DEPARTMENT OF ECONOMICS AND FINANCE

SCHOOL OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

Did China's Pilot Emissions Trading Scheme Reduce CO2 Emissions? Evidence from a Meta-Analysis

Yanxia Yu
Chenfei Qu
Tom Coupé
Mathilda Featherston-Lardeux
Andreas Loeschel
Arne R. Weiss
Da Zhang

WORKING PAPER

No. 4/2025

Department of Economics and Finance
UC Business School
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 4/2025

Did China's Pilot Emissions Trading Scheme Reduce CO2 Emissions? Evidence from a Meta-Analysis

Yanxia Yu[1,2]
Chenfei Qu[3]
Tom Coupé[1]
Mathilda Featherston-Lardeux[4]
Andreas Loeschel[5]
Arne R. Weiss[6,7]
Da Zhang[3]

March 2025

Abstract: How effective are Emission Trading Schemes (ETS) in reducing carbon emissions? This paper addresses this question using a meta-analysis of 80 primary studies and 732 effect size estimates assessing the impact of China's pilot ETS, introduced in seven provinces and cities in 2011. Our findings suggest that China's pilot ETS led to small to medium-sized reductions in CO_2 emissions, with estimated reductions ranging from 5.5% to 16.6% compared to a no-ETS counterfactual. Meta-regression and Bayesian Model Averaging analyses further show that the estimated impacts of the ETS vary with the empirical methods employed and the control variables used. The range of estimates we obtain from synthesizing the literature provides policy-makers with an evidence-based expectation of the potential scale of emissions reductions in the initial phase of developing an ETS.

Keywords: Meta-analysis, Emission trading scheme

JEL Classifications: C18, Q58

Data Availability Statement: All the data and codes need to generate the results in this paper are posted at https://osf.io/7x924/.

Acknowledgements: We gratefully acknowledge the helpful input from Bob Reed.

[1] Department of Economics and Finance & UCMeta, University of Canterbury, NEW ZEALAND.
[2] School of Economics, Lanzhou University of Finance and Economics, CHINA
[3] Institute of Energy, Environment, and Economy, Tsinghua University, CHINA.
[4] Center for Evaluation and Development, GERMANY
[5] Ruhr University, GERMANY
[6] University of Alicante, SPAIN
[7] Wyss Academy for Nature, University of Bern, SWITZERLAND

† Corresponding author: Tom Coupé. Email: tom.coupe@canterbury.ac.nz

1. Introduction

Recent assessments highlight the increasing severity of climate change impacts, underscoring the urgent need for effective measures to reduce carbon emissions (IPCC, 2024). Emission Trading Schemes (ETS), recognized as a market-based solution, are being adopted widely across the globe. In 2024, there are 36 operational ETS worldwide, covering 18% of global emissions (ICAP, 2024). Theoretically, an ETS with a sufficiently low cap can reduce emissions as it gives firms whose marginal abatement costs are lower than the market-determined price of allowances an incentive to reduce emissions and sell extra allowances in the market. Payoffs generated through carbon trading can provide firms the means to develop and deploy cleaner fuels and cleaner energy technologies (Kong & Freeman, 2013).

A significant trend in recent years is the rapid development of ETS in emerging economies, which constitutes a major source of global ETS expansion (ICAP, 2024). These emerging economies which either have already introduced an ETS (e.g., China and Indonesia) or initiated plans to establish one (e.g., Turkey, India and Brazil), are already among the top global emitters.

Among the operational ETS in emerging economies, China's regional ETS pilots stand out as a particularly notable example. In 2011, China started to implement ETS in six pilot provinces and one city, namely Beijing, Tianjin, Shanghai, Chongqing, Hubei, Guangdong and Shenzhen., covering 7% of China's total CO_2 emissions (Zhang et al., 2014). This pilot program signaled a significant shift in China's environmental policy, from regulatory interventions to market-based instruments, to achieve the ambitious Paris pledge of a 60–65% reduction in CO_2 emissions by 2030, compared to the 2005 level (Goulder et al., 2017; Jotzo & Löschel, 2014).. The pilot program served as an important experiment that shaped the national ETS, launched in 2021 in the power sector, which is the world's largest in terms of coverage. China's pilot ETS thus can provide evidence as to how effective a national carbon trading program could be

(Almond & Zhang, 2021; Zhang et al., 2014).

China's pilot ETS has some unique features that distinguish it from established ETS such as the EU ETS. These features are likely to be shared by ETS planned by other emerging economies (such as Turkey, India or Brasil), particularly during the initial phrases of operation, where regulatory frameworks and market mechanisms are still being established (Lo, 2016). For example, China's ETS pilots lack an absolute cap in the sense that the allowances are based on output levels and emission-to-output ratios, rather than the absolute level of emissions (Goulder et al., 2017). Since the cap is relative, total emissions can still grow if the output increases. In addition, the pilots have faced issues such as over-allocation of emissions permits and the need for robust measurement and verification processes (Jotzo & Löschel, 2014). Allowances also were primarily allocated for free, and carbon prices have remained low (5-10 US dollars per ton), signaling unambitious relative caps. Thus, China serves as a particularly relevant case for these emerging countries, as the lessons learned from China's pilot ETS can guide them in having realistic expectations about what emission reductions they initially can expect.

Given these characteristics, one should indeed wonder how effective such ETS can be in reducing CO_2 emissions. Many papers have tried to answer this question and have estimated the effectiveness of China's pilot ETS, exploiting its quasi-experimental nature. Some of these studies find large and statistically significant reduction effects, among which the largest estimate of reduction indicates that China's ETS reduced CO_2 emissions in the pilot regions by as much as 71.7% (Luo & Qu, 2023). Other studies report small negative estimates: Han et al. (2022) estimated reductions of around 2%, and the average reduction found by Dong et al. (2022) was 4%. And some studies even reported positive estimates like Wu and Zhu (2023) and Zhu (2017).

This variation in sign and size of the estimated effect sizes leads to two questions: what is the 'overall' estimate across this literature and what causes the estimates to vary across studies. This is exactly where meta-analysis can come and help.

In this study, we conduct meta-analyses to synthesize the estimated effect of China's Pilot ETS on CO_2 emissions that have been reported in the literature. Our dataset includes 732 estimates from 80 studies. As this sample might be distorted by selection effects in what gets published, we use meta-analytical techniques to obtain a bias-adjusted overall mean effect and estimate the "true" effect on CO_2 emissions of China's ETS. The heterogeneity across estimates is further explained using unweighted meta-regression and Bayesian Model Averaging.

Two recent meta-analyses have already studied the effectiveness of carbon pricing schemes (including both carbon taxes and ETS). Döbbeling-Hildebrandt et al. (2023) base their analysis on 101 primary studies, among which 46 studies with 179 estimates targeted China's pilot emission trading scheme. Ahmad et al. (2024) collected 81 studies, with 29 on China's pilot ETS program.

Our meta-analysis improves on the above two meta-analyses in two ways. First, in terms of China's pilot ETS, our meta-analysis is much more comprehensive than the existing literature, both in data and methods. We cover substantially more studies and estimates – 80 primary studies, with a total amount of 732 estimated effect sizes. We also use a more comprehensive list of bias correction techniques as suggested in Irsova, Doucouliagos, et al. (2023). These methods include PET-PEESE (Stanley & Doucouliagos, 2012), selection models documented in Coburn and Vevea (2015), the MAIVE estimator (Irsova, Bom, et al., 2023), which deals with the endogeneity of standard errors in PET-PEESE, and RoBMA-PSMA (Bartoš et al., 2023).

Second, by solely focusing on China's ETS, our meta-analysis avoids a drawback of the existing meta-analyses which merge studies from various carbon pricing regimes and countries. In addition to the difference between carbon taxes and ETS, structural differences among countries and the ETS they implemented can lead to differences in the effectiveness of an ETS in these countries. In short, findings on effect sizes, influencing factors and the presence of publication bias from a meta-analysis which merges studies covering different countries might not be relevant to any of the countries or carbon pricing regimes included.

Consistent with Döbbeling-Hildebrandt et al. (2023), but in contrast to Ahmad et al. (2024), our analysis finds strong evidence of the presence of publication bias. We further show that after correcting for such publication bias, the overall mean effect, measured as a Fisher's z transformed partial correlation coefficient, of China's ETS on CO_2 emissions ranges between -0.101 and -0.057, representing a "small" negative effect of ETS on emissions (Xue et al., 2023). Restricting the sample to estimates that are from log-linear specifications, we estimate China's ETS to decrease CO_2 emissions by between 5.5% and 16.6%. Our estimates thus tend to be higher than Ahmad et al (2024), who report an overall across-countries emissions reduction of -0.025 (measured in partial correlation coefficient) and Döbbeling-Hildebrandt et al. (2023), who estimate a bias-adjusted average treatment effect of -8.66% for China's pilot ETS. Together, these studies thus suggest that, in the context of an emerging economy, introducing an ETS with relatively low carbon prices, can have a meaningful, but moderate (compared to for example the Paris pledge of a 60–65% reduction in CO_2 emissions by 2030), negative effect on emissions.

Our study further shows that differences in assumptions, which lead to differences in empirical methods used and covariates included, contribute to the heterogeneity of the estimated effect of China's ETS reported in the literature. Researchers who aim to estimate the effect of ETS therefore should present results across a wide range of techniques and specifications to highlight the robustness, or lack thereof, of their results.

This paper proceeds as follows: Section 2 describes the literature search and data construction process; Section 3 introduces the meta-analytic models used in this paper; Section 4 reports the publication bias-unadjusted overall mean effects; Section 5 tests for the presence of publication bias and uses different approaches to adjust for it; Section 6 conducts Bayesian Model Averaging and unweighted meta-regression to identify study characteristics that influence the estimated effect sizes; Section 7 applies the same procedure in Section 4, 5 and 6 to a subsample with untransformed effect sizes to obtain an overall mean effect with interpretable economic meaning; Section 8 concludes.

2. Literature Search and Data Construction

2.1 Identifying relevant literature

Central to any literature search is the choice of keywords to search on. Our approach began by selecting five representative studies that estimated the effect of China's ETS on emissions. The five papers we selected were Chen et al. (2020); Gao et al. (2020); Hu et al. (2020); Tang et al. (2021); Zhang et al. (2019). These five papers meet the inclusion criteria for our meta-analysis (as described below) and have been well-cited and hence are influential. At the time we conducted the search on March 8th, 2023, Hu et al. (2020) had been cited 219 times in Google Scholar, whereas the least cited paper, Tang et al. (2021), had 38 citations.

We then searched the reference lists of these five papers, extracted and listed the keywords of all the reference papers, and sorted the most-frequently used keywords into three groups, which correspond to "Place", "Policy" and "Targets":

Keywords list 1 (Place):

("China" OR "ETS pilot*" OR "Beijing" OR "Tianjin" OR "Shanghai" OR "Hubei" OR "Chongqing" OR "Guangdong" OR "Shenzhen")

Keywords list 2 (Policy):

("emission* trading system" OR "emission* trading scheme" OR "carbon trading" OR "cap and trade" OR "ETS" OR "carbon market*" OR "market-based instruments" OR "carbon pricing")

Keywords list 3 (Targets):

("carbon dioxide" OR "carbon emission*" OR "greenhouse gas*" OR "carbon intensity"

OR “emission* reduction*” OR “CO_2 emission*” OR “air pollution” OR “mitigation” OR “efficiency”)

The combination of the above keyword lists was used to search for English-language papers that study the relationship between ETS and carbon emissions in China. We used the following search engines: Google Scholar, Scopus and RePEc. The searching captures published papers, working manuscripts, reports, books, doctoral dissertations, and master’s theses.

As indicated in the PRISMA diagram in Figure 1, initially, the search resulted in 3713 records. After merging records from the three databases and removing duplicates, 2744 papers were left for screening.

The paper screening process consisted of two steps. In the first step, we reviewed the list of papers sorted by relevance based on the search algorithm. We inspected the title and abstract of each paper to identify relevant studies. The screening process concluded when we stopped encountering relevant papers. Under this screening procedure, 1272 papers were screened, among which 119 papers were identified as relevant. In the second step, we examined the full text of these 119 papers.

The second step can be described as the coding process, in which two coders worked simultaneously to record key information needed for conducting meta-analysis, including estimated effect sizes, standard errors, and study characteristics. After scrutinizing these papers, 80 studies were selected to form the final meta-analytic dataset. The inclusion criteria we used were:

1) The study must be *ex post* research that uses observational data to conduct regression analysis. This criterion rules out studies using Computable General Equilibrium (CGE) modelling or Data Envelopment Analysis (DEA) (e.g. Zhang et al. (2020)).

2) The dependent variable measures carbon emissions (or carbon intensity[1]), and the policy of interest is China's pilot ETS program. This criterion excludes studies that analyze the effect of other policies on carbon emissions, such as the SO_2 Pilot Emission Trading Scheme (e.g. Shen et al. (2023)) and the Low-Carbon Provinces and Cities (LCPC) program (e.g. Liu et al. (2022)). It also excludes studies focusing on the impact of ETS on other outcome variables, such as energy efficiency, e.g. Zhou et al. (2022), and carbon emission efficiency, for example, Chen et al. (2021); Du et al. (2023).

3) Since all the regression analyses were conducted in a Difference-in-Differences (DID) framework, the treatment variable ETS should always be a dummy shown in the following general specification:

$$Emissions_{it} = \beta_0 + \beta_1 ETS_{it} + \boldsymbol{X\alpha} + \beta_2 Treated_i + \beta_3 Post_t + \varepsilon_{it}, \qquad (1)$$

where $ETS_{it} = Treated_i \times Post_t$. $Treated_i$ equals one for the pilot regions that implemented ETS, and $Post_t$ equals one for the post-treatment period. $\boldsymbol{X}$ stands for a vector of control variables.

Accordingly, studies using the trading volumes and/or the market price of allowances as the key independent variables are excluded from the sample (e.g. Wu (2022)). Regressions including the interaction term of ETS and the other variables are also excluded from the dataset, because of the difficulty of calculating reliable standard errors for the marginal effects (e.g. Xue et al. (2023)).

4) Papers need to report the standard errors of the estimates and the degree of freedom. Otherwise, they would be excluded. The standard errors of the estimates are required in both the fixed-effect and random-effects meta-analytic models to perform weighted regressions and are further used in Egger's regression to test for publication bias. Though many studies applied Synthetic Control Method (SCM) to

[1] "Carbon intensity" here is defined as the ratio of carbon emissions to regional gross domestic product, also called "CO_2 emissions per GDP" hereafter.

assess the emissions-mitigating effect of China's pilot ETS program, they are excluded from the dataset for this reason. In addition, papers that didn't report the degrees of freedom, which is required to standardize and transform the original effect sizes, also dropped out.

Apart from the above criteria, two papers were excluded due to likely mistakes in estimation or with respect to reported effect sizes.[2] In the end, the full dataset includes a total of 80 papers and 732 estimates.

2.2 Summary Statistics

Table 1 summarizes the study characteristics of our final dataset, illustrating the variations in research decisions that could lead to heterogeneity in the estimated effect sizes.

Outcome Measures: Among the 80 studies, four types of dependent variables are used, including total CO_2 emissions (account for two-thirds of all the estimates), CO_2 emissions per capita (6.3% of all estimates), CO_2 emissions per GDP (26.1% of all estimates) and other measures (less than 1% of the estimates, such as CO_2 emissions per unit area and CO_2 emissions divided by per capita GDP (Shi et al., 2022)). Compared to total CO_2 emissions, CO_2 emissions per capita adjusts for provincial population size, whereas CO_2 emissions per GDP adjusts for the level of economic output.

While "total CO_2 emissions" seems to be rather straightforward, why would studies use the other two outcome measures? In other words, why are these adjustments necessary?

[2] 'We excluded 2 studies: H. Zhang, M. Duan and P. Zhang (2019) controls for a treatment group dummy and unit fixed effects simultaneously, which should lead to perfect multi-collinearity (we did not receive a response when we approached the authors).
Yu and Xu (2023) reports extremely large estimates of the impact of ETS on provincial CO2 emissions (in log), using a log-linear equation. The estimated coefficient of -15.77 (from Table 4, Panel 1) indicates the ETS reduced provincial CO2 emissions in the pilot regions by 100%, which is implausible. In addition, one estimate in Tian et al. (2022) (Table 4, Column 3) was regarded as an outlier because of unlikely precision: the estimated effect size from a log-linear equation is -0.246, with a standard error of 0.001.

One explanation is that they reduce variations in the emissions levels that are brought about by population size, levels of economic development and other social-economic factors other than ETS. CO_2 emissions per GDP is also called “carbon intensity”. It measures the amount of CO_2 emissions emitted to produce one extra unit of GDP. As it is central government’s goal to mitigate carbon emissions without hurting economic growth, using carbon intensity as the environmental effectiveness measure rather than the absolute level of CO_2 could make more sense.

Functional Forms (Log-linear versus. Linear): Almost three-quarters of the estimates are from log-linear specifications. To obtain the approximate percentage change in CO_2 emissions induced by the implementation of ETS for these specifications, we need to take the coefficient of the ETS dummy and perform the transformation $e^{\widehat{\beta}_{ETS}} - 1$.

Carbon Emissions Data Sources: There are three data sources for CO_2 emissions used in primary studies. Since there is no official carbon emissions data by provinces in China, over half of the papers (63.4%) directly calculate CO_2 emissions by following the IPCC (2006) guidelines and use energy consumption data from the China Energy Statistical Yearbook, for example, Chen et al. (2020) and Chen et al. (2022). A quarter of the papers use CO_2 emissions data from the Carbon Emissions Accounts and Datasets (CEADs, 2023). And 11.2% of the estimates are based on other CO_2 emission data sources.

Scopes of Carbon Emissions: There are three scopes of emissions involved: direct carbon emissions from fossil fuel combustion, process emissions from cement production, and indirect emissions from consuming imported electricity. Accordingly, effect sizes can be categorized into those that 1) only account for direct emissions, which consists of half of total effect sizes; 2) account for both direct and process emissions (29.7%); 3) account for both direct and indirect emissions (9.5%); and 4) the range of carbon emissions is unknown, which corresponds to papers using other emission data sources (10.8%). Why is the emission scope important? Counting indirect and/or process emissions alongside the direct emissions captures the full picture of the

ETS-induced emission mitigation. It would be interesting to know whether regressions using data with different scope of emissions reach different conclusions.

Time of Treatment: The existing literature also has different ways of defining the time of policy intervention. The pilot ETS program was announced in 2011, with five emission trading markets (in Shanghai, Beijing, Guangdong, Tianjin, and Shenzhen) opening in 2013 and two (Chongqing and Hubei) in 2014. To take into account the potential anticipating effect, if a study sets 2011 or 2012 as the treatment year, we define that as a "treatment based on announcement". If a study sets 2013 or 2014 as the treatment year for all the pilots, we define that as "treatment based on trading".

Less commonly, some studies designate multiple treatment years according to the actual operation time of the emission trading market. In our data set, the percentages of effect sizes with "treatment based on announcement", "treatment based on trading" and multiple treatment years are 24.0%, 57.8% and 18.1%, respectively.

Late-adopters of ETS: In the literature, studies also differ with respect to the ways of dealing with observations from the two late-adopters of ETS, Fujian and Sichuan, as both provinces started carbon emissions trading in December, 2016 (Wang et al., 2022). Since three-quarters of the estimates in our dataset use data after 2016, the ways they dealt with late-adopters could potentially make a difference. To address the changes in treatment status, studies either drop observations for these two provinces after 2016 or regard Fujian and Sichuan as the treatment units. In our meta-analytic dataset, we distinguish between whether Fujian alone is taken into account (10.4% of the total observations) or both late-adopters are taken into account (4.5%), using either of the above methods. The majority of the effect sizes (85.1%) are from regression models that treat Fujian and Sichuan as control units even for years after 2015, though it might be inappropriate to do that. By distinguishing between these three cases, we can see whether the estimated impact of ETS is influenced by excluding late-adopters from the control units.

Research Levels: As all the papers use panel data, we further record at what level the analyses were conducted. The percentages of effect sizes from firm, city, province, and provincial-sector level studies are 14.2%, 22.5%, 52.5% and 10.8%, respectively. We also use a ETS sector dummy to indicate studies that focus on emissions from the ETS-covered industrial sectors in the pilot regions, rather than emissions from the whole economy within a pilot region. The summary statistics shows that 15.7% of regressions in our dataset are using emissions only from sectors covered by the ETS.

Empirical Methods: There are roughly seven types of empirical methods used in the literature. The most frequently used one is the standard difference-in-differences (DID) method, generating almost 60% of all the estimates. Apart from that, triple DID (1.8%) is used in the provincial-sector level studies. The "Instrumental" method group (1.2%) includes IV estimator and GMM to address endogeneity due to sample selection. "Matching" (22.8%) includes DID combined with Propensity Score Matching (PSM-DID) and Synthetic DID, which conducts DID using the control units chosen by the Synthetic Control Method. In both cases, the treated units are matched with control units that have similar characteristics. Spatial models (4.5%) take into account spatial correlations between regions and account for possible spillover effects.

In addition, two variations of the standard DID model are also estimated in the literature. One is the "one-way fixed effect model" (4.4%), in which the key independent variable is the ETS dummy, yet the model controls for either the year fixed effect or the unit fixed effect but not both. The other is the "interaction only" model (6.4%), which only includes the ETS dummy with no time or unit fixed effect. These two models seem to be miss-specified (omitted variables) and yet produced more than 10% of the estimates. We generate two dummies to indicate the estimates from these two models.

Control Variables: Table 1 also lists 13 control variables and shows how often regression models include them. Among these variables, the top 3 most used variables are "GDP", "industrial structure" and "trade openness". We pay particular attention to two related policies, which are Five-Year-Plan carbon/energy intensity mitigation targets (FYP-targets) and the Low-Carbon Provinces and Cities (LCPC) pilot program. These two policies not only serve as indicators of environmental policy stringency (Zhang et al., 2019), but can directly lead to emission reduction. Failure to control for them might cause omitted variable bias and overestimation of the impact of ETS.

Publication Status: Finally, two variables concerning publication status are listed in Table 1. The dummy "Published" indicates that 90.7% of the estimates are from journal articles, whereas less than 10% of estimates are from working papers (including dissertations). The publications are all quite recent, with the mean publication year being 2021.

2.3 Effect Size Transformation

The original effect sizes from the 80 primary studies are not directly comparable. While three-quarters of the studies applied log-linear equations, the rest used emissions level rather than logged emissions as the dependent variable. The different dependent variables used in primary studies also lead to heterogenous effect sizes. As a result, effect size transformation is required before comparing the estimates. While partial correlation coefficients (PCC) seem to be a commonly used effect size transformation in many meta-analyses, it is correlated with its standard errors by definition (Stanley et al., 2024).

A better effect size to use is the Fisher's z transformed PCCs ("Fisher's z" for short). The sampling variance of Fisher's z does not depend on Fisher's z itself (Irsova, Doucouliagos, et al., 2023; van Aert, 2023). Moreover, Fisher's z more closely approximates a normal distribution compared to the PCC, especially when the PCC is different from zero (van Aert, 2023; Xue et al., 2023). The formula used to transform PCC into Fisher's z is based on van Aert (2023) (see Appendix).

The distribution of Fishers' z is shown in Figure 2. According to the size categories of PCC documented in Doucouliagos (2011), Xue et al. (2023) further calculated corresponding threshold values for "small", "medium" and "large" Fisher's z values, which are ± 0.1, ± 0.23, and ± 0.41, respectively. Thus, in Figure 2, the blue and red dashed lines are set at ± 0.1 and ± 0.23 to indicate "small" and "medium" effects. Figure 2 shows that the mean and median z values of our sample are -0.146 and -0.112, corresponding to medium effect sizes. Although most z values are within the small or medium size range, there are still many data points outside the range, especially in the negative domain (emission reductions).

3. Meta-analytic Models

3.1 Conventional Meta-analytic Models

This section briefly introduces two conventional 2-Level meta-analytic models assuming one estimate per study and two multilevel models that account for multiple estimates within a study and effect size dependency.

The fixed-effect (FE) model assumes that all effect sizes stem from a single, homogeneous population. It assumes that effect sizes deviate from the population mean only because of the sampling error. While the FE model is a commonly used meta-analytic mode, its assumption of homogeneity may not be realistic in many real-world applications (Harrer et al., 2021). An example would be different outcomes of interest, such as CO_2 emissions per capita and CO_2 emissions per GDP in our dataset, that likely lead to heterogeneous estimated effect sizes.

The random-effects (RE) model acknowledges heterogeneity in effect sizes by assuming each study has a true effect, and all the study-level true effects are drawn from a normal distribution. Its model specification can be written as:

$$ES_i = \beta_0 + \mu_i + \varepsilon_i,$$
$$\mu_i \sim N(0, \tau^2), \varepsilon_i \sim N(0, se(ES_i)^2), i = 1, \dots, S, \tag{2}$$

where i stands for ith study; S is the total number of studies; μ_i is the deviation of study-level true effect from the population mean β_0, and ε_i is the sampling error.

Note that the fixed-effect model is a special case of random-effects models: When μ_i

equals zero, the RE model reduces to the FE model. Since both FE and RE models assume one estimate per study, in Equation (3) there is no subscript j – as for jth estimate in a study.

While both the FE and RE meta-analytic models produce a weighted overall mean effect, they differ in how the weights are assigned. While FE weighs effect size k based on precision (inverse-variance) with the weight being $w_k^{FE} = \frac{1}{SE_k^2}$, RE includes the between-study heterogeneity when determining the inverse-variance weight of each effect size and has as weight $w_k^{FE} = \frac{1}{SE_k^2 + \tau^2}$. Here, τ^2 is between-study variance, which quantifies the heterogeneity of the true effect sizes across studies. A large τ^2 would result in an estimated overall mean effect close to the unweighted overall mean effect from OLS.

In the economics literature, it is common for a single study to provide multiple estimates. In our final data set, the average number of effect sizes per study is 9.1. Hedges et al. (2010) document three cases in which estimates from the same study could be correlated: 1) same unit of observation, different measures of outcome; 2) same unit of observation, effect sizes measured at different points in time; 3) same control group, several treatment groups. Some studies, such as Cui et al. (2021) and Dong et al. (2022), fit into case (1) because they base the analysis on the same dataset and use both carbon emissions and carbon intensity as the outcome variables. Some studies fit into case (2). For example, Dong et al. (2022) sets 2012 as the time of policy intervention and further switches to 2014 for robustness check. Zhang and Cheng (2021) is an example of the third case because it estimates the emission reduction effects for Beijing and Shanghai separately, using the same control units.

For the above reasons, multilevel meta-analytic models that acknowledge the nesting structure of multiple effect sizes within the same study seem more appropriate, given the data we have. The first one is the hierarchical model (“HIER” for short), which is also called the “3-Level (3L) model”. It has three levels in the sense that it adds an extra layer upon the RE model: The HEIR model assumes effect sizes within the same study follow a normal distribution with the mean being the study-level true effect. According to Cheung (2014) and Yang et al. (2022), the HIER model can be written as:

$$ES_{ij} = \beta_0 + \mu_i + \mu_{ij} + \varepsilon_{ij},$$
$$\mu_i \sim N(0, \tau_b^2), \mu_{ij} \sim N(0, \tau_w^2), \varepsilon_{ij} \sim N\left(0, se\left(ES_{ij}\right)^2\right), i = 1, \dots, S, \tag{3}$$

where μ_i stands for deviation of study-level true effects from the population overall mean β_0, and μ_{ij} stands for the deviation of effect size j within study i from its study-level true effect, μ_i . τ_b^2 and τ_w^2 are the sampling variances of the two normal distributions, which are called "between-study heterogeneity" and "within-study heterogeneity".

While the HIER model addresses the correlation between multiple effect sizes from the same study by assuming a normal distribution, it assumes the sampling errors to be uncorrelated with each other. The correlated and hierarchical (CHE) model takes a step further and assumes the within-study sampling errors are correlated. Accordingly, with the CHE model, the off-diagonal elements in the within-study variance-covariance (VCV) matrix are not zero. In practice, to build this VCV matrix with non-zero off-diagonal elements, we follow Xue et al. (2023) and assume the level of correlation to be 0.5 (i.e. $corr\left(\mu_{ij}, \mu_{ik}\right) = 0.5$).

3.2 Test and Correction for Publication-bias

Publication bias exists when the probability of a study getting published is affected by its results (Harrer et al., 2021). If the decision on whether or not to publish a study depends on the presence of statistically significant results, the resulting sample of published studies would not be representative of the population of all studies, which introduces a bias in the estimate of the true population effect. With publication selection based on statistical significance, studies with small effect sizes therefore are more likely to remain unpublished. As a result, the estimated overall mean effect based on the published studies can be inflated.

In the meta-analysis literature, the most common way of testing for publication bias is a visual inspection of the funnel plot combined with the Egger's regression test. Funnel asymmetry is key to identifying whether a given area of research suffers from publication bias (Stanley & Doucouliagos, 2012). Alternatively, in Egger's regression test, a statistically significant impact in a univariate regression of standard errors on the effect sizes is also considered to indicate the presence of publication bias (Stanley & Doucouliagos, 2012).

In the presence of publication bias, there are two general approaches to obtaining bias-adjusted estimates of the overall mean effect. One approach uses funnel-based models, whereas the other employs selection models. Irsova, Doucouliagos, et al. (2023) suggests that since both approaches have pros and cons, a meta-analyst should always use at least one procedure from each approach. In addition, Carter et al. (2019) also suggests that no single bias-correcting meta-analytic method (i.e., methods from both families) consistently outperforms all the others. Thus, in this paper, we use methods from the above two families in addition to two additional methods to correct for publication bias.

<u>PET-PEESE:</u> The first method is the PET-PEESE model from the funnel-based method family to address publication bias. It builds on Egger's regression, where observed effect sizes are regressed on a constant and their standard errors. In this framework, PET (Precision-Effect Test) focuses on determining whether a genuine effect exists beyond potential bias. Specially, after running Egger's regression, if we set standard error equals to zero, we'll obtain the estimated overall mean effect in the absence of publication bias, which is also called "effect beyond bias". Note that this bias-adjusted overall mean effect from PET applies when the constant in the Egger's regression is statistically insignificant.

When the intercept is statistically different from zero, it might be more appropriate to fit a non-linear model between the effect sizes and standard errors. That's when PEESE (which stands for "Precision-Effect Estimate with Standard Errors") comes in, which regresses effect sizes on a constant and standard error square. PEESE provides a better effect size approximation in the presence of an effect (Bartoš et al., 2022).

Selection Model: The selection model used in this paper was first documented in Vevea and Hedges (1995). It assumes that the effect sizes are normally distributed and follows a random-effects model in the absence of publication selection. Yet publication selection distorts this normal distribution by introducing a relationship between the probability of getting published and the p-values (Vevea & Hedges, 1995). Usually effect sizes with p larger than 0.05 are less likely to be published, compared to the effect sizes statistically significant at the 5% level. Thus, the selection model weighs effect sizes from a set of prespecified p-intervals differently to adjust for publication bias.

Estimation of the above selection model returns the mean μ and variance σ^2, and a single weight, w, for the interval with $0.025 < p < 1.00$, relative to that of $p < 0.025$. The weight associated with positive and significant estimates is by default set equal to one, while w is estimated for those observations with $0.025 < p < 1.00$. The magnitude of w indicates how likely insignificant or wrong-signed estimates to be observed as the significant estimates. This model is commonly called the "Three Parameter Selection Model" (3PSM), because it returns μ, σ^2, and w. The researcher can specify several p-intervals if each interval contains enough data points.

RoBMA-PSMA: The above two approaches produce a single "best" estimate based on/ their respective assumptions. Robust Bayesian meta-analysis publication selection model-averaging (RoBMA-PSMA) fits the data with a set of 36 pre-specified models and produces a weighted overall mean effect using Bayesian Model Averaging, with weights decided by the fits of these models. The 36 models in the model space of RoBMA-PSMA are defined to accommodate three groups of situations: 1) The presence/absence of effect size ($\mu = 0,\ \mu \neq 0$);2) The presence/absence of publication bias; 3) The presence/absence of between-study heterogeneity: ($\tau = 0, \tau \neq 0$).

Due to the lack of *a priori* information, RoBMA initially assigns equal prior model probabilities (0.5) to models assuming the presence of an effect size and models assuming effect size equaling to zero. The underlying assumption is that before we know anything about what's behind the "thick black curtain", the assumptions are equally likely to be true. In other words, models assuming the effect size is different from zero have a 50% chance of being the "true" model. The same holds for the other two assumptions.

When there is publication bias, eight methods (six selection models, PET and PEESE) are used to correct for publication bias. Each publication bias-adjusted model has a prior probability of $\frac{0.5}{8} \approx 0.13$. After fitting the data with the models, Bayesian model averaging updates the probability of each model being the "true" model and returns the posterior model probability. Eventually, a weighted overall mean effect is obtained, with the weights based on posterior model probabilities indicating how well these models fit the data.

<u>MAIVE</u>: The Meta-Analysis Instrumental Variable Estimator (MAIVE) estimator documented in Irsova, Bom, et al. (2023) is another way of correcting for publication bias. With PET-PEESE, standard errors of the effect sizes are treated as given and exogenous. Yet in the primary studies, given an effect size, researchers can p-hack the standard errors to achieve statistical significance, for example, by using another estimator or choosing not to cluster. This type of p-hacking leads to endogeneity of standard errors.

To address this problem, Irsova, Bom, et al. (2023) proposes to use the inverse of the sample size an instrumental variable (IV) for the standard error/variance in PET-PEESE. A two-step procedure is recommended to obtain the bias-adjusted overall mean effect. In the first step, the square of the standard error is regressed on the inverse of the sample size (as shown in Equation A.7 in the Appendix). In the second step, the fitted value $SE(\tilde{\alpha})_i^2$ from the first step is used to replace $SE(\hat{\alpha})_i^2$ before estimating the PEESE model.

4. Estimates of the Overall Mean Effect

Table 2 produces both weighted and unweighted estimates of the overall mean effect from the full sample. The effect size is Fisher's z. Columns (1)-(3) in Table 2 report estimates from OLS, fixed-effect (FE) and random-effects (RE) models, assuming within-study independence between effect sizes. Column (4) reports the results from the hierarchical model (HIER) and Column (5) from the correlated and hierarchical (CHE) model. The standard errors reported in parentheses are clustered at the study level, using cluster robust standard error CR2 (Pustejovsky, 2016).

As shown in Table 2, the bias-unadjusted overall mean effects are consistently negative across different estimators, varying between -0.157 and -0.041 (in Fisher's z) and are statistically significant[3].

Table 3 describes the calculated weights that the FE and RE models assigned to each study. In the FE model, the top 3 studies with the highest weights account for more than two-thirds (68.6%) of total weights in estimating the overall mean effect. In the RE model, the sum of weights for the top 3 studies is only 5.01%. The study weights in Table 3 serve as a caution. With so much weight given to only three studies using the FE estimator, one needs to be concerned that the resulting overall mean is unrepresentative of the sample as a whole. Apart from that, both the FE and RE models ignore dependency within studies and hence are miss-specified.

This is why we focus on 3-Level model. As we turn to the multilevel meta-analytic models, the HIER model in Column (4) and the CHE model in Column (5) of Table 2 give us an estimated overall mean effect of -0.157 and -0.141, respectively. Our preferred estimate comes from the CHE model as it allows the within-study sampling errors to be correlated.. It indicates that the bias-unadjusted overall mean effect (in Fisher's z scale) of the impact of China's ETS on CO_2 emissions is of medium size (-

[3] Fisher's z scores are not very 'economically' intuitive, so we conduct a subsample analysis using the original effect sizes taken from the log-linear regressions in section 7.

0.141) and is statistically significant.

From the discussion in the methodology section, we know that total heterogeneity across estimates can be sourced to between-study (τ_b^2) and within-study heterogeneity (τ_w^2). Following Assink and Wibbelink (2016), two separate log-likelihood-ratio tests were conducted to determine whether the between-study and within-study variances are significant. The null hypothesis (H_0: $\tau^2 = 0$) was rejected at the 1% level, indicating there are significant within- and between-study variance in the data set. Formulas from Cheung (2014) were used to calculate the within- and between-study I^2, with the results shown in Table 2.

5. Correction for Publication Bias

5.1 Results from FAT-PET-PEESE

The contour-enhanced funnel plot based on the FE model is shown in Figure 3. Figure 3 suggests more negative effect sizes than positive ones. It also becomes visible that fewer effect sizes fall within the insignificant area (the white area) than within the significant area. The asymmetric funnel plot serves as the primary evidence of selective publication. Furthermore, one can notice a diagonal trend in the distribution of effect sizes along the contour line. In the left side of Figure 3 with negative effect sizes, a larger SE corresponds to a more negative effect size, and vice versa. This represents the primary indication of publication bias, which we now test for statistically.

Table 4 reports the results from Egger's regression. Columns (1)-(5) are based on univariate regressions with the standard error being the only explanatory variable. Columns (6)-(10) are based on multivariate regression, in which additional control variables are included. Since the estimated coefficients of $SE(z_i)$ are statistically significant across all 10 models in Table 4, we can conclude that there is strong evidence of the presence of publication bias.

The first row of Table 4 reports the "effect beyond bias". For columns (1)-(5), it is the constant term in the univariate regressions, which show us what the overall mean effect would be if there is no publication bias (i.e., the predicted estimate of an infinitely precise study, with $SE(z_i) = 0$). Columns (6)-(10) reports "effect beyond bias" from multivariate regressions, obtained in post-estimation prediction by setting $SE(z_i) = 0$ and the values of control variables at their sample means.

Table 4 shows that the estimated "effects beyond bias" from multivariate models are on average larger than those from the univariate models. The "effect beyond bias" estimated from the CHE model is -0.040 in univariate regression and -0.062 in multivariate regression. These estimates are both less than 0.10 in absolute value, thus, are categorized as "small" effect sizes according to Xue et al. (2023). The fact that they are also less than half the size of the unadjusted overall mean effect (-0.141) suggests that publication bias considerably inflates the overall un-adjusted mean effect.

5.2 Alternative Ways of Addressing Publication Bias

Table 5 reports bias-adjusted estimates from alternative methods[4]. According to the univariate CHE model (cf. Column (5), Table 4), the coefficient of standard errors is significant, and the constant is significantly different from zero. Thus, we can go one step further and fit the PEESE model. Table 5 Column (1) shows that the bias-adjusted effect from the multivariate CHE-PEESE model is -0.111 and statistically significant at the 1% level.

Table 5 Column (2) reports the results from the one-sided selection model, assuming selection is based on both sign and significance. To be specific, negative and statistically significant effect of China's pilot ETS on CO_2 emissions is more likely to be published. To fit the selection model, estimated effect sizes are divided into two groups: estimates with a negative sign and significant at the 5% level ($p > 0.975$); and all other estimates. The resulting bias-adjusted, estimated overall mean effect is -0.057

[4] The newly-developed R packages RoBMA and MAIVE do not have an option to add explanatory variables. For consistency, no moderators are thus included in any of the meta-analytic models across all four columns in Table 5. Only by holding the moderator constant, we can analyze how different methods affect estimates.

and statistically significant.

In Table 5 Column (3), we report the bias-adjusted estimate of the overall mean effect from RoBMA, which is -0.101 with a 95% credible interval of [-0.110, -0.092]. Built into RoBMA are six selection models as well as PET and PEESE to correct for publication bias. In our case, RoBMA selects just two of these models. The associated estimate is a weighted average of these two one-sided selection models, with the weights based on posterior model probabilities. Both models assume the overall mean effect is different from zero and have cross-study heterogeneity. They also assume publication selection based on significance and sign. The first one-sided selection model uses three p-value cutoffs to distinguish effect sizes that are positive and significant ($0 < p < 0.025$), positive and marginally significant ($0.025 < p < 0.05$), positive and insignificant ($0.05 < p < 0.5$), and negative ($0.5 < p < 1$). The posterior model probability indicates that the probability of this model being the "true" model underlying the data is 94.1%. The other one-sided selection model uses two p-value cutoffs, 0.05 and 0.5, which does not distinguish positive effect size that are significant at the 5% level and marginally significant at the 10% level. Its posterior model probability is 5.9%.

Column (4) applies the MAIVE estimator. The resulting bias-adjusted estimate of the true effect is -0.080 and is statistically significant. In the subsequent Hausman test, the null hypothesis of no endogeneity is rejected, likely due to reverse causality between the effect size and its standard error in the PEESE model, or omitted variable bias. For this reason, we prefer the estimates from Columns (2)-(4) in Table 5 over the PET-PEESE estimates, which assume exogeneity of the standard error variable.

Our conclusion about the publication bias-adjusted overall mean effect is based on bias correction techniques other than PET-PEESE, given the endogeneity of standard errors of effect sizes indicated by the MAIVE estimator. According to the selection model, RoBMA and the MAIVE estimator, the bias-adjusted estimates of the overall mean vary between -0.057 and -0.101, corresponding to "small" effect sizes. The estimates from these methods tell a consistent story, showing that the bias-adjusted effect of China's ETS on CO_2 emissions is negative and significant, and smaller than the unadjusted overall mean effect.

6. Explaining Heterogeneity

So far, we have focused on publication bias and finding the "best" overall mean estimate. In this section, we focus on explaining why different studies find different estimates in the first place. With this aim, the moderators listed in Table 1 are included in the meta-regression to examine the influence of research characteristics on the estimated effect sizes. To address publication bias, we include the standard error variable in the specification. The endogeneity of standard errors should not be a problem if the omitted moderators are the only source of the endogeneity. The unweighted least squares (OLS) meta-regression estimates are reported in Columns (1)-(2), Table 6.

According to Columns (1)-(2) in Table 6, the unweighted meta-regression shows that the coefficient of the standard errors of Fisher's z (-2.559) is strongly significant. Apart from that, the coefficients of "Instrumental" (0.056) and "Population" (-0.039) are significant at the 10% level. It means that if – all else equal – a primary study uses an IV estimator, the absolute effect size is reduced by 0.056, compared to the DID estimates. Meanwhile, controlling for population size in the regression model would increase the estimated reduction effect by 0.039. The variable "CO_2 per capita" has a coefficient of 0.065, indicating that using CO_2 emissions per capita as the dependent variable would result in a smaller reduction effect, compared to using the "Other" kind of dependent variable. Yet this coefficient is not statistically significant.

Undoubtedly, the estimates from OLS meta-regression reported in columns (1)-(2), Table 6, are affected by multicollinearity. Bayesian Model Averaging (BMA) is one way of addressing this problem. Instead of basing inference on the full model with all the moderators, BMA averages all possible models of variable combinations using their posterior model probabilities as weights.

The BMA results are shown in Table 6, Columns (3)-(5). Given strong evidence of the presence of publication bias from the preceding analysis, we add the Fisher's z standard error variable into all model specifications. The choice of model priors and coefficient priors follows the default in Zeugner and Feldkircher (2015).

Column (3), Table 6 reports the posterior inclusion probability (PIP) for each moderator, indicating its probability of belonging to the "true" model underlying the observed effect sizes. In identifying important moderators, the standard PIP threshold is 0.50. Moderator with low PIP (less than 0.5) suggests weaker evidence for the moderator's importance in explaining the observed effect sizes. Note that the Fisher's z SE variable has a PIP equaling to 1, since we forced it to be included in all model specifications. Otherwise, only "Population" has a PIP larger than 0.5, while all the other variables have PIPs less than 0.50.

The posterior mean of coefficients in Column (4), Table 6 denotes the weighted average marginal effect of a study characteristic on the reported effect sizes (in Fisher's z scale) across multiple models. Table 6 indicates that the weighted average coefficient for Fisher's z SE is -2.71. The coefficient for "Population" is -0.032, close to that from OLS regression. This suggests that in the primary studies, population size, growth or intensity is correlated with both CO_2 emissions and the treatment status, possibly because ETS pilots were more likely to be implemented in developed regions with large populations. Thus, regressing CO_2 emissions on ETS while omitting "Population" from the model, decreases the size of the effect of ETS. The coefficient estimates of the other moderators are all very small.

In summary, the unweighted, OLS meta-regression and the BMA analysis indicate that the only research characteristic that plays a significant role in explaining heterogeneity in the estimated effect of China's ETS on CO_2 emissions is "Population", namely whether a study controls for population size, growth or density.

7. Economic Significance and Subsample Analysis

Following Irsova, Doucouliagos, et al. (2023)'s advice, we conduct subsample analysis using the original effect sizes taken from the log-linear regressions in the sample, which make up almost three quarters of the estimates. By doing so, we can retrieve economically meaningful results in terms of relative carbon emission reduction. This subsample consists of 540 estimates from 57 primary studies. We apply the same techniques as in the full sample analysis to the subsample to check the robustness of the main results (i.e., whether our conclusion that China's pilot ETS significantly reduced CO_2 emissions holds in the subsample) and obtain an economic interpretation of the overall mean effect. The results are discussed in the following paragraphs, and the detailed results are presented in the appendix.

7.1 Estimating the Overall Mean Effect in Subsample

Table A1 reports bias-unadjusted overall mean effects of the subsample from five estimators, which range from -0.048 to -0.188. Since the effect size here is the regression coefficients from log-linear models, we can further transform it into the percentage change in CO_2 emissions using $e^{\widehat{\beta}_{ETS}} - 1$. The transformed overall mean effects range thus between -4.7% and -17.1%.

Our preferred estimate is the one transformed from the CHE model in Column (5), which shows an overall mean reduction of 14.3% in CO_2 emissions. Our decision is based on: 1) The high level of heterogeneity as indicated by the I^2 values corresponding to the four meta-analytic models; 2) The impression that the FE model estimates the population mean based on a few studies in Table A2, which casts doubt on the reliability of the FE estimates and votes for the other meta-analytic estimators that can avoid this drawback; 3) The fact that the studies have multiple effect sizes.

7.2 Test and Correcting for Publication Bias in Subsample

Table A3 reports the results from Egger's regression. Except for the OLS regressions in Columns (1) and (6), we find strong evidence of the presence of publication bias in the other models.

Table A3 also reports the bias-adjusted overall mean effect from both univariate regressions (Columns (1)-(5)) and multivariate regressions (Columns (6)-(10)). Our preferred estimate from the multivariate CHE model indicates that China's ETS reduces CO_2 emissions in the pilot regions by 12.3%. Again, the estimated overall absolute mean effects from the multivariate models are larger in size (more negative) than those from univariate regressions.

Bias-adjusted estimates of the overall mean effect from alternative approaches for the subsample are reported in Table A4. No control variables were added to these models. The first three columns in Table A4 estimate negative and significant bias-adjusted overall mean effects ranging from -5.5% to -14.4%.

While the estimates from PET-PEESE, selection model and RoBMA show a consistent picture, the MAIVE estimator produces a different estimate as reported in Column (4), Table A4. It estimates a positive overall mean effect of ETS on CO_2 emissions, 31.3% ($e^{0.272} - 1$), which is large in size but is statistically insignificant. Since the test statistics of the Hausman test is 0.548, compared to a critical value of 3.841, we fail to reject the null hypothesis of no endogeneity. As a result, we favor the OLS estimate which is BLUE rather than the IV estimate. Thus, Column (5) further reports the OLS estimate of the PEESE model which does not involve weighting based on precision, just as the MAIVE estimate in Column (4). The resulting bias-adjusted OLS estimate of the overall mean effect is an emission reduction of 16.6% ($e^{-0.181} - 1$).

Given all these methods, it's hard to decide which one gives us the preferred bias-adjusted estimate because they all depend on different assumptions that are not directly testable. We therefore conclude that the bias-adjusted overall mean effect ranges between -16.6% (from the OLS estimate of the PEESE model) and -5.5% (from the one-sided selection model).

7.3 Explaining Heterogeneity in Subsample

Table A5 shows the results from Bayesian Model Averaging and OLS meta-regression. The moderators are almost the same as in the full sample analysis, with two slight changes. Firstly, the dummy indicating whether the effect size is from a log-linear model is excluded. Secondly, due to multicollinearity, one emission scope indicator "Direct Emissions Only" is also dropped from the variable list. To capture the impact of publication bias, we force standard errors of the effect size into every model specification when conducting BMA.

Factors causing between-study and within-study heterogeneity in the estimated effect sizes are identified according to Table A5. To begin with, the standard error variable has a relatively large impact, with its posterior mean being -0.25. Variables with posterior inclusion probability (PIP) larger than 0.5 include "Direct_Indirect" (which means accounting for both direct and indirect emissions in emission calculation), "Instrumental", "Matching", "Fixed Asset Ratio", and "Published". These variables all have high probabilities of being in the "true" model that explains the variation in the estimated effect sizes. In the OLS meta-regression, only "Matching", "Urbanization" and "Fixed Asset Ratio" have estimated coefficients significant at 10% level. Since multicollinearity in the OLS regression might inflate standard errors and cause difficulties in testing individual coefficients, one should refer to the results from BMA for contributing factors of the estimated effect sizes.

8. Conclusions

Using meta-analysis, this paper estimates how much China's pilot ETS program has reduced CO_2 emissions. Knowing the extent to which China's ETS pilots have been effective is important not just for China but also for other emerging markets, that are considering implementing ETS with characteristics similar to China's ETS.

We apply both 2-Level and 3-Level meta-analytic models to a sample that consists of 80 studies and 732 estimated effect sizes. To estimate the overall mean effect, we transform these effect sizes into Fisher's z scores. In addition, because Fisher's z scores are not so easy to interpret, we also conduct a meta-analysis on a subsample of effect sizes from log-linear model specifications, which make up 540 estimates from 57 studies out of 80. The effect sizes in the subsample have a clear economic meaning and measure the percentage change in CO_2 emissions induced by the implementation of the ETS program.

While there seems to be a tradeoff between the number of observations in the meta-analytic dataset and the interpretability of effect sizes in the full sample and subsample, their meta-analytic results are quite consistent. Based on our preferred models, both the full sample and subsample analysis obtain negative and statistically significant overall mean effects. The size of the overall mean effect in the full sample can be considered as a "medium" effect size (approximately -0.14 in both in Fisher's z scale and as partial correlation coefficient). In the subsample, the overall mean effect indicates an emission reduction of 14.3% on average in the pilot regions.

Like most meta-analyses in economics, strong evidence of the presence of publication bias was detected in both the full sample and the subsample. Correspondingly, we base our conclusion on the publication bias-adjusted estimates. Due to the endogeneity of the standard error variable, we further prefer the estimates from the selection model, RoBMA-PSMA (the weighted average of two selection models) and the MAIVE estimator over the estimates from the PET-PEESE model. The resulting bias-adjusted overall mean effects range between -0.101 and -0.057 (in Fisher's z scale). The fact that these numbers are smaller than the bias-unadjusted estimates suggests that publication bias does inflate the estimated effect of China's ETS in the literature. In the subsample, the bias-adjusted overall mean effects range from -16.6% to -5.5%.

Our analysis further suggests that what methods a study uses and what variables a study controls for affect the estimated effect size: for example, including population size results in a larger estimated effect of ETS in the full sample, while using a matching methodology is associated with smaller effect sizes in the sub-sample. Researchers' choices thus affect effect sizes found in the primary study. This indicates it is important for researchers to both clearly motivate their methodological choices as well as highlight the variation in estimates as one changes methodological assumptions.

Overall, this paper, combined with two other meta-analyses (Döbbeling-Hildebrandt et al. (2023), Ahmad et al. (2024)) of carbon pricing systems worldwide, provides robust evidence that carbon pricing regimes can significantly reduce CO_2 emissions. Our overall means show emission reductions of up to -16.6%. This suggests that an ETS, like the one implemented in China, can make a sizeable dent in the total emissions of emerging economies, even without a fixed cap and despite low carbon prices. Those aiming for even larger declines in emissions, however, will need to sharpen how emission trading systems are applied or complement an ETS with additional emission-reducing measures.

References

Ahmad, M., LI, X., & Wu, Q. (2023). Carbon Taxes and Emission Trading Systems: Which One is More Effective in Reducing Carbon Emissions?—A Meta-Analysis. *A Meta-Analysis (September 30, 2023)*.

Ahmad, M., Li, X. F., & Wu, Q. (2024). Carbon Taxes and Emission Trading Systems: Which One Is More Effective in Reducing Carbon Emissions?—A Meta-analysis. *Journal of Cleaner Production*, 143761.

Almond, D., & Zhang, S. (2021). Carbon-trading pilot programs in China and local air quality. AEA Papers and Proceedings,

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154-174.

Bartoš, F., Maier, M., Quintana, D. S., & Wagenmakers, E.-J. (2022). Adjusting for publication bias in JASP and R: Selection models, PET-PEESE, and robust Bayesian meta-analysis. *Advances in Methods and Practices in Psychological Science*, *5*(3), 25152459221109259.

Bartoš, F., Maier, M., Wagenmakers, E. J., Doucouliagos, H., & Stanley, T. (2023). Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research synthesis methods*, *14*(1), 99-116.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115-144.

CEADs. (2023). *China CO2 Inventory 1997-2021 (IPCC Sectoral Emissions)*. Retrieved December 12 from https://www.ceads.net/data/nation/

Chen, L., Liu, Y., Gao, Y., & Wang, J. (2021). Carbon emission trading policy and carbon emission efficiency: an empirical analysis of China's prefecture-level cities. *Frontiers in Energy Research*, *9*, 793601.

Chen, L., Wang, D., & Shi, R. (2022). Can China's carbon emissions trading system achieve the synergistic effect of carbon reduction and pollution control? *International Journal of Environmental Research and Public Health*, *19*(15), 8932.

Chen, S., Shi, A., & Wang, X. (2020). Carbon emission curbing effects and influencing mechanisms of China's Emission Trading Scheme: The mediating roles of technique effect, composition effect and allocation effect. *Journal of Cleaner Production*, *264*, 121700.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological methods*, *19*(2), 211.

Coburn, K. M., & Vevea, J. L. (2015). Publication bias as a function of study characteristics. *Psychological methods*, *20*(3), 310.

Cui, J., Wang, C., Zhang, J., & Zheng, Y. (2021). The effectiveness of China's regional carbon market pilots in reducing firm emissions. *Proceedings of the National*

Academy of Sciences, *118*(52), e2109912118.

Döbbeling-Hildebrandt, N., Miersch, K., Khanna, T., Bachelet, M., Bruns, S., Callaghan, M., Edenhofer, O., Flachsland, C., Forster, P., & Kalkuhl, M. (2023). Effectiveness of carbon pricing–A systematic review and meta-analysis of the ex-post literature.

Dong, Z., Xia, C., Fang, K., & Zhang, W. (2022). Effect of the carbon emissions trading policy on the co-benefits of carbon emissions reduction and air pollution control. *Energy Policy*, *165*, 112998.

Doucouliagos, C. (2011). How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics.

Du, M., Wu, F., Ye, D., Zhao, Y., & Liao, L. (2023). Exploring the effects of energy quota trading policy on carbon emission efficiency: Quasi-experimental evidence from China. *Energy Economics*, 106791.

Gao, Y., Li, M., Xue, J., & Liu, Y. (2020). Evaluation of effectiveness of China's carbon emissions trading scheme in carbon mitigation. *Energy Economics*, *90*, 104872.

Goulder, L. H., Morgenstern, R. D., Munnings, C., & Schreifels, J. (2017). China's national carbon dioxide emission trading system: An introduction. *Economics of Energy & Environmental Policy*, *6*(2), 1-18.

Han, Y., Tan, S., Zhu, C., & Liu, Y. (2022). Research on the emission reduction effects of carbon trading mechanism on power industry: plant-level evidence from China. International Journal of Climate Change Strategies and Management, 15(2), 212-231.

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, *1*(1), 39-65.

Hu, Y., Ren, S., Wang, Y., & Chen, X. (2020). Can carbon emission trading scheme achieve energy conservation and emission reduction? Evidence from the industrial sector in China. *Energy Economics*, *85*, 104590.

ICAP. (2024). *Emissions Trading Worldwide: Executive Summary*. https://icapcarbonaction.com/system/files/document/icap-2024-status-report-executive-summary_en_240517.pdf

IPCC. (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Retrieved December 13 from https://www.ipcc-nggip.iges.or.jp/public/2006gl/

IPCC. (2024). *Climate Change 2023: Synthesis Report Summary for Policymakers*. https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_FullVolume.pdf

Irsova, Z., Bom, P. R., Havranek, T., & Rachinger, H. (2023). Spurious precision in meta-analysis.

Irsova, Z., Doucouliagos, H., Havranek, T., & Stanley, T. (2023). Meta-Analysis of Social Science Research: A Practitioner's Guide.

Jotzo, F., & Löschel, A. (2014). Emissions trading in China: Emerging experiences and international lessons. In (Vol. 75, pp. 3-8): Elsevier.

Kong, B., & Freeman, C. (2013). Making sense of carbon market development in China. *Carbon & Climate Law Review*, 194-212.

Liu, X., Li, Y., Chen, X., & Liu, J. (2022). Evaluation of low carbon city pilot policy effect on carbon abatement in China: An empirical evidence based on time-varying DID model. *Cities*, *123*, 103582.

Lo, A. Y. (2016). Challenges to the development of carbon markets in China. *Climate Policy*, *16*(1), 109-124.

Luo, H., & Qu, X. (2023). The impact of carbon emissions trading pilot on the carbon intensity of China's energy-intensive manufacturing industry.

Pustejovsky, J. E. (2016). *Clustered standard errors and hypothesis tests in fixed effects models*. Retrieved September 20 from https://jepusto.com/posts/clubSandwich-for-CRVE-FE/index.html

Shen, J., Tang, P., Zeng, H., Cheng, J., & Liu, X. (2023). Does emission trading system reduce mining cities' pollution emissions? A quasi-natural experiment based on Chinese prefecture-level cities. *Resources Policy*, *81*, 103293.

Shi, B., Li, N., Gao, Q., & Li, G. (2022). Market incentives, carbon quota allocation and carbon emission reduction: evidence from China's carbon trading pilot policy. *Journal of environmental management*, *319*, 115650.

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. routledge.

Stanley, T. D., Doucouliagos, H., & Havranek, T. (2024). Meta-analyses of partial correlations are biased: Detection and solutions. *Research synthesis methods*, *15*(2), 313-325.

Tang, K., Zhou, Y., Liang, X., & Zhou, D. (2021). The effectiveness and heterogeneity of carbon emissions trading scheme in China. *Environmental Science and Pollution Research*, *28*, 17306-17318.

van Aert, R. C. (2023). Meta-analyzing partial correlation coefficients using Fisher's z transformation.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*, 419-435.

Wang, K., Li, S., Lu, M., Wang, J., & Wei, Y. (2022). *A Review of China's Emission Trading Market (2022)*.

Wu, L., & Zhu, Q. (2023). Has the Emissions Trading Scheme (ETS) promoted the end-of-pipe emissions reduction? Evidence from China's residents. *Energy*, *277*, 127665.

Wu, Q. (2022). Price and scale effects of China's carbon emission trading system pilots on emission reduction. *Journal of environmental management*, *314*, 115054.

Xue, X., Reed, W. R., & van Aert, R. C. (2023). Social capital and economic growth: A meta-analysis. *Journal of Economic Surveys*.

Yang, Y., Macleod, M., Pan, J., Lagisz, M., & Nakagawa, S. (2022). Advanced methods and implementations for the meta-analyses of animal models: current practices and future recommendations. *Neuroscience & Biobehavioral Reviews*, 105016.

Zeugner, S., & Feldkircher, M. (2015). Bayesian model averaging employing fixed and flexible priors: The BMS package for R. *Journal of statistical software*, *68*, 1-

37.
Zhang, D., Karplus, V. J., Cassisa, C., & Zhang, X. (2014). Emissions trading in China: Progress and prospects. *Energy Policy*, *75*, 9-16.
Zhang, H., Duan, M., & Deng, Z. (2019). Have China's pilot emissions trading schemes promoted carbon emission reductions?–the evidence from industrial sub-sectors at the provincial level. *Journal of Cleaner Production*, *234*, 912-924.
Zhang, Y.-J., & Cheng, H.-S. (2021). The impact mechanism of the ETS on CO2 emissions from the service sector: evidence from Beijing and Shanghai. *Technological Forecasting and Social Change*, *173*, 121114.
Zhang, Y.-J., Liang, T., Jin, Y.-L., & Shen, B. (2020). The impact of carbon trading on economic output and carbon emissions reduction in China's industrial sectors. *Applied Energy*, *260*, 114290.
Zhou, A., Xin, L., & Li, J. (2022). Assessing the impact of the carbon market on the improvement of China's energy and carbon emission performance. *Energy*, *258*, 124789.
Zhu, L. (2017). *Does China's Carbon Emission Trading Scheme Reduce Emissions? Analyzing the Impact of the ETS Pilots*. Georgetown University.

Table 1. Description of Variables

Variable	Description	Mean	Min	Max
Dependent Variable				
1) Total_CO_2	=1, if Dep. Variable is total CO_2 emissions	0.675	0	1
2) CO_2_per_capita	=1, if Dep. Variable is CO_2 emissions per capita	0.063	0	1
3) CO_2_per_GDP	=1, if Dep. Variable is CO_2 emissions per GDP	0.260	0	1
--- Other_Dep*	=1, if Dep. Variable is other measure	0.005	0	1
4) Log	=1, if Dep. Variable is in natural Logarithm	0.738	0	1
Emissions Data Source				
5) Source_Selfcalc	=1, if CO_2 emission is calculated by the author(s)	0.634	0	1
6) Source_CEADs	=1, if CO_2 emission is from the CEADs database	0.254	0	1
--- Source_Other*	=1, if CO_2 emission is from other source	0.108	0	1
7) Direct_Process	=1, if "CO_2 emissions" includes both direct emission from fossil fuel combustion and process emissions from cement production	0.297	0	1
8) Direct_Indirect	=1, if "CO_2 emissions" includes both direct emission from fossil fuel combustion and indirect emissions from electricity consumption	0.095	0	1
--- Direct_Emissions*	=1, if "CO_2 emissions" only includes direct emission from fossil fuel combustion	0.499	0	1
Treatment Year & Units				
9) Treat_Trading	=1, if treatment year is 2013 or 2014	0.578	0	1
10) Treat_Multiperiod	=1, if treatment year is after 2012 and is set differently for different treated unit	0.181	0	1
--- Treat_Announce*	=1, if treatment year is 2011 or 2012	0.240	0	1
11) FJ_treated	=1, if Fujian is regarded as the treated unit	0.104	0	1
12) FJ_SC_treated	=1, if both Fujian and Sichuan are regarded as the treated units	0.045	0	1
--- Both_nontreated	=1, if neither Fujian nor Sichuan is regarded as the treated unit	0.851	0	1
Study Level				
13) Firm_level	=1, if data are firm level	0.142	0	1
14) City_level	=1, if data are city (or county) level	0.225	0	1
--- Province_level*	=1, if data are province level	0.525	0	1
15) Sector_level	=1, if data are provincial-sector level	0.108	0	1
16) ETS_Sector	=1, if data only includes one or several ETS-covered industrial sectors instead of the whole economy	0.157	0	1
Empirical Models & Methodology				
17) Triple DID	=1, if data is fitted by a triple DID model	0.018	0	1
18) One-way FE	=1, if regression model includes a Post*Treated dummy and controls for either time or unit FE	0.044	0	1
19) Instrumental	=1, if address for endogeneity	0.012	0	1

20) Spatial Model	=1, if data is fitted by a spatial model	0.045	0	1
21) Matching	=1, if treated and control units are matched before regression	0.228	0	1
22) Interaction only	=1, if regression model includes only a Post*Treated dummy (without neither time or unit FE), or is regressed using a random-effect model	0.064	0	1
--- Standard DID*	=1, if data is fitted by a standard DID model	0.589	0	1
Control Variables				
23) GDP	=1, if controls for GDP or total output	0.612	0	1
24) GDP_square	=1, if controls for squared GDP or squared total output	0.090	0	1
25) Population	=1, if controls for total population, population growth or population density	0.386	0	1
26) Urbanization	=1, if controls for level of urbanization	0.180	0	1
27) Industrial_Structure	=1, if controls for indices of industrial structure	0.535	0	1
28) Energy_Consump	=1, if controls for energy intensity (energy consumption/GDP) or energy structure (coal/total energy consumption)	0.266	0	1
29) Envi_Regulations	=1, if controls for the level/stringency of environmental regulation	0.232	0	1
30) Openness	=1, if controls for FDI, export (or import)/GDP	0.430	0	1
31) R&D	=1, if controls for level of R&D, No. of patents, etc.	0.246	0	1
32) Fixed_Asset_Ratio	=1, if controls for the ratio of fixed asset to GDP (or gross output)	0.072	0	1
33) State_Owned_Ratio	=1, if controls for the ratio of state-owned asset to GDP (or gross output)	0.010	0	1
34) FYP_Targets	=1, if controls for the Five-Year-Plan carbon intensity (or energy intensity) mitigation targets or FYP dummy	0.037	0	1
35) LCPC	=1, if controls for the Low-Carbon Provinces and Cities pilot program	0.011	0	1
Publication Status				
36) Published	=1, if the study is a published journal article	0.907	0	1
37) Publication Year	The year of publication	2021	2019	2023

Note: The asterisk * marks the baseline category among a group of dummies stemming from the same categorial variable.

Table 1. Estimating Overall Mean Effect

	OLS (1)	FE (2)	RE (3)	HIER (4)	CHE (5)
	-0.146*** (0.015)	-0.041** (0.010)	-0.128*** (0.013)	-0.157*** (0.011)	-0.141*** (0.010)
I^2	-	93.4%	98.8%	Within: 58.7% Between: 40.1%	Within: 67.8% Between: 31.0%
Estimates	732	732	732	732	732
Studies	80	80	80	80	80

Note: The dependent variable is Fisher's z. Standard errors are estimated using the "CR2" cluster robust standard error estimator and are reported in parentheses. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively.

Table 3. Study Weights from FE and RE Models

FE			**RE**		
Author	No. of Estimates	Study weight	Author	No. of Estimates	Study weight
Shi et al. (2022)	25	37.5%	Cui et al. (2021)	46	1.70%
Cui et al. (2021)	46	23.2%	Shi et al. (2022)	25	1.66%
Gao et al. (2020)	24	7.9%	Yang et al. (2023)	24	1.65%
Sum of Top 3		68.6%	Sum of Top 3		5.01%

Note: This table reports the top 3 studies with the highest weights (%) in the fixed-effect model (FE) and random-effects model (RE).

Table 4. Bias and Effect beyond Bias from the PET Model

	OLS (1)	FE (2)	RE (3)	HIER (4)	CHE (5)	OLS (6)	FE (7)	RE (8)	HIER (9)	CHE (10)
Constant (Effect beyond bias)	-0.009 (0.012)	-0.006 (0.004)	-0.010 (0.009)	-0.048*** (0.015)	-0.040** (0.013)	-0.020 (0.022) [a]	-0.021 (0.016) [a]	-0.027 (0.025) [a]	-0.064*** (0.031) [a]	-0.062*** (0.031) [a]
SE (Bias)	-2.782*** (0.295)	-2.864*** (0.185)	-2.755*** (0.239)	-2.193*** (0.294)	-2.286*** (0.275)	-2.559*** (0.458)	-2.545*** (0.301)	-2.418*** (0.433)	-1.710*** (0.442)	-1.718** (0.465)
Estimates	732	732	732	732	732	732	732	732	732	732
Studies	80	80	80	80	80	80	80	80	80	80

Note: The dependent variable is Fisher's z. Standard errors are estimated using the "CR2" cluster robust standard error estimator and are reported in parentheses. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively. Columns (1)-(5) corresponds to estimates from the univariate regression; columns (6)-(10) multivariate regression.

Table 5. Alternative Bias-adjusted Estimates of the Overall Mean Effect

	CHE (PEESE) (1)	**Selection Model (2)**	**RoBMA (3)**	**MAIVE (4)**
μ (Bias-adjusted estimate of the overall mean)	-0.111*** (0.010)	-0.057*** (0.009)	-0.101*** [-0.110, -0.092]	-0.080*** (0.008)
Summary	Negative, significant	Negative, significant	Negative, significant	Negative, significant

Note: The dependent variable is Fisher's z. Column (3) reports 95% confidence intervals of the estimate. In Column (1), (2) and (4), standard errors are estimated using the "CR2" cluster robust standard error estimator and are reported in parentheses. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively. The estimates in all four columns are from univariate regression.

Table 6. Bayesian Model Averaging

	OLS		BMA		
	Coeff	**SE**	**PIP**	**Post Mean**	**Post SD**
Fisher's z SE	-2.559***	0.458	1.000	-2.710	0.146
Total_CO_2	0.024	0.033	0.024	0.000	0.002
CO_2_per_capita	0.065	0.048	0.313	0.014	0.024
CO_2_per_GDP	0.016	0.033	0.021	0.000	0.002
Log	0.001	0.022	0.027	0.000	0.002
Source_Selfcalc	-0.016	0.043	0.020	0.000	0.002
Source_CEADs	-0.012	0.053	0.018	0.000	0.002
Direct_Process	0.026	0.067	0.019	0.000	0.001
Direct_Indirect	0.045	0.065	0.033	0.001	0.004
Direct_only	0.021	0.053	0.019	0.000	0.001
Treat_Trading	-0.003	0.022	0.020	0.000	0.002
Treat_Multiperiod	0.017	0.035	0.019	0.000	0.002
FJ_treated	-0.027	0.041	0.025	0.000	0.003
FJ_SC_treated	-0.002	0.039	0.019	0.000	0.003
Firm_level	0.036	0.036	0.030	0.000	0.004
City_level	0.037	0.031	0.021	0.000	0.002
Sector_level	-0.009	0.048	0.021	0.000	0.002
ETS_Sector	0.027	0.033	0.025	0.000	0.003
Triple DID	-0.002	0.027	0.017	0.000	0.005
Instrumental	0.056*	0.027	0.038	0.002	0.012
Interaction only	-0.012	0.040	0.030	-0.001	0.004
Matching	-0.029	0.026	0.074	-0.001	0.006
One-way FE	-0.029	0.062	0.037	-0.001	0.007
Spatial Model	-0.004	0.026	0.018	0.000	0.003
GDP	0.020	0.024	0.093	0.002	0.006
GDP_square	0.043	0.032	0.260	0.010	0.018
Population	-0.039*	0.020	0.927	-0.032	0.014
Urbanization	-0.016	0.021	0.028	0.000	0.003
Industrial_Structure	-0.011	0.020	0.020	0.000	0.002
Energy_Consump	0.031	0.023	0.102	0.002	0.006
Envi_Regulations	0.000	0.024	0.019	0.000	0.002
Openness	0.017	0.024	0.029	0.000	0.002
R&D	-0.027	0.027	0.164	-0.004	0.010
Fixed_Asset_Ratio	-0.046	0.029	0.032	-0.001	0.005
State_Owned_Ratio	-0.022	0.073	0.017	0.000	0.007
FYP_Targets	-0.022	0.036	0.018	0.000	0.003
LCPC	-0.013	0.019	0.016	0.000	0.006
Published	-0.017	0.030	0.020	0.000	0.002
Publication Year	-0.010	0.010	0.053	0.000	0.002

Note: The dependent variable is Fisher's z. The column headings *Post Mean, Post SD* and *PIP* stand for Posterior Mean, Posterior Standard Deviation, and Posterior Inclusion Probability. The OLS standard errors in Column (5) are estimated using the "CR2" cluster robust standard error estimator. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively.

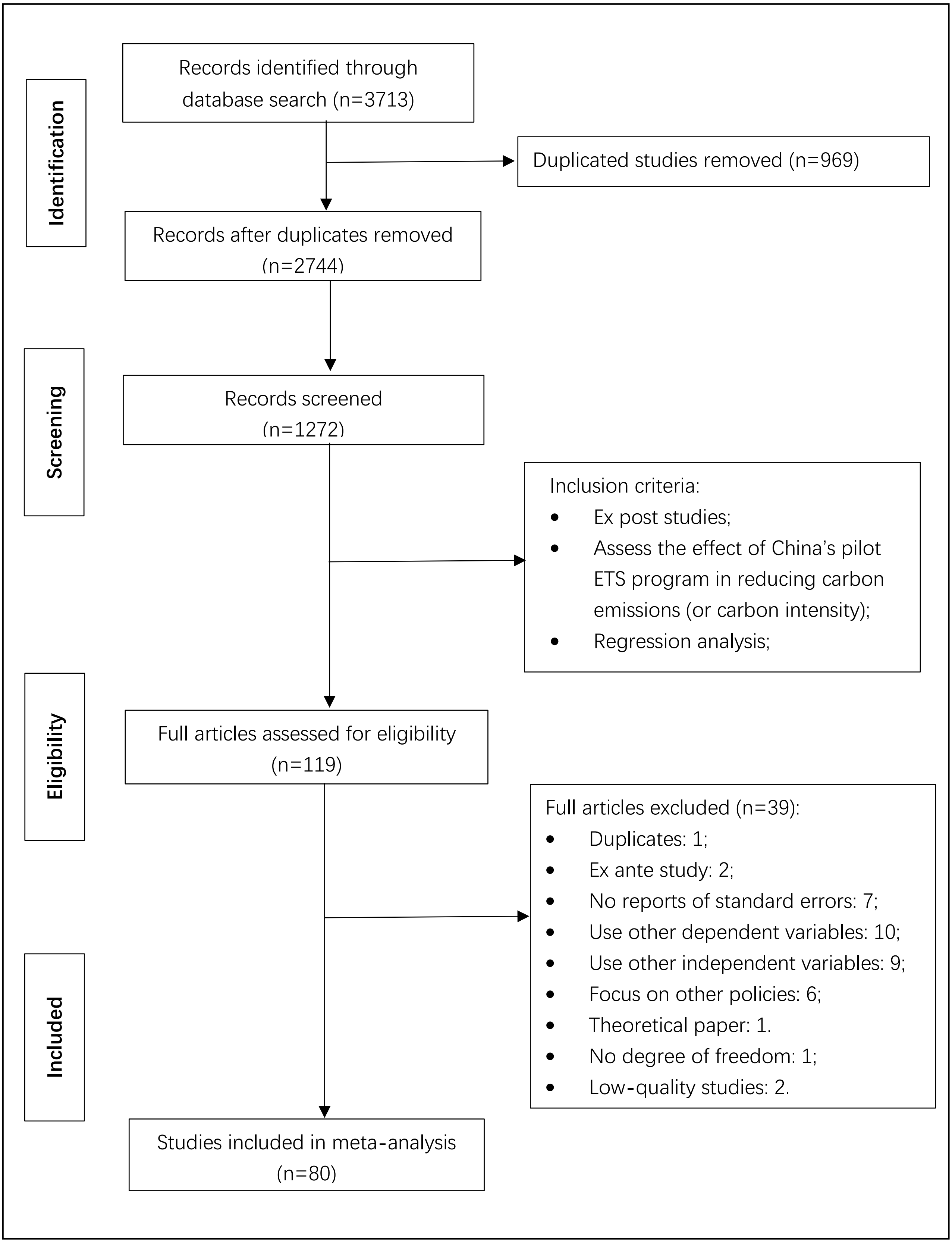


Figure 1. PRISMA Flow Diagram

Figure 2. Distribution of Fisher's z

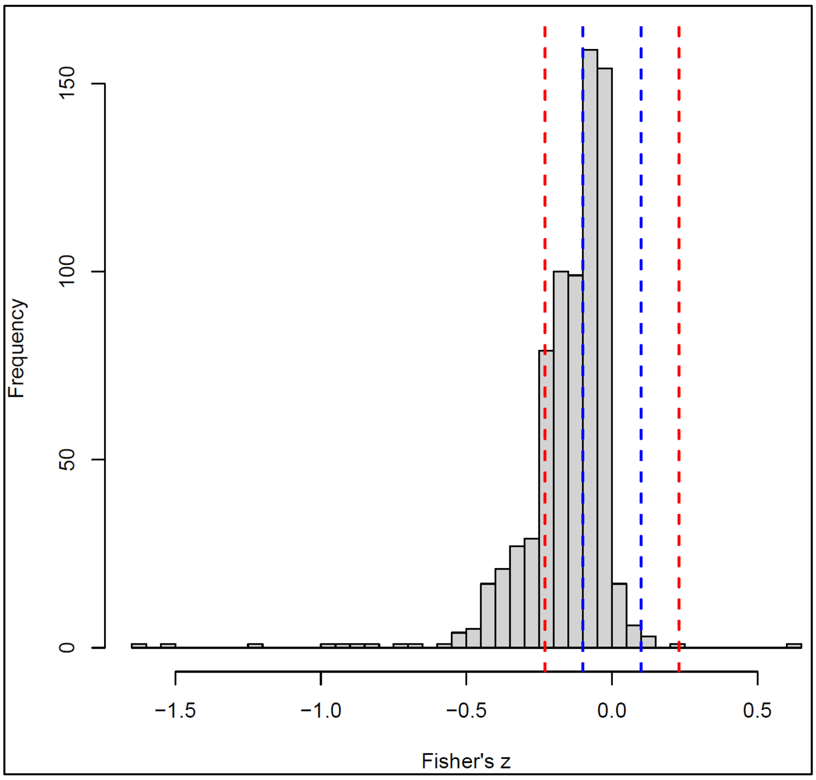
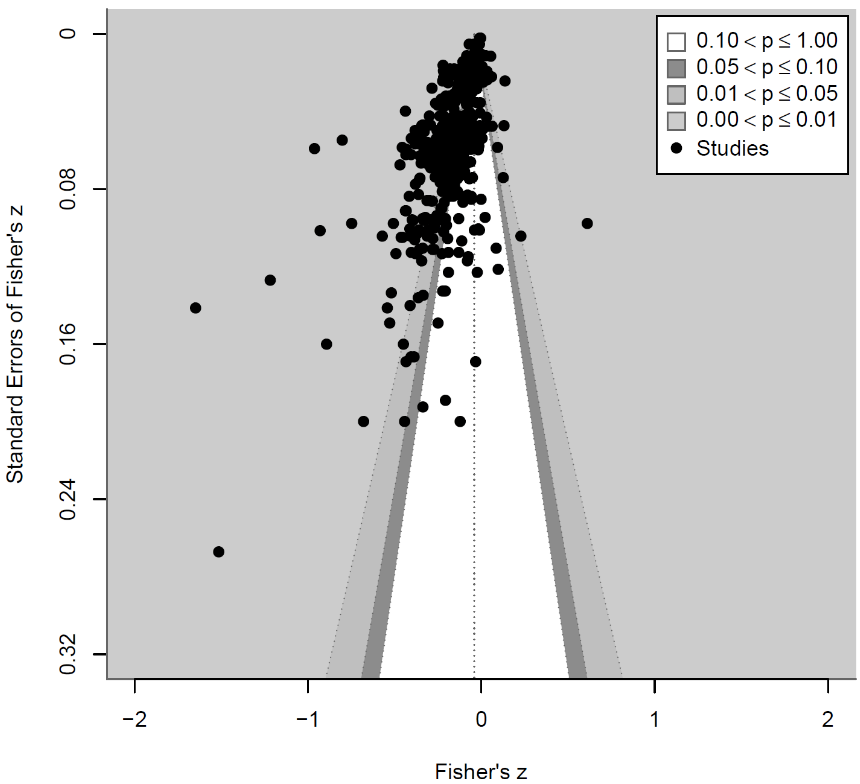


Figure 3. Funnel Plot from the FE Model (Full Sample)



Appendix

Part A: Effect Sizes Transformation

1. Transforming regression coefficients into partial correlation coefficients (PCC) (Xue et al., 2023):

$$PCC_i = \frac{t_i}{\sqrt{t_i^2 + df_i}}, \quad \text{A.1}$$

$$se(PCC_i) = \frac{1 - PCC_i^2}{\sqrt{df_i}} \quad \text{A.2}$$

Here, t_i and df_i are the *t*-statistic and degrees of freedom associated with the respective effect size.

2. Transforming PCC into Fisher's z (van Aert, 2023):

$$z_i = 0.5 \times log\,(\frac{1 + PCC_i}{1 - PCC_i}) \quad \text{A.3}$$

$$se(z_i) = \frac{1}{\sqrt{df_i - 3}} \quad \text{A.4}$$

Here, z_i stands for Fisher's z.

Part B: Meta-analytic Models

1. PET-PEESE

PET: $$\hat{\alpha}_i = \alpha_0 + \beta SE(\hat{\alpha})_i + \varepsilon_i \quad \text{A.5}$$

PEESE: $$\hat{\alpha}_i = \alpha_0 + \beta SE(\hat{\alpha})_i^2 + \varepsilon_i \quad \text{A.6}$$

2. The MAIVE estimator

$$SE(\hat{\alpha})_i^2 = \psi_0 + \psi_1 \left(\frac{1}{N_i}\right) + \upsilon_i \quad \text{A.7}$$

Here, i stands for ith study. $\hat{\alpha}_i$ stands for effect sizes and $SE(\hat{\alpha})_i$ is the standard errors of the effect size.

Part C: Results from Subsample Analysis

Table A1. Estimating overall mean effect (Subsample)

	OLS (1)	FE (2)	RE (3)	HIER (4)	CHE (5)
constant	-0.188*** (0.023)	-0.048** (0.018)	-0.153*** (0.017)	-0.161*** (0.017)	-0.154*** (0.017)
I^2	-	98.1%	98.1%	Within: 15.4% Between: 82.9%	Within: 29.1% Between: 69.4%
Estimates	540	540	540	540	540
Studies	57	57	57	57	57

Note: The dependent variable measures the change in carbon emissions (in log) caused by the implementation of ETS. Standard errors are estimated using the "CR2" cluster robust standard error estimator and are reported in parentheses. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively.

Table A2. Study Weights (Subsample)

FE			**RE**		
Author	No. of Estimates	Study weight (%)	Author	No. of Estimates	Study weight
Han et al. (2023)	12	46.62	Cui et al. (2021)	46	2.24
Li & Wang (2022)	6	9.49	Yang et al. (2023)	24	2.19
Yang et al. (2023)	24	7.86	Dong et al. (2022)	25	2.17
Sum of Top 3		63.97	Sum of Top 3		6.6

Note: This table reports the top 3 studies with the highest weights (%) in the fixed-effect model (FE) and random-effects model (RE) in the subsample analysis.

Table A3. Bias and Effect beyond Bias from the PET Model (Subsample)

	OLS (1)	FE (2)	RE (3)	HIER (4)	CHE (5)	OLS (6)	FE (7)	RE (8)	HIER (9)	CHE (10)
Constant (Effect beyond bias)	-0.149*** (0.028)	-0.019 (0.008)	-0.073*** (0.016)	-0.119*** (0.019)	-0.119*** (0.023)	-0.164*** (0.023)	-0.052*** (0.020)	-0.093*** (0.028)	-0.123*** (0.040)	-0.131*** (0.055)
SE (Bias)	-0.390 (0.295)	-2.522*** (0.282)	-1.481*** (0.358)	-0.750** (0.274)	-0.897** (0.352)	-0.240 (0.138)[a]	-1.843*** (0.343)[a]	-1.188*** (0.317)[a]	-0.698** (0.270)[a]	-0.811** (0.350)[a]
Estimates	540	540	540	540	540	540	540	540	540	540
Studies	57	57	57	57	57	57	57	57	57	57

Note: The dependent variable measures the change in carbon emissions (in log) caused by the implementation of ETS. Standard errors are estimated using the "CR2" cluster robust standard error estimator and are reported in parentheses. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively. Columns (1)-(5) corresponds to estimates from the univariate regression; columns (6)-(10) multivariate regression.

Table A4. Alternative Bias-adjusted Estimates of the Overall Mean Effect (Subsample)

	CHE (PEESE) **(1)**	**Selection Model** **(2)**	**RoBMA** **(3)**	**MAIVE** **(4)**	**OLS(PEESE)** **(5)**
μ	-0.155*** (0.017)	-0.057*** (0.012)	-0.113*** [-0.125, -0.102]	0.272 (0.431)	-0.181*** (0.023)
Summary	Negative, significant	Negative, significant	Negative, significant	Positive, insignificant	Negative, significant

Note: The dependent variable measures the change in carbon emissions (in log) caused by the implementation of ETS. Column (3) reports 95% confidence intervals of the estimate. In Column (1), (2), (4) and (5), standard errors are estimated using the "CR2" cluster robust standard error estimator and are reported in parentheses. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively. The estimates in all four columns are from univariate regression.

Table A5. Results from Bayesian Model Averaging (Subsample)

	OLS		BMA		
	Coeff	**SE**	**PIP**	**Post Mean**	**Post SD**
Standard Error	-0.240	0.138	1.000	-0.250	0.068
Total_CO_2	0.034	0.037	0.023	0.000	0.005
CO_2_per_capita	0.018	0.076	0.014	0.000	0.005
CO_2_per_GDP	0.076	0.043	0.027	0.001	0.006
Source_Selfcalc	-0.013	0.088	0.019	0.000	0.004
Source_CEADs	0.104	0.160	0.021	0.000	0.005
Direct_Process	-0.148	0.137	0.027	-0.001	0.006
Direct_Indirect	0.099	0.086	0.658	0.072	0.061
Treat_Trading	-0.057	0.059	0.119	-0.005	0.016
Treat_Multiperiod	-0.033	0.066	0.060	0.002	0.012
FJ_treated	0.009	0.056	0.064	0.004	0.017
FJ_SC_treated	-0.003	0.086	0.015	0.000	0.006
Firm_level	0.006	0.064	0.023	0.001	0.007
City_level	0.030	0.057	0.059	0.002	0.011
Sector_level	0.061	0.061	0.015	0.000	0.007
ETS_Sector	-0.094	0.087	0.197	-0.014	0.033
Triple DID	-0.093	0.088	0.015	-0.001	0.023
Instrumental	-1.132	0.503	1.000	-1.129	0.108
Interaction only	-0.035	0.077	0.017	0.000	0.007
Matching	-0.138*	0.075	0.960	-0.096	0.036
One-way FE	-0.057	0.039	0.021	-0.001	0.010
Spatial Model	-0.113	0.078	0.340	-0.045	0.071
GDP	-0.017	0.039	0.044	-0.001	0.009
GDP_square	0.006	0.075	0.018	0.000	0.006
Population	-0.036	0.043	0.023	0.000	0.005
Urbanization	0.113*	0.054	0.229	0.018	0.037
Industrial_Structure	-0.032	0.031	0.140	-0.007	0.019
Energy_Consump	0.039	0.031	0.022	0.000	0.004
Envi_Regulations	0.028	0.053	0.018	0.000	0.004
Openness	0.004	0.032	0.017	0.000	0.003
R&D	-0.003	0.053	0.019	0.000	0.005
Fixed_Asset_Ratio	-0.180*	0.072	0.995	-0.161	0.043
State_Owned_Ratio	0.094	0.100	0.025	0.002	0.022
FYP_Targets	-0.126	0.070	0.048	-0.006	0.037
LCPC	-0.167	0.102	0.042	-0.006	0.037
Published	0.088	0.074	0.580	0.054	0.053
Publication Year	0.003	0.014	0.016	0.000	0.002

Note: The dependent variable measures the change in carbon emissions (in log) caused by the implementation of ETS. The column headings *Post Mean, Post SD* and *PIP* stand for Posterior Mean, Posterior Standard Deviation, and Posterior Inclusion Probability. The OLS standard errors in Column (5) are estimated using the "CR_2" cluster robust standard error estimator. *, **, and *** indicate statistical significance at the 10-, 5-, and 1-percent level, respectively.

Part D: Studies included in this meta-analysis

[1] Chai, S., Sun, R., Zhang, K., Ding, Y., & Wei, W. (2022). Is emissions trading scheme (ETS) an effective market-incentivized environmental regulation policy? Evidence from China's eight ETS pilots. International Journal of Environmental Research and Public Health, 19(6), 3177.

[2] Chen, G., & Hibiki, A. (2022). Can the carbon emission trading scheme influence industrial green production in China? Sustainability, 14(23), 15829.

[3] Chen, J., Gui, W., & Huang, Y. (2023). The impact of the establishment of carbon emission trade exchange on carbon emission efficiency. Environmental Science and Pollution Research, 30(8), 19845-19859.

[4] Chen, L., Wang, D., & Shi, R. (2022). Can China's carbon emissions trading system achieve the synergistic effect of carbon reduction and pollution control? International Journal of Environmental Research and Public Health, 19(15), 8932.

[5] Chen, S., Shi, A., & Wang, X. (2020). Carbon emission curbing effects and influencing mechanisms of China's Emission Trading Scheme: The mediating roles of technique effect, composition effect and allocation effect. Journal of Cleaner Production, 264, 121700.

[6] Chen, Y., Xu, Z., Zhang, Z., Ye, W., Yang, Y., & Gong, Z. (2022). Does the carbon emission trading scheme boost corporate environmental and financial performance in China? Journal of Cleaner Production, 368, 133151.

[7] Cong, J., Wang, H., Hu, X., Zhao, Y., Wang, Y., Zhang, W., & Zhang, L. (2023). Does China's Pilot Carbon Market Cause Carbon Leakage? New Evidence from the Chemical, Building Material, and Metal Industries. International Journal of Environmental Research and Public Health, 20(3), 1853.

[8] Cui, J., Wang, C., Zhang, J., & Zheng, Y. (2021). The effectiveness of China's regional carbon market pilots in reducing firm emissions. Proceedings of the National Academy of Sciences, 118(52), e2109912118.

[9] Dong, Z., Xia, C., Fang, K., & Zhang, W. (2022). Effect of the carbon emissions trading policy on the co-benefits of carbon emissions reduction and air pollution control. Energy Policy, 165, 112998.

[10] Feng, R., Lin, P., Hou, C., & Jia, S. (2022). Study of the effect of China's emissions trading scheme on promoting regional industrial carbon emission reduction. Frontiers in Environmental Science, 10, 947925.

[11] Feng, Y. (2020). Does China's carbon emission trading policy alleviate urban carbon emissions? IOP Conference Series: Earth and Environmental Science,

[12] Gao, Y., Li, M., Xue, J., & Liu, Y. (2020). Evaluation of effectiveness of China's carbon emissions trading scheme in carbon mitigation. Energy Economics, 90, 104872.

[13] Han, Y., Tan, S., Zhu, C., & Liu, Y. (2022). Research on the emission reduction effects of carbon trading mechanism on power industry: plant-level evidence from China. International Journal of Climate Change Strategies and Management, 15(2), 212-231.

[14] He, Y., & Song, W. (2022). Analysis of the impact of carbon trading policies on

carbon emission and carbon emission efficiency. Sustainability, 14(16), 10216.
[15] Hu, Y., Li, R., Du, L., Ren, S., & Chevallier, J. (2022). Could SO2 and CO2 emissions trading schemes achieve co-benefits of emissions reduction? Energy Policy, 170, 113252.
[16] Hu, Y., Ren, S., Wang, Y., & Chen, X. (2020). Can carbon emission trading scheme achieve energy conservation and emission reduction? Evidence from the industrial sector in China. Energy Economics, 85, 104590.
[17] Lan, J., Li, W., & Zhu, X. (2022). The road to green development: How can carbon emission trading pilot policy contribute to carbon peak attainment and neutrality? Evidence from China. Frontiers in Psychology, 13, 962084.
[18] Li, L., Dong, J., & Song, Y. (2020). Impact and acting path of carbon emission trading on carbon emission intensity of construction land: Evidence from pilot areas in China. Sustainability, 12(19), 7843.
[19] Li, S., Liu, J., Wu, J., & Hu, X. (2022). Spatial spillover effect of carbon emission trading policy on carbon emission reduction: Empirical data from transport industry in China. Journal of Cleaner Production, 371, 133529.
[20] Li, X., Shu, Y., & Jin, X. (2022). Environmental regulation, carbon emissions and green total factor productivity: A case study of China. Environment, Development and Sustainability, 24(2), 2577-2597.
[21] Li, Z., & Liu, B. (2023). Understanding Carbon Emissions Reduction in China: Perspectives of Political Mobility. Land, 12(4), 903.
[22] Li, Z., & Wang, J. (2021). Spatial emission reduction effects of China’s carbon emissions trading: Quasi-natural experiments and policy spillovers. Chinese Journal of Population, Resources and Environment, 19(3), 246-255.
[23] Li, Z., & Wang, J. (2022). Spatial spillover effect of carbon emission trading on carbon emission reduction: Empirical data from pilot regions in China. Energy, 251, 123906.
[24] Li, Z., Wang, J., & Che, S. (2021). Synergistic effect of carbon trading scheme on carbon dioxide and atmospheric pollutants. Sustainability, 13(10), 5403.
[25] Lin, B., & Huang, C. (2022). Analysis of emission reduction effects of carbon trading: Market mechanism or government intervention? Sustainable Production and Consumption, 33, 28-37.
[26] Ling, Y., & Zhou, W. (2023). Verification of the effectiveness of the carbon emissions trading pilot system in China toward the Paris Agreement reduction target.
[27] Ling, Y., Zhou, W., & Qian, X. (2021). Design and Analysis of a Carbon Emissions Trading System for Low-Carbon Development in China. East Asian Low-Carbon Community: Realizing a Sustainable Decarbonized Society from Technology and Social Systems, 273-288.
[28] Liu, K., Zhao, L., & Li, Y. (2022). Research on the Emission Reduction Effect and Mechanism of China's Carbon Emission Trading Pilot. Available at SSRN 4202306.
[29] Liu, L., Feng, T., & Kong, J. (2023). Can carbon trading policy and local public expenditures synergize to promote carbon emission reduction in the power industry? Resources, Conservation and Recycling, 188, 106659.

[30] Liu, Y., Yuan, L., & Tan, Y. (2023). Synergistic Governance Effects of Carbon Emission Trading on Pollution and Carbon Reduction in China: Quasi-Natural Experiments and Policy Spillovers. Yaxin, Synergistic Governance Effects of Carbon Emission Trading on Pollution and Carbon Reduction in China: Quasi-Natural Experiments and Policy Spillovers.
[31] Liu, Z., Cai, L., & Zhang, Y. (2023). Co-benefits of China's carbon emissions trading scheme: Impact mechanism and spillover effect. International Journal of Environmental Research and Public Health, 20(5), 3792.
[32] Luo, H., & Qu, X. (2023). The impact of carbon emissions trading pilot on the carbon intensity of China's energy-intensive manufacturing industry.
[33] Luo, K., Xu, A., Ye, R., & Li, W. (2022). Monitoring the CO2 emission trajectory and reduction effects by ETS and its market performances for pre-and post-pandemic China. Frontiers in Public Health, 10, 848211.
[34] Ma, Q., Yan, G., Ren, X., & Ren, X. (2022). Can China's carbon emissions trading scheme achieve a double dividend? Environmental Science and Pollution Research, 29(33), 50238-50255.
[35] Ma, Y., Feng, H., Meng, Y., & Yue, L. (2023). Analysis of the spatio-temporal evolution of sustainable land use in China under the carbon emission trading scheme: A measurement idea based on the DID model. Plos one, 18(5), e0285688.
[36] Nie, X., Chen, Z., Wang, H., Wu, J., Wu, X., Lu, B., Qiu, L., & Li, Y. (2022). Is the "pollution haven hypothesis" valid for China's carbon trading system? A re-examination based on inter-provincial carbon emission transfer. Environmental Science and Pollution Research, 29(26), 40110-40122.
[37] Qi, S., Cheng, S., & Cui, J. (2021). Environmental and economic effects of China's carbon market pilots: Empirical evidence based on a DID model. Journal of Cleaner Production, 279, 123720.
[38] Ren, F., & Liu, X. (2023). Evaluation of carbon emission reduction effect and porter effect of China's carbon trading policy. Environmental Science and Pollution Research, 30(16), 46527-46546.
[39] Shahid, R., Li, S., Gao, J., Altaf, M. A., Jahanger, A., & Shakoor, A. (2022). The carbon emission trading policy of China: does it really boost the environmental upgrading? Energies, 15(16), 6065.
[40] Shao, Q., & Zhang, Z. (2022). Carbon mitigation and energy conservation effects of emissions trading policy in China considering regional disparities. Energy and Climate Change, 3, 100079.
[41] Shen, J., Tang, P., & Zeng, H. (2020). Does China's carbon emission trading reduce carbon emissions? Evidence from listed firms. Energy for Sustainable Development, 59, 120-129.
[42] Shi, B., Li, N., Gao, Q., & Li, G. (2022). Market incentives, carbon quota allocation and carbon emission reduction: evidence from China's carbon trading pilot policy. Journal of environmental management, 319, 115650.
[43] Tang, A., & Xu, N. (2023). The impact of environmental regulation on urban green efficiency—Evidence from carbon pilot. Sustainability, 15(2), 1136.
[44] Tang, K., Liu, Y., Zhou, D., & Qiu, Y. (2021). Urban carbon emission intensity

under emission trading system in a developing economy: evidence from 273 Chinese cities. Environmental Science and Pollution Research, 28, 5168-5179.
[45] Tang, K., Zhou, Y., Liang, X., & Zhou, D. (2021). The effectiveness and heterogeneity of carbon emissions trading scheme in China. Environmental Science and Pollution Research, 28, 17306-17318.
[46] Tao, M., & Goh, L. T. (2023). Effects of Carbon Trading Pilot on Carbon Emission Reduction: Evidence from China's 283 Prefecture-Level Cities. The Chinese Economy, 56(1), 1-24.
[47] Tian, G., Yu, S., Wu, Z., & Xia, Q. (2022). Study on the emission reduction effect and spatial difference of carbon emission trading policy in China. Energies, 15(5), 1921.
[48] Wang, C., Shi, Y., Zhang, L., Zhao, X., & Chen, H. (2021). The policy effects and influence mechanism of China's carbon emissions trading scheme. Air Quality, Atmosphere & Health, 14(12), 2101-2114.
[49] Wang, J., Wang, Y., & Song, J. (2023). The policy evaluation of China's carbon emissions trading scheme on firm employment: A channel from industrial automation. Energy Policy, 178, 113590.
[50] Wang, L. (2023). Low carbon management of China's hotel tourism through carbon emission trading. Sustainability, 15(5), 4622.
[51] Wang, Q., Gao, C., & Dai, S. (2019). Effect of the emissions trading scheme on CO2 abatement in China. Sustainability, 11(4), 1055.
[52] Wang, X., Huang, J., & Liu, H. (2022). Can China's carbon trading policy help achieve Carbon Neutrality?—A study of policy effects from the Five-sphere Integrated Plan perspective. Journal of environmental management, 305, 114357.
[53] Wang, Z., Liang, L., Cheng, D., Li, H., & Zhang, Y. (2022). Emission reduction benefits and economic benefits of China's pilot policy on carbon emission trading system. Computational Intelligence and Neuroscience, 2022(1), 5280900.
[54] Weng, Z., Liu, T., & Cheng, C. (2022). Reduction effect of carbon markets: A case study of the Beijing-Tianjin-Hebei region of China. Frontiers in Environmental Science, 10, 1013708.
[55] Wu, L., & Zhu, Q. (2023). Has the Emissions Trading Scheme (ETS) promoted the end-of-pipe emissions reduction? Evidence from China's residents. Energy, 277, 127665.
[56] Wu, Q., Tambunlertchai, K., & Pornchaiwiseskul, P. (2021). Examining the impact and influencing channels of carbon emission trading pilot markets in China. Sustainability, 13(10), 5664.
[57] Wu, Y., Peng, B., & Lao, Y. (2023). The Emission Reduction Effect of Financial Agglomeration under China's Carbon Peak and Neutrality Goals. International Journal of Environmental Research and Public Health, 20(2), 950.
[58] Xia, Q., Li, L., Dong, J., & Zhang, B. (2021). Reduction effect and mechanism analysis of carbon trading policy on carbon emissions from land use. Sustainability, 13(17), 9558.
[59] Xu, W. (2021). The impact and influencing path of the pilot carbon emission trading market——Evidence from China. Frontiers in Environmental Science, 9,

787655.
[60] Xuan, D., Ma, X., & Shang, Y. (2020). Can China’s policy of carbon emission trading promote carbon emission reduction? Journal of Cleaner Production, 270, 122383.
[61] Yang, F., & Wang, C. (2023). Clean energy, emission trading policy, and CO2 emissions: Evidence from China. Energy & Environment, 34(5), 1657-1673.
[62] Yang, X., Zhang, J., Bi, L., & Jiang, Y. (2023). Does China’s carbon trading pilot policy reduce carbon emissions? Empirical analysis from 285 cities. International Journal of Environmental Research and Public Health, 20(5), 4421.
[63] Yang, Z., Yuan, Y., & Zhang, Q. (2022). Carbon emission trading scheme, carbon emissions reduction and spatial spillover effects: quasi-experimental evidence from China. Frontiers in Environmental Science, 9, 824298.
[64] Yi, L., Bai, N., Yang, L., Li, Z., & Wang, F. (2020). Evaluation on the effectiveness of China’s pilot carbon market policy. Journal of Cleaner Production, 246, 119039.
[65] Yu, W., & Luo, J. (2022). Impact on carbon intensity of carbon emission trading—evidence from a pilot program in 281 cities in China. International Journal of Environmental Research and Public Health, 19(19), 12483.
[66] Zhang, H., Duan, M., & Deng, Z. (2019). Have China's pilot emissions trading schemes promoted carbon emission reductions?–the evidence from industrial sub-sectors at the provincial level. Journal of Cleaner Production, 234, 912-924.
[67] Zhang, H., & Liu, Y. (2022). Can the pilot emission trading system coordinate the relationship between emission reduction and economic development goals in China? Journal of Cleaner Production, 363, 132629.
[68] Zhang, H., & Wu, J. (2022). The energy saving and emission reduction effect of carbon trading pilot policy in China: evidence from a quasi-natural experiment. International Journal of Environmental Research and Public Health, 19(15), 9272.
[69] Zhang, H., Zhang, R., Li, G., Li, W., & Choi, Y. (2019). Sustainable feasibility of carbon trading policy on heterogenetic economic and industrial development. Sustainability, 11(23), 6869.
[70] Zhang, J., & Zhang, Y. (2020). Could the ETS reduce tourism-related CO2 emissions and carbon intensity? A quasi-natural experiment. Asia Pacific Journal of Tourism Research, 25(9), 1029-1041.
[71] Zhang, J., & Zhang, Y. (2021). Why does tourism have to confront the emissions trading scheme? Evidence from China. Tourism Management Perspectives, 40, 100876.
[72] Zhang, K., Qian, J., Zhang, Z., & Fang, S. (2023). The impact of carbon trading pilot policy on carbon neutrality: Empirical evidence from Chinese cities. International Journal of Environmental Research and Public Health, 20(5), 4537.
[73] Zhang, W., Li, G., & Guo, F. (2022). Does carbon emissions trading promote green technology innovation in China? Applied Energy, 315, 119012.
[74] Zhang, W., Li, J., Li, G., & Guo, S. (2020). Emission reduction effect and carbon market efficiency of carbon emissions trading policy in China. Energy, 196, 117117.
[75] Zhang, W., Zhang, N., & Yu, Y. (2019). Carbon mitigation effects and potential cost savings from carbon emissions trading in China's regional industry.

Technological Forecasting and Social Change, 141, 1-11.
[76] Zhang, Y., & Cheng, H. (2021). The impact mechanism of the ETS on CO2 emissions from the service sector: evidence from Beijing and Shanghai. Technological Forecasting and Social Change, 173, 121114.
[77] Zhang, Y., Li, S., Luo, T., & Gao, J. (2020). The effect of emission trading policy on carbon emission reduction: Evidence from an integrated study of pilot regions in China. Journal of Cleaner Production, 265, 121843.
[78] Zheng, Y., Tan, R., & Zhang, B. (2023). The joint impact of the carbon market on carbon emissions, energy mix, and copollutants. Environmental Research Letters, 18(4), 045007.
[79] Zhou, B., Zhang, C., Song, H., & Wang, Q. (2019). How does emission trading reduce China's carbon intensity? An exploration using a decomposition and difference-in-differences approach. Science of the total environment, 676, 514-523.
[80] Zhu, L. (2017). Does China's Carbon Emission Trading Scheme Reduce Emissions? Analyzing the Impact of the ETS Pilots. Georgetown University.

Declaration of generative AI and AI-assisted technologies in the writing process.

During the preparation of this work the authors used tools like ChatGPT in order to improve the readability and language of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the article.

Declaration of Conflicts of Interest

We have nothing to declare.