A Machine Learning Approach for Investigating the Impact of Circular Economy on Taiwan's Economy and the Environment

YUZE TSAI

May 20, 2025

Contents

Abstract

This study innovatively investigates the impact of circular economy practices on Taiwan's economic development and environmental performance through both traditional econometric and modern machine learning approaches. Utilising a self-prepared panel dataset spanning 22 cities/-counties from 2014 to 2023, we analyse how those key circular economy indicators—including waste recycling rates, renewable energy capacity, and environmental budget allocation—relate to three critical outcomes: household income, per capita electricity usage, and housing vacancy rates. Furthermore, the SHAP value is introduced to interpret non-linear models' predictions and assess feature importance. Results suggest that environmental budget allocation and renewable energy capacity positively influence income levels, while food waste is associated with increased electricity use and housing inefficiency. Additionally, we found out that the Machine learning models consistently outperform traditional methods in predictive accuracy, underscoring their value in policy-oriented socioeconomic analysis. The findings highlight the multifaceted effects of circular economy strategies and provide some clairvoyance guidance for evidence-based policymaking in sustainable development.

1 Introduction

1.1 What is the Circular Economy?

The concept of circular economy arose in 1966, when Kenneth Boundling, the American economist, proposed the idea of "Spaceship Economy", considering that people must find the way of recycling the resources to achieve sustainability on the planet, or we will doom once the paucity of resource leverage to an irrecoverable level. Since that proposal, this concept is becoming more and more popular nowadays, with the emerging conscious and international consensus on protecting the environment. Although there may exist some nuances on the definition of circular economy, we may have a look at the gross performance of the Ellen MacArthur Foundation (EMF), which has concentrated on the devotion to global transformation of circular economy. According to their statement, circular economy can be defined as establishing a renewable and recyclable economy system and minimizing the waste we create. And from Weetman's handbook (2016) talking about the circular economy, he summarized the six necessary factors of the circular economy framework, which translate the four principles of this concept, shown in the figure 1

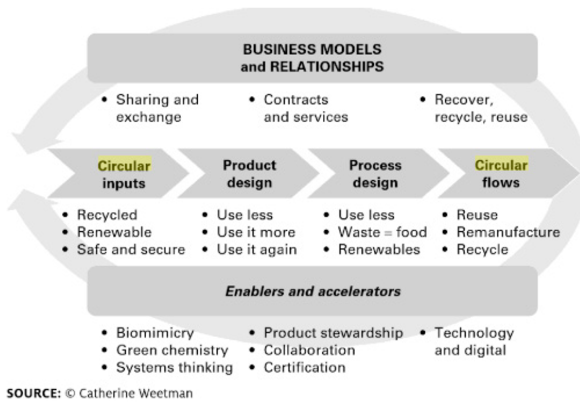


Figure 1: Caption

, which effectively give us a large amount of flavoring on constructing the idea of circular economy.

Upon mentioning those theoretical concepts, what exactly did the governments do worldwide? In 2015, the European Commission announced its first circular economy action plan. It included measures to help stimulate Europe's transition towards a circular economy, boost global competitiveness, foster sustainable economic growth, and generate new jobs. After the first, the European Commission adopted the new circular economy action plan (CEAP) in March 2020, their new action plan promoting initiatives throughout the life cycle of products. It targeted how products are designed, promotes circular economy processes, encourages sustainable consumption, and aims to ensure that waste is prevented and that the resources used are kept in the EU economy for as long as possible. Except for EU countries, our government launched the "5+2 Industrial Innovation Plan" in 2016, with the "circular economy" and "new agriculture" being key strategies for Taiwan's national transformation. Additionally, the Resource Circulation Administration, Ministry of Environment stated (2023) that Taiwan's total material input has reached 300 million metric tons, with over 70% relying on imports, and after deducting processed and re-exported portions, the actual domestic consumption exceeds 200 million metric tons, averaging around 11 metric tons of material per person per year. Meanwhile, Taiwan generates approximately 30 million metric tons of waste annually. Given the limitations of resources and environmental burdens, the key challenge for both the Environmental Protection Administration and various industries is how do we reintegrate materials from the "vein industry" back into the "artery industry" for circular reuse, ensuring that waste resources are returned to industrial production processes.

1.2 Variables of Interest

Then how do we select the variable? In Tanatu, A., et al. (2018)'s study on the dependencies between those indicators for a Circular Economy in the European Union, their conclusion suggested that resource productivity and domestic material consumption exist a strong statistical relationship with municipal waste recycling, and showed that an increment is associated with a positive change in municipal waste recycling. Additionally, The Ex'TAX project thought tax systems that subsidize the fossil fuels and other polluters are a major barrier on the transformation from linear economy to the circular economy, and thus they called for a price (tax) on natural resource use and pollution, which would provide incentives to save resources and the natural world. And their 2022 study on the added value and additional costs of circular construction at the Floriade, Dutch, drew the conclusion that tax shift can accelerate the transition to circular in construction industry. Last but not the least, Radivojević, V., et al. (2024) examined the impact of circular economy indicators on gross domestic product (GDP) per capita in the EU, and they obtained the results aligned with their assumptions, which indicated a strong positive correlation between circular economy indicators and GDP per capita. Also, their results also revealed a positive and statistically significant impact of Resource Productivity (RP), Generation of municipal waste per capita (MWpc), and Recycling rate of municipal waste (RRMW) on economic growth (GDPpc). Since this study will focus on cross-city, in contrast to the previous sections that emphasized national-level differences, we build upon the theoretical foundations discussed earlier and conduct an extensive review of existing outstanding studies. Based on these frameworks, we replace the measurement of GDP and resource productivity with the total income of each city and electricity consumption, respectively, to better correspond to the assessment of economic development and resource utilization in a particular city.

1.3 Hypothesis

Last but not the least, let us give some intuitive clairvoyance about our analysis results, those hypothesis combine with intuitive guessing and the empirical results made from others, stating as follows:

- **Hypothesis 1:** Environmental budgets, regulations(findings)and recycling efforts have a positive effect with the electricity efficiency(i.e. negative on per capita electricity usage).
- **Hypothesis 2:** Higher levels of landfill usage and incineration rate, but being poor on recycling practices are negatively associated with household income.
- **Hypothesis 3:** Per capita food waste and Per capita generation of general waste are positive correlated to the household income and electricity usage per capita (i.e. less electricity efficiency), at the same time, they are negative correlated to Vancancy rate.
- **Hypothesis 4:** Per capita of capacity of renewable energy are positive correlated to household income and electricity efficiency(i.e. negative correlation with per capita electricity usage).

2 Data Collections

For the present research, I utilize a dataset covering multiple cities/counties, comprising 19 variables related to economic development, energy consumption, environmental policies, and pollution indicators. The dataset is structured as a panel, including observations over 10 years (from 2014 to 2023), allowing for both cross-sectional and temporal analysis.

2.1 Data Sources

The data were collected from various official sources, including government statistical reports, environmental agencies, and energy-related institutions. The key sources include:

- **Economic and energy data:** Taiwan Power Company
- **Environmental policy and taxation data:** Directorate-General of Budget, Accounting and Statistics, Executive Yuan, R.O.C (Taiwan)
- **Pollution and waste management data:** Directorate-General of Budget, Accounting and Statistics, Executive Yuan, R.O.C (Taiwan)

2.2 Variables Description

The dataset consists of dependent variables measuring economic and energy consumption outcomes, and various independent variables capturing environmental policies, waste management, and related factors.

2.2.1 Dependent Variables

- **Household Income from Total Earnings** (in thousand NTD) – Represents the overall economic output of each administrative region.
- **Per Capita Electricity Usage** (kWh per person) – Measures energy usage efficiency at the local level.
- **Vacancy Rate of House** (percentage) – Used as a proxy for assessing housing market dynamics and regional economic conditions.

2.2.2 Independent and Policy-Related Variables

To assess the impact of circular economy policies on economic development and productivity, the following key indicators are included:

- **Waste Management and Recycling Indicators:**
  - General waste recycling rate (%)
  - Per capita incinerated waste volume (kg)
  - Per capita landfilled waste volume (kg)
  - Per capita resource recovery by local authorities (kg)
  - Per capita generation of general waste (kg)
  - Per capita food waste (kg)
- **Environmental Expenditures:**
  - Budget structure ratio for community development and environmental protection (%)
- **Environmental Penalties:**
  - Proportion of environmental fines (%)
- **Renewable Energy Infrastructure:**
  - Per capita installed capacity of renewable energy (MW)

2.3 Summary Table of Variables

The following tables provide a statistical summary of the key variables used in the analysis, including count, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum values. The variables are split into two tables for clarity.

	Household_income	General_waste_recycling_rate	Budget_ratio_for_environmental	Proportion_of_environmental_fines
count	200.0	220.0	220.0	220.0
mean	1 158 675.0	58.0	4.4	5.6
std	247 155.9	6.7	2.5	5.2
min	746 981.0	44.5	0.4	0.1
25%	963 350.5	54.0	2.1	3.0
50%	1 099 166.0	57.8	4.8	4.8
75%	1 300 000.0	61.3	6.4	7.0
max	1 851 383.0	78.2	9.4	61.6

	p-c_electricity_usage	p-c_incinerated_waste_kg	p-c_landfilled_waste_kg	p-c_resource_recovery_by_local_authorities_kg
count	220.0	220.0	147.0	220.0
mean	8842.4	153.3	15.2	210.1
std	6208.9	57.7	29.6	59.9
min	1197.5	8.6	0.0	84.6
25%	5109.6	117.2	0.3	164.3
50%	6914.5	158.1	1.8	210.6
75%	11 784.4	184.4	10.1	251.4
max	61 171.0	295.0	158.7	346.4

	p-c_generation_of_general_waste_kg	p-c_t_capacity_of_renewable_energy	Vancancy_rate	p-c_food_waste_kg
count	220.0	220.0	220.0	220.0
mean	413.0	0.4	11.0	29.5
std	101.9	0.5	2.6	29.2
min	189.3	0.0	6.2	9.2
25%	339.1	0.0	9.3	15.6
50%	404.4	0.2	10.6	23.7
75%	479.9	0.5	12.3	31.5
max	714.3	2.6	18.2	248.2

2.4 Data Processing

To ensure data consistency and suitability for statistical and machine learning models, the dataset undergoes several preprocessing steps:

- **Handling missing values:** Interpolation or mean imputation is applied where necessary.
- **Standardization and normalization:** Key variables are standardized to facilitate model convergence.
- **Panel data structuring:** The dataset is formatted to align with panel regression requirements, incorporating both city and year fixed effects where applicable.

The analysis will cover the period from **2014 to 2023** to capture the changes and impacts following Taoyuan's administrative restructuring of cities and counties in 2014, allowing for the observation of both the short-term and long-term effects of circular economy policies. This time frame also provides a comprehensive view of the evolving trends in economic performance, energy consumption, and environmental indicators.

3 Methods

For this study, we will dive into different kinds of machine learning methods, both linear and non-linear, and analyze which methods included could be the best predictor.

3.1 Linear Approach – Panel Data Regression Model

First of all, we employ a panel data regression model. Panel data combines longitudinal (time-series) and cross-sectional data, allowing us to control for unobservable individual heterogeneity and improve estimation accuracy. However, for those unobservable variables, which might introduce correlation issues, we need to specify which type of panel data models will be applied in our test.

3.1.1 Fixed Effects Model

Assumes that individual characteristics are time-invariant and correlated with the explanatory variables. Therefore, it placates individual heterogeneity through the demeaning method. We express it mathematically:

$$E(c_i \mid X'_{it}) \neq 0 \tag{1}$$

Consider the following model:

$$y_{it} = a + X'_{it}\beta + c_i + e_{it} \tag{2}$$

where we define:

- $i = 1, \ldots, N$, where i represents a county/city, and we have $N = 22$).
- $t = 2014, \ldots, 2023$.
- y_{it} is the dependent variable.
- X'_{it} are the independent variables of K-dimension.
- a is the intercept.
- β is the coefficient vector of K-dimension.
- c_i is the individual effect.
- e_{it} is the error term, and $e_{it} \sim \mathcal{N}(0, \sigma_e^2)$.

3.1.2 Random Effects Model

In the random effects model, we assume that the individual effect c_i is a random variable which doesn't relate to the other independent variables X'_{it}, and we can express it mathematically:

$$E(c_i \mid X'_{it}) = 0 \tag{3}$$

Thus the model will become:

$$y_{it} = a + X'_{it}\beta + c_i + e_{it} \tag{4}$$

where we define:

- c_i is the random effect, and $c_i \sim \mathcal{N}(0, \sigma_c^2)$
- The others remain the same as the fixed effects model.

3.1.3 Hausman Test

To determine the most suitable model, we conduct a Hausman test and construct the null hypothesis.

$$H_0 : E(c_i \mid X'_{it}) = 0$$
$$H_1 : E(c_i \mid X'_{it}) \neq 0$$

And $H_{\text{Statistic}} \sim \chi^2(K)$, where K represents the degree of freedom. We test the hypothesis at a 5% significance level. If the test results indicate that the p-value is less than 0.05, suggesting that individual effects are correlated with the explanatory variables, we will apply the FE model. In contrast, if the p-value does not reveal its significance, we then apply the RE model. Despite the fact that we have to conduct this test to select which model we will apply, in most empirical methods, however, we may have the proclivity of choosing the FE model in practice.

3.2 Non-Linear Approaches

In addition to the linear panel data regression models, we also employ machine learning algorithms such as Support Vector Machine (SVM), XGBoost, and Random Forest to capture more complex, non-linear relationships in the data. These models do not make assumptions about the form of the underlying relationship between variables and can handle high-dimensional data effectively. Below is a description of these models and their application to our analysis.

3.2.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning model used for both classification and regression tasks. It works by finding a hyperplane that best separates the data into different classes (in classification) or fits the data (in regression). The objective of SVM is to maximize the margin between the data points and the hyperplane, thereby improving generalization and reducing the risk of overfitting.

In regression tasks, SVM seeks to find a function that deviates from the actual data points by no more than a specified threshold. SVM is particularly useful when the data is not linearly separable in its original space. The mathematical formulation for the hard-margin SVM classification problem is given by:

$$\max_{\mathbf{w}} \left\{ \frac{2}{\|\mathbf{w}\|} \right\} \quad \Longrightarrow \quad \min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^\top \mathbf{w} \tag{5}$$

$$\text{subject to} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, n \tag{6}$$

Where:

- y_i is the true label of the i-th sample.
- $\mathbf{x}_i$ represents the input features.
- $\mathbf{w}$ is the weight vector.
- b is the bias term.

But that is an optimal case of SVM; sometimes we may not be able to find such a hyperplane that perfectly divides those data. In this scenario of the training dataset, we can tolerate some data lying "under the hyperplane". This method is known as "Soft-Margin SVM". Therefore, we can rewrite the mathematical optimization as:

$$\min_{\mathbf{w},\boldsymbol{\xi}} \frac{1}{2}\mathbf{w}^\top \mathbf{w} + C\sum_{i=1}^{n} \xi_i \text{subject to} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, n \tag{7}$$

Where:

- $\mathbf{w}^\top \mathbf{w}$: Represents the squared margin, which is minimized to increase the margin width.
- $\sum_{i=1}^{n} \xi_i$: Represents the total amount of margin violation across all samples.
- C: A regularization parameter that balances margin maximization and classification error.
- ξ_i: Slack variables that allow some flexibility by permitting samples to be within the margin or misclassified.

To solve this optimization problem, we can apply Lagrangian multipliers, take differentiation, and the remaining parts are trivial to tackle. We had better stop diving into the unfathomable mathematical world; also, we can save some space for the following introduction.

3.2.2 XGBoost (Extreme Gradient Boosting)

XGBoost is an efficient and ameliorated implementation of gradient boosted decision trees, widely used in machine learning competitions and real-world applications for both classification and regression tasks. It builds an additive ensemble model, where new decision trees are added sequentially to correct the residual errors made by the previous trees. The predicted output for a given input x is defined as:

$$\hat{y} = \sum_{k=1}^{K} f_k(x), \quad f_k \in \mathcal{F} \tag{8}$$

where:

- $\hat{y}$ is the final prediction score.
- K is the total number of trees.
- $f_k(x)$ is the function (i.e., decision tree) from the function space $\mathcal{F}$ learned in the k-th iteration.

XGBoost optimizes by minimizing the following regularized objective function:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(f_k) \tag{9}$$

where:

- $l(y_i, \hat{y}_i^{(t)})$ is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i^{(t)}$ and our actual label y_i.
- $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2$ is the regularization term, where T is the number of leaves in the tree and w represents the leaf weights.
- γ and λ are regularization parameters that control model complexity (L1 and L2 regularization, respectively).

3.2.3 Random Forest

Random Forest is an ensemble learning method; unlike the previous XGBoost, which builds trees in a sequential manner, Random Forest constructs each tree independently and then aggregates their predictions (using averaging for regression or majority voting for classification). The mathematical expression for the Random Forest regression model is:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x) \tag{10}$$

where:

- $\hat{y}$ is the predicted output.
- N is the number of trees in the forest.
- $T_i(x)$ represents the prediction from the i-th tree.

Each tree in the forest is built by randomly selecting a subset of the data and features (bootstrapping) for training, thus ensuring the model remains robust against overfitting and is able to generalize well on unseen data.

In a Random Forest, each decision tree is built by repeatedly splitting the dataset into two at each node. The goal is to select the feature and threshold that produce the purest possible child nodes. And purity is typically measured by using either **Gini Impurity** or **Entropy**, introduced as follows.

Gini Impurity. The Gini impurity for a node t is defined as:

$$\text{Gini}(t) = 1 - \sum_{k=1}^{K} p_k^2 \tag{11}$$

where:

- K: Number of classes.
- p_k: Proportion of samples in node t that belong to class k.

Entropy. The entropy for a node t is defined as:

$$\text{Entropy}(t) = -\sum_{k=1}^{K} p_k \log_2 p_k \tag{12}$$

where p_k is defined as above.

Information Gain. Given a candidate split that partitions the data into left and right child nodes t_L and t_R, the **Information Gain** is computed as:

$$\text{Gain} = \text{Impurity}(t) - \left(\frac{n_L}{n} \cdot \text{Impurity}(t_L) + \frac{n_R}{n} \cdot \text{Impurity}(t_R)\right) \tag{13}$$

where:

- $\text{Impurity}(t)$ can be either Gini or Entropy.
- n: Number of samples in the parent node.
- n_L, n_R: Number of samples in the left and right child nodes.

The algorithm searches for the split that maximizes our Information Gain, which is tantamount to minimizing the weighted impurity after the split.

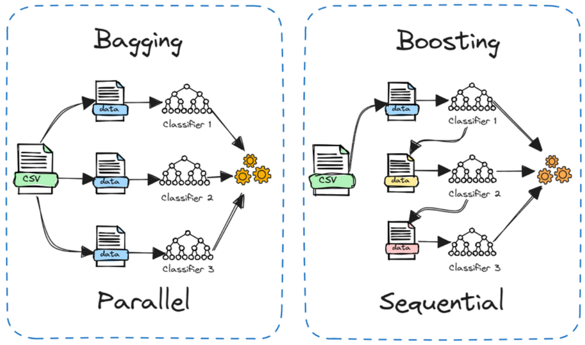


Figure 2: The difference in both ensemble methods (The left is Bagging, and the right is Boosting)

3.2.4 CatBoost (Category Boosting)

CatBoost is a gradient boosting algorithm specifically designed to handle categorical features effectively (which still works with continuous variables), reducing the need for extensive preprocessing. It introduces an ordered boosting mechanism to mitigate overfitting and employs a symmetric tree structure for balanced splits, enhancing robustness. Because of the similarity of the mathematical structure with XGBoost, the author will leave it as trivial(or as a practice).

With ordered boosting, CatBoost is particularly suitable for our relatively small datasets with mixed feature types; additionally, it performs well in tabular data tasks compared to XGBoost and LightGBM, which will be mentioned afterward.

3.2.5 LightGBM

LightGBM is a highly efficient gradient boosting framework, offering superior speed and memory usage compared to traditional implementations like XGBoost. It employs a histogram-based algorithm to bucket continuous feature values into discrete bins, reducing memory consumption and accelerating training. Paramountly, LightGBM uses a leaf-wise tree growth strategy, which prioritizes splitting the leaf with the maximum loss reduction, enabling faster convergence but requiring careful regularization to prevent overfitting. I'll provide some illustrations below. The mathematical intuition resembles XGBoost as mentioned above, thus I leave some space for further, more intriguing analysis.

Worth mentioning is that, in this study, due to our inherently limited dataset, which may engender the issue of overfitting. To tackle this potential problematic occurrence, I'll introduce 10-fold cross-validation in all of my nonlinear methods.

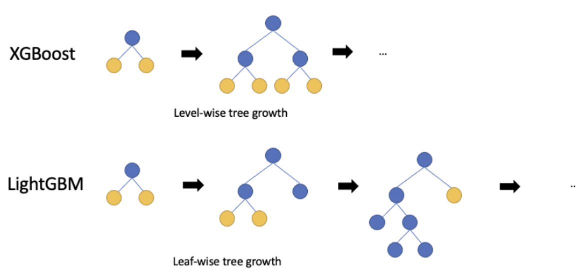


Figure 3: The difference in both Boosting methods (Photo credited by ithelp.ithome.com.tw)

3.2.6 Model Evaluation and Comparison

For the SVM, XGBoost, Random Forest, CatBoost, and LightGBM models, we evaluate the model performance using various metrics, such as the mean squared error (MSE), R-squared (R^2), and so on, to evaluate their predictive capabilities.

We also perform classical 10-fold cross-validation to ensure the robustness of the results and avoid overfitting. Hyperparameter tuning (e.g., number of nodes in each tree, learning rate) is performed to optimize the performance of each model. The results from these models are then compared to those obtained from the panel data regression models to assess the improvements in prediction accuracy and model performance.

3.2.7 SHAP (SHapley Additive exPlanations) for Model Interpretation

With a view to interpret the above-mentioned nonlinear machine learning models beyond mere predictive performance, we will employ SHAP (SHapley Additive exPlanations), a theoretic approach based on cooperative game theory. SHAP assigns each feature an importance value for a particular prediction by computing its marginal contribution, averaged over all possible feature coalitions. Unlike those traditional feature importance metrics (such as gain or split count), SHAP values rendre both the magnitude and direction of a feature's influence on the model output. And this approach offers several advantages:

1. Allow individual-level interpretation (i.e., how each feature affects in a particular city/year sample)

2. Support global feature importance ranking based on aggregated effects

3. Reveal nonlinear patterns or threshold effects in complex models. In this study, SHAP (TreeExplainer)values are calculated for all ensemble models (Random Forest, LightGBM, CatBoost, and XGBoost)

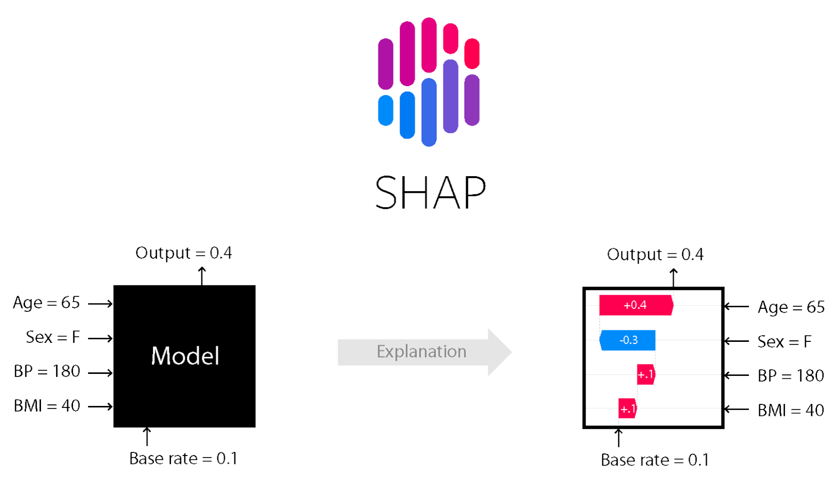


Figure 4: How SHAP value works in model interpretation(https://github.com/shap/shap)

4 Results

4.1 Correlation Heat Map

First of all—we take a glimpse of the correlation matrix and heat map of independent variables, which is illustrated as follows:

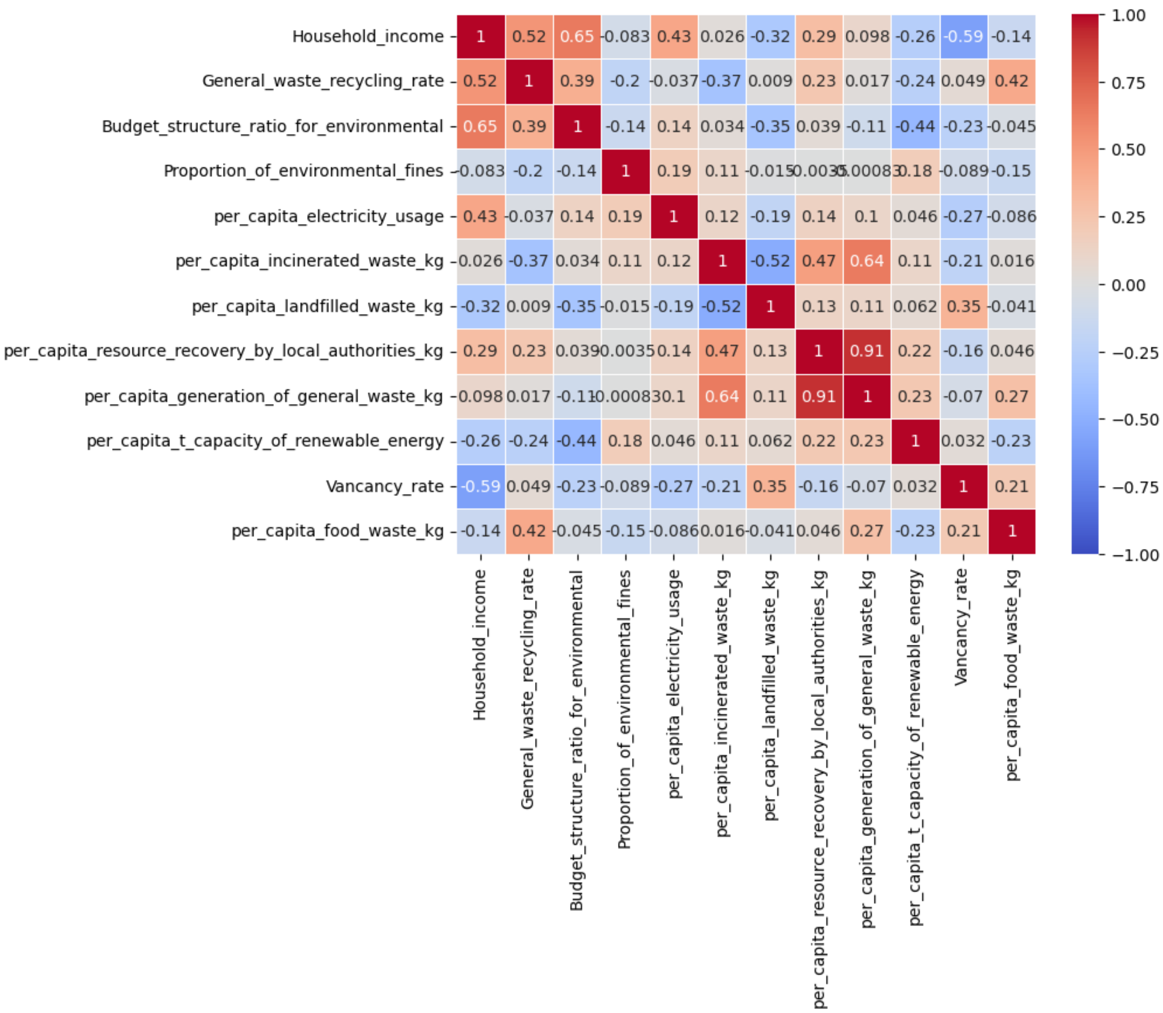


Figure 5: Correlation Heat Map of Independent Variables

4.2 Panel data Regression

Table 1: Model Comparison of Household Income

	Fixed Effects	**Random Effects**
Dep. Variable	Household_income	Household_income
R-squared	0.5755	0.5269
R-Squared (Within)	0.5755	0.5754
R-Squared (Between)	-7.664e-05	0.0008
R-Squared (Overall)	0.0132	0.0141
F-statistic	16.718	15.717
P-value (F-stat)	0.0000	0.0000
General_waste_recycling_rate	1.656e+04 (0.3543)	1.565e+04 (0.3247)
Budget_structure_ratio_for_environmental	-2.424e+04 (-1.5064)	-2.416e+04 (-1.4588)
Proportion_of_environmental_fines	-1656.4 (-0.3080)	-1520.4 (-0.2741)
per_capita_incinerated_waste_kg	-1.459e+04 (-1.0786)	-1.509e+04 (-1.0815)
per_capita_landfilled_waste_kg	-8152.9 (-0.7027)	-8504.2 (-0.7109)
per_capita_generation_of_general_waste_kg	9.208e+04 (0.9104)	9.113e+04 (0.8736)
per_capita_t_capacity_of_renewable_energy	2.426e+04 (2.6360)***	2.446e+04 (2.5775)***
per_capita_food_waste_kg	-6.258e+04 (-1.0021)	-6.412e+04 (-0.9958)
per_capita_resource_recovery_by_local_authorities_kg	-1.877e+04 (-0.1813)	-1.786e+04 (-0.1673)
Effects	Entity	

Then we concentrated on scrutinising our first regression table, in which the Dependent variable is Household Income. In this model, we compared both Fixed Effects (FE) and Random Effects (RE) estimators. Albeit the results of Hausman test isn't reject the hypothesis of individual effect having no correlation with the other independent variables, given the panel structure of the data and our primary focus on within-unit variation (i.e., how changes in circular economic variables affect household income over time within each region), the Fixed Effects model is our preferred specification, and thus the following part we are going to focus on only the FE model. The within R-squared for the FE model is 0.5755, indicating that approximately 57.6% of the variation in household income within regions over time can be explained by the included predictors. Among the variables, only the per capita installed capacity of renewable energy being statistically significant and positive correlation of household income, suggesting a potential linkage between regional investments in green energy and improved household economic well-being, and thus serves as a moderated evidence to support the establishing of 4-th hypothesis. Here, we illustrated the feature importance

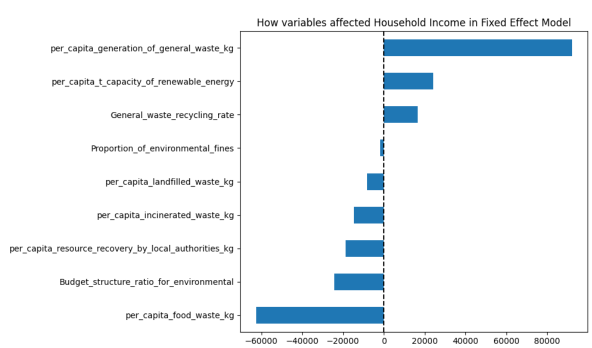


Figure 6: The Impact Variables on Household Income: Fixed Effect Model

Table 2: Model Comparison of Per Capita Electricity Usage

	Fixed Effects	**Random Effects**
Dep. Variable	p-c_electricity_usage	p-c_electricity_usage
R-squared	0.0538	0.0430
R-Squared (Within)	0.0538	0.0527
R-Squared (Between)	-0.0322	-0.0167
R-Squared (Overall)	-0.0776	-0.0570
F-statistic	0.7588	0.6891
P-value (F-stat)	0.6545	0.7178
General_waste_recycling_rate	-1748.9 (-1.0245)	-1630.1 (-0.9591)
Budget_structure_ratio_for_environmental	1469.0 (1.5002)	1385.3 (1.4779)
Proportion_of_environmental_fines	-56.664 (-0.1633)	-9.4697 (-0.0272)
per_capita_incinerated_waste_kg	-446.12 (-0.5155)	-573.76 (-0.6610)
per_capita_landfilled_waste_kg	-246.51 (-0.3294)	-371.66 (-0.4985)
per_capita_generation_of_general_waste_kg	-3261.6 (-0.7386)	-2803.4 (-0.6335)
per_capita_t_capacity_of_renewable_energy	374.06 (0.6450)	415.61 (0.7202)
per_capita_food_waste_kg	5519.8 (1.7388)*	4844.5 (1.5677)
per_capita_resource_recovery_by_local_authorities_kg	4079.9 (0.9360)	3577.6 (0.8217)
Effects	Entity	

Secondly, let us focus on our regressing results with the per capita of electricity usage as our dependent variable. In this model, the within R-squared for the FE model is 0.0538, saying that less than 6 % of the variation in dependent variable within regions over time can be explained by the included predictors, which has the large difference with our day-dreaming estimation results. Despite being not that comfortable, we still can be inferred from this regression table. Among the variables, only the per capita food waste stands out statistically significant and positive correlation of dependent variable under the 90% of confidence interval, suggesting a potential linkage between wasting food and also energy, and thus 3-th may be proved; worth-noting is that, the more the per capita of capacity of renewable energy, the more the per capita usage of electricity, which is contradict to our 4-th hypothesis on electricity efficiency, indicaing that the consciousness of saving electricity may not go with the increment of usage in renewable energy.

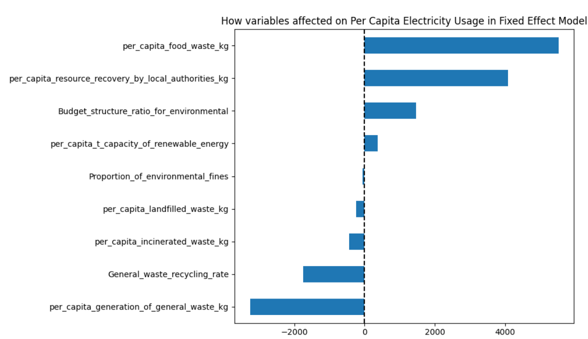


Figure 7: The Impact Variables on Per Capita Electricity Usage: Fixed Effect Model

Table 3: Model Comparison of Vacancy rate

	Fixed Effects	**Random Effects**
Dep. Variable	Vacancy_rate	Vacancy_rate
R-squared	0.3796	0.3371
R-Squared (Within)	0.3796	0.3795
R-Squared (Between)	0.0117	0.0125
R-Squared (Overall)	-0.0008	0.0002
F-statistic	8.1573	7.7976
P-value (F-stat)	0.0000	0.0000
General_waste_recycling_rate	1.1738 (4.8452)***	1.1819 (4.7618)***
Budget_structure_ratio_for_environmental	0.1066 (0.7670)	0.1001 (0.7037)
Proportion_of_environmental_fines	0.0098 (0.1999)	0.0108 (0.2138)
per_capita_incinerated_waste_kg	-0.1464 (-1.1922)	-0.1513 (-1.2023)
per_capita_landfilled_waste_kg	-0.3178 (-2.9923)***	-0.3194 (-2.9346)***
per_capita_generation_of_general_waste_kg	2.6392 (4.2111)***	2.6646 (4.1490)***
per_capita_t_capacity_of_renewable_energy	0.1055 (1.2815)	0.1081 (1.2824)
per_capita_food_waste_kg	-0.7543 (-1.6744)*	-0.7786 (-1.6881)*
per_capita_resource_recovery_by_local_authorities_kg	-2.7926 (-4.5143)***	-2.8214 (-4.4511)***
Effects	Entity	

Last but not the least, we examined the regression results where the dependent variable is Vacancy Rate, potentially reflecting urban usage intensity or housing inefficiency, and the economic situation within a city/county. The within R-squared of the FE model is 0.3796, implying that about 38% of the within-region variation over time in vacancy rates is explained by those predictors. In this model, however, several predictors showed significant associations, comepared to the previous two. Notably: General waste recycling rate is positively associated with vacancy rates which could suggest that areas with more recycling activity also experience higher vacancy, potentially due to population or consumption patterns. Per capita landfilled waste and resource recovery by local authorities are both negatively associated with vacancy rates, with strong statistical significance, possibly indicating that better waste management is linked to more efficient space usage. These findings suggest that certain circular economic metrics, particularly waste management indicators, are meaningfully related to urban vacancy dynamics.

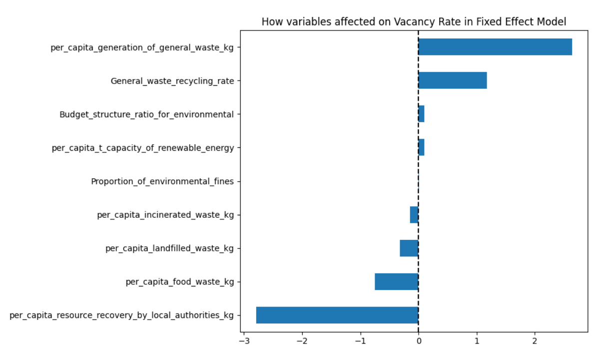
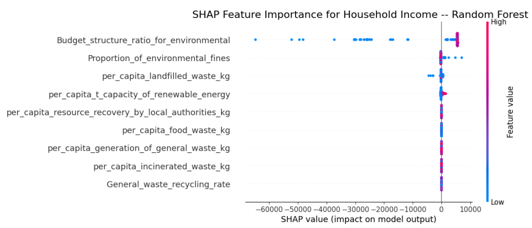


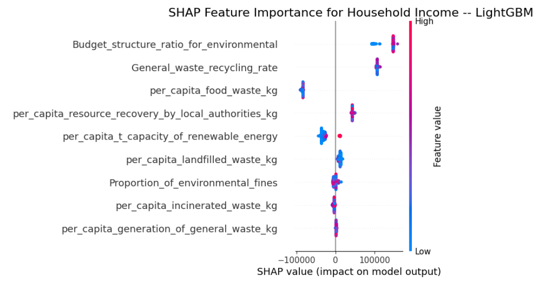
Figure 8: The Impact Variables on Per Capita Electricity Usage: Fixed Effect Model

4.3 Nonlinear Approach Cross-Comparison

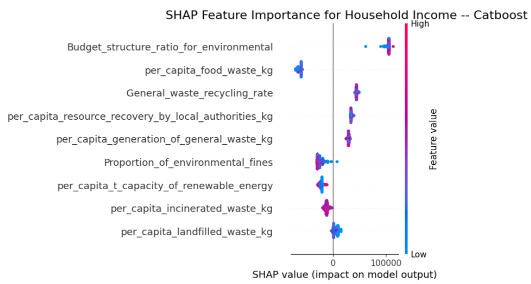
4.3.1 Results for Household Income



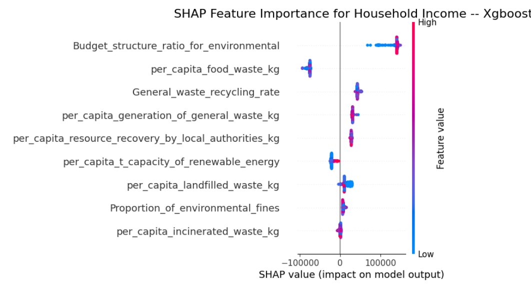
(a) Random Forest



(b) LightGBM



(c) CatBoost



(d) XGBoost

Figure 9: SHAP Value Comparison of Ensemble Methods

Table 4: SHAP Average Value Summary

Feature	shap_avg_xgb	shap_avg_rf	shap_avg_cat	shap_avg_lgbm
Budget_structure_ratio_for_environmental	135392	184066	103778	140778
General_waste_recycling_rate	43613	56781	44104	106733
per_capita_generation_of_general_waste_kg	30847	11152	29307	1692
per_capita_resource_recovery_by_local_authorities	27689	30928	34221	43616
per_capita_landfilled_waste_kg	16090	30967	4279	11109
Proportion_of_environmental_fines	8109	4410	-24839	-2747
per_capita_incinerated_waste_kg	298	-10202	-11706	-3360
per_capita_t_capacity_of_renewable_energy	-20539	-16840	-22892	-28026
per_capita_food_waste_kg	-76820	-72989	-63354	-84387

To better understand the contribution of miscellaneous circular economic indicators to prediction out first target variable, we calculated SHAP values across four nonlinear models, worth metioned is that, for the reasons why we only include four model instead of five? Frankly speaking, which conducted with those following data modelisation, some unexpected outcomes arosed – R-squared value of SVM model was negative, which could be implied as using SVM as our model has terrible compatability and peccable performance of the variables interpretation, thus based on the reason, the author decided to relinquish the following illustrations on SVM model, and concentrate only on those four models.

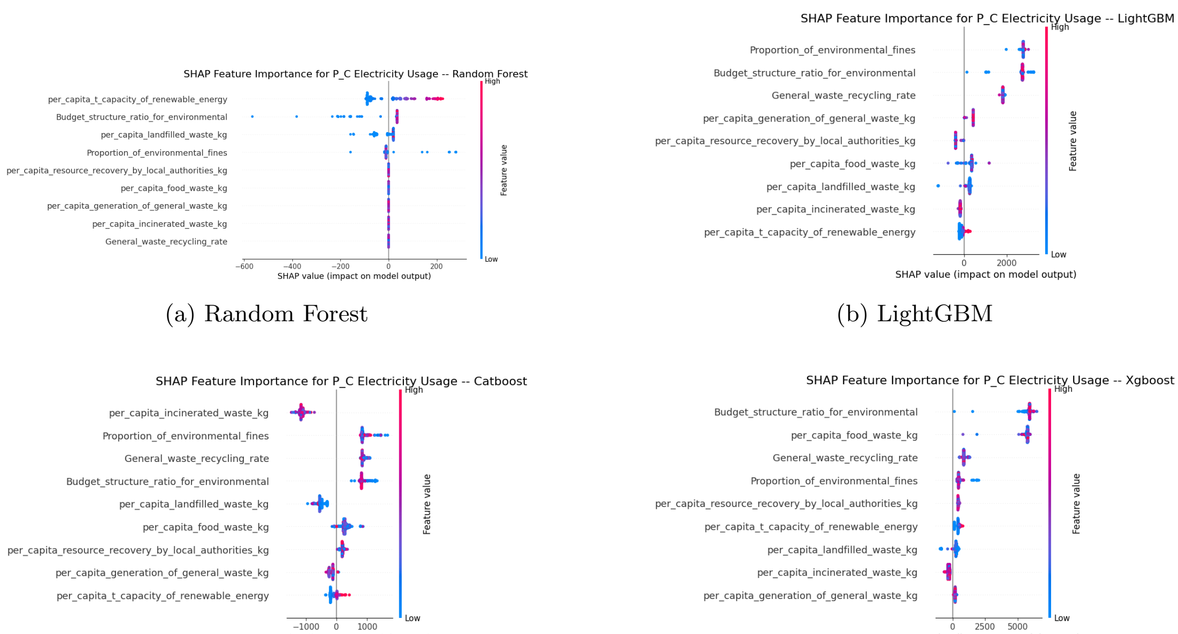
Let us back to the interpretation session, overall, the Budget structure ratio for environmental proved to be the most significant and stable feature across all models, with an average SHAP value exceeding 100,000(NTD) and consistently contributing positively. This indicates that a higher proportion of the budget allocated to environmental concerns tends to lead to higher predicted values. Other features, such as General waste recycling rate and per capita generation of general waste, also showed stable and positive contributions, but the value in LightGBM model deviates from others much. This suggests that societal or economic structures related to high recycling rates and per capita waste generation may indirectly influence the prediction target, implying that the third assumption we made is proved.

Intriguely to mentioned is that, per capita food waste and per capita of total capacity of renewable energy consistently had negative SHAP values across all models. This suggests a negative relationship with the prediction targets, contradicting to our 4-th hypothesis. The results indicating that higher food waste or inefficient renewable energy infrastructure may harm our income, the reason behind this may attribute to the richer the people, the higher the level of their concerns on food waste, as for the other, however, is worth dicussing in our further analysis.

Additionally, variables such as "Per capita incinerated waste (kg)" exhibited consistently negative SHAP values, Per capita landfilled waste (kg) is in converse, however. This implies an outcome that runs partially contrary to our second hypothesis.

Lastly, there was considerable variation in the contribution assessment for some features, such as Proportion of environmental fines. While CatBoost showed a strong negative contribution, Random Forest and XGBoost illustrated positive contributions. This suggests that there may be complex nonlinear interactions between these features and the target variable, which may be considered in our further investigation.

4.3.2 Results for Per Capita Electricity Usage



(a) Random Forest

(b) LightGBM

(c) CatBoost

(d) XGBoost

Figure 10: SHAP Value Comparison of Ensemble Methods

Table 5: SHAP Average Value Summary

Feature	shap_avg_xgb	shap_avg_rf	shap_avg_cat	shap_avg_lgbm
Budget_structure_ratio_for_environmental	5818	3	872	2683
per_capita_food_waste_kg	5680	0	292	307
General_waste_recycling_rate	903	0	880	1825
Proportion_of_environmental_fines	552	-1	902	2782
per_capita_resource_recovery_by_local_authorities	443	0	196	-387
per_capita_t_capacity_of_renewable_energy	417	8	-95	-108
per_capita_landfilled_waste_kg	247	-6	-509	223
per_capita_generation_of_general_waste_kg	190	0	-173	400
per_capita_incinerated_waste_kg	-283	0	-1143	-195

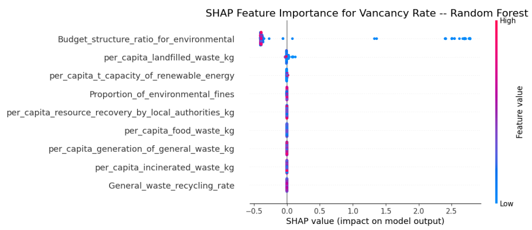
Let's going further, to investigate the contribution of various circular economy-related indicators to the prediction of the second variable, there we also computed SHAP (SHapley Additive exPlanations) values across four nonlinear models—namely, XGBoost, Random Forest, CatBoost, and LightGBM.

Diving into the results, "Budget structure ratio for environmental" emerged as one of the first impactful feature across all models, with SHAP values of particularly high magnitude—especially in XGBoost and LightGBM. This suggests that cities or regions allocating a higher proportion of their budget to environmental policies are more likely to achieve favorable outcomes in the prediction target, reinforcing our first assumption about policy investment indeed playing a key role. Other features like the "General waste recycling rate" and "Per capita generation of general waste" also demonstrated positive contributions across most models, thus serve as a concrete evidence our third assumption on electricity efficiency. Interestingly, in LightGBM, the SHAP value for recycling rate was significantly higher than in other models, hinting at model-specific sensitivity or possibly nonlinear interactions with other variables.

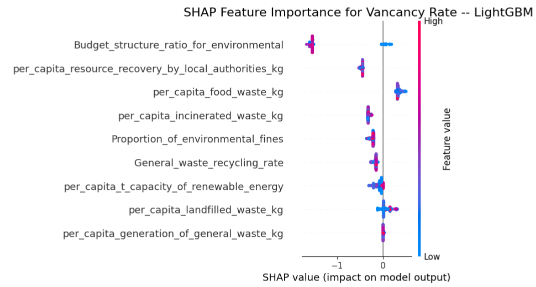
On contrary, per capita of total capacity of renewable energy had negative SHAP values in the last two models, which fited our expectations, but positive in XGBoost and Random Forest model(even the most significantly positive one). This results, Hardly can we interpret the validity of our 4-th hypothesis.

One possible interpretation for positive value is that, for renewable energy capacity warrants more electricity, it could reflect the misconception on overutilization or misallocation of the energy, thus causing additional waste on electricity. Interestingly, mostly all features show significant variation compared to Random Forest model. For example, the recycling rate has minimal impact in Random Forest but is much more influential in LightGBM (1825). This discrepancy suggests that tree-based ensemble models capture feature interactions differently, and model choice can materially affect interpretability outcomes.

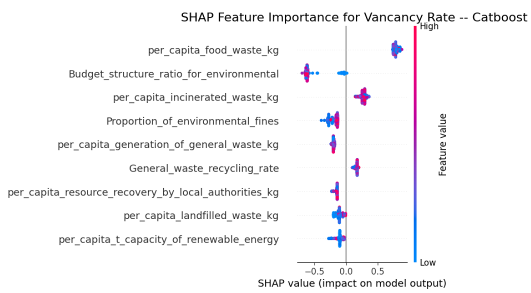
4.3.3 Nonlinear Approach Cross-Comparison– Results for Vacancy rate



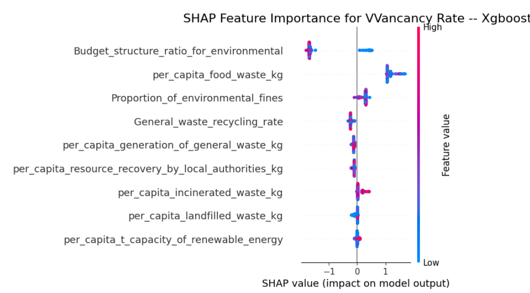
(a) Random Forest



(b) LightGBM



(c) CatBoost



(d) XGBoost

Figure 11: SHAP Value Comparison of Ensemble Methods

Table 6: SHAP Average Value Summary

Feature	shap_avg_xgb	shap_avg_rf	shap_avg_cat	shap_avg_lgbm
per_capita_food_waste_kg	1.146	0.000	0.786	0.334
Proportion_of_environmental_fines	0.273	0.000	-0.201	-0.229
per_capita_incinerated_waste_kg	0.093	0.000	0.268	-0.320
per_capita_t_capacity_of_renewable_energy	0.009	-0.000	-0.095	-0.087
per_capita_landfilled_waste_kg	-0.032	0.004	-0.109	0.051
per_capita_resource_recovery_by_local_authorities	-0.113	0.000	-0.149	-0.457
per_capita_generation_of_general_waste_kg	-0.127	0.000	-0.197	0.003
General_waste_recycling_rate	-0.226	0.000	0.164	-0.166
Budget_structure_ratio_for_environmental	-1.418	0.017	-0.547	-1.327

Then about the last target variable, among all features, Per Capita Food Waste (kg) stands out with consistently high SHAP values, especially in XGBoost (1.146) and CatBoost (0.786), indicating that food waste per person is a strong and stable predictor across models. This does not align with our third assumption on Vancancy rate should demonstrate negative correlation with food waste.

In contrast, Budget Structure Ratio for Environmental—which was previously dominant in raw SHAP value terms—now shows strong negative SHAP values in most models (e.g., -1.418 in XGBoost and -1.327 in LightGBM). This suggests that a higher share of budget allocated to environmental purposes might be associated with lower predictions of the target variable, perhaps reflecting higher spending in environment serve as an external benefit of economic performance(people are more willing to move into the city which care more about the environment).

Several features show more model-specific behavior. For instance, General Waste Recycling Rate has a slightly positive impact in CatBoost (0.164) but negative in both XGBoost and LightGBM, partially correspond to our third assumption.

Finally, it is notable that Random Forest contributes almost no variation in SHAP values across features—most are effectively zero. This could reflect lower model sensitivity to individual predictors or possible over-smoothing due to tree averaging.

5 Discussion and Implication

5.1 Model Comparison

Table 7: Summary R-squared Value Across Different Models

Feature	XGBoost	Random Forest	LightGBM	CatBoost	FE Panel Regression
Household Income	0.720	0.688	0.699	0.752	0.578
Per Capita Electricity Usage	0.427	0.479	0.474	0.420	0.054
Vacancy rate	0.736	0.659	0.600	0.743	0.380

Moving forward to the cross comparison, showing as above, Table 7 presents a comparative analysis of R-squared values across five different modeling approaches for three dependent indicators. Here, the author rendre some pivitol observations, which include:

- **Household Income Prediction:**
  - CatBoost demonstrates superior performance (R^2=0.752), followed closely by XGBoost (0.720)
  - The 23.1% improvement of CatBoost over FE Panel Regression (0.578) highlights the advantage of gradient boosting methods for income-related predictions
- **Per Capita Electricity Usage:**
  - All models show relatively weak predictive power ($R^2 < 0.5$), suggesting inherent complexity in electricity consumption patterns
  - Random Forest achieves the highest performance (0.479), other non-linear methods follows its predicability, in general they are above 0.42, while FE Panel Regression fails to capture meaningful variance (R^2=0.054)
- **Vacancy Rate Estimation:**
  - CatBoost again leads (0.743), with XGBoost showing comparable performance (0.736)
  - The significant gap between machine learning models and FE Panel Regression (0.380) suggests traditional methods may oversimplify vacancy dynamics

5.2 Key Findings

Here, we draw several conclusions based on the observations and results from the report.

- **Algorithm Performance:**
  1. Boosting algorithms, particularly **CatBoost and Xgboost**, consistently outperform other modeling approaches in terms of predictive accuracy and robustness. This performance edge is evident across various dependent variables and tasks, highlighting the algorithm's capacity to handle structured tabular data with complex interactions.

- **Prediction Difficulty:**

  1. Among the three outcome variables, **per capita electricity usage** emerges as the most challenging to predict. This may reflect higher variability, unobserved behavioural factors, or external influences such as climate and infrastructure not fully captured by the model.
  2. By contrast, **household income and vacancy rate** show relatively higher predictability, suggesting that socioeconomic indicators are more responsive to the selected explanatory variables. This implies that household and urban characteristics tend to be more systematically influenced by policy-related and environmental variables than energy usage patterns.

- **Methodological Implications:**

  1. The results reinforce the substantial **advantages of machine learning methods** for modeling complex socioeconomic outcomes. These approaches effectively capture nonlinear relationships and high-dimensional interactions that traditional econometric models may miss.
  2. However, there remains a **trade-off between interpretability and predictive power**. While fixed effects panel regression provides clear causal inference and policy-relevant coefficients, its limited flexibility may reduce predictive accuracy. Thus, the selection of method should align with the research objective—whether the goal is to explain (interpret) or to predict.

Table 8: Variable Significance Across Different Models: Household Income

Variable	XGBoost	Random Forest	CatBoost	LightGBM	Linear (FE)
Budget structure ratio for environmental	positive (+)	positive (+)	positive (+)	positive (+)	not significant
General waste recycling rate	not significant	positive (+)	positive (+)	positive (+)	not significant
Per capita generation of general waste (kg)	not significant	not significant	not significant	not significant	not significant
Per capita resource recovery by local authorities	not significant	not significant	not significant	not significant	not significant
Per capita landfilled waste (kg)	not significant	not significant	not significant	not significant	not significant
Proportion of environmental fines	not significant	not significant	negative (-)	not significant	not significant
Per capita incinerated waste (kg)	not significant	not significant	not significant	not significant	not significant
Per capita capacity of renewable energy	not significant	not significant	not significant	not significant	positive (+)
Per capita food waste (kg)	negative (-)	negative (-)	negative (-)	negative (-)	not significant

Table 9: Variable Importance and Direction of Impact across Methods: Per Capita of Electricity Usage

Variable	XGBoost	Random Forest	CatBoost	LightGBM	Linear (FE)
Budget structure ratio for environmental	Positive (+)	Positive (+)	Positive (+)	Positive (+)	Not significant
Per capita food waste (kg)	Positive (+)	Not significant	Not significant	Not significant	Positive (+)
General waste recycling rate	Not significant	Not significant	Positive (+)	Positive (+)	Not significant
Proportion of environmental fines	Not significant	Not significant	Positive (+)	Positive (+)	Not significant
Per capita resource recovery by local authorities	Not significant	Not significant	Not significant	Not significant	Not significant
Per capita t capacity of renewable energy	Not significant	Positive (+)	Not significant	Not significant	Not significant
Per capita landfilled waste (kg)	Not significant	Negative (-)	Not significant	Not significant	Not significant
Per capita generation of general waste (kg)	Not significant	Not significant	Not significant	Not significant	Not significant
Per capita incinerated waste (kg)	Not significant	Not significant	Negative (-)	Not significant	Not significant

Table 10: Variable Importance and Direction of Impact across Methods: Vacancy Rate

Variable	XGBoost	Random Forest	CatBoost	LightGBM	Linear (FE)
Per capita food waste (kg)	Positive (+)	Not significant	Positive (+)	Not significant	Negative (-)
Proportion of environmental fines	Not significant	Not significant	Not significant	Not significant	Not significant
Per capita incinerated waste (kg)	Not significant	Not significant	Not significant	Not significant	Not significant
Per capita t capacity of renewable energy	Not significant	Not significant	Not significant	Not significant	Negative (-)
Per capita landfilled waste (kg)	Not significant	Positive (+)	Not significant	Not significant	Not significant
Per capita resource recovery by local authorities	Not significant	Not significant	Not significant	Negative (-)	Negative (-)
Per capita generation of general waste (kg)	Not significant	Not significant	Not significant	Not significant	Positive (+)
General waste recycling rate	Not significant	Not significant	Not significant	Not significant	Positive (+)
Budget structure ratio for environmental	Negative (-)	Positive (+)	Negative (-)	Negative (-)	Not significant

Variable	Household Income	Pc Electricity Usage	Vacancy Rate
Budget structure ratio for environmental	+4	+4	-2
General waste recycling rate	+3	+2	+1
Per capita generation of general waste (kg)	0	0	+1
Per capita resource recovery by local authorities	0	0	-2
Per capita landfilled waste (kg)	0	-1	0
Proportion of environmental fines	-1	+2	0
Per capita incinerated waste (kg)	0	-1	0
Per capita capacity of renewable energy	+1	+1	-1
Per capita food waste (kg)	-4	+2	+1

Table 11: Summary of Variable Scores across Models

At the final stage of our analysis, we evaluate the relative significance of each explanatory variable across the three dependent indicators using five different modeling approaches. To determine the significance of a variable in the nonlinear models, we compute the average absolute SHAP value for all variables within each model. A variable is deemed as significant if its SHAP value exceeds the average, with the direction of impact indicated by the sign of the SHAP value—positive or negative, as shown from Table 8 to Table 10. For the linear fixed-effects panel regression, significance is simply determined based on the presence of statistical significance markers (e.g., asterisks) in the regression output. To quantify the overall importance of each variable, we assign a score based on its significance in each method: +1 for significant positive impact, -1 for significant negative impact, and 0 if not significant. These scores are then aggregated across the five models for each of the three indicators. The summarized results are illuminated in Table 11.

5.3 Summary of Explanatory Variables' Effects

As in Table 11, we conclude with interpretation of each explanatory variables, and worth methioning is that, larger positive values indicate a stronger positive relationship (e.g., more electricity usage), while larger negative values indicate a stronger negative effect. Note that a higher **Vacancy Rate** is interpreted as a *negative economic outcome* (i.e., more vacant homes imply weaker urban development).

- **Budget structure ratio for environmental:** Strong positive effect on household income and electricity usage, suggesting cities with higher environmental budget shares tend to have higher income levels and energy use. It has a moderate negative effect on vacancy rate, implying improved economic activity and lower vacancy.
- **General waste recycling rate:** Positively affects all three indicators, especially for household income. This suggests that a high recycling rate is associated with better environmental management and may contribute to economic vitality, but on the contrary, it slightly pull up the electricity usage.
- **Per capita generation of general waste:** In general, this explanatory variable has no significant effect on income or electricity usage, with a slightly positive association with vacancy rate—possibly reflecting inefficient consumption patterns in areas with more vacant properties.
- **Per capita resource recovery by local authorities:** Shows a moderate negative effect on vacancy rate, indicating that better local recycling systems help reduce their house vacancy rate. However, there is no significant impact on income or electricity usage.
- **Per capita landfilled waste:** Slight negative effect on electricity usage, possibly suggesting that areas relying more on landfilling have lower energy demand, but in general it is not a critical explanatory variable, implying the priority of policies making is lower than the others.

- **Proportion of environmental fines:** Has a moderate positive effect on electricity usage, possibly due to more industrial or urban activities. However, it shows a negative effect on income, which may reflect poorer governance or environmental issues.
- **Per capita incinerated waste:** Slightly reduces electricity usage. No notable effects on other indicators. Tantamount to **Per capita landfilled waste**, it may not be the prior concerns when it comes to policies decision making.
- **Per capita capacity of renewable energy:** Slightly increases household income and electricity usage, while reducing vacancy rate. This suggests that renewable energy capacity is linked to more vibrant, economically stable cities. However, it deficiency of significance is not enough to support our **Assumption 4**.
- **Per capita food waste:** Strong negative effect on household income—potentially associated with income inequality or lower-income groups. However, it increases electricity usage and slightly increases vacancy, indicating resource inefficiency in consumption. Additionally, it serves as a strong evidence of proving the contradiction of **Assumption 3**, providing us an innovative respective.

5.4 Evaluations of The Assumptions

Note: A higher value of per capita electricity usage indicates more energy consumption (i.e., lower energy efficiency), and a higher vacancy rate reflects more unoccupied housing, which is generally interpreted as a sign of weaker economic conditions. Therefore, a ***negative effect*** *on electricity usage or vacancy rate implies a* ***positive economic or environmental outcome****.*

Based on the empirical results from both linear (Fixed Effects) and nonlinear machine learning models (XGBoost, LightGBM, CatBoost, Random Forest), the following evaluations of our initial hypotheses are presented:

- **Hypothesis 1:** Environmental budgets, regulations, and recycling efforts have a positive relationship with electricity efficiency(i.e. negative on per capita electricity usage).
  Status: Inversely supported
  Environmental budget ratios exhibit a consistently positive effect on electricity usage (General waste recycling rate exhibit less but positive), which actually implies **higher** energy consumption (i.e., **lower** energy efficiency), not necessarily the intended policy outcome. Therefore, while a relationship exists, the direction is opposite to efficiency expectations.
- **Hypothesis 2:** Higher levels of landfill usage and incineration, combined with poor recycling practices, are negatively associated with household income.
  Status: Partially supported
  The signs of coefficients for landfill and incineration are mostly negative (as expected), but lack statistical significance in linear models. Some support is observed in linear regression outcomes and SHAP-based nonlinear results.
- **Hypothesis 3:** Per capita food waste and the generation of general waste are positively correlated with household income and electricity usage (i.e. less electricity efficiency), and negatively correlated with vacancy rate.
  Status: Partially supported
  Food waste is negatively associated with household income and positively associated with electricity usage (less electricity efficiency) and vacancy rate—suggesting inefficiency and weaker economic outcomes. These patterns contradict the initial assumption.
- **Hypothesis 4:** Per capita capacity of renewable energy is positively correlated with household income and electricity efficiency (i.e., negatively with electricity usage).

Status: Partially supported
A significant and positive effect on income is observed in linear models. However, the relationship with electricity usage is inconsistent and often positive—indicating more, not less, energy consumption. The efficiency benefit remains unverified.

Table 12: Summary of Hypotheses Evaluation

Hypothesis	Statement	Conclusion
1	Environmental budgets and recycling improve electricity efficiency	**Inversely supported**
2	Landfill/incineration harm income, poor recycling lowers income	**Partially supported**
3	Waste positively linked to income and electricity usage, negatively to vacancy	**Partially supported**
4	Renewable energy capacity improves income and electricity efficiency	**Partially supported**

6 Future Aspects

This study differs from previous research on the circular economy by conducting a comprehensive analysis of its impacts on both economic and environmental outcomes. In addition to employing traditional linear panel data regression models, this study also incorporates 5 ways of machine learning techniques to perform variable importance analysis, aiming to identify key circular economy indicators that influence crucially economic performance and environmental quality.

Based on the preliminary and (relatively) rudimentary results, several ambitious directions are proposed for the future research:

First, there are many additional variables that could better capture the degree of circular economy implementation. Including such as but not limited to more circular economy indicators—such as the City/County Circular Material Use Rate, Secondary Raw Materials Usage, proportion of Outstanding Enterprises in Resource Circulation, could enrich our analysis. Furthermore, as many countries are beginning to adopt circular economy practices, robustness checks and comparative studies across countries or regions could provide deeper insights.

Second, while this study primary focuses on key circular economy variables affecting economic and environmental outcomes, there are many other potential influencing factors. Incorporating these into the analysis could allow for the development of predictive models using machine learning algorithms to forecast variables such as CO_2 emissions or Air Quality Index (AQI) Value, or even House Pricing, will be much more intriguing.

Lastly, aside from implementation of machine method to give summary results, apply it to prediction, causal inference (In particular, we can conduct a Regression Discontiuinity on the proportion of environment budget. It is juat a sketch, however, we may be more cautious on examining the feasabilities), or apply to environmental Kuznets curve are also worth exploring, can help clarify more and being more insigtful when formating policies.

References

Boundling, K. E. (1966). The economics of the coming spaceship Earth. In H. Jarrett (Ed.), *Environmental quality in a growing economy* (pp. 3–14). Johns Hopkins University Press.

Ellen MacArthur Foundation. (2015). *Towards the circular economy: Economic and business rationale for an accelerated transition*. Retrieved from `https://www.ellenmacarthurfoundation.org`

European Commission. (2020). *Circular economy action plan: For a cleaner and more competitive Europe*. Retrieved from `https://ec.europa.eu/environment/circular-economy/`

Ex'TAX Project. (2022). *Tax as a force for good: Aligning tax systems with the SDGs and the circular economy*. Retrieved from `https://www.ex-tax.com`

Radivojević, V., Jovanović, L., & Milinković, D. (2024). The impact of circular economy indicators on GDP per capita: Evidence from the European Union. *Sustainability*, *16*(4), 1145. https://doi.org/10.xxxx/su16041145

Tanatu, A., Vasilescu, M. D., & Dobre, R. (2018). Dependencies between circular economy indicators in the European Union. *Romanian Journal of Economic Forecasting*, *21*(4), 89–101.

Weetman, C. (2016). *A circular economy handbook for business and supply chains*. Kogan Page Publishers.