ARTICLES
https://doi.org/10.1038/s41558-018-0394-4
nature climate change

Integrity of firms' emissions reporting in China's early carbon markets

Da Zhang[1,2], Qin Zhang[1], Shaozhou Qi[3,4], Jinpeng Huang[3], Valerie J. Karplus[2,5*] and Xiliang Zhang[1*]

The integrity of greenhouse gas emissions data is essential to assess progress towards countries' pledges under the Paris Agreement on climate change. Building credible systems for emissions measurement, reporting and verification is challenging, especially in developing countries. Using a unique dataset from two of China's pilot emissions trading systems (Beijing and Hubei), we compare firms' self-reported CO_2 emissions with emissions verified by third parties ('verifiers'). In Beijing, we find that the average discrepancy fell by statistically significant levels (from 17% in 2012 to 4% in 2014 and 2015), while in Hubei it started lower and showed a statistically insignificant decrease (from 6% in 2014 to 5% in 2015). We observe no evidence of deliberate misreporting in these two pilots, and show that improvements in firms' reporting capacity are associated with discrepancies of decreasing magnitude in Beijing. The results suggest that the administrative and firm capabilities required to support emissions trading systems in developing countries will require substantial time and effort to build.

Determining the impact of the Paris Agreement on climate change will require confidence in the ability of nations to accurately report emissions of greenhouse gases, most notably carbon dioxide (CO_2) from industrial sources. This task is not trivial: even in developed nations, reporting requirements are still relatively new, while in many developing countries, few managers have heard of CO_2 accounting. It is therefore of global interest to study the progress of developing domestic capabilities in measurement, reporting and verification (MRV) for CO_2 emissions, which can inform national paths to establishing effective systems. These paths will be vital to increase transparency and accountability—objectives that are now front and centre in the United Nations Framework Convention on Climate Change process.

The challenge of developing reliable domestic emissions accounts is not new, but it is often underestimated. The second Conference of Parties in 1996 adopted the first reporting guidance for countries not covered by the Kyoto Protocol[1]. Twenty years later, how to design these requirements flexibly, recognizing differences in capacity, remains a major sticking point in global climate negotiations.

The implementation of these systems has been another matter altogether. Most countries rely on data reported by emitting firms themselves. Studies in developing countries have revealed that misreporting of environmental data can be substantial[2,3]. On the one hand, firms may not have the capacity to accurately measure and report emissions, while on the other, data manipulation may be more widespread when governments lack the funding and expertise to monitor the accuracy of self-reported data. Given that these countries account for a growing share of total global emissions, even modest improvements in accounting practices can have a big impact on the quality of global emissions inventories. Here, we study the efforts of China—the world's largest CO_2 emitter and a developing nation—to establish CO_2 accounting in firms participating in the nation's emissions trading pilots. We find no evidence of deliberate misreporting, although we do observe that discrepancies between self-reported and verified emissions data decreased over time, consistent with improvements in emissions accounting capabilities.

Role of reliable data

The nature of the Paris Agreement, as a collection of national contributions designed to reduce countries' greenhouse gas emissions, has sharpened the accounting imperative. A total of 177 countries that have submitted their first nationally determined contributions will be asked to demonstrate progress towards their own self-imposed goals. Credibility will hinge on the quality of emissions reporting.

The questionable accuracy of CO_2 emissions data in China is a well-known problem. Several years ago, the gap between annualized national emissions totals and the national sum based on provincial statistics was around 20%—equivalent to the annual CO_2 emissions of Japan[4]. Recognizing data quality challenges, the national government regularly updates historical statistics on national energy use based on detailed economic surveys. A recent update revised annual coal use statistics upwards for the years 1998 to 2014, with the largest annual increase in coal consumption reaching 14%[5,6]. Inaccuracies in Chinese data have historically arisen because energy and emissions data are self-reported by companies and assembled by local governments, serving as a basis for higher levels of government to evaluate performance. These reports were subject to limited independent scrutiny[7].

The Chinese government has invested heavily in a national emissions trading system (ETS) as a cornerstone of its climate change mitigation strategy, building on its experience with regional pilot programmes starting in 2013. The national system is expected to gradually cover large entities that account for about half of total domestic CO_2 emissions, based on information released in connection with its official launch in December 2017. Given that an ETS turns CO_2 into a commodity that can be traded to meet covered

[1]Research Center for Contemporary Management, Institute of Energy, Environment and Economy, Tsinghua University, Beijing, China. [2]Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Climate Change and Energy Economics Study Center, Economics and Management School, Wuhan University, Wuhan, China. [4]Center of Hubei Cooperative Innovation for Emissions Trading System, Hubei University of Economics, Wuhan, China. [5]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA.
*e-mail: vkarplus@mit.edu; zhang_xl@tsinghua.edu.cn

firms' compliance obligations, a functioning system is contingent on confidence in the emissions levels that they report. Recognizing this, architects of the system have emphasized emissions verification.

Changes in reporting accuracy

China's ETS pilots covered 7% of national CO_2 emissions when they were established in 5 cities and 2 provinces starting in 2013[8]. These systems offer an opportunity to assess the design and performance of MRV mechanisms in a large developing country. We focus on the experience of two ETS pilots: Beijing (a relatively developed urban area) and Hubei (a less developed inland province with substantial heavy industry).

The Beijing and Hubei ETS pilots have several distinctive features that are important for the interpretation of our results. The Beijing ETS was among the first of the pilots to launch. Initially, 415 organizations participated in 2013 (the first compliance year). For brevity, we refer to these organizations as firms throughout the paper, although they include universities, hospitals and government agencies in addition to industrial firms. Historical annual emissions from these firms (from 2009 to 2012 by year) were reported and verified in 2013, and emissions for 2013 and later years were reported and verified one year later. Of these 415 firms, 403 stayed in the market for 4 consecutive years and reported complete data between 2012 and 2015. These 403 firms are included as the main sample in our study for Beijing. The 12 firms that prematurely exited the system only represented 2.9% of firms and 0.5% of total verified emissions in 2013, so they are very unlikely to affect our results. The Hubei ETS launched in 2014 and initially included 164 industrial firms.

Figure 1 provides an overview of both systems that incorporates data from 403 firms in Beijing and 164 firms in Hubei, using verified emissions data. Data collection and cleaning procedures are described in the Methods.

Firm eligibility criteria for inclusion in the ETS in Beijing and Hubei reflected differences in industry composition. For instance, the emissions threshold applied to determine firm eligibility was much lower in Beijing. In 2012, most of the firms in Beijing had emissions ranging from 10,000 to 1 million tons of CO_2 equivalent, while a substantial share of Hubei firms emitted more than 1 million tons of CO_2 equivalent. The share of firms in the industrial sector in the Beijing pilot ETS is much lower (43%), while all covered firms in Hubei are industrial firms. In Beijing, the power and heat sector is the dominant industrial emissions source, and aggregate emissions have decreased over time. Industrial emissions contributors are more diversified in Hubei, and there has not been a clear decreasing trend in aggregate emissions, suggesting that Hubei's industrialization is still in progress.

We show self-reported and third-party verified emissions in Fig. 2, based on the 403 firms in Beijing and 63 firms in Hubei for which we were able to collect both self-reported and verified emissions over the study period.

Discrepancies in reported CO_2 associated with a particular source (emissions type) are defined as the difference between self-reported and verified emissions for a firm, divided by the firm's total verified emissions (see equation (1) in the Methods). Several observations are noteworthy. First, the discrepancies grew smaller over time in Beijing, especially between compliance years 2013 and 2014. Meanwhile, the gap in Hubei started off smaller and showed a statistically insignificant drop in 2015 compared with 2014. Second, most of the box plots are centred at 0, and the medians overlap with the 25th and 75th percentiles for each emissions type, indicating that at least half of the self-reported emissions match verified emissions. Third, reported emissions from coal were initially subject to large deviations in both Beijing and Hubei. Fourth, and surprisingly, emissions associated with gas and electricity use, which are metred and therefore presumably easier to track, were also initially subject to large deviations.

Reasons for misreporting

People who are familiar with Chinese energy and environmental data will rightly worry whether the observed reduction in the discrepancy between self-reported and verified emissions is due to learning to circumvent rules or learning how to count correctly. While it is impossible to know for sure, we explore several explanations that may have contributed to the narrowing of the reported gap.

In order to test for evidence of possible data manipulation, we study how the magnitude of discrepancies changed over time and in response to a change from state assigned and funded verification to firm-funded verification with independent oversight. In 2016, after funding third-party verification costs and assigning verifiers to firms for three years, the Beijing government required firms to bear verification costs for the 2015 compliance year in order to free up government resources to support historical emissions verification for new programme entrants (the Hubei government covered verification costs in both years). Following this change, verifiers may have had stronger incentives to match the emissions totals reported by their paying clients.

Concerned about possible adverse selection and collusion, we set out to test for quantitative evidence. First, we test whether firms that change verifiers selected lenient or low-quality verifiers. We run a two-level fixed-effects regression to estimate firms' and verifiers' fixed effects where the dependent variable was the discrepancy in firms' self-reported CO_2 emissions (see the Methods for details). We focus on the fixed effects associated with a particular verifier, which captures whether the verifier is relatively strict or lenient. A strict verifier (lower fixed effect) would tend to provide firms with unfavourable results, meaning that on average, it found that firms' self-reported emissions higher than verified emissions in the year 2012, when benchmarks for grandfathering free emissions allowances were set, and lower than verified emissions in later years. Since there is almost no significant correlation between firm characteristics (for example, sector and ownership, with the exception of the marginally significant 'other manufacturing' sector dummy) and the magnitude of firms' fixed effects, as shown in Supplementary Table 7 (column 1) and Supplementary Fig. 3, we focus on verifiers' fixed effects, which are shown in Fig. 3.

We found heterogeneity in our fixed-effects measure of strictness among the 8 major verifiers that verified emissions for more than 50 firms (the market shares for these verifiers were 97% in 2012 and above 80% in 2015, as assigned by the government on the basis of their capacity and experience). For example, one verifier (V3, a relatively strict one) has a fixed effect of −0.06, meaning that if a firm switches its verifier from V1 (the one with the highest market share) to V3, the discrepancy will change by 6 percentage points on average in an unfavourable direction for the firm. Estimated verifier fixed effects are robust to the application of alternative size thresholds for excluding verifiers with small market shares, as shown in Supplementary Fig. 1 and Supplementary Table 2. In Supplementary Fig. 2, we show a negative correlation between verifiers' fixed effects and the average number of errors identified in firms' emissions reports (hereafter referred to as reporting errors, discussed later), which indicates that stricter verifiers tended to find more errors.

Next, we examined the relationship between verifiers' fixed effects, verifier changes and past discrepancies in self-reported emissions (see the Methods for details). Supplementary Table 3 shows that there is no significant correlation between lagged adjusted discrepancies and verifier changes. In other words, there is no evidence that (unfavourable) verification results caused a firm to change verifiers in the next year. We further find that firms did not change to relatively lenient verifiers in 2015 (column 3, Supplementary Table 4a), reducing concerns about adverse selection. We further tested for evidence of collusion, by examining whether the event of changing verifiers was, on average, significantly correlated with

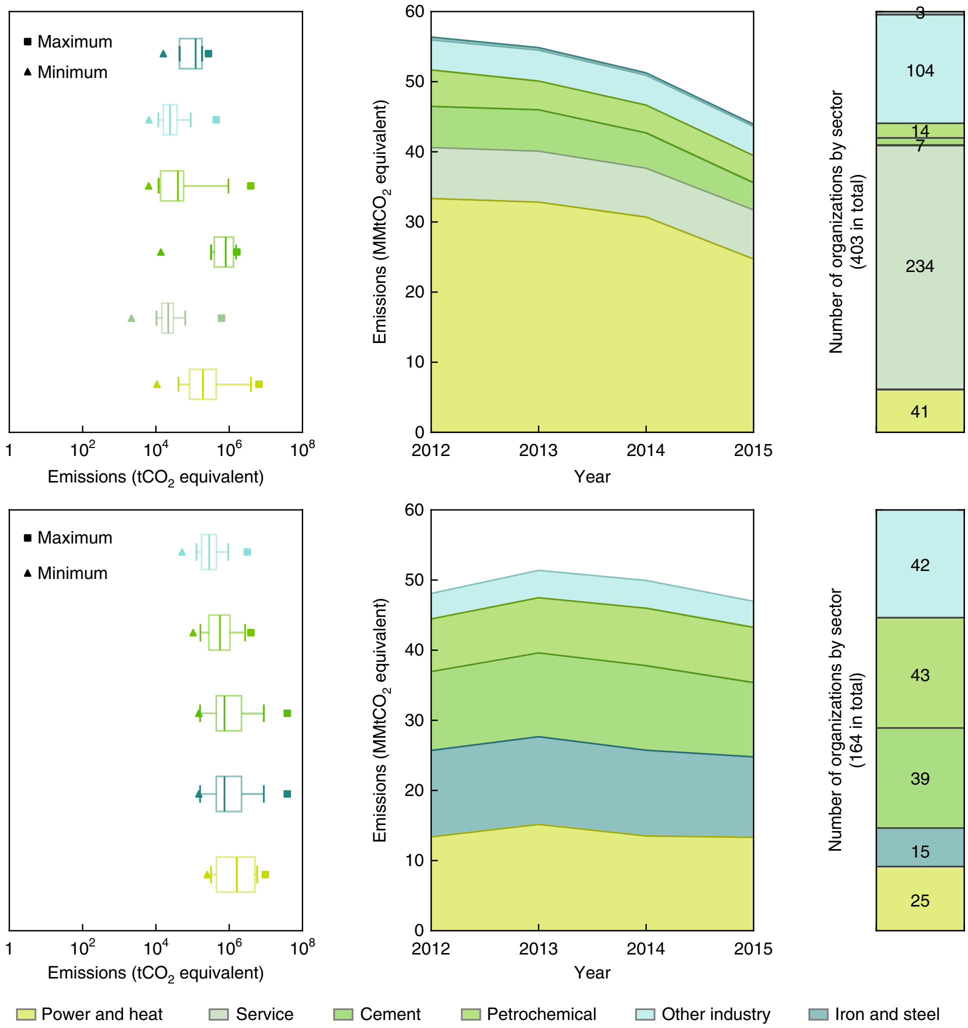

Fig. 1 | Summary statistics for the sample firms in Beijing and Hubei. Top, Beijing. Bottom, Hubei. The figure shows verified emissions based on firm-level data. On the left, box plots show the distribution of firm-level emissions in 2012 by sector. In each box plot, the lower and upper bounds of the boxes indicate the 25th and 75th percentiles, respectively. The thick bar in the middle represents the median. The whiskers extending on either side of the box indicate the 5th and 95th percentiles. The graphs in the centre show emissions by sector for 2012–2015. The graphs on the right show the breakdown of firms by sector.

any change in discrepancies. We found that verifier changes in 2015 were uncorrelated with changes in discrepancies, as shown in column 3 of Supplementary Table 5.

In Supplementary Table 6, we test for the presence of a significant correlation between the discrepancies and an indicator for whether a verifier was used by a firm in multiple years. The correlation is insignificant (column 1). We ran an additional regression with a dummy variable for whether or not a firm used the same verifier over the four years (2012–2015), and its effect was insignificant (column 2). Taken together, we found no evidence to support the hypothesis of adverse selection and collusion between firms and verifiers.

Second, substantial reductions in discrepancies predated the change in the source of funding for verification, an observation consistent with firm learning over time. Moreover, the favourability of the discrepancy to the firm did not seem to influence the magnitude of the reduction. In Table 1, we first show the results from a fixed-effects regression with the absolute value of adjusted discrepancies as the dependent variable. Here, the absolute value of discrepancies was chosen because we aimed to test whether or not the magnitude of all discrepancies shrank over time by assigning the same weight to two types of discrepancies (our goal here was not to estimate the relative strictness/leniency of the verifiers). The first column shows that year fixed effects are negative and significant, indicating that the magnitude of discrepancies between self-reported and verified emissions became smaller starting in 2013. The second column uses a time trend instead of year fixed effects and shows that the average annual reduction in the magnitude of discrepancies from 2013 to 2015 compared with 2012 was about 4% per year. In the third column, we focus only on the absolute value of favourable discrepancies as the dependent variable (unfavourable discrepancies are set to zero). The magnitude of favourable discrepancies also decreased in 2014 and 2015, suggesting no bias towards increasingly favourable results for firms over time. The fourth column uses the number of reporting errors as a dependent variable, and finds a positive and significant association. We computed the number of reporting errors in each firm's emissions report in each year by reviewing 1,612 confidential verification reports from the 403 covered firms in Beijing. In verification reports, verifiers are supposed to provide information about why there were emissions discrepancies, as well as mistakes firms made that did not comply with reporting standards. This result suggests that the more reporting errors firms made in their emissions reports, the larger the discrepancies that were found. Year fixed effects in this regression are still negative and significant. The fifth column shows that the year fixed effects in 2014 and 2015 are negative and significant if we treat the number of reporting errors in firms' emissions reports as the dependent

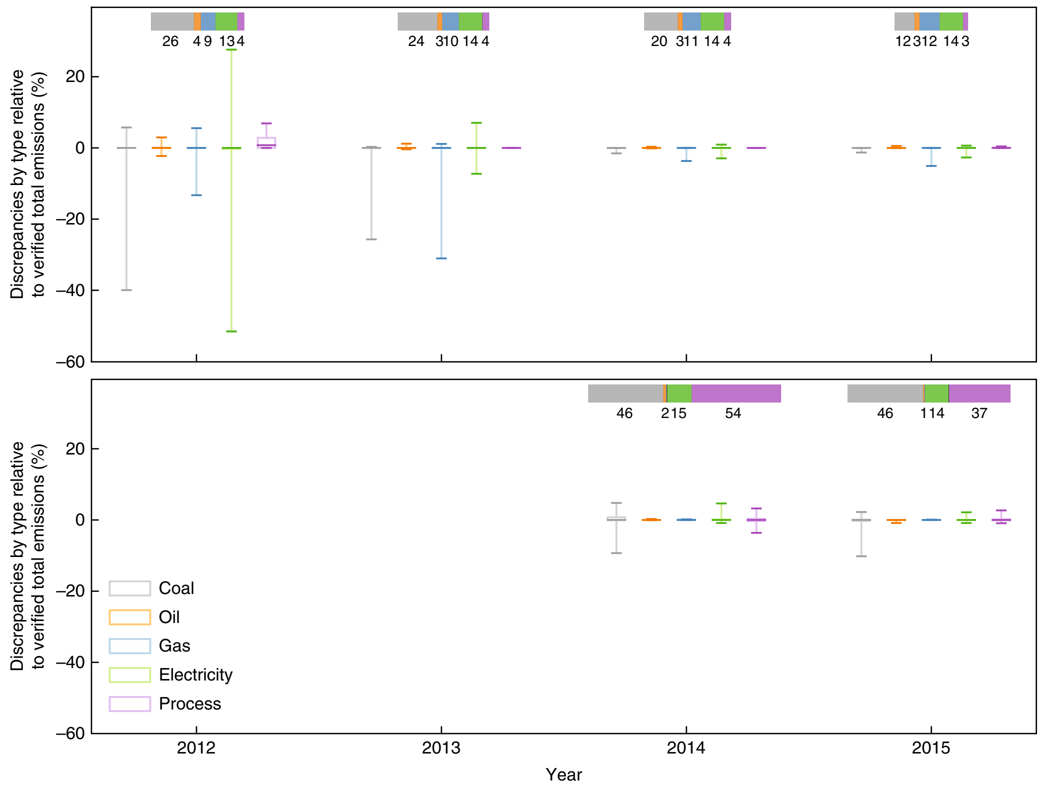

Fig. 2 | Discrepancy between self-reported and verified emissions by type, normalized by total verified emissions in Beijing and Hubei. Box plots show the discrepancy between self-reported and verified emissions by type, normalized by total verified emissions, in Beijing (top) and Hubei (bottom) in each year for which data are available. We exclude firms that have no emissions from the corresponding source category in the box plots. For example, some firms may not have emissions from natural gas consumption in a certain year, so discrepancies shown in the box plot for natural gas in that year do not include observations for these firms. In each box plot, the lower and upper bounds of the boxes indicate the 25th and 75th percentiles. In some cases, discrepancies are low in general, resulting in overlap between the 25th and 75th percentiles. The thick bar in the middle represents the median. The whiskers extending above and below the boxes span the 5th to 95th percentiles. Bars in each panel show the composition of aggregate emissions in this sample by emissions type (units for the numbers shown are $MMtCO_2$ equivalent).

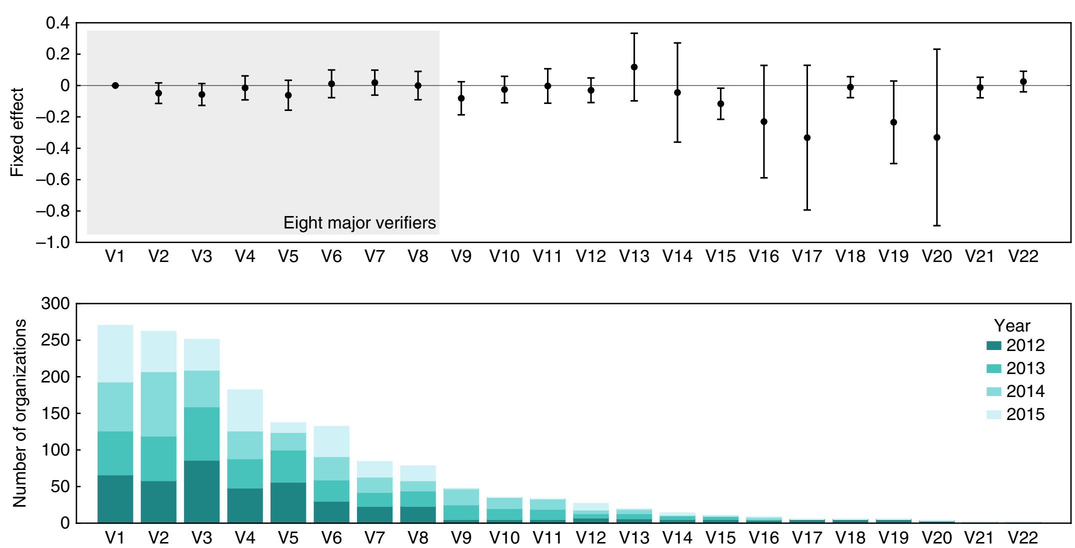

Fig. 3 | Estimated fixed effects of verifiers and the number of firms that each verified in Beijing from 2012 to 2015. Top panel, estimated verifier fixed effects. The fixed effect of the largest verifier (V1) was normalized to 0 (expected verified emissions for this verifier were 6.9% higher in 2012, 0.4% higher in 2013, 0.2% higher in 2014 and 0.8% lower in 2015 than the average self-reported emissions for a random firm selected for verification). Error bars represent the estimated standard errors of the verifier fixed effects. Bottom panel, number of organizations verified by each verifier by year (2012–2015).

Table 1 | Factors correlated with discrepancies and reporting errors in firms' self-reported emissions

	(1) Absolute value of adjusted discrepancies	(2) Absolute value of adjusted discrepancies	(3) Absolute value of favourable discrepancies	(4) Absolute value of adjusted discrepancies	(5) Number of reporting errors	(6) Absolute value of adjusted discrepancies
Number of reporting errors				0.021*** (0.004)		
Number of reporting errors due to inattention						0.018*** (0.006)
Number of reporting errors due to misunderstanding of the rules						0.024*** (0.005)
Number of reporting errors due to other reasons						0.011 (0.018)
Year dummy (year = 2013)	−0.077*** (0.017)		−0.002 (0.014)	−0.072*** (0.016)	−0.224 (0.141)	−0.072*** (0.016)
Year dummy (year = 2014)	−0.129*** (0.016)		−0.029** (0.012)	−0.096*** (0.016)	−1.549*** (0.148)	−0.096*** (0.016)
Year dummy (year = 2015)	−0.128*** (0.017)		−0.033*** (0.011)	−0.084*** (0.016)	−2.097*** (0.153)	−0.083*** (0.016)
Year trend		−0.043*** (0.005)				
Constant	0.128*** (0.022)	0.155*** (0.023)	0.029 (0.017)	0.070*** (0.023)	2.812*** (0.204)	0.070*** (0.022)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Verifier fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,609	1,609	1,609	1,609	1,609	1,609
R^2	0.101	0.094	0.038	0.136	0.305	0.137

*$P < 0.10$, **$P < 0.05$, ***$P < 0.01$. Heteroskedasticity-adjusted robust standard errors clustered at the firm level are shown in parentheses.

variable. A reduction in the number of reporting errors over time may help to explain the narrowing discrepancies.

Third, we analysed the origins of reporting errors by carefully reading all of the verification reports. Discrepancies between self-reported and verified emissions may have at least partially resulted from firms' inattention or misunderstanding of the accounting rules, while narrowing discrepancies over time reflect improvements along these dimensions. For example, these reporting errors include inaccurate meter readings, incorrect scope of emissions accounting, and missing energy use data for some equipment. We categorize the causes of reporting errors into three groups: inattention; misunderstanding the rules; and other reasons (see details in Supplementary Table 8). The average numbers of reporting errors per firm decreased from 3.7 to 1.9 from 2012 to 2015, with reporting errors from the first 2 groups (inattention and misunderstanding the rules) dropping most significantly, from 1.2 and 1.6 in 2012 to 0.7 and 0.7, respectively, in 2015. This is consistent with previous literature on the quality of reporting, which finds that respondents report more accurately after one or two survey rounds, as they may better understand the concepts being asked[9]. In the sixth column of Table 1, we show that the numbers of reporting errors due to inattention and misunderstanding the rules have a significant impact on the discrepancies between self-reported and verified emissions. One additional reporting error due to inattention or misunderstanding of the rules increases the predicted magnitude of discrepancy by 1.8 or 2.4% (compared with the mean value, 9.0%), respectively. Reporting errors that we categorize under 'other reasons' show no statistically significant relationship with the discrepancy. This may reflect deviations from recommended internal practices that bear no direct relationship to measurement accuracy.

Finally, the Beijing government hired experts to review all verification reports and rate the quality based on detailed guidelines (see Supplementary Table 9), starting from the compliance year 2015. Each report is rated separately by three independent experts, and firms that have reports with lower scores are subject to another 'backcheck' by an additional outside verifier referred as a 'fourth-party' verifier. If deemed neccessary, this process is fully funded by the Beijing government, with the objective of preserving verifier independence and thereby increasing the chances that deliberate manipulation or error would be discovered. Verifiers with low scores face state-imposed restrictions on market access in later years. We acquired quality scores for all verification reports issued for the compliance year 2015. In column 3 of Supplementary Table 4b–d, we show that firms that changed verifiers in 2015 did not change to low-quality verifiers. The Beijing government also found no evidence of data manipulation or collusion from these backchecks. Starting in 2017, the Beijing government requires firms to select a new third-party verifier every three years, to limit repeat interaction with firms.

To conclude, our argument that firms in Beijing have elevated their reporting capacity over time is consistent with all of the evidence we present, although it cannot be proven with certainty. Other factors, such as clearer political messaging on the seriousness of climate policy, may have also contributed to increased firm attention and falling discrepancies.

Lessons and future challenges

Regardless of the chosen policy approach, all countries with Paris pledges will be expected to develop credible CO_2 emissions accounting systems. This will not be easy, given heterogeneity in institutions and experiences. The increased consistency in self-reported and verified emissions in Beijing—where governance is relatively strong—is a reason to be optimistic. Importantly, we find no evidence that firms deliberately manipulated reported CO_2 emissions. However, our findings are also consistent with limited initial emissions management and reporting capabilities among a large share of participating firms. Moreover, the encouraging pattern of improvement in Beijing may not be easily replicated nationwide, or in other developing countries that establish an ETS for CO_2.

In Beijing, we find no evidence that the integrity of reporting changed after the government stopped funding the audits. In early years, government assignment of verifiers may have limited collusion, and the fourth-party backchecks may have successfully deterred misreporting in later years. In our setting, public funding for independent backchecks appears to be an effective way to encourage reporting integrity that is far less costly than assigning and funding all verifiers. However, this observation needs to be more comprehensively tested. So far, we are only able to observe one year of firms' self-selection of verifiers. Our findings are consistent with prior work that finds independent verification is an effective mechanisms for encouraging truthful reporting [10–12] and with the need for credible data for an emissions trading system to function properly[13].

Building credible compliance regimes will require both funds and attention from ETS architects. Funding could come from a combination of central and local government sources, including perhaps the revenues from ETS allowance auctioning. Insights from this study are relevant to a broad range of developing country settings, which will benefit from initiatives to improve the transparency and accuracy of emissions data (such as the Non-State Actor Zone for Climate Action[14]) and leave these countries better prepared for the international consultation and analysis process adopted by the United Nations Framework Convention on Climate Change. Multilateral entities, such as the Green Climate Fund or development banks, could supply resources to support verification costs associated with these programmes.

In short, effective governance is as important as dedicated policy, but the former necessitates engaging a more diverse set of actors in a country than would be required for establishing policy alone. Ensuring that China's provincial governments are as comprehensive as Beijing in laying foundations for a functional national ETS will be a major challenge. The stakes are high: without strong MRV capabilities, fictitious reduction credits generated in lenient provinces will undermine abatement in all provinces, especially those with stricter enforcement of reporting accuracy. Under these conditions, ETS initiatives will be far less effective, and fail to reveal the marginal cost of reducing CO_2. Strong law enforcement will be necessary to punish collusion between the verifiers and firms, should it be detected[15]. While no case of collusion or deliberate falsification of emissions data has surfaced in China's pilot CO_2 ETSs so far, the ability to detect such cases remains largely untested. In China and beyond, reporting systems that demonstrate accountability at the firm level will be critical to convince the world of the effectiveness of a post-Paris climate regime.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41558-018-0394-4.

Received: 17 May 2017; Accepted: 19 December 2018;
Published online: 28 January 2019

References

1. *Handbook on Measurement, Reporting and Verification for Developing Country Parties* (United Nations Framework Convention on Climate Change, 2014).
2. Duflo, E., Greenstone, M., Pande, R. & Ryan, N. Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India. *Q. J. Econ.* **128**, 1499–1545 (2013).
3. *Global Corruption Report: Climate Change* (ed. Transparency International) (Routledge, London, 2013).
4. Guan, D., Liu, Z., Lindner, S. & Hubacek, K. The gigatonne gap in China carbon dioxide inventories. *Nat. Clim. Change* **2**, 672–675 (2012).
5. *China Energy Statistical Yearbook 2016* (China Statistics Press, Beijing, 2016).
6. *Recent Statistical Revisions Suggest Higher Historical Coal Consumption in China* (EIA, 2015); https://www.eia.gov/todayinenergy/detail.php?id=22952
7. Ghanem, D. & Zhang, J. `Effortless perfection': Do Chinese cities manipulate air pollution data?. *J. Environ. Econ. Manage.* **68**, 203–225 (2014).
8. Zhang, D., Karplus, V. J., Cassisa, C. & Zhang, X. Emissions trading in China: progress and prospects. *Energy Policy* **75**, 9–16 (2014).
9. Mckenzie, D. Beyond baseline and follow-up: the case for more T in experiments. *J. Dev. Econ.* **99**, 210–221 (2012).
10. Dasgupta, S., Laplante, S., Mamingi, N. & Wang, H. Inspections, pollution prices, and environmental performance: evidence from China. *Ecol. Econ.* **36**, 487–498 (2001).
11. Rahman, D. But who will monitor the monitor? *Am. Econ. Rev.* **102**, 2767–2797 (2012).
12. Turnbull, J. C., Keller, E. D., Morris, M. W. & Wiltshire, R. W. Independent evaluation of point source fossil fuel CO_2 emissions to better than 10%. *Proc. Natl Acad. Sci. USA* **113**, 10287–10291 (2016).
13. Schmalensee, R. & Stavins, R. N. Lessons learned from three decades of experience with cap and trade. *Rev. Environ. Econ. Policy* **11**, 59–79 (2017).
14. Hsu, A., Cheng, Y., Weinfurter, A., Xu, K. & Yick, C. Track climate pledges of cities and companies. *Nature* **532**, 303–306 (2016).
15. Gray, W. B. & Shimshack, J. P. The effectiveness of environmental monitoring and enforcement: a review of the empirical evidence. *Rev. Environ. Econ. Policy* **5**, 3–24 (2011).

Acknowledgements

We acknowledge the support of the National Science Foundation of China (project number 71690244) and Ministry of Science and Technology, China (grant number 2017YFA0605304). We further thank the Energy Information Administration of the US Department of Energy for supporting this work under a cooperative agreement (grant number DE-EI0003030). This research received further support from an MIT Energy Initiative Seed Fund Grant and the MIT Joint Program on the Science and Policy of Global Change, which is funded through a consortium of industrial sponsors and federal grants, including the US Department of Energy (grant number DE-FG02-94ER61937). We are grateful to J. Caron (HEC Montreal), J. Jacoby (MIT), S. Li (Cornell University), B. Pizer (Duke University), R. Schmalensee (MIT), R. Stavins (Harvard University), S. Tanaka (Tufts University), D. Victor (UCSD) and X. Zhou (Harvard University) for helpful comments, and to L. Sun (McGill University) for advice on data visualization.

Author contributions

V.J.K., D.Z. and X.Z. conceived the research. S.Q. and X.Z. provided the data. J.H., V.J.K., D.Z. and Q.Z. performed the analysis. D.Z. and Q.Z. drew the figures. V.J.K., S.Q., D.Z. and X.Z. wrote the paper and contributed to the interpretation of the findings.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41558-018-0394-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to V.J.K. or X.Z.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Collection of firm CO_2 emissions data. Our analysis relies on two distinct sets of CO_2 emissions data: self-reported emissions provided by covered firms and third-party audited emissions for firms, which are submitted by verifiers. At the beginning of each compliance year, firms self-report their CO_2 emissions by energy type and industrial process in an emissions report. A third-party verifier then verifies the emissions report following guidelines published by the local government (an example from the Beijing City Government can be found in Supplementary Note 1). The verifier first reviews the emissions report and drafts a list of issues to be reviewed during the site visit. The verifier is required to prepare a plan for the site visit and communicates with the firm in advance to make sure that key firm personnel could be debriefed on site. Tasks that the verifier conducts on site involve a comprehensive review of compliance documentation, including verification of energy use records, coal procurement invoices, and gas and electricity bills. The verifier is also required to conduct some additional checks. For example, a verifier examines whether energy measurement devices installed at the firm operate normally and whether the firm's monitoring system for energy use is complete and effective. Also, for firms that use coal, a verifier evaluates whether the firm correctly followed the standard to measure the heating value (used to determine the emissions factor) and the weight of the coal. In Beijing, the vast majority of site visits (1,392 out of 1,605) last for 1 day, and the average duration is 1.2 days. However, 67 site visits last for 3 or more days. These visits were concentrated in firms that have many branches in the city, such as banks and supermarkets. After the site visit, the verifier notifies the firm of any errors and inconsistencies in the firm's emissions report, accompanied by a list of recommendations to correct them. The verifier also needs to compile a verification report (according to separate guidelines). The report must undergo internal technical review by another team of experts within the verifier's organization before submission to the government. The government then evaluates the verification reports and may perform any additional inquiries required to finalize and certify them. The Beijing ETS includes an additional step after verification for the year 2015: each report is reviewed and the quality is rated by three independent experts hired by the government, and firms with low-quality reports or deemed to need a second audit (for example, firms verified by a verifier that had newly entered the market) are visited by another verifier different from their original verifier (approximately one-third of the total firms)—a mechanism intended to encourage the original third-party verifier to perform an honest and thorough audit. These 'fourth-party' backchecks are paid by the Beijing government.

For Beijing, we acquire and analyse emissions reports and corresponding emissions verification reports for the 403 firms that stayed in the market for 4 consecutive years and reported complete data between 2012 and 2015. Although the Hubei ETS required firms to submit self-reported emissions, a significant number of firms did not submit them due to inattention, weak internal capacity or limited enforcement by the local government. Therefore, we were only able to collect and analyse data for the 63 Hubei firms that have both self-reported and verified emissions available for both compliance years, 2014 and 2015.

Definition of emissions discrepancies. In Fig. 2, the discrepancy $D_{i,e}$ in emissions type e for firm i is defined as the deviation of self-reported emissions $E^{S}_{i,e}$ from verified emissions $E^{V}_{i,e}$, divided by the firm's total verified emissions, as shown in equation (1):

$$D_{i,e}=\frac{E^{S}_{i,e}-E^{V}_{i,e}}{\sum_{e}E^{V}_{i,e}} \tag{1}$$

We then calculate the 5th, 25th, 50th, 75th and 95th percentiles of discrepancies $D_{i,e}$ for all of the firms in Beijing or Hubei by year, and visualize them as the box plots shown in Fig. 2.

Estimating verifiers' fixed effects. Our main regression model estimates both verifiers' and firms' fixed effects that correlate with discrepancies in firms' self-reported emissions for Beijing's 403 firms from 2012 to 2015. We use estimated verifiers' fixed effects as a basis for later analysis because this measure represents whether verifiers are relatively strict or lenient. We define a relatively strict verifier as one that on average delivers unfavourable verification results to firms. This means that the verifier on average finds firms' self-reported emissions higher than verified emissions in 2012 (emissions in 2012 were used to determine the amount of free emissions allowances allocated to firms in later years, so lower emissions in 2012 will lead to fewer emissions allowances, which is unfavourable for a firm) and self-reported emissions lower than verified emissions in 2013, 2014 and 2015 (an audit by a strict verifier would require a firm to submit more emissions allowances in these compliance years).

Similar regression model formulations have been widely used in labour and health economics to estimate the effects of units (for example, teachers, firms or hospitals; verifiers in our setting) on the individuals that are attached to the units (for example, students, employees or patients; firms in our setting). Econometric methods to consistently estimate large numbers of fixed effects in two dimensions have been readily developed[16,17] and applied in influential empirical work[18,19]. The model is shown in equation (2):

$$y_{it}=\varphi_{j(i,t)}+\theta_i+\delta_t+\varepsilon_{it} \tag{2}$$

where the dependent variable y_{it} is the adjusted discrepancy (approximately equal to the discrepancy $D_{i,e}$ defined in equation (1), as explained below) in firm i's self-reported emissions in year t (except for the year 2012, for which we use the negated adjusted discrepancies). The index $j(i,t)$ denotes the matching between firm i and verifier j in year t, so $\varphi_{j(i,t)}$ represents the fixed effect of verifier j. θ_i is the fixed effect of firm i, δ_t is the fixed effect of year t, and ε_{it} is the mean zero error term. The largest verifier is used as the reference (with its fixed effect set to zero) when estimating verifiers' fixed effects. We note that obtaining fixed effects precisely requires the assumption that matching status is not correlated with ε_{it}. This assumption is often violated when matching is endogenous[20,21] (for example, for verifier changes in 2015). We conducted multiple tests to examine our data for evidence of selection, and find no support for a role for strategic choice or adverse selection. We also find that firm characteristics exhibit a limited correlation with the magnitude of discrepancies, as shown in Supplementary Table 7 (column 2), except that larger firms had smaller discrepancies.

The dependent variable, y_{it}, is the hyperbolic tangent log transformation of the ratio of firm i's self-reported emissions to its verified emissions by verifier j in year t (except for the year 2012, for which we use the negated value of the transformation). We introduce this transformation to keep all of the observations in our regression and avoid arbitrarily dropping outliers. Although discrepancies in firms ' self-reported emissions are in general small (see the range of the 5th and 95th percentile in Fig. 2), there are some extreme values in each year showing that firms ' self-reported emissions are significantly higher (up to 1,000 times) or lower than their verified emissions, mainly due to mistakes in unit conversion. The hyperbolic tangent log transformation limits the extreme values of the dependent variable to the range −1 to 1 and provides a meaningful interpretation for most of the observations with the ratio of a firm 's self-reported emissions to its verified emissions close to 1, which is just the discrepancy between self-reported emissions and verified emissions normalized by verified emissions, consistent with our definitions in equation (1). The hyperbolic tangent log transformation has been adopted in the trade literature (for example, when calculating relative world market share, which is a ratio index for calculating the relative advantage or disadvantage of a certain country in a certain class of goods or services as evidenced by trade flow ratios with extreme values[22]). Similar hyperbolic (for example, the inverse hyperbolic sine) transformations have been applied in labour economics literature to transform variables with extreme values (for example, income or wealth[23,24]). Descriptive statistics of the adjusted discrepancies are shown in Supplementary Table 1. Both the mean and standard deviation of discrepancies fall in magnitude over time. A subset of firms changed their verifiers in each year. In 2013 and 2014, verifiers were allocated by the Beijing government, so these changes were exogenously decided beyond the firms' control. In 2015, firms were allowed to choose verifiers, so these changes were endogenously determined by the firms (although less than one-third of firms changed their verifiers). Using the regression specifications presented below, we find little evidence of adverse selection or collusion (see Supplementary Tables 3–6).

We implement a likelihood ratio test to ensure that including verifiers' fixed effects adds to the model's explanatory power. The likelihood test rejects the null hypothesis that the one-level fixed-effects model (with only a firm fixed effect) and two-level fixed-effects model (with both firm and verifier fixed effects) show no difference in explaining the variation in the dependent variable (adjusted discrepancies).

Tests for adverse selection. In Supplementary Table 3, we use a logit model to estimate the correlation between the lagged adjusted discrepancy and verifier changes. We run the model for each year from 2013 to 2015 separately. The model is:

$$y_{it}=\beta_t x_{i,t-1}+\varepsilon_{it},\ \text{where}\ t=2013, 2014\ \text{or}\ 2015 \tag{3}$$

The dependent variable y_{it} is a 0/1 variable that represents whether a firm i changes ($y_{it}=1$) or does not change ($y_{it}=0$) its verifier in year t, $x_{i,t-1}$ is the lagged adjusted discrepancy in firm i's self-reported emissions in year $t-1$, and ε_{it} is the mean zero error term.

In Supplementary Table 4, we estimate ordinary least squares (OLS) models that include a dummy variable for firms that changed their verifiers to explore the correlation between verifier changes and changes in verifiers' strictness (verifiers' fixed effects estimated earlier, shown in panel a) or quality (a dummy variable that indicates whether a verifier is a relative low-quality verifier, explained as below and in panels b–d) in 2013, 2014 and 2015. We focus on the regressions for 2015 (column 3 for all of the panels) because this is the first year in which we expect adverse selection could happen. The model is:

$$\Delta y_{j(i,t)t}=\beta D_i+\varepsilon_{it} \tag{4}$$

where the dependent variable $\Delta y_{j(i,t)t}$ is the change in the verifier's fixed effect or a low-quality indicator for the verifier that verified firm *i*'s self-reported emissions in year *t* (verifier *j*) compared to the verifier that verified firm *i* in the previous year. To obtain the low-quality indicator, we first calculate the average score of a verifier using all report quality scores for firms verified in 2015. Under different definitions, we treat a verifier as 'low quality' if its score is lower than the median, first tertile or first quartile of the scores from all of the verification reports. We use these three different definitions (the results shown in panels b–d) to test the robustness of the results. Since we only have access to the quality scores for the year of 2015, here we assume verifiers' quality is constant over time—a plausible assumption that we also applied when estimating verifier fixed effects for strictness. In addition, D_i is a dummy variable that equals 1 for firms that change their verifiers, and ε_{it} is the mean-zero error term. The coefficient of the verifier change dummy, β, will be significant and positive if a firm changes to a more lenient or lower-quality verifier.

Tests for collusion. In Supplementary Table 5, we run OLS models that include a dummy variable for firms that changed their verifiers to explore the correlation between verifier changes and discrepancies between self-reported and verified emissions for 2013, 2014 and 2015. The model is:

$$\Delta y_{it} = \beta D_i + \varepsilon_{it} \quad (5)$$

where the dependent variable Δy_{it} is the change of adjusted discrepancy in firm *i*'s self-reported emissions in year *t* relative to the previous year, D_i is the dummy variable that equals 1 for firms that change their verifiers, and ε_{it} is the mean zero-error term. If discrepancies are significantly correlated with the change, the coefficient of the verifier change dummy, β, will be significant.

In Supplementary Table 6, the model corresponding to the regression estimates presented in column 1 is:

$$y_{it} = \beta x_{j(i,t)t} + \varphi_{j(i,t)} + \theta_i + \delta_t + \varepsilon_{it} \quad (6)$$

where the dependent variable y_{it} is the adjusted discrepancy in firm *i*'s self-reported emissions in year *t*, $\varphi_{j(i,t)}$, θ_i and δ_t represent verifier fixed effects, firm fixed effects and year fixed effects, respectively, and $x_{j(i,t)t}$ represents how many times firm *i* has used verifier *j* by year *t*.

The model corresponding to column 2 of Supplementary Table 6 is:

$$y_{it} = \beta D_{j(i,t)} + \varphi_{j(i,t)} + \theta_i + \delta_t + \varepsilon_{it} \quad (7)$$

where $D_{j(i,t)}$ is a dummy variable that equals 1 only in 2015 if firm *i* has the same verifier in 2015 as it had in the previous three years and equals 0 otherwise. All other variables are similarly defined as in equation (6).

Decreasing discrepancies and errors. In Table 1, the model estimated in columns 1 and 3 is:

$$y_{it} = \varphi_{j(i,t)} + \theta_i + \delta_t + \varepsilon_{it} \quad (8)$$

where the dependent variable y_{it} is the absolute value of adjusted discrepancies or adjusted favourable discrepancies ($y_{it}=0$ if a discrepancy is unfavourable) in firm *i*'s self-reported emissions in year *t*, and all of the other variables are similarly defined as in equation (2).

The model estimated in column 2 is:

$$y_{it} = \varphi_{j(i,t)} + \theta_i + t + \varepsilon_{it} \quad (9)$$

where the dependent variable y_{it} is the absolute value of adjusted discrepancy in firm *i*'s self-reported emissions in year *t*, *t* is the time trend ($t=1$, 2, 3 and 4 for the years 2012, 2013, 2014 and 2015, respectively), and all the other variables are similarly defined as in equation (2).

The model estimated in columns 4 and 6 is:

$$y_{it} = \beta x_{it} + \varphi_{j(i,t)} + \theta_i + \delta_t + \varepsilon_{it} \quad (10)$$

where the dependent variable y_{it} is the absolute value of the adjusted discrepancy in firm *i*'s self-reported emissions in year *t*, where x_{it} is the number of reporting errors (total number or number by each category) in firm *i*'s emissions report in year *t*, and all of the other variables are similarly defined as in equation (2).

The model estimated in column 5 is:

$$y_{it} = \varphi_{j(i,t)} + \theta_i + \delta_t + \varepsilon_{it} \quad (11)$$

where the dependent variable y_{it} is the number of reporting errors in firm *i*'s emissions report in year *t*, and all of the other variables are similarly defined as in equation (2).

Role of firm characteristics. In Supplementary Table 7, the OLS regression model estimates the coefficients on different covariates (for example size, ownership type dummies and sector dummies) on firms' fixed effects or the magnitude of discrepancies. The model is:

$$y_i = \beta \mathbf{X} + \varepsilon_i \quad (12)$$

where the dependent variable y_i is the fixed effect of firm *i* estimated by regression equation (2) (shown in column 1) or the average absolute value of adjusted discrepancies of firm *i* (shown in column 2), $\mathbf{X}$ is the vector of covariates containing selected firm characteristics of firm *i*, and ε_i is the mean zero error term.

Data availability

The original self-reported and verified emissions datasets were developed and are maintained by offices within the Chinese government, and can be accessed only with official permission. We make available all scripts that replicate the results presented in this study at https://github.com/zhangda1021/ChinaMRV. Additional data that support the findings of this study are available from the corresponding authors upon request.

References

16. Abowd, J., Kramarz, F. & Margolis, D. N. High wage workers and high wage firms. *Econometrica* **67**, 251–333 (1999).
17. Davis, P. Estimating multi-way error components models with unbalanced data structures. *J. Econ.* **106**, 67–95 (2001).
18. Bertrand, M. & Schoar, A. Managing with style: the effect of managers on firm policies. *Q. J. Econ.* **118**, 1169–1208 (2003).
19. Card, D., Heining, J. & Kline, P. Workplace heterogeneity and the rise of West German wage inequality. *Q. J. Econ.* **128**, 967–1015 (2013).
20. Rothstein, J. Teacher quality in educational production: tracking, decay, and student achievement. *Q. J. Econ.* **125**, 175–214 (2010).
21. Fee, C. E., Hadlock, C. J. & Pierce, J. R. Managers with and without style: evidence using exogenous variation. *Rev. Financ. Stud.* **26**, 567–601 (2013).
22. Lang, F. P. & Ohr, R. *Openness and Development: Yearbook of Economic and Social Relations 1996* (Physica, Heidelberg, 1996).
23. Johnson, N. L. Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–176 (1949).
24. John, B. B., Lonnie, M. & Robb, A. L. Alternative transformations to handle extreme values of the dependent variable. *J. Am. Stat. Assoc.* **83**, 123–127 (1988).