Using random forest to find the discontinuity points for carbon efficiency during COVID-19

Yingchi Qu  
Chongqing University  

Ming K. Lim  
University of Glasgow  

Mei Yang  
Chongqing University  

Du Ni (✉ mt15nn@mail.wbs.ac.uk)  
NUPT: Nanjing University of Posts and Telecommunications  https://orcid.org/0000-0002-7070-5255  

Zhi Xiao  
Chongqing University

Research Article

Keywords: carbon efficiency, fuzzy regression discontinuity, machine learning, random forest, COVID-19

Posted Date: September 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1943604/v1

License: Creative Commons Attribution 4.0 International License. Read Full License
Using random forest to find the discontinuity points for carbon efficiency during COVID-19

Yingchi Qu¹, Ming K. Lim², Mei Yang¹, Du Ni³, Zhi Xiao¹,*

¹School of Economics and Business Administration, Chongqing University, Chongqing 400044, PR China
²Adam Smith Business School, University of Glasgow, Glasgow G14 8QQ, United Kingdom
³School of Management, Nanjing University of Posts and Telecommunications, Jiangsu, 210003, PR China

Abstract

As there is a constant trade-off between carbon dioxide emissions against economic growth for every government, carbon efficiency is a key indicator to guide sustainable development. However, the energy crisis and COVID-19 recovery could affect carbon efficiency. Therefore, this paper combines the fuzzy regression discontinuity and random forest algorithm to estimate the discontinuity of the energy crisis and COVID-19 recovery on carbon efficiency. The results show that there are two cutoffs between carbon efficiency and coal prices. The positive treatment effect at cutoff 1 proves that the “zero-tolerance” policies effectively promote carbon efficiency. Besides, the negative treatment effect at cutoff 2 proves that electricity rationing has not always improved carbon efficiency during the energy crisis.

Keywords
carbon efficiency, fuzzy regression discontinuity, machine learning, random forest, COVID-19
1. Introduction

There is a constant trade-off between carbon dioxide emissions against economic growth for every government. China is not an exception (Sheng et al., 2020). Over 40 years, China’s economic growth was accompanied by high energy consumption, which led to severe pollution with low-carbon efficiency (Liu et al., 2016). On September 22, 2020, China announced that it would make every effort to achieve a peak carbon emission by 2030, and achieve the carbon-neutral goal by 2060 (Cai & Ye, 2022). Carbon efficiency, an indicator, reveals the companies’ carbon emission level to produce a given output (Cui & Li, 2015). Generally, carbon efficiency can be affected by transportation (Zhao et al., 2022), technology progress (Zhang & Fu, 2022), and carbon regulatory policies (Tan et al., 2020; Zhang et al., 2021). Among these, carbon regulatory policies are the most efficient ways to reduce emissions in the short term. Studies have shown that the carbon regulatory policy was able to improve carbon efficiency by a significant 1.7%, reducing carbon dioxide emissions by approximately 8.84 million tons in China from 2003 to 2018 (Yu & Zhang, 2021). Meanwhile, based on the data on industrial carbon emissions in 30 provinces of China from 2008 to 2016, carbon regulatory policies increase the economic dividend (13.6%) in the long run (Zhang et al., 2020). Therefore, carbon regulatory policy has drawn the attention of scholars for a long time.

However, electricity rationing, the mandatory carbon regulatory policy China has taken during the energy crisis, and “zero-tolerance” policies under the COVID-19 recovery are lacking discussion in previous studies. The energy crisis and COVID-19 recovery both would have a significant influence on carbon efficiency. During the COVID-19 pandemic, the rapid spread of COVID-19 has led to a massive collapse in economic activities and energy demand (Haxhimusa & Liebensteiner, 2021), while carbon emissions have been significantly reduced by the lock-down policy (Yin et al., 2022). However, COVID-19 stopped raging with the “zero-tolerance” policies, and transportation and manufacturing industries started to recover. The recovery of COVID-19 can increase carbon emissions compared with the COVID-19 pandemic. Thus, there is currently a lack of research to prove whether carbon efficiency will increase or decrease after the COVID-19 recovery.

In addition, as the energy crisis always goes along with soaring energy prices, it can possibly become another reason to affect the positiveness between carbon regulatory policies and carbon efficiency (Pu & Yang, 2022). Intuitively, rising energy prices can lower the energy demand and improve carbon efficiency, which gives consumers hesitation before consuming energy (Wang et al., 2021). However, there is no consensus on the positiveness between energy prices and carbon efficiency. For example, some researchers found that a high energy price positively impacted efficiency. Antonietti and Fontini (2019) measured 120 countries, discovering that the average oil price increased by 322% between 1980 and 2013, and the corresponding average increase in worldwide energy efficiency has amounted to roughly 2%.
The increase in the energy price contributes to a slight upward rise in energy efficiency. Notwithstanding, the others reveal that energy prices can negatively influence energy efficiency. Tajudeen (2021) analyzed 32 Organizations of Economic Cooperation and Development (OECD) countries, proving that the oil price could reduce the efficiency index. Sha et al. (2021) investigated the carbon efficiency of provinces in China and found that the effects of coal, oil, and gas on carbon efficiency were −0.049, −0.077, and −0.09, respectively. Thus, the energy price can negatively affect carbon efficiency.

Hence, a tool to study the impact of the energy crisis and COVID-19 recovery is required in such a scenario. Many scholars have used fuzzy regression discontinuity (FRD) to analyze the causal effect of policies or crises. Zhang et al. (2017) estimated electricity consumption fluctuations in response to the block pricing policy in China and found an approximately 40% increase in marginal price, inducing an approximately 35% decrease in electricity consumption (284 kWh per month). Kawaguchi et al. (2021) identified the causal effects of government subsidies related to the COVID-19 pandemic. Without the subsidy, the number of surviving small businesses would have dropped from 3.63 million to 3.29 million in Japan. These studies prove that FRD is a possible and effective approach to our research. Although the increasing popularity in the economics of regression discontinuity applications (McCrary, 2008) shows the superiority of FRD. However, because FRD analysis is sometimes not accurate enough, it increases the uncertainty of the results. And the recent popularity of machine learning has been shown to improve accuracy (Aria et al., 2021). As one of the widely used methods in machine learning, Random Forest (RF) algorithm guarantees high prediction accuracy, flexibility, and immediacy. Therefore, this study considers the integration method of FRD and RF algorithms to improve the analysis accuracy.

The contribution of our research is twofold. First, while previous studies had a conflicting view about the relationship between energy prices and carbon efficiency, we find that rising coal prices negatively impact carbon efficiency after the energy crisis burst, and COVID-19 recovery positively influences carbon efficiency. This could offer a new theoretical suggestion for the improvement of carbon efficiency. Second, this paper proposes a machine learning-based method, Fuzzy Regression Discontinuity-Random Forest, to evaluate the treatment effect of the energy crisis and COVID-19 recovery and reduce estimation error.

The remainder of this paper is structured as follows. Section 2 provides detailed descriptions of the datasets and indicators used in this study. Section 3 demonstrates the random forest-regression discontinuity method. Section 4 presents the empirical result and discuss the possible reasons for the different treatment effect. Finally, a conclusion is given in Section 5.
2. Data collection

Data used in this paper was from the CSMAR database (China Stock Market & Accounting Research Database), including the SSE (Shanghai Stock Exchange) 180 carbon efficiency Index and Daily Quotes of Coke Futures from October 8, 2019, to November 5, 2021. Since carbon efficiency is to explore how to meet the growing energy demand while minimizing carbon emissions (Lin & Li, 2021; Cantore et al., 2021), SSE 180 carbon efficiency Index effectively evaluates 180 representative low-carbon companies’ performances. SSE 180 carbon efficiency Index is based on the carbon footprint data of China-listed companies. After excluding stocks with high carbon footprints [exceeding 1000 (tonne CO\textsubscript{2}/ USD mln) in the past year], the remaining stocks are weighted according to:

\[ w_i = \frac{1/carbonfootprint_i}{\sum_i^n 1/carbonfootprint_{is}} \times w_s \]

Where, \( i \) denotes the representative company, \( i = 1, 2, \ldots, n \). \( s \) represents the industry, \( w_s \) is the weight of \( s \) industry in the SSE 180 index. Then, the SSE 180 carbon efficiency Index can be given as:

\[ index = \frac{\sum_i^m (p_i \times s_i \times f_i \times c_i)}{divisor} \]

\( p_i \) is the stock price of company \( i \), \( s_i \) is the total equity, \( f_i \) is the weighted proportion, and the \( divisor \) is the total market value of the sampled shares on the base date (June 8, 2013). \( c_i \) is calculated by:

\[ c_i = \frac{w_i}{p_o \times s_o \times f_i} \]

Where, \( p_o \) is the latest 5 working days closing price of the company \( i \) before the adjustment taking effect. \( s_o \) is the latest 5 working days closing total equity of company \( i \) before the adjustment taking effect. Meanwhile, the coal price deserves discussion during the energy crisis. The coal price is used as the assignment variable \( X_i \). The original data obtained from the CSMAR database is Daily Quotes of Coke Futures, and the daily data include 12-month future prices. For the empirical analysis, we processed the original data by weighted coke's daily opening price with the transaction amount, given by the following:

\[ daily \ coal \ price = \sum_{t}^{12} w_t p_t \]

where \( t \) is the month, \( w_t = \frac{transaction \ amount_t}{\sum_{t}^{12} transaction \ amount_t} \). \( p_t \) is the future price for each month.
From Figure 3, we can see that coal prices were relatively stable before October 2020, fluctuating within 1,500 - 2,000 yuan and showing an upward trend after October. The inflated coal price was induced by floods in Shanxi, a key coal-producing province, worsening the short-term supply crunch (CNBC, 2021). Furthermore, since April 2021, coal prices have risen rapidly from around 2,200 yuan, reaching 4,300 yuan in October 2021. It reflects that abnormally high coal prices occurred during the energy crisis. Figure 4 shows carbon efficiency. It can be easily seen that in July 2020, carbon efficiency surged from around 1,700 to close to 2,000 yuan. After that, it fluctuated in the range of 1,800 – 2,100 yuan. The economic recovery from COVID-19 and carbon regulatory policies might contribute to this surge.

From Figure 3 and Figure 4, it can be seen that the overall carbon efficiency increases with the rise of coal prices, but it seems to decrease slightly from 2021 onwards. However, it is not clear enough to draw conclusions from observations. Therefore, FRD is used to quantify the relationship between carbon efficiency and coal price, which can shed some light on the implication of policies. The descriptive statistics of the assignment variable (coal prices) and the outcome variable (carbon efficiency) are given in Table 1. Data used in this paper include 467 sample sizes from October 8, 2019, to November 5, 2021.
Table 1 Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Coal prices (yuan)</th>
<th>SSE 180</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>467</td>
<td>467</td>
</tr>
<tr>
<td>Mean</td>
<td>2299.99</td>
<td>1822.42</td>
</tr>
<tr>
<td>Std</td>
<td>529.82</td>
<td>142.01</td>
</tr>
<tr>
<td>Min</td>
<td>1631.73</td>
<td>1508.60</td>
</tr>
<tr>
<td>25%</td>
<td>1856.88</td>
<td>1683.22</td>
</tr>
<tr>
<td>50%</td>
<td>2145.64</td>
<td>1869.55</td>
</tr>
<tr>
<td>75%</td>
<td>2627.50</td>
<td>1928.84</td>
</tr>
<tr>
<td>Max</td>
<td>4315.75</td>
<td>2108.02</td>
</tr>
</tbody>
</table>

Table 1 describes the characteristics of the assignment variable and outcome variable. The sample includes 467 observations from October 8, 2019, to November 5, 2021. There is a wide fluctuation in coal prices, its standard deviation was 529.82, and the maximum coal price was 4315.75 yuan on October 19, 2021. The carbon efficiency index ranged from 1508.60 to 2108.02 yuan, and its standard deviation was 142.01.

3. Proposed RF-FRD model

In this study, the effect of the energy crisis, with electricity rationing, is discussed in order to reveal the relationship between carbon efficiency and coal prices. Therefore, coal price is an assignment variable \( X_i \), carbon efficiency is an outcome variable \( Y_i \), and energy crisis is a treatment variable. However, economic recovery from COVID-19 is chosen as another treatment variable. As illustrated in Figure 5, the beginning of the economic recovery from COVID-19 and the energy crisis, in July 2020 and August 2021, are selected as cutoff \( C_1 \) and cutoff \( C_2 \), respectively. cutoff \( C_1 \) and cutoff \( C_2 \) divide the data into three intervals according to time, represented by different colored ellipses. Consequently, the treatment states \( D_i \), which is determined by time, represents whether the individual \( i \) receives the impact of economic recovery or energy crisis. In this case, \( D_i = 1 \) denotes that individual \( i \) is influenced by economic recovery from COVID-19 \( (D_0) \), \( D_i = 2 \) denotes that individual \( i \) is affected by the energy crisis and following electricity rationing \( (D_1) \). Therefore, the treatment state of individuals is not strictly related to the cutoff. The individual \( i \) could be “crossover” at the cutoff, which means individuals who actually belong to the treatment group could be on the left side of the cutoff. So, FRD can provide an unbiased estimate of the local average treatment effect (LATE) (Jacob et al., 2012).

![Figure 3 Fuzzy regression design with two cutoffs.](image-url)
Analytically, the estimation of the treatment effect in an FRD is often carried out by the two-stage least squares (TSLS) method. The first-stage equation and second-stage equation are expressed as equation (5) and equation (6):

\[ D_i = g(x_i) + \gamma T_i + \epsilon_i \]  
\[ Y_i = f(x_i) + \beta D_i + \mu_i \]

Where, \( D_i \) is the real treatment state. \( T_i \) denotes the state predetermined by the cutoff. For individuals near cutoff \( C \), \( T_i = 1 \) if individual \( i \) is on the right side of the threshold, \( T_i = 0 \) if individual \( i \) is on the left side of the threshold. Moreover, \( T_i \) can be used as an instrumental variable for \( D_i \) in TSLS. \( g(\cdot) \) and \( f(\cdot) \) are smooth functions of \( x_i \) described by low-order polynomials on the two sides of the cutoff \( C \). \( \epsilon_i \) and \( \mu_i \) are random errors.

The treatment probability for cutoff \( C_1 \) and cutoff \( C_2 \) can be expressed as follow (Imbens & Lemieux, 2008):

\[ P(D_i = 1|x_i) = \begin{cases} 
  g_0(x_i), & \text{if } T_i = 0 \\
  g_1(x_i), & \text{if } T_i = 1 
\end{cases} \] (3)

\[ P(D_i = 2|x_i) = \begin{cases} 
  g_1(x_i), & \text{if } T_i = 1 \\
  g_2(x_i), & \text{if } T_i = 2 
\end{cases} \] (4)

In FRD, the treatment state \( D_i \) is not completely determined by \( x_i \), so the jump at the cutoff is a jump of treatment probability, which can be described as follows:

\[ \lim_{\Delta \to 0^+} P[D_i = 1|x_i = C_1 + \Delta] \neq \lim_{\Delta \to 0^-} P[D_i = 1|x_i = C_1 - \Delta] \] (5)

\[ \lim_{\Delta \to 0^+} P[D_i = 2|x_i = C_2 + \Delta] \neq \lim_{\Delta \to 0^-} P[D_i = 2|x_i = C_2 - \Delta] \] (6)

Where \( C_1 \) and \( C_2 \) denote the value of the assignment variable at the cutoff. The LATE at the cutoff \( C_1 \) and \( C_2 \) can be estimated by Equations (11) and (12):

\[ \tilde{\tau}_{FRD1} = \frac{\lim_{x \to C_1^+} E[Y_i|x_i = C_1] - \lim_{x \to C_1^-} E[Y_i|x_i = C_1]}{\lim_{x \to C_1^-} E[D_i = 1|x_i = C_1] - \lim_{x \to C_1^+} E[D_i = 1|x_i = C_1]} \] (7)

\[ \tilde{\tau}_{FRD2} = \frac{\lim_{x \to C_2^+} E[Y_i|x_i = C_2] - \lim_{x \to C_2^-} E[Y_i|x_i = C_2]}{\lim_{x \to C_2^-} E[D_i = 2|x_i = C_2] - \lim_{x \to C_2^+} E[D_i = 2|x_i = C_2]} \] (8)

Therefore, on the basis of drawing analysis, this paper employs the TSLS method to estimate the treatment effect, and the estimation method is provided by Calonico et al. (2017). The results show that the relationship between coal price and carbon efficiency at \( C_1 \) and \( C_2 \) has changed, caused by different treatment \( D_i \). Equation (9) and (10) is used to prove the existence of the jump, and the treatment effect is calculated by equation (11) and (12) at the threshold, which measures the impact of an external treatment on the relationship between coal prices and carbon efficiency. Hence, section 5.1 and section 5.2 show the different treatment effects under economic recovery from COVID-19 and the energy crisis, respectively, analyzing potential reasons for these jumps with relevant energy policies. However, bandwidth selection is essential in FRD estimation. An appropriate bandwidth can avoid overfitting, which directly
affects the validity of the experiment. Therefore, the optimal bandwidth selection is based on
the cross-validation method (Imbens & Kalyanaraman, 2012), selecting the bandwidth with the
minimum mean squared error as the optimal bandwidth.

Although the increasing popularity in the economics of regression discontinuity
applications (McCrary, 2008) shows the superiority of FRD. This study takes an integration
method of FRD and RF into consideration because FRD analysis is sometimes not precise
enough, Random Forest can improve the accuracy, as it ensures high predictive precision,
flexibility, and immediacy. As one of the widely used approaches in machine learning, the lack
of interpretability limits its use in economics (Aria et al., 2021). The combination of FRD and
RF can achieve complementary advantages.

Random Forests regression (Breiman, 2001) is formed by growing trees depending on a
random vector $\theta$ such that the tree predictor $h(X_i, \theta)$. The output value of Random Forests is
$Y_i$. $Y_i = h(X_i, \theta)$. And the mean-squared generalization error for tree predictor $h(x_i)$ is:

$$MSE = E_{X_i, Y_i}(Y_i - h(x_i))^2$$

RF-FRD is based on the Random Forests tree predictor, Equations (14) and (15) denote
that the outcome variable is determined by tree $h(X_i, D_i)$, with the two-stages estimation.

$D_i = h_1(x_i, T_i) + \epsilon_i$  \hspace{1cm} (10)

$Y_i = h_2(x_i, D_i) + \mu_i$ \hspace{1cm} (11)

Where the value of treatment $D_i$ follows Equation (7) and (8), at cutoff $C_1$ and cutoff $C_2$,
respectively. $\epsilon_i$ and $\mu_i$ are random errors. In FRD, the treatment state $D_i$ is not completely
determined by $x_i$, so the jump at the cutoff is a jump of treatment probability, which can be
described as follows:

$\lim_{\Delta \to 0^+} P[D_i = 1|x_i = C_1 + \Delta] \neq \lim_{\Delta \to 0^-} P[D_i = 1|x_i = C_1 - \Delta]$ \hspace{1cm} (12)

$\lim_{\Delta \to 0^+} P[D_i = 2|x_i = C_2 + \Delta] \neq \lim_{\Delta \to 0^-} P[D_i = 2|x_i = C_2 - \Delta]$ \hspace{1cm} (13)

When the condition of Equation (16) and (17) is satisfied proving the existence of the
jump, the average treatment effect at every cutoff $C$ can be expressed as:

$$\bar{\tau}_{RF-FRD1} = \frac{\lim_{x \to C_1^+} E[Y_i|h(X_i) = C_1] - \lim_{\Delta \to C_1^-} E[Y_i|h(X_i) = C_1]}{\lim_{x \to C_1^+} E[D_i=1|h(X_i) = C_1] - \lim_{x \to C_1^-} E[D_i=1|h(X_i) = C_1]}$$ \hspace{1cm} (14)

$$\bar{\tau}_{RF-FRD2} = \frac{\lim_{x \to C_2^+} E[Y_i|h(X_i) = C_2] - \lim_{\Delta \to C_2^-} E[Y_i|h(X_i) = C_2]}{\lim_{x \to C_2^+} E[D_i=2|h(X_i) = C_2] - \lim_{x \to C_2^-} E[D_i=2|h(X_i) = C_2]}$$ \hspace{1cm} (15)
4. Results and discussion

This study uses the RF-FRD model to explore the impact of economic recovery and energy crisis on carbon efficiency, and discuss whether there is a possible jump under different coal price levels, proving that the relationship on the two sides of the threshold has changed.

4.1 The treatment effect of economic recovery and energy crisis

There is a possible jump where the coal price equals 1950 yuan. Figure 6 illustrates the results where the \( c_{\text{utoff}} C_1 = 1950 \), and the lines represent the estimated value of \( Y_i \). The scatter points with different colors denote the sample with or without treatment \( D_0 \). The blue points represent the control group, and the salmon points represent the treatment group.

![Figure 4 The effect of economic recovery from COVID-19.](image)

Note: (1) The line of dashes denotes the \( c_{\text{utoff}} (1950) \), and the curves at two sides of the cutoff denote the FRD results. (2) It is noteworthy that the local regression is estimated by Ordinary Least Squares, and polynomial order is determined by AIC (Akaike Information Criterion). (3) Horizontal axis represents coal prices, and the vertical axis represents carbon efficiency, and these scatter points show the relationship between carbon efficiency and coal prices.

The treatment effects with TSLS estimation are demonstrated in Table 2. By varying the polynomial order of \( X_i \), the experimental results can be checked for robustness. Two types of kernel, epanechnikov and triangular have been tested. Columns 1, 3, and 5 are results of epanechnikov kernel estimation, and columns 2, 4, and 6 are results of triangular kernel estimation.

<table>
<thead>
<tr>
<th>Polynomial order</th>
<th>(1)</th>
<th>(1)</th>
<th>(2)</th>
<th>(2)</th>
<th>(3)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment effect</td>
<td>290.12***</td>
<td>266***</td>
<td>303.62</td>
<td>266.53***</td>
<td>273.35</td>
<td>249.62***</td>
</tr>
<tr>
<td>Standard error</td>
<td>37.69</td>
<td>38.43</td>
<td>100.75</td>
<td>39.88</td>
<td>47.95</td>
<td>41.40</td>
</tr>
<tr>
<td>P value of first-stage</td>
<td>0.047</td>
<td>0.051</td>
<td>0.447</td>
<td>0.087</td>
<td>0.127</td>
<td>0.074</td>
</tr>
</tbody>
</table>
From Table 2, the treatment effect $\tau_{FRD1}$ is around 266, in triangular kernel estimation, which testify the carbon efficiency on the right side of the threshold is higher than that on the left side. The growing carbon efficiency and positive $\tau_{FRD}$ prove that the economic recovery from COVID-19 has a positive impact on carbon efficiency.

There is a possible jump where the coal price equals 2870 yuan, as shown in Figure 7. The lines represent the estimated value of $Y_i$, and the scatter points with different colors denote the sample with or without treatment $D_1$. The blue points represent the control group, and the salmon points represent the treatment group. The treatment of the control group is $d_1 = 1$, and the treatment group is $d_2 = 2$.

![Figure 7 The effect of electricity rationing.](image)

Note: (1) The line of dashes denotes the cutoff (2870), and the curves at two sides of the cutoff denote the FRD results. (2) It is noteworthy that the local regression is estimated by Ordinary Least Squares, and polynomial order is determined by AIC (Akaike Information Criterion). (3) Horizontal axis represents coal prices, and the vertical axis represents carbon efficiency, and these scatter points show the relationship between carbon efficiency and coal prices.

The treatment effects with TSLS estimation are demonstrated in Table 3. By varying the polynomial order of $X_i$, the experimental results can be checked for robustness. Two types of kernel, epanechnikov and triangular have been tested. Columns 1, 3, and 5 are results of epanechnikov kernel estimation, and columns 2, 4, and 6 are results of triangular kernel estimation. Treatment effects under different bandwidths also have been tested, presented in section 6, indicating that the experimental results are valid.
Table 3 TSLS estimate results of treatment effect ($C_2$)

<table>
<thead>
<tr>
<th>Polynomial order</th>
<th>(1)</th>
<th>(1)</th>
<th>(2)</th>
<th>(2)</th>
<th>(3)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment effect</td>
<td>-182.69**</td>
<td>-170.05**</td>
<td>-161.42**</td>
<td>-146.08**</td>
<td>-154**</td>
<td>-155.08**</td>
</tr>
<tr>
<td>Standard error</td>
<td>91.407</td>
<td>79.57</td>
<td>69.98</td>
<td>64.23</td>
<td>70.43</td>
<td>73.02</td>
</tr>
<tr>
<td>P value of first-stage</td>
<td>0.005</td>
<td>0.002</td>
<td>0.003</td>
<td>0.004</td>
<td>0.006</td>
<td>0.01</td>
</tr>
<tr>
<td>P value of second-stage</td>
<td>0.046</td>
<td>0.033</td>
<td>0.021</td>
<td>0.023</td>
<td>0.029</td>
<td>0.034</td>
</tr>
<tr>
<td>Optimal bandwidth</td>
<td>235.52</td>
<td>257.01</td>
<td>285.45</td>
<td>299.93</td>
<td>446.5</td>
<td>472.34</td>
</tr>
<tr>
<td>Observations</td>
<td>75</td>
<td>86</td>
<td>95</td>
<td>103</td>
<td>142</td>
<td>146</td>
</tr>
</tbody>
</table>

Note: (1) ** is the significance level of 5%. (2) The optimal bandwidth is selected by minimum mean squared error. (3) Kernel Epa. and Tri. denote Epanechnikov and Triangular, respectively. (4) P>|z| is the p value of second-stage (treatment effect) estimation.

It can be seen from Table 3 that the treatment effect $\tau_{FRD2}$ of quadratic polynomial estimation is around -150, which testify the carbon efficiency on the right side of the threshold is lower than that on the left side. The negative treatment effect reveals that the energy crisis decreases carbon efficiency, and the growing energy prices and carbon regulatory policies do not always improve carbon efficiency. Therefore, the economic performance and carbon regulatory policies, especially electricity rationing, are demonstrated to discover the possible reasons for this phenomenon. Compared with the positive impact of economic recovery from COVID-19 on carbon efficiency, the economic fluctuation cannot contribute to carbon efficiency. Because the Shanghai Composite Index rose from 3397.36 on July 31, 2021, to 3547.34 on October 31, 2021, generally, the upward economy improves carbon efficiency. Moreover, since the National Development and Reform Commission announced the reduction of energy consumption in 34 provinces on August 17, 2021, the government has adopted carbon regulatory policies during the energy crisis, as shown in Table 4.

Table 4 Policies adopted by the Chinese government

<table>
<thead>
<tr>
<th>Event date</th>
<th>Name</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>2021/8/17</td>
<td>A barometer of the completion of dual control targets for energy consumption in each region in the first half of 2021.</td>
<td>Suspend the approval of energy conservation for high energy consumption and high emission projects in 9 regions with increasing energy intensity.</td>
</tr>
<tr>
<td>2021/9/7</td>
<td>Establishment of a market mechanism for clean energy consumption.</td>
<td>Green power trading was launched for the first time, with 7.935 billion kWh of electricity traded. This transaction is expected to reduce the burning of standard coal by 2.436 million tons and the emission of carbon dioxide by 6.0718 million tons.</td>
</tr>
<tr>
<td>2021/9/11</td>
<td>Improving the dual control scheme for energy consumption intensity and total amount.</td>
<td>Control high-emission projects and encourage clean energy consumption.</td>
</tr>
<tr>
<td>2021/10/12</td>
<td>Notice on deepening the market-oriented reform of coal-fired power generation on-grid tariffs.</td>
<td>The fluctuation range of the trading price of the coal-fired power generation market has been expanded to no more than 20%. However, the market trading price of high energy-consuming companies is not subject to a 20% rise.</td>
</tr>
</tbody>
</table>
Note: This table only includes national policies, and the specific policies of each province will be slightly different.

These policies can encourage the commitment of China to carbon emission reduction targets. These policies mainly consist of two parts: restricting high-energy-consuming production and promoting renewable energy development. However, after the outbreak of the energy crisis, some local governments adopted electricity rationing policies to reduce electricity load and carbon emissions. The iron and steel industry, a high-carbon emission, and the biggest energy-consuming manufacturing industry were affected by electricity rationing the most. In Hebei province, iron and steel companies have implemented a 30%-40% electricity rationing and staggered productions, since September 2021. In Guangxi province, four steel plants started off-peak production, and four steel plants were shut down in August. In Jiangxi province, construction steel decreased by 2.3 million tons from September to October. Furthermore, Guangdong province implemented “intermittent production” (work started for two days and stopped for five days) on September 16. High-energy-consuming companies have been curtailed for one week since September 22.

Contrary to previous literature, carbon regulatory policies do not always promote carbon efficiency since the treatment effect is negative. Therefore, carbon regulatory policies that were originally conducive to carbon efficiency did not improve carbon efficiency during the energy crisis. But the difference between the treatment effects of the two thresholds $\Delta \tau_{FRD} = \tau_{FRD1} - \tau_{FRD2}$ is positive, which proves that the energy crisis did not make carbon efficiency worse than it was before COVID-19 recovery.

4.2 RF-FRD analysis

Although Section 4.1 and 4.2 can prove the existence of jumps, the accuracy of the FRD method is not high enough. As shown in Table 5, the RMSE (Root Mean Squared Error) of \( \text{cutoff } C_1 \) and \( \text{cutoff } C_2 \) is 48.16 and 58.33, respectively. The purpose of putting forward the RF-FRD method is to improve the prediction accuracy of $\hat{Y}_i$. Besides, the Random Forest has the advantage that it is not sensitive to multiple collinearities. In this section, the bandwidth used to test the RF-FRD method is the same as the optimal bandwidth of quadratic polynomial TSLS estimation. The RF-FRD estimate results are listed in Table 5.

<table>
<thead>
<tr>
<th>Treatment effect</th>
<th>cutoff $C_1$</th>
<th>cutoff $C_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment effect</td>
<td>232.84</td>
<td>-131.3</td>
</tr>
<tr>
<td>RMSE (FRD)</td>
<td>48.16</td>
<td>58.33</td>
</tr>
<tr>
<td>RMSE (RF-FRD)</td>
<td>28.53</td>
<td>30.89</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>142</td>
<td>300</td>
</tr>
</tbody>
</table>

The treatment effect of the RF-FRD method is 232.84 at cutoff $C_1$. The positive treatment effect shows that the “zero-tolerance” policies can improve carbon efficiency, consistent with
Section 5.1. Meanwhile, the RMSE dropped to 28.53, which has a 40.7% improvement. The treatment effect of the RF-FRD method is -131.3 at cutoff $C_2$. The negative treatment effect testifies that carbon regulatory policies and electricity rationing do not always improve carbon efficiency, consistent with section 5.2 conclusion. Meanwhile, the RMSE decreased to 30.89, which has a 47% improvement.

### 4.3 Robustness test

This section contains two parts to test the robustness of FRD. McCrary discontinuity tests are used to avoid the endogenous grouping, ensuring the assignment variable is not manipulated. And we check the FRD with different bandwidths. The premise of FRD is that the conditional density is continuous at $X_i = C$. Therefore, it is necessary to verify whether the variable grouping has endogenous grouping. Endogenous grouping means that the sample individuals know the grouping principle in advance and then choose to enter the processing group through their selection. Such endogenous grouping will lead to the failure of FRD. McCrary discontinuity tests (Figure 8) are used to test its setting and check whether the assignment variable is manipulated.

![Figure 8 McCrary discontinuity tests for cutoffs $C_1$ and $C_2$.](image)

Note: (1) The horizontal axis denotes the assignment variable coal price. (2) The left figure illustrates the McCrary discontinuity tests where the coal price equals 1950, and the right one shows the McCrary discontinuity tests where the coal price equals 2870.

It can be seen in Figure 8 that most of the confidence intervals of the estimated values of the density function on both sides of the threshold overlap. So there is no significant difference in the density function on both sides of the threshold, and the non-manual-manipulation assumption is valid. A robust FRD result requires less sensitivity to the bandwidth since the bandwidth can affect the significance of FRD results. Sections 4.1 demonstrate the FRD estimation results under the optimal bandwidth, with minimum mean squared error results. Hence, these robustness checks and falsification tests are focused on FRD testing with different bandwidths. Table 6 represents the FRD results with different bandwidths at cutoff $C_1$. The optimal bandwidth is 92.191, with the first-order polynomial. At bandwidths of 75, 100, 150, and 200, the FRD results are significant at the 1% level, which means FRD is robust.
Table 6 FRD results of \( \textit{cutoff} \ C_1 \) (coal price = 1950)

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>75</th>
<th>100</th>
<th>150</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment effect</td>
<td>289.48***</td>
<td>287.83***</td>
<td>317.26***</td>
<td>260.18***</td>
</tr>
<tr>
<td>Standard error</td>
<td>48.226</td>
<td>37.809</td>
<td>52.132</td>
<td>73.02</td>
</tr>
<tr>
<td>Polynomial order</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Kernel</td>
<td>Tri.</td>
<td>Tri.</td>
<td>Tri.</td>
<td>Tri.</td>
</tr>
<tr>
<td>Observations</td>
<td>85</td>
<td>117</td>
<td>140</td>
<td>169</td>
</tr>
</tbody>
</table>

Note: *** denotes the significance level of 1%.

Table 7 represents the FRD results with different bandwidths at \( \textit{cutoff} \ C_2 \). The optimal bandwidth is 257.01, with the first-order polynomial. At bandwidths of 200, 300, and 325, the FRD results are both significant at a 5% level, which means FRD is robust.

Table 7 FRD results of \( \textit{cutoff} \ C_2 \) (coal price = 2870)

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>200</th>
<th>250</th>
<th>300</th>
<th>325</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment effect</td>
<td>-158.25**</td>
<td>-168.93**</td>
<td>-171.64**</td>
<td>-164.86**</td>
</tr>
<tr>
<td>Standard error</td>
<td>76.556</td>
<td>79.783</td>
<td>76.367</td>
<td>71.63</td>
</tr>
<tr>
<td>Polynomial order</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>Kernel</td>
<td>Tri.</td>
<td>Tri.</td>
<td>Tri.</td>
<td>Tri.</td>
</tr>
<tr>
<td>Observations</td>
<td>70</td>
<td>83</td>
<td>103</td>
<td>115</td>
</tr>
</tbody>
</table>

Note: ** denotes the significance level of 5%.

5. Conclusions

As economies grow, energy consumption is inevitably increasing. The major contribution of this study is to explore the impact of the energy crisis and COVID-19 economic recovery on carbon efficiency, discussing two typical policies in China from a short-term period. We build a machine learning method called RF-FRD. RF-FRD can reduce the interference of multicollinearity between samples and promote the predicted accuracy. With the RF-FRD method, RMSE of \( \textit{cutoff} \ C_1 \) and \( \textit{cutoff} \ C_2 \) reduce by 40.7% and 47%, respectively, which provides a precise carbon efficiency prediction. During the energy crisis and COVID-19 economic recovery, the government implemented electricity rationing and “zero-tolerance” policies namely. The model results and robustness tests prove that the treatment effect at cutoff \( C_1 \) is positive. This means the “zero-tolerance” policies help the economy recover from COVID-19, improving carbon efficiency. The model results at cutoff \( C_2 \) show that the treatment effect is negative, so electricity rationing during an energy crisis cannot improve carbon efficiency as “zero-tolerance” policies.

Some limitations of this work must be listed. First, this study discusses the impact of the energy crisis and COVID-19 economic recovery on carbon efficiency at the macro level. However, the impact on each company at the micro-level is also worth exploring. Second, only 180 representative companies are chosen for this work, and further research can be done if
more specific data are available. In addition, this research only considers the coal price as the assignment variable since current natural gas and electricity prices are macroeconomic regulatory in China. The price of other fossil fuels deserves consideration. Finally, more machine learning algorithms can be considered for economic calculations to help reduce analytical errors.

**Ethical Standards statements**

**Ethical approval**

This work does not contain any studies with human participants or animals performed by any of the authors.

**Conflict of interest**

The authors declare no conflict of interest.

**Funding details**

This work was supported by the National Natural Science Foundation of China [grant numbers 71671019, 72071021] and the graduate research and innovation foundation of Chongqing, China [grant number CYS21047]. The authors declare no competing interests.

**Informed Consent**

Informed consent was obtained from all individual participants included in the study.

**Authorship contributions**

Yingchi Qu: Conceptualization, Investigation, Writing - original draft, Validation, Visualization.

Ming K. Lim: Conceptualization, Project administration, Validation. Mei Yang: Writing, Validation, Visualization. Du Ni: Conceptualization, Investigation, Validation. Zhi Xiao: Conceptualization, Methodology, Project administration.

**Data Availability**

The datasets are available from the corresponding author on reasonable request.

**References**


Julia, H., 2021. Europe was already facing a winter gas crisis. The risks just got even bigger. CNN. Retrieved December 15, 2021, from


https://www.newsweek.com/europe-energy-crisis-just-got-even-worse-1661136


Feng, E., 2021. Why China has to ration electricity and how that could affect everyone. NPR. https://www.npr.org/2021/10/01/1042209223/why-covid-is-affecting-chinas-power-rations


https://doi.org/10.1016/j.jjie.2021.101138


