# Content



# Supplementary Notes

## Measures

Urban carbon emissions originate from direct energy consumption, such as gas and liquefied petroleum gas, and electricity and heat consumption. The relevant conversion factors provided by IPCC2006 can be used to calculate the carbon emissions of direct energy consumption.

We draw on Glaeser and Kahn's approach to calculate carbon emissions from electricity consumption [1], in which only one emission factor exists for each regional grid. The power network in China is divided into six regional grids: North China (0.7119 tCO2/MWh), Northeast China (0.66125 tCO2/MWh), East China (0.58955 tCO2/MWh), Central China (0.57205 tCO2/MWh), Northwest China (0.66645 tCO2/MWh), and South China (0.50885 tCO2/MWh). We use baseline emission factors of regional power grids and urban electricity consumption to calculate carbon emissions from city electricity consumption.

Urban thermal energy mainly includes heat from boiler rooms and cogeneration plants, and its raw materials are primarily raw coal. The China Urban Construction Statistical Yearbook provides statistical data on centralized heat supply. The minimum standard of thermal efficiency of coal-fired industrial boilers stipulated in GB/T 15317-2009 Coal-fired Industrial Boiler Energy Saving Monitoring is between 65% and 78%. Considering that centralized heating boilers in China are mainly middling and small coal boilers, we adopt a thermal efficiency value of 70% for calculation. The amount of raw coal needed can be calculated according to heat supply, thermal efficiency, and the calorific value coefficient of raw coal. Furthermore, the energy consumed for heating can be calculated using the standard coal coefficient of raw coal conversion (0.7143 kg of standard coal per kg). According to IPCC 2006, the carbon emission factor per kg of raw coal is 2.53 kg CO2/kg, and the calculation of carbon emissions from centralized heating can be derived from the amount of raw coal consumed for heat energy.

# Heterogeneity analysis

We use the assignment of the municipal party secretary in the current year to measure official promotion pressure (lopp) [2]. According to the study of these municipal party secretaries, the samples were divided into two groups: the first three years of the assignment and near the end of the assignment, representing low promotion pressure and high promotion pressure. The results display that the effect of vertical management auditing reform on carbon productivity is significant only in the group where officials have low promotion pressure. We use the mobile terminal

'environmental pollution' Baidu search index to measure the public scrutiny pressure (psp) on local governments regarding environmental governance [3]. A higher Baidu search index means the local government receives more public scrutiny. Grouped by a median, those with a Baidu search index greater than or equal to the median were categorized as the group with higher public scrutiny pressure. The results show that vertical management auditing reform can increase carbon productivity only in the group with higher public scrutiny pressure. Therefore, the reform can take effect more significantly in areas with solid public scrutiny pressure (see Supplementary Tab. 2).

# Placebo test

We conducted a placebo test by randomly setting up an experimental group to exclude the effect of the non-random selection of policy pilot areas. Because a total of 74 cities in the sample implemented vertical management auditing reform, we randomly selected 74 cities as pilot cities from a total sample of 281 cities to obtain a pseudo-processing group. Then, the above process is randomly sampled 500 times. As shown from the kernel density distribution and p-values of the 500 coefficient estimates, the regression coefficients follow a normal distribution (see Supplementary Fig. 4). They are concentrated around the zero point, most of whose p-values are insignificant. The actual coefficient estimates in the benchmark regression are located at the red vertical dashed line, which is inconsistent with the kernel density distribution of the spurious regression coefficients. The results rule out the possibility that the regression results are influenced by other unobservable factors, having robustness to the preceding conclusion.

# Alter fixed effects and clustered levels

In the two-way fixed effect model, clustering can identify trends in time and similarities and differences between individuals. Therefore, we further controlled the province's fixed effect, the interactive clustering standard error of the city and the year, the interactive clustering standard error of the city and the year, and the clustering standard error of the city to alleviate the problem of missing variables at high latitudes. The findings demonstrate that the above conclusions are valid (see Supplementary Fig. 5).

# Exclude policy interferences

Carbon productivity may also be affected by other environmental policies. We consider the following locational guidelines. First, carbon trading pilots have been proven to promote carbon

emission reduction [4]. In the sample of the experimental group, eight cities were also carbon trading pilot cities established before the auditing reform. Secondly, central environmental protection supervision significantly improves air pollution [5]. The central environmental protection supervision and inspection group began to conduct environmental inspections on local governments and relevant departments in 2016, with the timing of the inspections overlapping with the audit pilot period. Thirdly, China's Supreme People's Procuratorate issued the procuratorial organs filed public interest litigation pilot program in 2015. The public can file a lawsuit against the behavior of polluting the environment, which can improve public participation in environmental affairs and the rationality of environmental governance decision-making [6].

We further control the above policies by adding dummy variables based on the benchmark regression model to eliminate the possible interference of these parallel environmental policies. After excluding the impact of the carbon trading pilot, central environmental supervision, and public interest litigation reform, the effect of the vertical management auditing reform on carbon productivity is still positive and significant (see Supplementary Tab. 3). Therefore, concurrent environmental policies will not impact the policy effects of vertical management auditing reform.

# Synthetic DID and counterfactual estimation

Reweighting and matching controlled units and the pre-treatment using synthetic difference-in-differences will weaken the estimators' dependence on the parallel trend assumption [7]. Therefore, the synthetic difference-in-differences method is utilized to evaluate further the impact of the vertical management auditing reform on carbon productivity. We conducted a placebo test to calculate standard errors. The average treatment effect is 0.153 without covariates, while the average treatment effect is 0.133 when considering covariates. The results are consistent with the baseline regression, implying that the auditing reform can significantly improve carbon productivity (see Supplementary Tab. 4).

We further estimate policy effects based on the empirical framework of counterfactual estimation to exclude the influence of unobservable time-varying confounders [8]. We select the matrix completion estimator (MC) and hyper-parameter of 0.001 through a cross-validation procedure. The result of the average treatment effect is 0.133, which is significant at the statistical level of 5%, showing that our conclusion holds (see Supplementary Tab. 4).

## Replace data and independent variable

We also conducted the robust test of replacing the data and the explanatory variable. The data on carbon emissions from EDGAR is replaced. We convert the original grid point data into raster data and then summarize prefecture-level cities' carbon emission panel data by region. In addition, we replace carbon productivity with carbon intensity [9]. The specific calculation formula follows: CI = carbon emissions (million tons) / regional GDP (billion yuan). The result indicates that the vertical management auditing reform significantly reduced carbon intensity (see Supplementary Tab. 5).

## Estimator for average treatment effects

We further employ the methods of regression adjustment (RA), inverse probability weighting (IPW), augmented inverse probability weighting (AIPW), and inverse probability weighted regression adjustment (IPWRA) to estimate the average treatment effect of the auditing reform. The average treatment effect is significantly positive at 1 %, supporting our prediction that vertical management auditing reform significantly contributes to carbon productivity (see Supplementary Tab. 6).

# Supplementary Tables

Supplementary Table1. Definition and measurement of main variables.

| Variable | Definitions and operations |
|---|---|
| Dependent variable | |
| CP | The logarithm of the ratio of gross regional product to carbon dioxide emissions in that year |
| Independent variables | |
| did | If a city is in the pilot year and later years, the assignment is 1. Otherwise, it is 0. |
| Mediating variable | |
| lner | The logarithm of the proportion of words related to environmental protection in local government reports |
| lntl | Theil index |
| perugrma | The number of green inventions of utility model per capita |
| lnenergyeffi | The logarithm of the ratio of total energy consumption to GDP |
| Control variables | |
| lngy | The logarithm of the gross product of the secondary sector |

| | |
|---|---|
| lnpeople | The logarithm of the regional registered population at the end of the year |
| lnedu | The logarithm of regional education expenditure |
| lnperGDP | The logarithm of the GDP to total population ratio at the end of the year. |
| lnfis | The logarithm of regional general public budget expenditure. |
| lnbus | The logarithm of the number of buses in operation at the end of the year. |

Supplementary Table 2. The results of subdivided sample analysis

| VARIABLES | CP | | | |
|---|---|---|---|---|
| | lopp | | psp | |
| | Low | High | low | high |
| | (1) | (2) | (3) | (4) |
| did | 0.133** | 0.119 | -0.022 | 0.174** |
| | (2.38 ) | ( 1.41 ) | (-0.26) | (2.57) |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes | Yes |
| Observations | 2488 | 1353 | 1,241 | 2,605 |
| R-squared | 0.782 | 0.747 | 0.834 | 0.743 |

Note: * p<0.10, ** p<0.05, *** p<0.01; Numbers in parenthesis are t values.

Supplementary Table 3. Estimation results of excluding policy interferences

| VARIABLES | CP | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| did | 0.134** | 0.129** | 0.132** |
| | (2.35) | (2.27) | (2.35) |
| CT | 0.119 | | |
| | (0.90) | | |
| CEI | | 0.298 | |
| | | (0.83) | |
| PW | | | 0.459*** |
| | | | (13.08) |
| Controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes |
| Observations | 3,931 | 3,931 | 3,931 |
| R-squared | 0.749 | 0.749 | 0.749 |

Note: * p<0.10, ** p<0.05, *** p<0.01; Numbers in parenthesis are t values.

Supplementary Table 4. Estimation results of synthetic DID and counterfactual estimation

| VARIABLES | CP | | |
|---|---|---|---|
| | sdid | | fect |
| | (1) | (2) | (3) |
| ATT | 0.153** | 0.122** | 0.140** |
| Controls | No | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes |

Note: * p<0.10, ** p<0.05, *** p<0.01.

Supplementary Table 5. Estimation results of replacing data and independent variable

| VARIABLES | CP1 | | CI |
|---|---|---|---|
| | (1) | (2) | (3) |
| did | 0.170*** | 0.135*** | -0.258** |
| | (4.29) | (4.91) | (-2.15) |
| Controls | No | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes |
| Observations | 3,913 | 3,848 | 3,848 |
| R-squared | 0.957 | 0.973 | 0.562 |

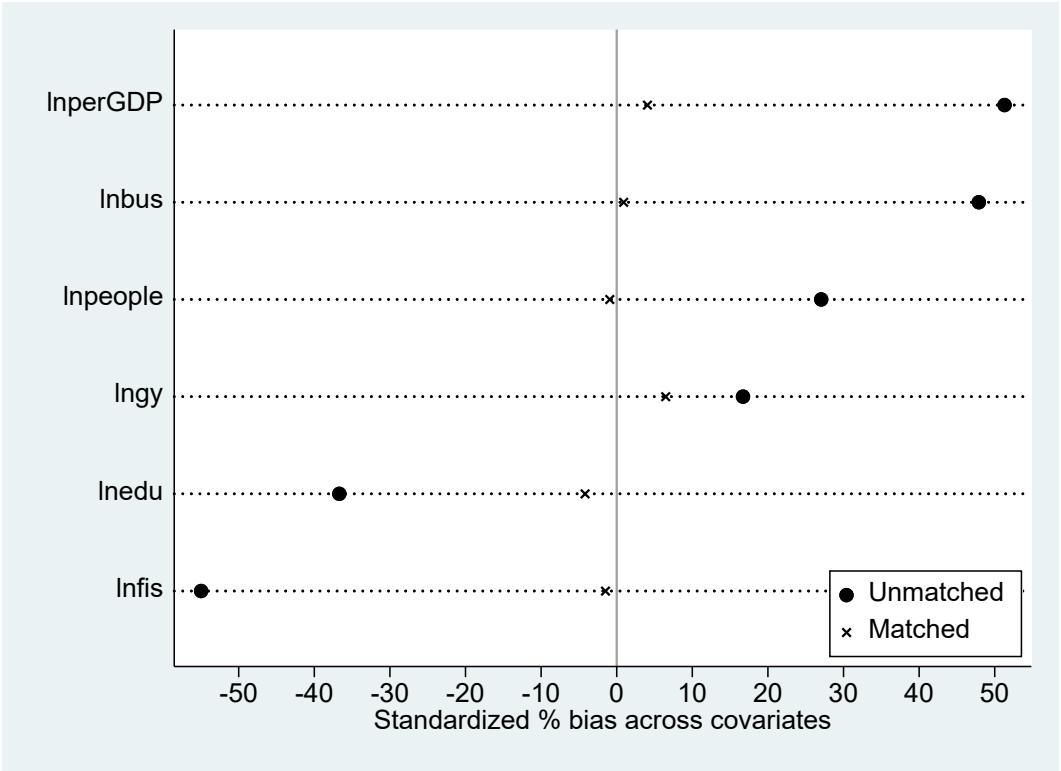
Note: * p<0.10, ** p<0.05, *** p<0.01; Numbers in parenthesis are t values.

Supplementary Table 6. Results of estimators for average treatment effects

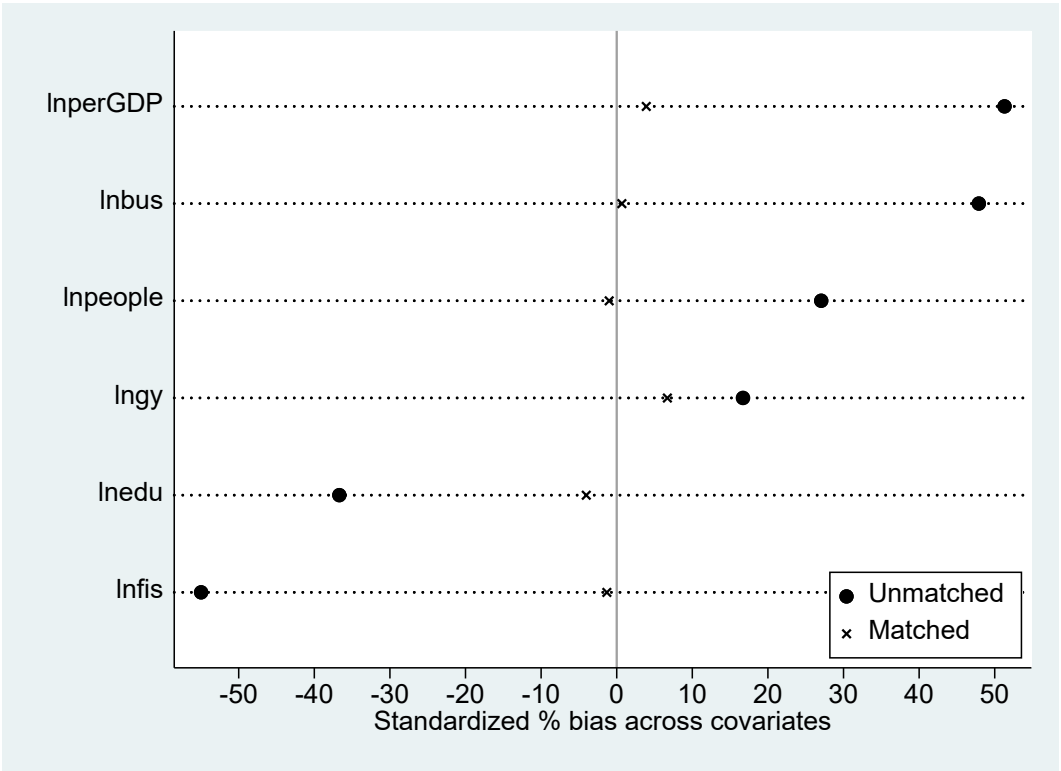
| VARIABLES | CP | | | |
|---|---|---|---|---|
| | RA | IPW | AIPW | IPWRA |
| | (1) | (2) | (3) | (4) |
| ATT | 0.294*** | 0.294*** | 0.294*** | 0.294*** |
| | (0.063) | (0.063) | (0.063) | (0.063) |
| Controls | Yes | Yes | Yes | Yes |
| Observations | 3,912 | 3,912 | 3,912 | 3,912 |

Note: * p<0.10, ** p<0.05, *** p<0.01.

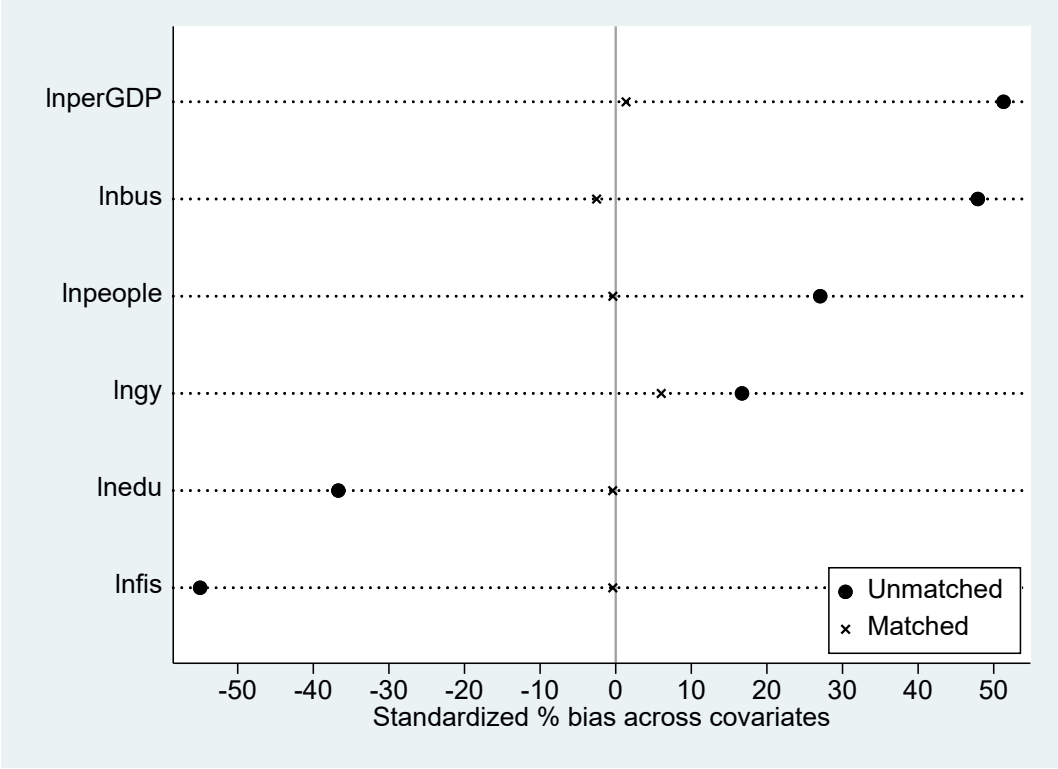
# Supplementary Figures



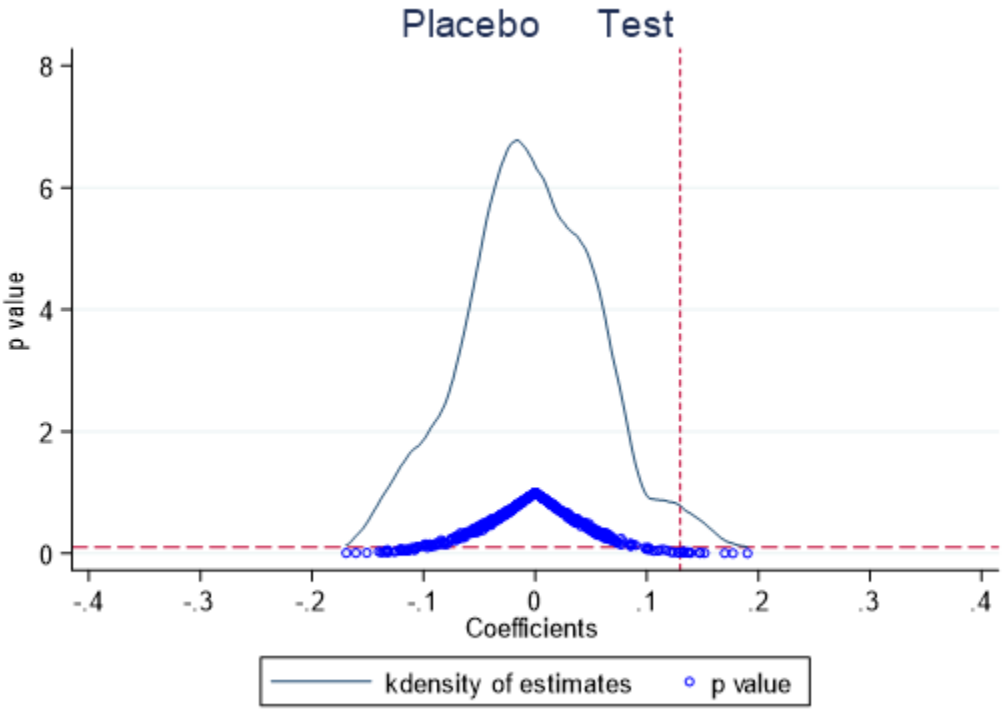

Supplementary Fig1. Balance Test of Calliper Matching



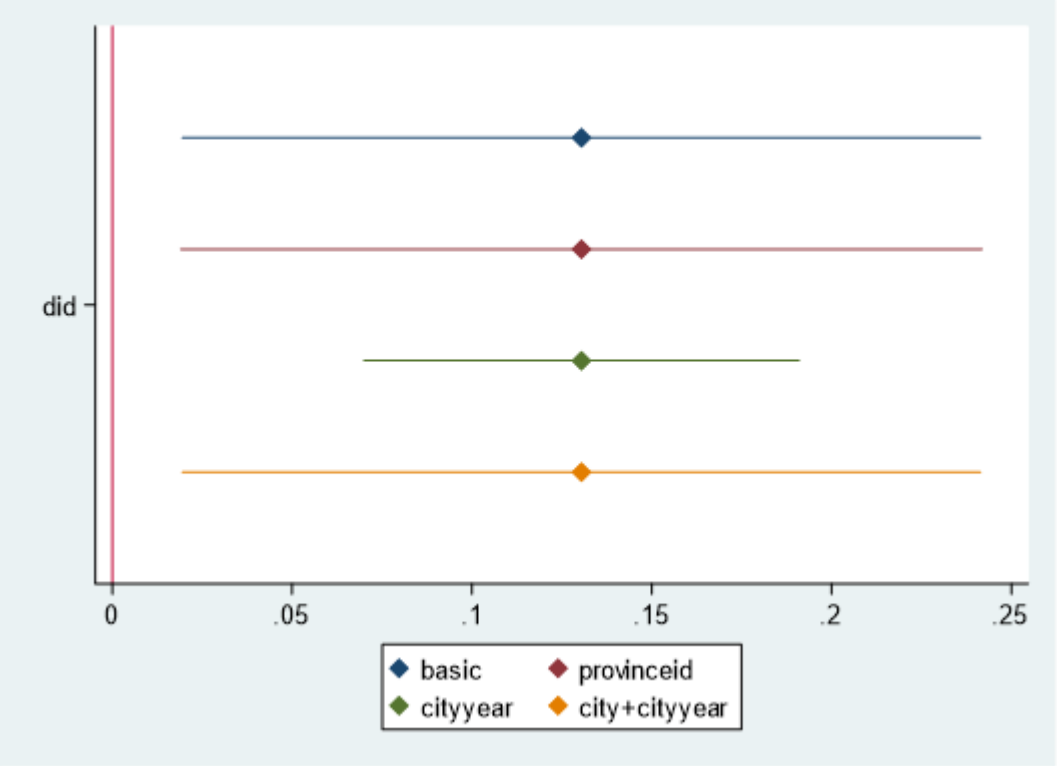

Supplementary Fig2. Balance Test of Kernel Matching




Supplementary Fig3. Balance Test of the Nearest Neighbor Matching (1:4)




Supplementary Fig 4. Placebo test results




Supplementary Fig 5. Effect with alternative fixed effects and clustered levels

# Supplementary References

1. GLAESER E. L., KAHN M. E. The greenness of cities: Carbon dioxide emissions and urban development. Journal of Urban Economics, **67** (3), 404, **2010**.
2. LIN X., CHEN S., CHENG X., WANG J. Local government audit and municipal debt risk: Evidence from audit reform in China. Finance Research Letters, 50, **2022**.
2. YU C., LONG H., ZHANG X., TAN Y., ZHOU Y., ZANG C., TU C. The interaction effect between public environmental concern and air pollution: Evidence from China. Journal of Cleaner Production, 391, **2023**.
3. ZHOU B., ZHANG C., SONG H., WANG Q. 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, **2019**.
4. WANG W., SUN X., ZHANG M. Does the central environmental inspection effectively improve air pollution?-An empirical study of 290 prefecture-level cities in China. Journal of Environmental Management, 286, **2021**.
5. BURTON P. Conceptual, Theoretical and Practical Issues in Measuring the Benefits of Public Participation. Evaluation, **15** (3), 263, **2009**.
6. ARKHANGELSKY D., ATHEY S., HIRSHBERG D. A., IMBENS G. W., WAGER S. Synthetic Difference-in-Differences. American Economic Review, **111** (12), 4088, **2021**.

7. LIU L., WANG Y., XU Y. A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. American Journal of Political Science, **2022**.
8. ROBERTS J. T., GRIMES P. E. Carbon intensity and economic development 1962–1991: A brief exploration of the environmental Kuznets curve. World Development, **25**, 191, **1997**.