Discussion Paper Series

IZA DP No. 18579

April 2026

The Power of Carbon Pricing: A Comment on Döbbeling-Hildebrandt et al. (2024) and Its Press Release

Elisa Piseddu
RWI and BNITM

Maximiliane Sievert
RWI

Abel Brodeur
University of Ottawa and IZA@LISER

Jörg Ankel-Peters
RWI and University of Passau

Julian Rose
RWI and LMU Munich

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18579 April 2026

The Power of Carbon Pricing: A Comment on Döbbeling-Hildebrandt et al. (2024) and Its Press Release*

Abstract

Döbbeling-Hildebrandt et al. (2024, DH2024) conduct a meta-analysis of the effectiveness of carbon pricing. DH2024's abstract concludes that 17 of 21 schemes evaluated in the literature produced substantial emissions reductions. A subsequent press release was headed: "Carbon pricing works". This comment revisits the meta-analysis and examines whether its empirical evidence supports the claims made in DH2024's abstract and, notably, the press release. We use DH2024's own approach of accounting for statistical power and potentially biased causal inference in the underlying studies. We show that when these criteria are applied simultaneously and conservatively – which we argue they should be – only nine effective schemes remain, eight in China and one regional US scheme. We emphasize that statistical power is a major issue in most carbon pricing evaluations, because most carbon prices are very low, leading to weak signal-to-noise ratios. We conclude that DH2024's policy implications and its press release therefore cannot be squared with its evidence base.

JEL classification

Q35, Q51, Q54

Keywords

carbon pricing, meta-analysis, statistical power

Corresponding author

Jörg Ankel-Peters
joerg.peters@rwi-essen.de

* *Acknowledgements*: This research received funding from Coefficient Giving.

The effectiveness of carbon pricing policies in reducing carbon emissions remains controversial (Ankel-Peters et al., 2025; Green, 2021; Lilliestam et al., 2021; Nesje et al., 2025; Van Den Bergh & Savin, 2021; Vrolijk & Sato, 2023). Döbbeling-Hildebrandt et al. (2024, henceforth DH2024) conduct a meta-analysis on the effectiveness of carbon pricing. The key finding, as reported in their abstract, is that the introduction of a carbon price "has yielded immediate and substantial emission reductions" for at least 17 of 21 schemes covered by the existing literature. The accompanying press release, titled "Carbon pricing works", states that the "conflict of beliefs over the core instrument of climate policy can be resolved with facts", based on the "17 real-world climate policies" in this meta-analysis (PIK, 2024).

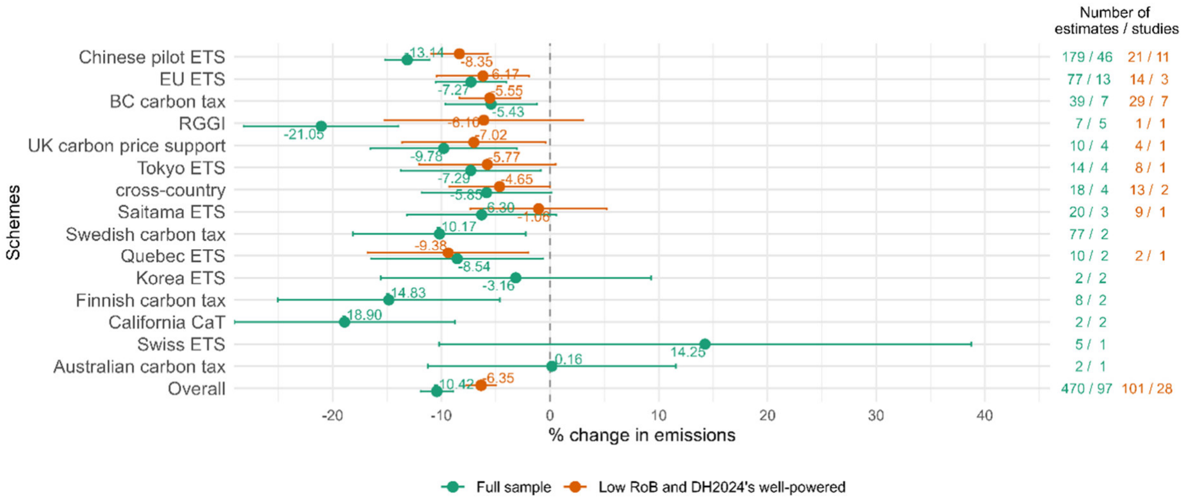
We show that the reviewed evidence is too weak to support these strong policy implications. The studies included in DH2024 meta-analysis (henceforth: the evidence base) are underpowered and at risk of bias – even more than accounted for in DH2024. In fact, statistical power is a breaking point for inference on DH2024's evidence base. We argue that this is inherent to the policies under evaluation: modest carbon prices as they are studied in DH2024, create only marginal reduction incentives, which are easily overshadowed by other concurrent policy measures, market fluctuations, and exogenous shocks. Statistically speaking, the signal-to-noise ratio is very low and, hence, it is almost impossible to isolate the causal impact from background variability. While not part of our analysis, it is important to note that the same noisiness also precludes the opposite claim that carbon pricing is *in*effective (Green, 2021).

We use two elements of DH2024's diligent meta-analysis to carve out our argument: their *risk of bias (RoB)* and *statistical power* assessment of the evidence base. DH2024 review 80 studies on 21 carbon pricing schemes in total, from which they extract 483 impact estimates.[1] For each impact estimate, DH2024 assess the RoB as low, medium or high, by examining the empirical design and the causal identification. Statistical power refers to the probability that a study detects a true effect if one exists. Underpowered studies cannot distinguish small true effect sizes from null effects. Crucially, underpowered studies are also more prone to exaggerating effect sizes. This is because the incentives to engage in questionable research practices, such

[1] DH2024 refer to these impact estimates as "effect sizes". We change this terminology to avoid confusion in this power-oriented comment.

as specification searches and p-hacking, which artificially inflate the significance and magnitude of a finding, may increase (Ioannidis et al., 2017).

Figure 1: Average emissions change – full sample vs. low RoB plus DH2024's well-powered sample

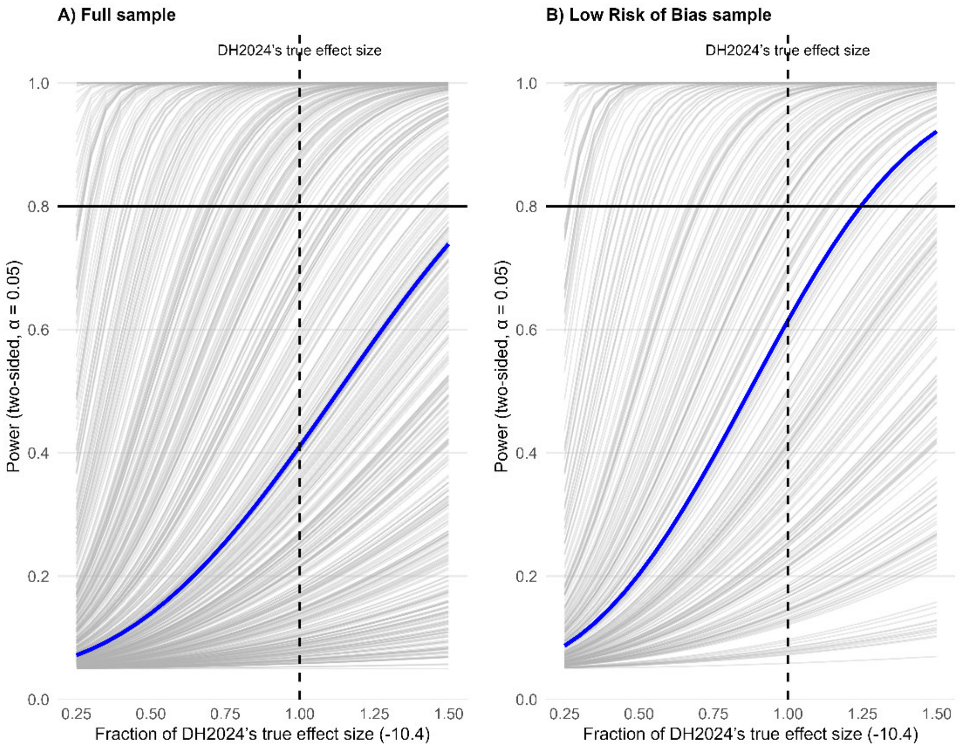


Notes: The figure shows weighted mean emission effects (with 95% confidence intervals) from multilevel random-effects models. Effects are expressed as percent changes in emissions relative to estimated counterfactual emissions, following DH2024's measurement conventions. The figure compares the full sample of impact estimates with the combined subsample including only estimates DH2024 rates as low RoB and well-powered, assuming DH2024's meta-average effect size as the true effect. Numeric labels indicate estimated mean effects. We follow the definitions of `Chinese Pilot ETS´ by DH2024, including all Chinese pilot schemes in a single variable.

DH2024 address both concerns, but separately. First, they exclude medium and high RoB estimates and analyze only estimates they rated as low RoB. Second, they analyze only impact estimates which they categorized as well-powered, based on an ex-post power analysis. We now apply both criteria simultaneously, which DH2024 do not do in their main paper. DH2024 report the combined analysis in their Supplementary Information. However, given the distorting potential of these two issues we contend that this combined subsample must be the basis for meta-analytical inference and especially policy conclusions. In Figure 1 we reproduce DH2024's main result figure ("full sample") and show our results for the combined subsample ("low RoB plus DH2024's well-powered"). This adjustment changes DH2024's abstract and press release summary to now 13 out of 21 schemes yielding "immediate and substantial emission reductions".[2]

[2] Note that the graph only displays 15 schemes as in DH2024. This is because the Chinese ETS pilot dummy pools all studies from the eight Chinese regional pilot schemes. DH2024 interpret the Chinese ETS pilot dummy as representing a total of up to eight schemes and report a total of 21 schemes, even though effectively they are only testing 15 scheme-dummies.

Figure 2: Sensitivity of estimate-level power to assumed true effect sizes

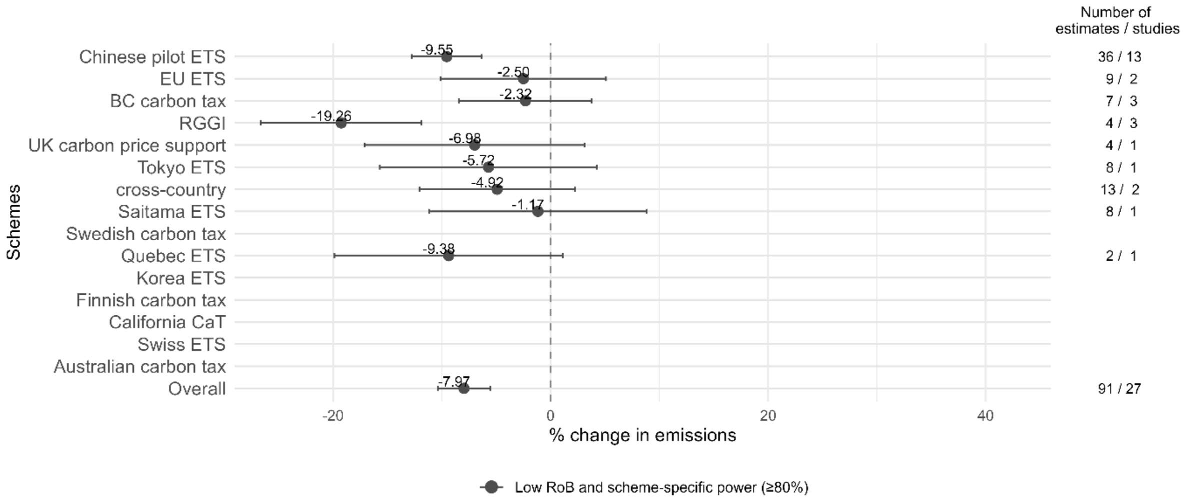


Notes: Power computed using the retrodesign framework with a two-sided significance level (α=0.05). The x-axis varies the true effect size as a fraction of DH2024's true effect size (10.4). Grey= individual impact estimates; blue=median. Horizontal black line=80% power threshold. Vertical dashed line=DH2024's true effect size.

Next, we probe into a pivotal assumption of ex-post power analysis, the assumed true effect size. DH2024 use the estimated meta-average effect size across all studies (-10.4%), a legitimate starting point. The meta-average effect size is also recommended by Ioannidis et al. (Ioannidis et al., 2017), as DH2024 highlight, but only when the reviewed literature can be considered unbiased. Ioannidis et al. (2017) explicitly warn, that "if there is selective reporting bias, […] [the meta-average] will tend to be too large and thereby give a power that is too high" (p243). In fact, DH2024 themselves document that selective reporting is likely, implying that their power calculations represent an optimistic upper bound. An alternative is to use a theory-based true effect size or to assess power under a range of plausible true effect sizes (Gelman & Carlin, 2014). From a theoretical perspective, we emphasize that an expected effect size of 10.4% is very unlikely, given the low carbon prices in the DH2024 evidence base: many schemes feature carbon prices below USD 10 per ton, with a median of USD 12 per ton. The mean carbon price is at USD 26 per ton because of three schemes above USD 100: Sweden and two building- and industry-focused schemes in Japan that target only large energy consumers.

Figure 2 demonstrates that DH2024's evidence base is highly sensitive to the assumed true effect size: we compute ex-post power for all estimates in DH2024's evidence base over a range of true effect sizes. The curves help to analyze what happens when other true effect sizes are assumed. Under DH2024's assumption of a 10.4% true effect, around 50% of estimates in the full sample have power of above 0.4 (blue line in Panel A) and only 34% of estimates are well-powered (power >0.8). Halving the assumed true effect size to 5.2% reduces the median ex-post power to 0.1. and only 14% of estimates are well-powered. The pattern is similar for the low RoB sample (Panel B). Here, the share of well-powered impact estimates shrinks from 42% for DH2024's 10.4% true effect size to 19% for the 5.2% true effect size. Note that even a 5.2% emission reduction is not particularly conservative, considering the median price in the sample of only USD 12 per ton.

Figure 3: Average emissions change – low RoB plus scheme-specific power sample



Notes: The figure shows weighted mean emission effects (with 95% confidence intervals) from multilevel random-effects models. Effects are expressed as percent changes in emissions relative to estimated counterfactual emissions, following DH2024's measurement conventions. The figure only uses the subsample of estimates assessed as low RoB and well-powered, assuming the scheme-specific meta-average effect size as the true effect. Numeric labels indicate estimated mean effects. We follow the definitions of `Chinese Pilot ETS´ by DH2024, including all Chinese pilot schemes in a single variable.

As a final step, we allow the assumed true effect size to vary across schemes, reflecting DH2024's argument throughout their paper that the different carbon pricing schemes differ fundamentally in their design characteristics. Therefore, expected effect sizes arguably vary substantially across schemes. One obvious heterogeneity is the price level, as mentioned above. Hence, assuming a single true effect size equal to the meta-average across all schemes ignores this heterogeneity across schemes. We propose using each *scheme's* observed metaeffect size, which is derived from all studies on that specific scheme, as a more accurate true effect size[3].

Figure 3 shows that results change considerably: only 91 out of 483 impact estimates are low RoB and well-powered. For 13 schemes, there is either no statistically significant impact estimate or no remaining impact estimate of sufficient quality. Hence, DH2024's abstract statement shrinks to only 9 schemes revealing "immediate and substantial emission reductions". What is more, one of these schemes is the US-American Regional Greenhouse Gas Initiative (RGGI) and the other 8 schemes are all Chinese regional ETS schemes. These schemes have particularly low external validity. According to DH2024, China had lower abatement costs and lower indirect carbon prices at baseline, "allowing for a higher marginal effect of the implementation of the ETS pilots in China" (DH2024, page 7). For the RGGI, DH2024 argue that "the policy implementation coincides with the shale gas boom, which drastically reduced the prices of natural gas in the USA […]. In face of these general price dynamics in the US energy sector, RGGI participating states reduced their emissions considerably stronger compared to non-regulated states while the carbon price was only US$ 3 on average."

To conclude, according to DH2024's evidence base, "carbon pricing works" in these two specific settings. However, using schemes in China and RGGI to derive authoritative policy implications for carbon pricing in other settings seems questionable.

[3] While we believe this approach is more accurate than the DH2024 approach, it is still an optimistic approach with respect to power because it equally relies on likely upward biased averages due to publication bias, as discussed above.

References

Ankel-Peters, J., Bensch, G., Dabadge, A., Munyehirwe, A., Rose, J., Sievert, M., Nshakira-Rukundo, E., & Lay, J. (2025). Tax carbon cautiously for sub-Saharan Africa. *Nature Climate Change*, *15*(1), 7–9. https://doi.org/10.1038/s41558-024-02213-w

Döbbeling-Hildebrandt, N., Miersch, K., Khanna, T. M., Bachelet, M., Bruns, S. B., Callaghan, M., Edenhofer, O., Flachsland, C., Forster, P. M., Kalkuhl, M., Koch, N., Lamb, W. F., Ohlendorf, N., Steckel, J. C., & Minx, J. C. (2024). Systematic review and meta-analysis of ex-post evaluations on the effectiveness of carbon pricing. *Nature Communications*, *15*(1), 4147. https://doi.org/10.1038/s41467-024-48512-w

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Green, J. F. (2021). Does carbon pricing reduce emissions? A review of ex-post analyses. *Environmental Research Letters*, *16*(4), 043004. https://doi.org/10.1088/1748-9326/abdae9

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, *127*(605), F236–F265. https://doi.org/10.1111/ecoj.12461

Lilliestam, J., Patt, A., & Bersalli, G. (2021). The effect of carbon pricing on technological change for full energy decarbonization: A review of empirical ex-post evidence. *WIREs Climate Change*, *12*(1). https://doi.org/10.1002/wcc.681

Nesje, F., Schmidt, R. C., & Drupp, M. A. (2025). Designing Carbon Pricing Policies Across the Globe. *Environmental and Resource Economics*. https://doi.org/10.1007/s10640-025-01036-3

PIK. (2024). *Carbon Pricing works—Major meta-study shows findings more comprehensively than ever*. Pressrelease. https://www.pik-potsdam.de/en/institute/departments/climate-economics-and-policy/mcc-archive/mcc-news-3187

Van Den Bergh, J., & Savin, I. (2021). Impact of Carbon Pricing on Low-Carbon Innovation and Deep Decarbonisation: Controversies and Path Forward. *Environmental and Resource Economics*, *80*(4), 705–715. https://doi.org/10.1007/s10640-021-00594-6

Vrolijk, K., & Sato, M. (2023). Quasi-Experimental Evidence on Carbon Pricing. *The World Bank Research Observer*, *38*(2), 213–248. https://doi.org/10.1093/wbro/lkad001