1. Unzip the compressed package to a certain path, such as: d:\Desktop

2. Run python code using anaconda powershell prompt

- In order to run python code make sure you have anaconda installed
  - You can download anaconda here, and you can refer to this link to install anaconda.
  - You can refer to this link to make sure to setup Anaconda path to environment variable.
- Perform the following steps to install the required packages
- i. Open anaconda powershell prompt and run the following code to create a new conda environment named mypy (Change the path of the code below to the path of `efficiency_tree`)
  - ```
    conda env create -f d:\Desktop\efficiency_tree\environment.yml
    ```
- ii. After the environment installation is complete, continue to run the following code to use `mypy` environment
  - ```
    conda activate mypy
    ```
- iii. download package (change path)
  - ```
    pip install d:\Desktop\efficiency_tree\lpsolve55-5.5.2.11-cp39-cp39-win_amd64.whl

    pip install tqdm

    conda install -n mypy ipykernel --update-deps --force-reinstall
    ```
- iv. Enter the path where the `efficiency_tree` file is located (change path)
  - ```
    cd d:\Desktop\efficiency_tree
    ```

3. Install stata package

- ```
  net install st0665 , all replace   from( http://www.stata-journal.com/software/sj22-1)
  ```
  ```

```
ssc install lmdi,replace


ssc install mipolate, replace
```

4. Run programs by sequence (It takes about 2 hours.)

i.

- ```
  python d:\Desktop\efficiency_tree\1.step1_Figure3.py
  python d:\Desktop\efficiency_tree\2.step2.py
  ```

ii.

- run `3.step3.do` with stata

iii.

- ```
  python d:\Desktop\efficiency_tree\4.Figure1_5_6_8_9_10_Table3.py
  ```

iv.

- run `5.Figure2_4_7_Table2.do` with stata

In the package, .\1.step1_Figure3.py, .\2.step2.py and .\3.step3.do are the code that contain the code for the estimation of efficiency tree and the data required for the subsequent figures and tables.

- We list how the data is produced, and mark the number of lines corresponding to the input data in the code.
- The root directory is `efficiency_tree`, which is replaced by `.`.
- `.\city_inefficiency_tree.xlsx`, `.\temp\fig1Pre.xlsx`, `.\step2Pre.xlsx` and `.\chinamap\chinacity40_db.dta` are the original data, see the manuscript for the data source.

code	input	output	note

code	input	output	note
.\1.step1_Figure3.py	.\city_inefficiency_tree.xlsx (line 238)	.\temp\zuihou.xlsx (line 409) .\result\Figure 3 Efficiency tree.pdf (line 347)	Generate Efficiency Tree and the the data for subsequent figures and tables
.\2.step2.py	.\temp\zuihou.xlsx (line 5) .\city_inefficiency_tree.xlsx (line 12) .\step2Pre.xlsx (line 132)	.\temp\to_ddf_by_group.xlsx (line 168) .\temp\fig4Pre.xlsx (line 148)	Generate data needed to calculate grouping efficiency
.\3.step3.do	.\temp\to_ddf_by_group.xlsx (line 5)	.\temp\CEEm_CEEregion_CEE_global_by_id_year.dta (line 173)	Generate data needed to generate plots and tables
.\4.Figure1_5_6_8_9_10_Table3.py	.\fig1Pre.xlsx (Lines 11, 24)	.\result\Figure 1 Comparison on Energy consumption and carbon emissions.png (line 73)	Generate Figure 1

code	input	output	note
.\4.Figure1_5_6_8_9_10_Table3.py	.\temp\CEEm_CEEregion_CEE_global_by_id_year.dta(Lines 94, 209, 339,449)	.\result\Figure 5. Technology gap inefficiency TGIm by group over years.svg (line 196) .\result\Figure 6. Carbon emission inefficiency CEIm and technology gap inefficiency TGIm.png (line 328) .\result\Figure 8. Carbon reduction potential and its components.png (line 434) .\result\Figure 9. CO2 emission reduction potential changes caused by managerial failure.png (line 603) .\result\Figure 10. CO2 emission reduction potential changes caused by technology gap.png (line 696) .\result\Table 3.Results on carbon emission efficiency.xlsx (line 511)	Generate Figures 5, 6, 8, 9, 10, Table 3
.\5.Figure2_4_7_Table2.do	.\temp\fig2Pre.xlsx (line 6) .\chinamap\chinacity40_db.dta (Lines 22, 108, 148) .\temp\fig4Pre.xlsx (line 89) .\city_inefficiency_tree.xlsx (line 206)	.\result\Figure 2 Regional gross output and carbon emission intensity across Chinese cities.png (line 80) .\result\Figure 4 Groups obtained from efficiency tree.png (line 128) .\result\Figure 7 Comparison on carbon emission efficiency.png (line 201) .\result\table2.xls (line 228)	Generate Figures 2, 4, 7, Table 2

We describe the application of each file in the table below and mark whether it is raw data or not

file	application	Is it raw data?
.\chinamap	All the files in the folder are used to draw the map of China	YES
.\result ".\result\Figure 3 Efficiency tree-redrawn from Figure3.pdf.png" ".\result\Figure 3 Efficiency tree-redrawn from Figure3.pdf.pptx"	Figure3 in png and ppt format are manually drawn according to the pdf format of figure3	NO
.\temp	It is used to save the calculation process	NO
.\1.step1_Figure3.py .\2.step2.py .\3.step3.do .\4.Figure1_5_6_8_9_10_Table3.py .\5.Figure2_4_7_Table2.do	They are the code used to generate the efficiency tree and calculation process	
.\fig1Pre.xlsx	It was used to draw figure 1. It is used to generate Figure 1. It is annual data collected in China Statistical Yearbook.	YES
.\step2Pre.xlsx	It is the correspondence between administrative division names and codes. It is used to draw maps.	YES
.\environment.yml	It is used to install python packages	
.\lpsolve_wrapper.py	It is a wapper used to compute linear programming and does not need to be executed.	
.\lpsolve55-5.5.2.11-cp39-cp39-win_amd64.whl	It is the package of lpsolve	

file	application	Is it raw data?
readme_EfficiencyTree.html readme_EfficiencyTree.pdf readme_EfficiencyTree.png	readme file	