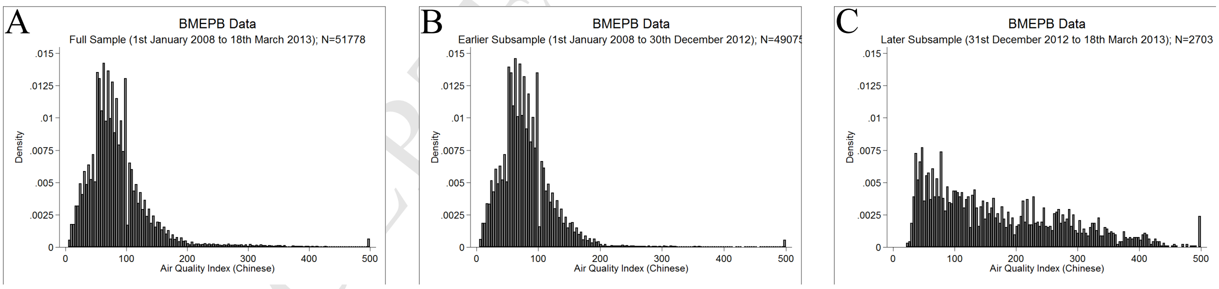


Fig. 1: Histogram of air pollution levels (BMEPB data). *Histograms of the BMEPB data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample. AQI values of 100 and less constitute* Blue Sky Days.

While misreporting seems to be a likely explanation, and was so interpreted by [10] who found the same anomaly until 2007, it is not the only one. Instead, Beijing officials defended themselves by attributing this anomaly to emergency measures such as temporary factory closures on borderline *Blue Sky Days* [28].

If emergency measures rather than misreporting explained the missing values in the BMEBP data, however, the US Embassy data should exhibit the same anomaly. The histograms for the US Embassy data show that this is not the case, neither for the full sample (Fig. 2, a) nor for the different subsamples (Fig 2, b and c). An analysis of the AERONET AOD data confirms these results (Fig. 3). This evidence suggests that misreporting by the BMEPB was prevalent until 2012 and then stopped.

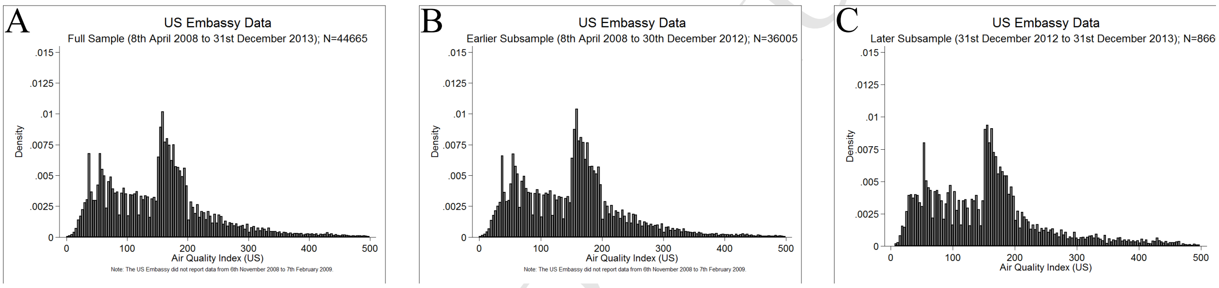


Fig. 2: Histogram of air pollution levels (US Embassy data). *Histograms of the US Embassy data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

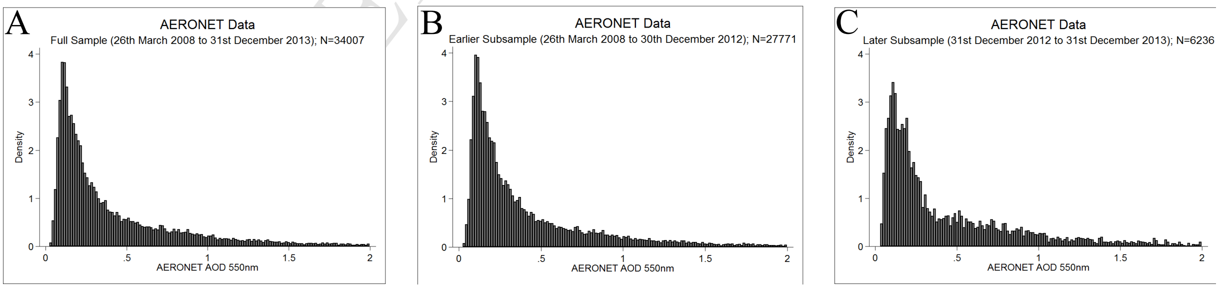


Fig. 3: Histogram of aerosol optical density (AERONET data). *Histograms of the AERONET ground-based AOD data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample. A small*

fraction of AERONET observations beyond 2 are not shown to allow for more detailed representation of observations in the relevant range.

4.2. Comparing the BMPEB and the US Embassy data

Yet, an important caveat is that the BMEPB and the US Embassy data 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 *Appendix B*)[d]. Benford's Law is therefore used to make both data sources comparable and find a direct indication for misreporting. As can be seen, the BMEPB data match Benford's Law poorly for the full sample (Fig. 4: a), suggesting misreporting. This misreporting seems to happen exclusively in the earlier subsample (Fig. 4: b). In the later subsample (Fig. 4: 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 5 shows that the fit of the US Embassy data is good and unchanged throughout all periods. Fig. 6 shows that the fit of the AERONET AOD data with Benford's Law is good and unchanged through all periods. The AERONET AOD data thus give the same conclusions as the US Embassy data, suggesting that the conclusions are not driven by a bias in the US Embassy data.

[d] 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). *Appendix D* provides further illustration of this point and shows the close comparability of Fig. 1c and Fig. 2c when using the same breakpoints.

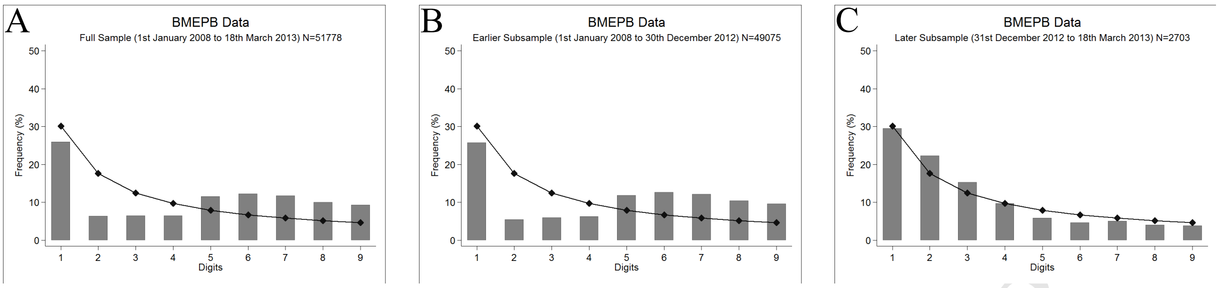


Fig. 4: Observed frequencies and Benford's Law (BMEPB data). *Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the BMEPB data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

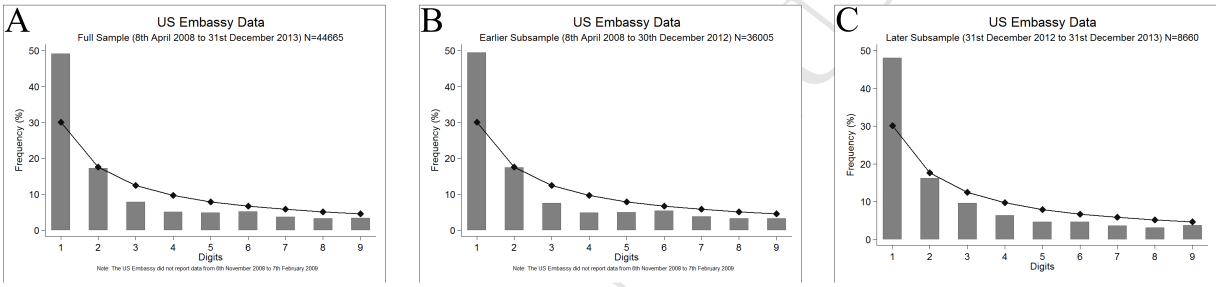


Fig. 5: Observed frequencies and Benford's Law (US Embassy data). *Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the US Embassy data.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

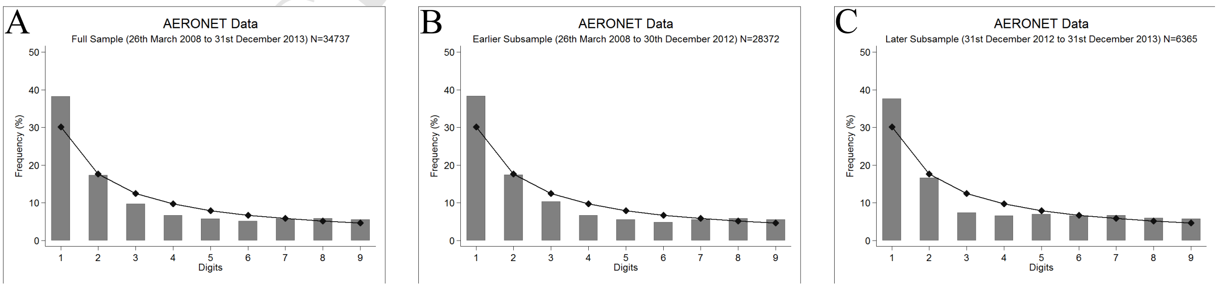


Fig. 6: Observed frequencies and Benford's Law (AERONET data). *Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law*

(black connected dots) for the AERONET ground-based AOD 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 7 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 8 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.

In summary, my findings suggest that emergency air control measures are unlikely to explain the anomaly in the BMEPB data, and that the BMEPB likely engaged in misreporting from 2008 to 2012. From 2013 onwards, however, misreporting appears to have stopped. At the same time, my analysis illustrates the usefulness of applying Benford's Law to detect misreporting in settings with independent measurements of air pollution.

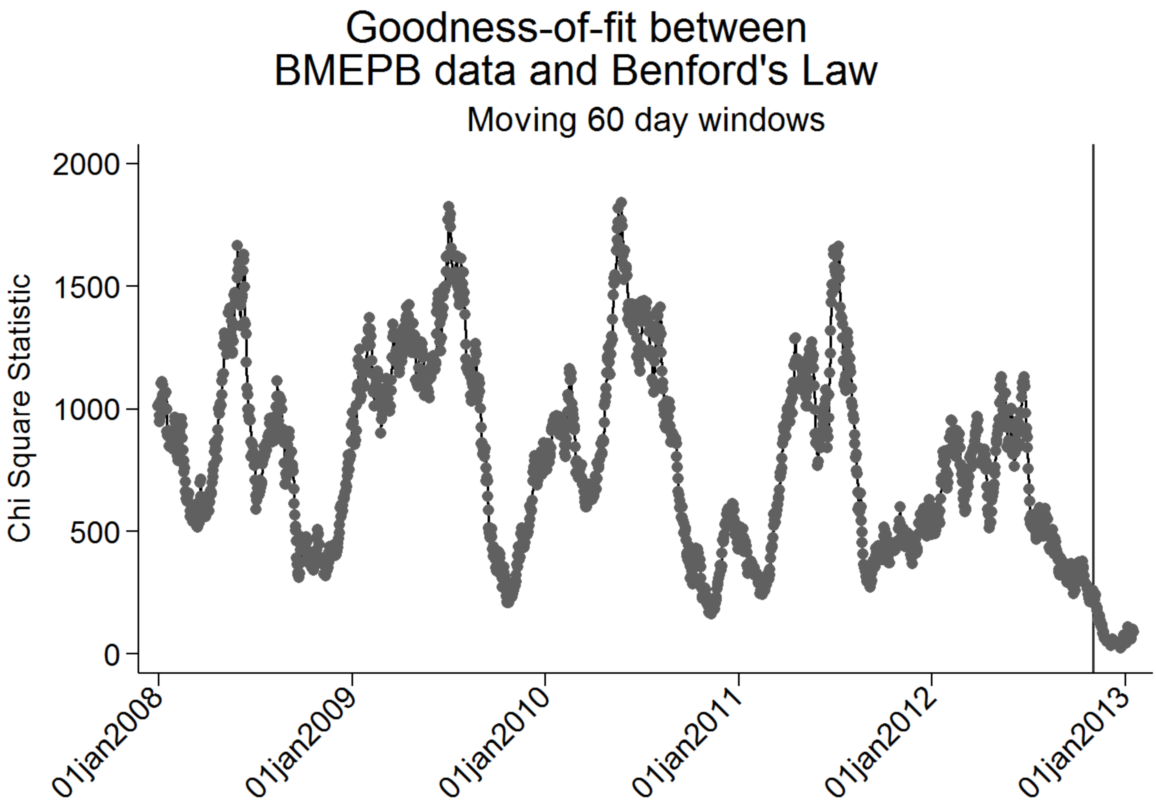


Fig. 7: Goodness-of-fit with Benford's Law over time (BMEPB data). *χ^2 statistic comparing the BMEPB data to Benford's Law. Computed over moving 60 day windows. The vertical black line marks the earliest time window during which the BMEPB started measuring $PM_{2.5}$ (from 31st December 2012 onwards).*

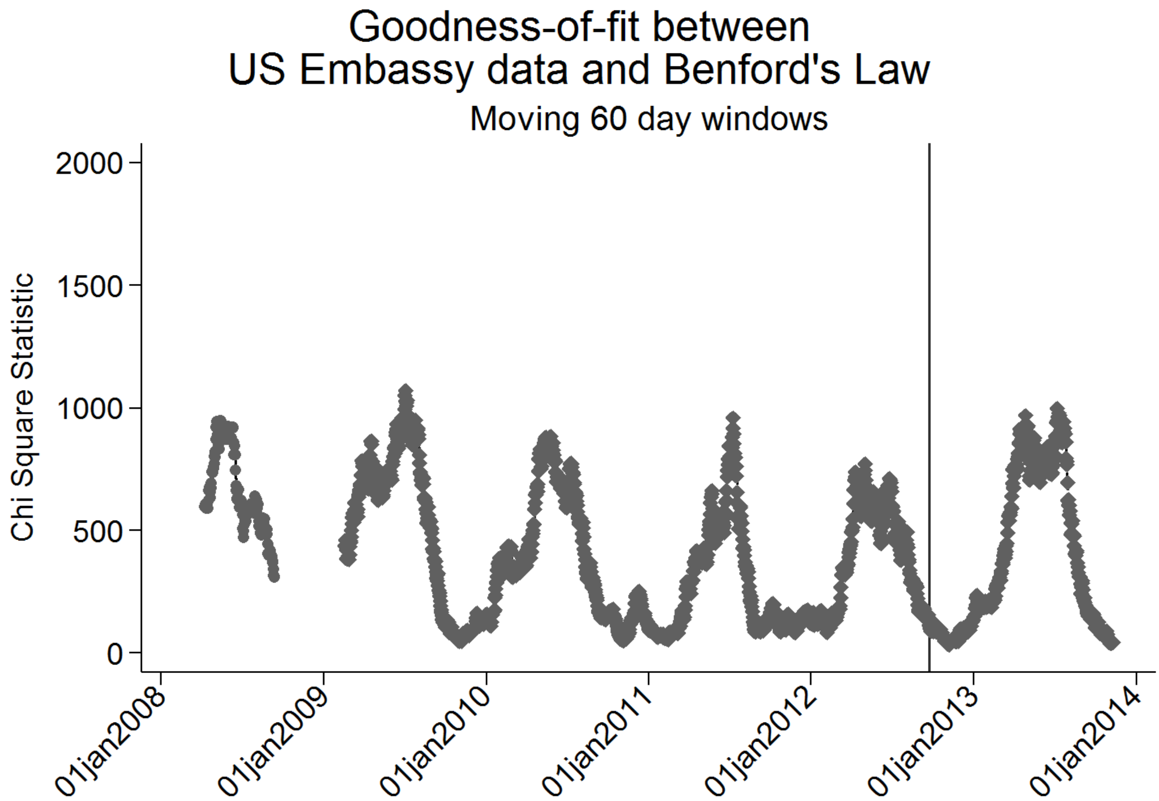


Fig. 8: Goodness-of-fit with Benford's Law over time (US Embassy data). *χ^2 statistic comparing the US Embassy data to Benford's Law. Computed over moving 60 day windows. The vertical black line marks the earliest time window during which the BMEPB started measuring $PM_{2.5}$ (from 31st December 2012 onwards). The US Embassy did not report data between 6th November 2008 and 7th February 2009.*

5. Discussion

The above results raise two questions. Firstly, is it possible that the air pollution control efforts taken during the Olympic Games 2008 might confound the results? The Olympics lasted from 8-24 August 2008, and emission control measures started on 20 July 2008 [29]. Subsequent evaluation based on both the official air pollution data and aerosol optimal depth from satellite imagery has shown that these control measures improved the average air pollution in the area of Beijing [30]. Traffic restrictions likely contributed to these improvements [31], but so did favourable meterological conditions in greater Beijing during the time of the Olympic Games [32-36]. The improvements in air quality lasted until about one year after the Olympic Games [30].

The time period of the Olympic Games 2008 and the associated improvements in air quality is part of my earlier subsample. In this subsample, I find likely misreporting in the BMPEB data. Given that my measure of misreporting is the scarcity of observations at the *Blue Sky Day* threshold and the excess of values just below it as well as the goodness of fit with Benford's Law, a change in the *mean* air pollution, as documented in the literature above, does not affect misreporting.

The Olympic Games 2008 drew attention to the problem of air pollution, however, possibly leading to a change in misreporting that is independent of the actual improvements. *Appendix E* shows that misreporting indeed improved for a short period in the immediate aftermath of the Olympics, only to return to its previous levels from 2009 onwards. Regarding the results of my analysis, the presence of the Olympics in my earlier subsample would go against me finding misreporting in the earlier timeperiod, however[e].

Secondly, given the disappearance of the anomaly in the reported air pollution measurements in the official data for Beijing at the end of 2012, a natural question to ask is what caused this change. To narrow down the reasons for the change, it is important to understand

[e] This consideration notwithstanding, *Appendix E* also repeats the analysis in Figures 1 and 3 for the BMEPB data for the earlier subsample excluding the data from the start of the control measures for the Olympic Games 2008 (20 July 2008) until one year after their end (24 August 2009). The results are unchanged.

whether this change is Beijing-specific or instead reflects a China-wide trend. To explore this question, I use air pollution data on two major Chinese cities that exhibit a *Blue Sky Day* anomaly similar to Beijing, but for whom the US Embassy does not measure air pollution. These cities are Anshan, in Liaoning Province, and Jining, Shandong Province. As shown in Figure E.1 in *Appendix E,* the mass of observations just below the cutoff is considerably smaller for both the comparison group and Beijing during most of the period of analysis (excepting the aftermath of the Olympic Games 2008). Misreporting seems to stop in Beijing at the end of the sample, however, while continuing in the two comparison cities. This indicates that the underlying reason for the likely end of misreporting in the BMEPB data is not the result of a China-wide trend.

Beijing differs in two crucial aspects. Firstly, the presence of the US Embassy has created pressure on misreporting throughout my sample period. Secondly, Beijing was amongst the first cities to be selected for trial of the new air pollution regulation HJ633-2012 at the end of 2012, when the change in misreporting occurred. While this regulation does not contain provisions related to the *Blue Sky Day* threshold, HJ633-012 introduced $PM_{2.5}$ into the list of pollutants to be measured, thus requiring the installation and calibration of new technology. It is plausible to presume that the process of setting up the new system might have led to increased inspections by the upper hierarchy. A press release from the Ministry of Environmental Protection dated 31st December 2012 confirms this, quoting MEP Chief Engineer Wan as specifically stressing quality control in the air quality monitoring as part of the implementation, and announcing intensified supervision and checks as part as the next step following the introduction of HJ633-2012 [37]. A situation of increased attention from the political hierarchy at the end of 2012, coupled with the presence of the US Embassy's measurements from 2008 onwards, might thus have led to the end of statistical corruption in Beijing's air quality data at the end of 2012.

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 [10] and other cities in China until 2010 [8-9] 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.

6. Concluding Remarks:

Political pressure to fix air pollution can result in "statistical corruption": government authorities respond by misreporting the desired data. 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. Using a novel way of assessing statistical misreporting via Benford's Law that makes both data sources comparable for the first time, this analysis suggests that the authorities in Beijing likely manipulated air quality from 2008 to 2012. My results thus suggest that the anomalies found in [8-10] are unlikely to be explained by emergency control measures as claimed by Beijing officials [28] but instead are likely to show true misreporting. My results furthermore show that misreporting of air pollution data in Beijing extended well beyond 2010 despite the recent increase in attention towards environmental policy making in China. From 2013 onwards, however, this has changed: despite ongoing suspicion regarding the quality of Chinese air pollution data [9], 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 [38], to improve health outcomes by reducing pollution [39], and to realize the co-benefits of reducing air pollution for climate change mitigation [40]. At the same time, research has shown that ambient air pollution in China is significantly impacted by government action and the structure of political incentives [41-42], 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.] Y. Chen, A. Ebenstein, M. Greenstone, H. Li, Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proc. Natl. Acad. Sci. U.S.A.* **10**(32), 12936-12941 (2013).

[3.] L. Chen, "Mayor of Beijing Told to Fix the Smog or 'Bring Your Head'" (Epoch Times, 18 March 2014; http://www.theepochtimes.com/n3/568271-mayor-of-beijing-told-fixthe-smogor-bring-your-head).

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

[5.] 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).

[6.] 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).

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

[8.] 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).

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

[10.] 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).

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

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

[13.] S.Q. Andrews, Seeing Through the Smog: Understanding the Limits of Chinese Air Pollution Reporting (China Environment Series 10: China Environment Forum, Wilson Center, 2008).

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

[15.] 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).

[16.] 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).

[17.] J. Li, B.E. Carlson, A.A. Lacis, How well do satellite AOD observations represent the spatial and temporal variability of $PM_{2.5}$ concentration for the United States?, *Atmos. Env.*, **102**, 260-273 (2015).

[18.] Y. Xue, X. He *et al.*, China Collection 2.0: The aerosol optical depth dataset from the synergetic retrieval of aerosol properties algorithm. *Atmos. Env.*, **95**, 45-58 (2014).

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

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

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

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

[23.] 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.

[24.] 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).

[25.] 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).

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

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

[28.] D. Cyranoski, "Beijing's clean air claims questioned" (*Nature* News, 17 June 2009; http://www.nature.com/news/2009/090617/full/news.2009.578.html).

[29.] J.P. Zhang, T. Zhu *et al.*, The impact of circulation patterns on regional transport pathways and air quality over Beijing and its surroundings. *Atmos. Chem. Phys*. **12**, 5031-5053 (2012).

[30.] Y. Chen, G.Z. Jin, N. Kumar, G. Shi, The promise of Beijing: Evaluating the impact of the 2008 Olympic Games on air quality. *J. Environ. Econ. Manage.* **66**, 424-443 (2013).

[31.] X. Wang, D. Westerdahl *et al.*, Evaluating the air quality impacts of the 2008 Beijing Olympic Games: On-road emission factors and black carbon profiles, *Atmos. Environ.* **43**, 4535-4543 (2009).

[32.] J.P. Zhang, T. Zhu *et al.*, The impact of circulation patterns on regional transport pathways and air quality over Beijing and its surroundings. *Atmos. Chem. Phys*. **12**, 5031-5053 (2012).

[33.] W. Wang, T. Primbs *et al.*, Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics, *Environ. Sci. Technol.* **43**, 5314-5320 (2009).

[34.] X. Tang, M. Shao, M. Hu, Z. Wang, and J. (J.) Zhang, Comment on "Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics", *Enviro. Sci. Technol.* **43**, 7588 (2009).

[35.] X. Yao, X. Xu, K. Sabaliauskas, and M. Fang, Comment on "Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics", *Enviro. Sci. Technol.* **43**, 7589 (2009).

[36.] S.L.M. Simonich, Response to Comments on "Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics", *Enviro. Sci. Technol.* **43**, 7590-7591 (2009).

[37.] People's Republic of China Ministry of Environmental Protection, "Air quality monitoring of the implementation of the first phase of the new standard task is completed in Beijing, Tianjin and other regions and 74 cities will take effect on January 1, 2013, to carry out surveillance" (31 December 2012, www.mep.gov.cn/zhxx/hjyw/201212/t20121231_244445.htm).

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

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

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

[41.] 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).

[42.] R. Xia, Pollution for Promotion (Unpublished Working Paper, 2014).

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. I thank Hong-Bin Chen and Philippe Goloub for establishing and maintaining the Beijing AERONET site and for their permission to use the AERONET data. 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.

Appendix to

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:

Appendix A: Data Sources

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

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

Appendix D: Robustness check: Figure 2 using BMEPB AQI breakpoints

Appendix E: Additional results for the Discussion section

Appendix F: Benford's Law for a generic air pollution dataset

Appendix A: Data sources

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.

AERONET (Aerosol Robotic Network) data interface for Beijing:
http://aeronet.gsfc.nasa.gov/cgi-in/type_one_station_opera_v2_new?site=Beijing&nachal=2&level=3&place_code=10
AERONET is a network of ground-based remote sensing aerosol measurement sites. The downloaded data are aerosol optical density (AOD) measurements for the Beijing site that is hosted by the Institute of Atmospheric Physics (PIs: Hongbin Chen and Philippe Goloub) for the time period from 26th March 2008 to 31st December 2013, with several measurements during different times of each day. For this study, I use the quality assured data (Level 2.0), which have a pre- and post-field calibration applied, are automatically cloud cleared, and manually inspected [i-ii]. There are no AOD measurements for around 20% of the days in this time period, with the missing days spaced equally across years. The reason for the missing values is twofold: these correspond to days during which the AOD measurements had to be discarded due to imprecise

measurements, or to days with excessive cloud coverage[f]. Both reasons explain roughly half of the missing values each.

The AERONET data for Beijing contain AOD measurements for the wavelengths 440nm and 675nm. I interpolate the data for AOD of a wavelength of 550nm, the wavelength that is typically used in the literature [iii], using the Ångström exponent, provided by AERONET, by rearranging formula (2) from [iv]:

$$\tau_{550} = \tau_{440}(\frac{550}{440})^{-\alpha},$$

where α is the Ångström exponent, and τ_λ is the AOD at wavelength λ: [g]

[f] Personal communication from PI Hongbin Chen of the AERONET Beijing site.

[g]To see this, rearrange $\alpha = -\frac{\ln(\frac{\tau_{550}}{\tau_{440}})}{\ln(\frac{550}{440})}$ to $-\alpha\ \ln\left(\frac{550}{440}\right) = \ \ln\left(\frac{\tau_{550}}{\tau_{440}}\right)$, which equals

$\exp[-\alpha\ \ln\left(\frac{550}{440}\right)] = \ \frac{\tau_{550}}{\tau_{440}}$, which in turn equals

τ_{440} [exp(ln $\left(\frac{550}{440}\right)$)] $(-\alpha) = \tau_{550}$ because $e^{a*b}=(e^a)^b$.

Appendix B: 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, $\mathrm{BP}_{\mathrm{High}}$ 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 $\mu g/m^3$).

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

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

<u>*Appendix C*: Robustness check by dropping BMEPB observations from non-particulate matter AQIs</u>

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.

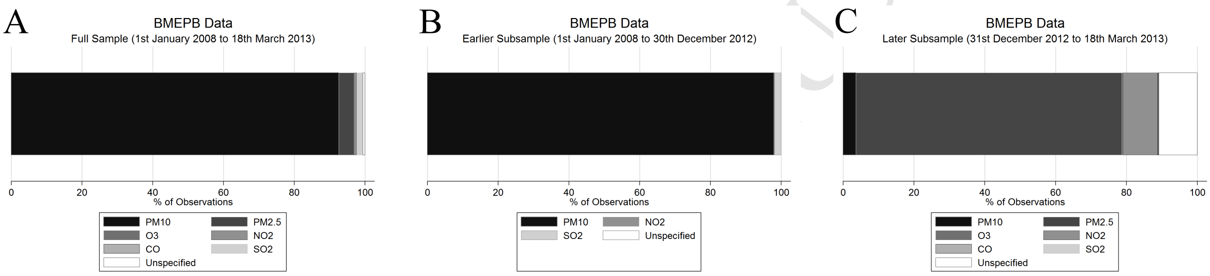


Fig. C.1: Pollutant composition of the BMEPB dataset. *Share of main pollutants for all observations with an AQI beyond 50.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

Figure C.1 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 4', 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).

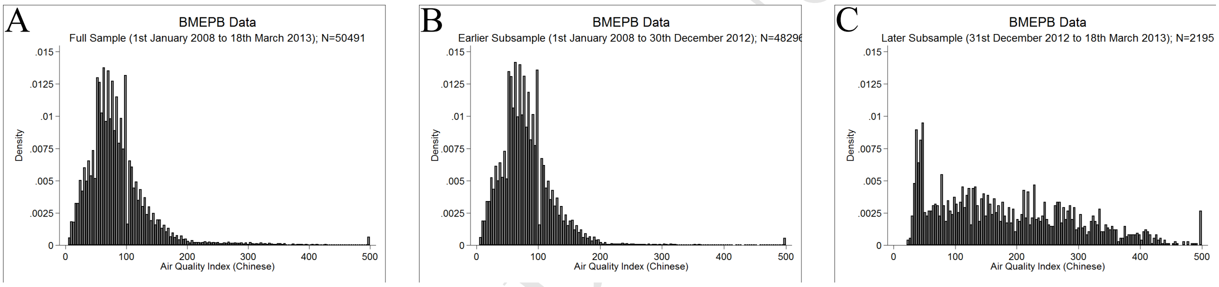


Fig. 1': Histogram of air pollution levels (Restricted BMEPB data). *Histograms of the BMEPB data after dropping observations identified as neither from PM_{10} nor $PM_{2.5}$.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample. AQI values of 100 and less constitute* 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. 4', a and b) and the fit markedly improves for the later subsample (Fig. 4', c).

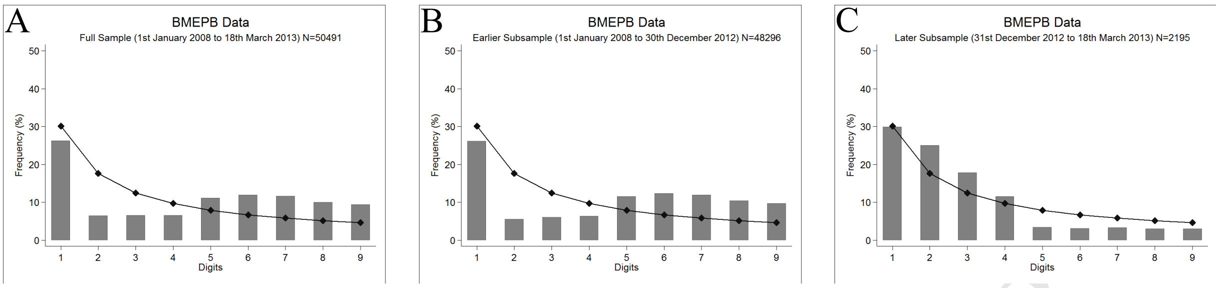


Fig. 4': Observed frequencies and Benford's Law (Restricted BMEPB data). *Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the BMEPB data after dropping observations identified as neither from PM_{10} nor from $PM_{2.5}$.* ***(a)*** *full sample,* ***(b)*** *earlier subsample,* ***(c)*** *later subsample.*

<u>*Appendix D*: Robustness check: Figure 2c using BMEPB AQI breakpoints</u>

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 *Appendix B*, recall that the US Embassy and the BMEPB use different breakpoints when converting PM2.5 concentrations into AQI values. As illustrated in Fig. D.1 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.

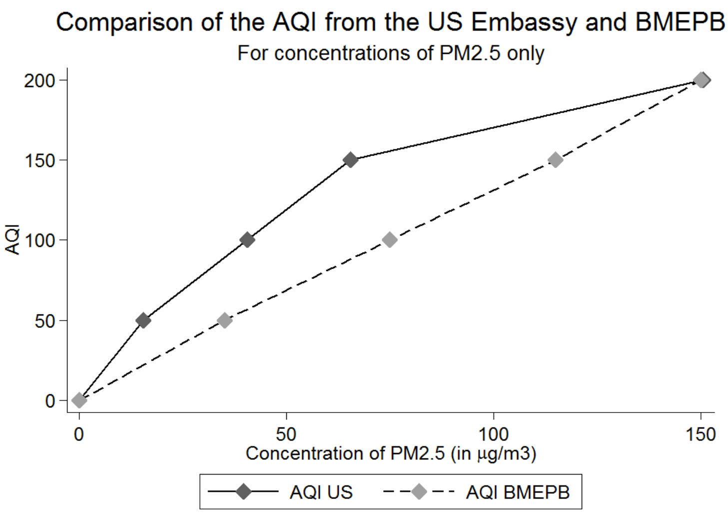
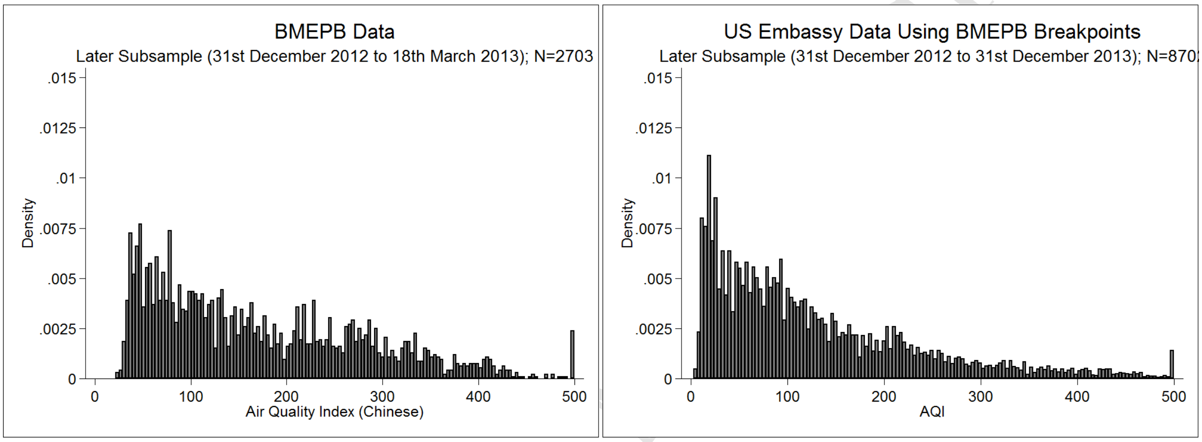


Fig. D.1: Illustration of the different PM2.5 breakpoints. *This graph illustrates the different breakpoints used by the US Embassy and the BMEPB in the later subsample for AQI values up to 200. The figure is based on the AQI definitions reported in* Appendix B.

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)**.*

Appendix E: Additional results for the Discussion section

This section contains additional results related to the Olympic Games 2008. Firstly, to track the development of the *Blue Sky Day* anomaly over time, I compute an index for the anomaly that I define as the frequency of observations just beyond the *Blue Sky Day* divided by the frequency of observations just below it. The reasoning behind this choice is that the likelihood of observing an AQI just beyond the threshold should be roughly equal to the likelihood of observing an AQI just below the schedule, giving the index a natural interpretation: index values far below one give an indication for mislabeling of values in this part of the distribution. Formally, I compute the anomaly index as the following ratio:

$$Anomaly\ Index = \frac{\#\ of\ observations\ for\ which\ 100 \leq AQI \leq 105}{\#\ of\ observations\ for\ which\ 95 \leq AQI < 100}$$

Fig. E.1 below graphs the time series of this index for both the pair of comparison cities and for Beijing. The data for Anshan and Jining are taken from the webpage of the Ministry for Environmental Protection of the People's Republic of China (http://english.mep.gov.cn/).

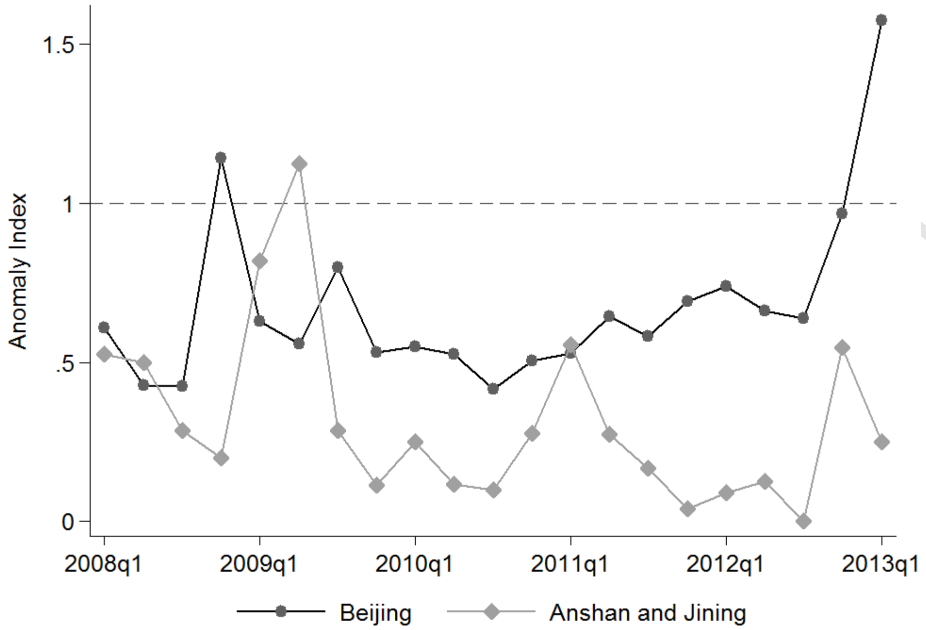
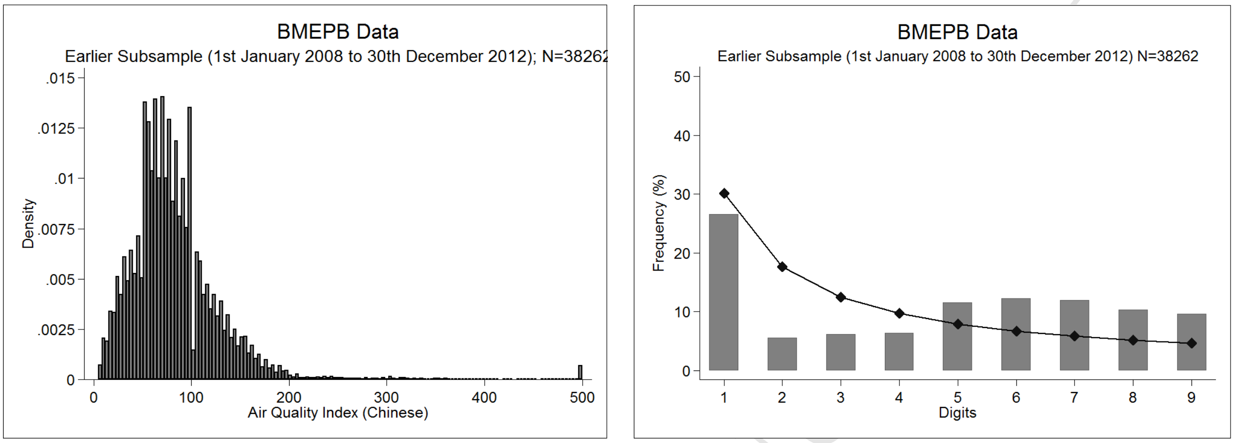


Fig. E.1: The *Blue Sky Day* Anomaly in Beijing and the Comparison Cities over time. *The anomaly index tracks the anomaly over time. One data point represents the average value of the anomaly index for a quarter of the natural year, for either Beijing or the mean of Anshan and Jining.*

Figure E.1 suggests that there was a temporary decrease in likely misreporting in Beijing following the Olympic Games until the end of 2008 , after which misreporting resumed to the previous level. If anything, this temporary end to misreporting makes it less likely for my analysis to detect misreporting for the earlier subsample as a whole.

To nonetheless show the robustness of the analysis to the choice of time period, Figures 1b' and 4b' show that my analysis is robust to dropping the entire period from the start of the control measures preceding the Olympic Games from 20 July 2008 onwards to the end of the temporary improvements in air quality to 24 August 2009, one year after the Olympics ended. The results are identical to the results from the earlier subsample based on the full dataset.



Figs. 1b' and 4b': Histogram of air pollution levels and observed frequencies and Benford's Law (Restricted BMEPB data). *Histogram* ***(left panel)*** *and comparison with Benford's Law* ***(right panel)*** *of the BMEPB data for the earlier subsample (1st January 2008 to 30th December 2012) after dropping observations from 20 July 2008 to 24 August 2009, the period of the Olympic Games 2008 and the related air quality improvements.*

Appendix F: Benford's Law for a generic air pollution dataset

This section shows that Benford's Law is a good description of a generic air pollution dataset. I use the historical dataset of air pollution in Europe. The source for these data is the full dataset from 1991-2012 from *AirBase*, a database on European air quality maintained by the European Environmental Agency (www.eea.europa.eu/themes/air/air-quality/map/airbase).

The fit of the air pollution data with Benford's Law is good both when combining the data from all European countries (Fig. F.1) and when looking at each country individually (Fig. F.2).

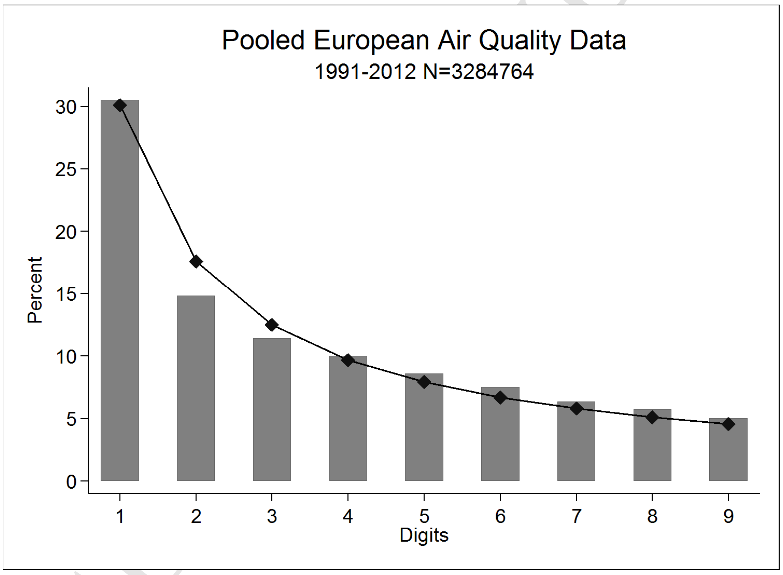


Fig. F.1: Observed frequencies and Benford's Law for a generic air pollution dataset.
Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the European air quality dataset Air Base, *pooled over all countries for the years 1991-2012.*

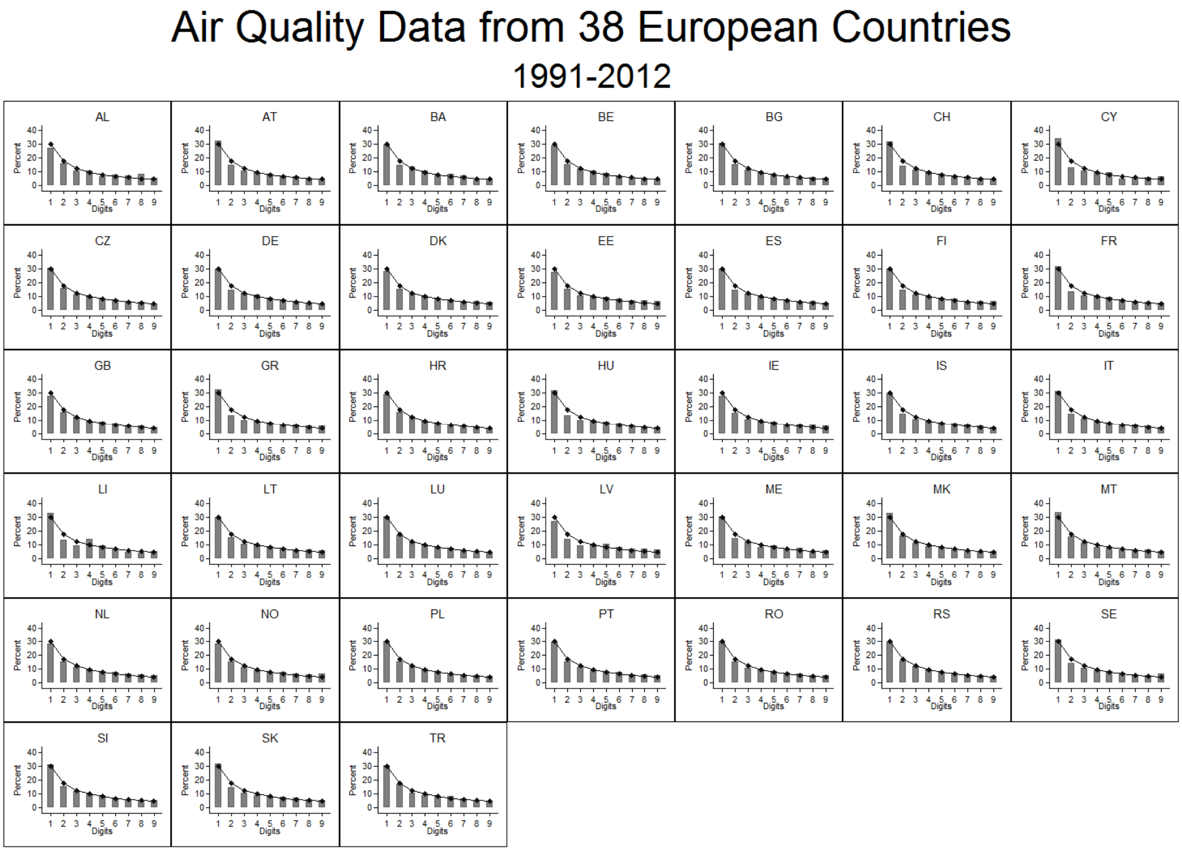


Fig. F.2: Observed frequencies and Benford's Law for a generic air pollution dataset.

Observed frequencies for the first significant digits (grey bars) against the frequencies predicted by Benford's Law (black connected dots) for the European air quality dataset Air Base*, shown for each individual countries for the years 1991-2012.*

References to the Appendix:

i. A. Smirnov, B.N. Holben, T.F. Eck, O. Dubovik, I. Slutsker, Cloud screening and quality control algorithms for the AERONET database, *Rem. Sens. Env.*, **73**, 337-349 (2000).

ii. B.N. Holben, T.F. Eck *et al.*, AERONET's Version 2.0 quality assurance criteria, *Proceedings of SPIE, the International Society for Optical Engineering*, **14**, 64080Q.1-64080Q.14 (2006).

iii. V. Cesnulyte, A.V. Lindfors *et al.*, Comparing ECMWF AOD with AERONET observations at visible and UV wavelenghts, *Atmos. Chem. Phys.*, **14**, 593-608 (2014).

iv. T.F. Eck, B.N. Holben *et al.*,Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *J. Geophys. Res.*, **104**(D24), 31,333-31,349 (1999).