JOURNAL OF
EMERGING INVESTIGATORS

Article

Distributional effects of residential energy tax credits: A machine learning approach

Aritra Ray[1], Alexandra Dill[2]

[1] Princeton High School, Princeton, New Jersey

[2] Sustainable Resource Management, SUNY College of Environmental Science and Forestry, Syracuse, New York

SUMMARY

The rollout of the Inflation Reduction Act in 2022 introduced new tax credits to accelerate adoption of renewable technologies. However, studies into how existing credits are distributed across income strata are sparse. We used a machine-learning based approach to analyze how residential energy tax credits are distributed across income brackets. We hypothesized that the level of tax incentives claimed would increase as the level of income increased. We created a logistic regression model to compare the log-odds of uptake and a linear regression model to compare the relative amounts of tax credits claimed across income brackets. A comparison of these models showed how uptake of credits differs across income brackets, with a non-linear relationship between income and usage of energy tax credits. While we observed a general trend of increasing tax credit utilization with higher income, the increase was not uniform throughout. Our results suggested that existing incentives may not effectively reach middle-income households, potentially due to eligibility limitations for lower-income programs and insufficient financial capacity to afford sustainable technologies. Further research is crucial to understand the specific barriers faced by this income group, which could include limited access to information about programs, the high upfront cost of technology adoption, or a lack of targeted support. Addressing the challenges faced by middle-income households, who may not qualify for low-income assistance but still face financial barriers to adopting sustainable technologies, is vital to maximizing the impact of such incentives and ensuring broader access to the benefits of a sustainable future for all income levels.

INTRODUCTION

The global push to transition towards a sustainable future is dependent on the widespread adoption of renewable energy technologies and the implementation of energy-efficient practices (1). However, this transition is not unfolding uniformly across society, with disparities emerging across income levels and thereby raising critical questions about equitable access to the benefits of a green economy (2). Both low-income households and middle-income households face significant barriers to adopting sustainable technologies, even though these technologies offer potential long-term cost savings and contribute to a healthier environment (3). Low-income households are defined by the U.S. Department of Housing and Urban Development as households having income less than 80% of an area's median income, while middle-income households are defined as having an income 80% to 150% of an area's median income (4). The upfront costs of solar panels, electric vehicles, and energy-efficient appliances can be prohibitive for these households, effectively excluding them from participating in the green transition and perpetuating a cycle of energy poverty (5).

Furthermore, the existing landscape of incentive programs, which is designed to encourage sustainable practices through tax credits, rebates, and other financial mechanisms, may not effectively reach or adequately support all communities (6). Historically, the U.S. federal government has relied on the Residential Energy Efficient Property Credit (Section 25D) and the Nonbusiness Energy Property Credit (Section 25C) as the main tools of renewable technology incentives. These programs support a variety of technologies, including solar panels, geothermal heat pumps, and fuel cells, as well as building improvements like high-efficiency insulation, windows, and efficient HVAC appliances (7). Administered through annual income tax filings with the Internal Revenue Service (IRS), the incentives allow homeowners to subtract a percentage of their equipment and installation costs directly from their total tax liability (8). These programs must often be paired with upfront investment. Consequently, the benefits of these incentives disproportionately accrue to higher-income households, exacerbating existing inequalities and hindering the overall progress towards national sustainability goals (9).

The Inflation Reduction Act (IRA), passed by the U.S. Congress in 2022, addresses these disparities with unprecedented investments in clean energy and provisions aimed at promoting equitable access to sustainable technologies (10). The IRA allocates substantial funding for programs designed to lower energy costs for low-income households, expand access to clean energy technologies, and create green jobs in disadvantaged communities (11). Disadvantaged communities targeted by the IRA include low- and middle-income communities, communities on Indigenous land, rural communities, and Black and Brown communities, among others (11). However, the effectiveness of these provisions in reaching both low-income and middle-income communities must be evaluated.

We therefore investigated the relationship between income and amount claimed of residential energy tax credits by leveraging data from the IRS's Statistics of Income (SOI) program, which provides a full dataset on tax returns in all U.S. zip codes (12). Specifically, we examined the uptake of energy-related tax credits across different income brackets to identify potential gaps in accessibility and effectiveness

https://doi.org/10.59720/25-070

within the existing incentive structures. We hypothesized that households with higher incomes would exhibit correspondingly higher levels of utilization of the Residential Energy tax credit because higher-income households have more disposable income to spend on upgrades for their home. Our analysis of IRS data revealed a non-linear relationship between income and the utilization of residential energy tax credits, with a notable decline in energy tax credit claims for middle-income brackets where individuals earned between $75,000 and $100,000. Our research is critical for informing policy interventions that promote equitable access to sustainable technologies. Ensuring that the benefits of a green economy are shared by all segments of society ultimately helps in accelerating the transition towards a more sustainable future for the U.S. (1). Understanding the relationship between income and incentives will allow for the creation of effective policies to drive sustainable technology adoption and foster the accelerated uptake of renewable technologies across economic strata.

RESULTS

To investigate the relationship between household income and energy tax credit claims, we first determined the distribution of total tax filers across six income categories (**Figure 1**). Household income was defined by Adjusted Gross Income (AGI), ranging from bracket 1 ($1 to $25,000) to bracket 6 (above $200,000). The number of filers in each income bracket varied widely, with income bracket 1 having the most filers, and income bracket 6 having the least (**Figure 1**). Generally, the number of filers decreased as income increased, with the exception of income bracket 5 (**Figure 1**). Across income brackets, the distribution of the number of credits claimed followed a similar pattern. To better understand these distributions, we generated histograms to visualize the distribution of non-zero credit claims for each of the six income brackets (**Figure 2**). Across income brackets, the distributions were skewed to the right, indicating that it is far more common for only a small number of households to claim energy tax credits in any given geographic and income stratum (**Figure 2**). For example, in income bracket 3, the most frequent number of claimants per zip code involved approximately 40–50 claimants, with instances of larger-scale adoption being significantly rarer (**Figure 2C**). This pattern held across all income levels, suggesting that widespread community usage of energy tax credits is relatively uncommon (**Figure 2**).

Given that the primary outcome of interest—whether a household claimed residential energy tax credits or not—was binary, a logistic regression model was used and revealed a non-linear relationship between income and likelihood of claiming a tax credit (**Figure 3**). While the likelihood of claiming a tax credit increased with income across income brackets 1, 2, and 3 with log-odds of -1.86, 0.39. and 0.66 respectively, we observed a notable decline in claims for households in income bracket 4, with a log-odds of 0.31, supporting the hypothesis of an accessibility gap (**Figure 3**). Using a logistic regression model, we found the highest positive log-odds in this model was 0.692 for income bracket 5 (**Figure 3**). Moreover, there was also a decrease in the model coefficient observed for income bracket 6, which we found to have a log-odds of -0.35 (**Figure 3**).

To understand how the magnitude of energy tax credits claimed varied across income, we examined the model coefficients for each bracket relative to a baseline (**Figure 4**). We found coefficients of -52.525 for income bracket 1 and -30.201 for income bracket 2, indicating that these groups claim minimal amounts compared to income bracket 3, which had a coefficient of 0. Notably, income bracket 4 deviated from the expected trend with a negative coefficient of -4.438, suggesting these filers claimed less than those in the bracket immediately below them (**Figure 4**). Similarly, income bracket

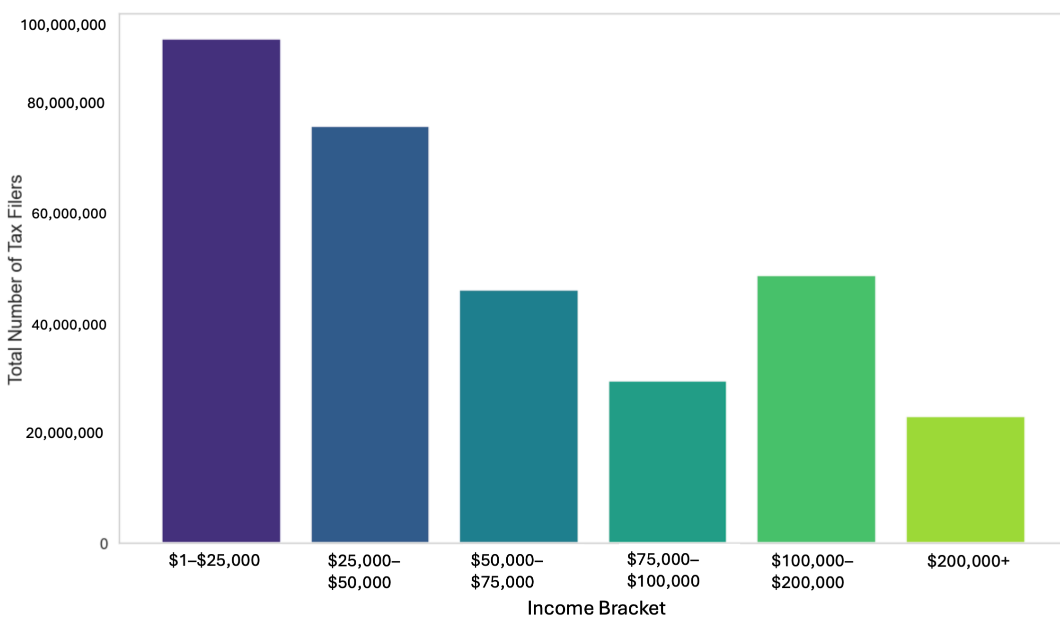

Figure 1: Distribution of tax filers by income bracket. The number of 2021 tax filers broken down by income bracket. Bracket 1 had the most filers, while Bracket 6 had the least filers. Bracket 1: $1–$25,000; bracket 2: $25,000–$50,000; bracket 3: $50,000–$75,000; bracket 4: $75,000–$100,000; bracket 5: $100,000–$200,000; bracket 6: $200,000 and up.

https://doi.org/10.59720/25-070

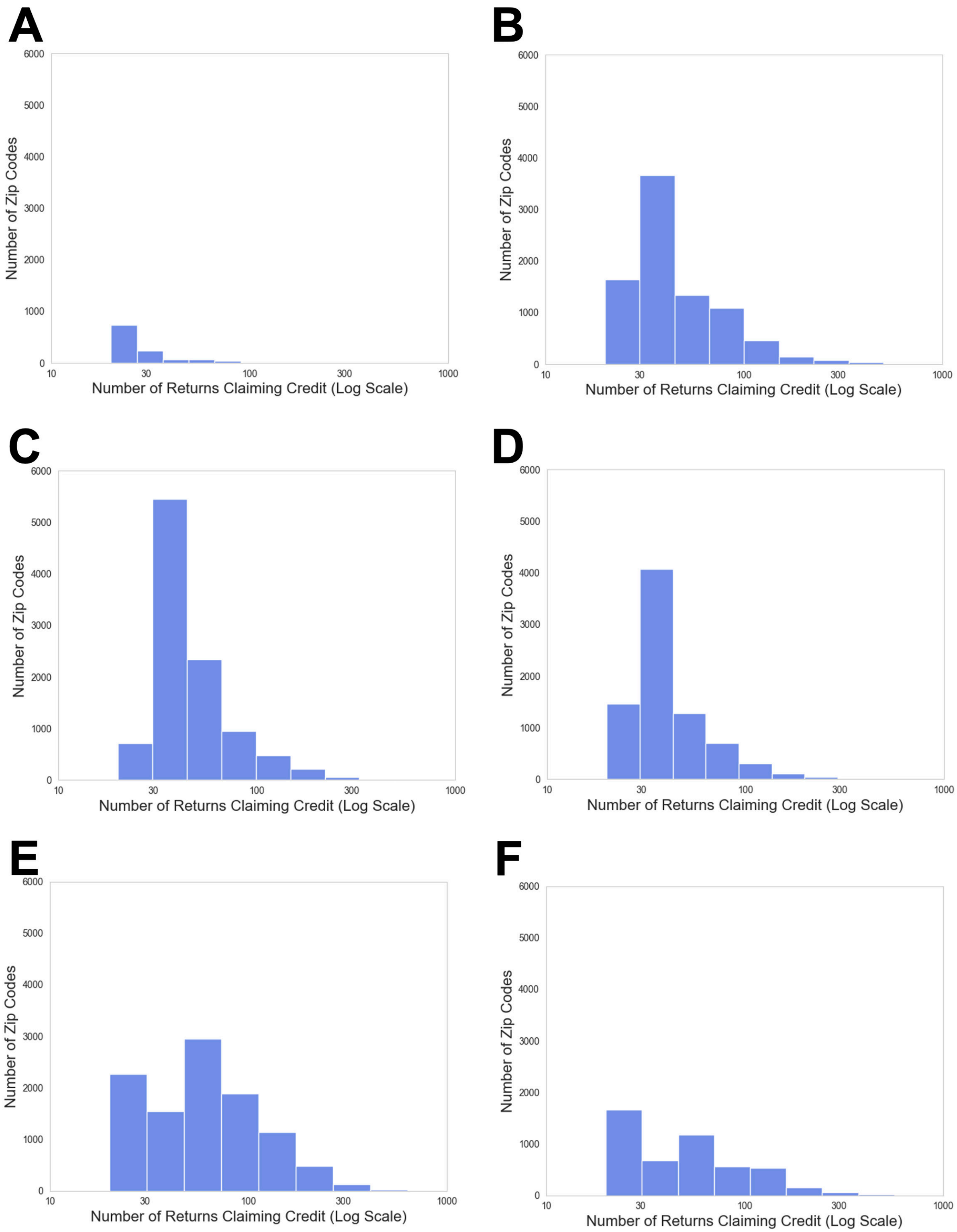

Figure 2: Distribution of number of resident energy tax credits claimed per zip code for each income bracket. A) income bracket 1, B) income bracket 2, C) income bracket 3, D) income bracket 4, E) income bracket 5, and F) income bracket 6. Because of the logarithmic transformation, the linear width of the 20 bins increases relative to the number of claimants, ranging from a width of 4.4 at the low end (Bin 1: 20.0 – 24.4) to 188.6 at the high end (Bin 20: 861.4 – 1050.0).

https://doi.org/10.59720/25-070

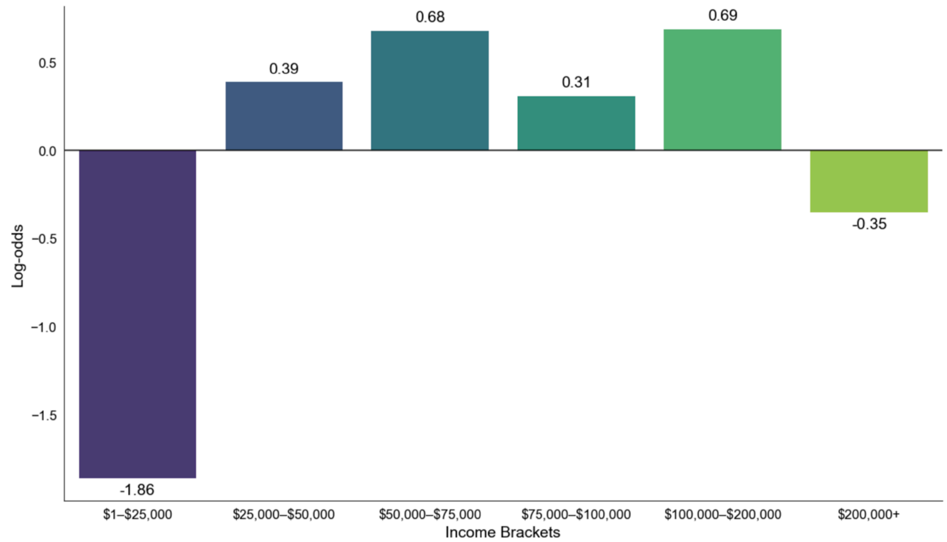

Figure 3: Log-odds for predicting the likelihood of using residential energy tax credits across income brackets. The log-odds from a logistic regression model where the dependent variable was whether a tax filer claimed the credit (1) or not (0), and the independent variable was the income brackets.

6 had a coefficient of 29.727, lower than the 75.621 coefficient for income bracket 5, suggesting that income bracket 6 was likely to claim lower amounts of residential energy tax credits, even if these credits were available for them (**Figure 4**).

Bracket 4 deviated from the increasing trend that was followed by income brackets 1, 2, and 3, where each income bracket claimed more in tax credits as income increased (**Figure 4**). This deviation is reminiscent of the decline in the likelihood of claiming a credit observed for bracket 4 (**Figure 3**). Similarly, the amount of tax credit claimed decreased in bracket 6 after its sharp peak in bracket 5 (**Figure 4**). While higher-income groups generally claimed larger credit amounts, income brackets 4 and 6 deviated from this positive correlation by claiming less than the preceding brackets 3 and 5, respectively (**Figure 4**).

DISCUSSION

We investigated the relationship between various household income levels and whether or not energy tax credits were claimed by these households. We aimed to understand the effectiveness of current incentive programs and identify potential disparities in access. Our initial hypothesis posited that households with higher incomes would utilize residential energy tax credits more frequently and to a higher extent than lower income households. Based on the analysis of IRS SOI data, our hypothesis was partially refuted. We observed a decrease in the utilization of energy tax credits among households in income bracket 4 relative to both income brackets 3 and 5, so the overall trend did not demonstrate a consistent linear relationship between income and utilization of tax credits. Specifically, we saw a coefficient of -4.438 for the magnitude of energy credit claimed for this income group, indicating a lower incentive amount claimed compared to other income brackets, such income bracket 5 with a coefficient of 75.621.

There was also a decrease in the likelihood of claiming the credit in income bracket 6 relative to all other brackets except income bracket 1, even though there is no income limitation for residential energy tax credits. This decrease may be explained by the fact that those making over $200,000 annually are more likely to have the capital needed to make investments into sustainable technologies without necessitating the need for claiming a tax credit (13). The relative financial impact of the credits may be less significant for those in income bracket 6, leading to less motivation to claim a tax credit, which could explain the lower likelihood of income bracket 6 claiming the credit (13).

These findings align with previous research highlighting the challenges faced by middle-income households in accessing and utilizing sustainable technology incentives (9). Filers in income bracket 4 with an annual income of $75,000–$100,000 comprise the missing middle, amounting to 29 million filers in 2021 (12). When a household's income falls within middle-income brackets, they exhibit these lower rates of incentive usage despite the availability of residential energy tax credits because they may not qualify for targeted low-income programs, while simultaneously facing financial constraints that prevent them from taking advantage of broader, less targeted incentive programs (14). The upfront costs of these technologies can be substantial even with incentives, potentially deterring the adoption of these technologies among households facing other financial pressures (5). Additionally, the lack of targeted outreach to the middle-income demographic could further contribute to lower credit usage rates, given the general complexity of navigating incentive programs (15). Moreover, middle-income groups often cannot afford tailored financial services that would guide them through the process of claiming tax incentives (16). While the IRA aims to broaden access to clean energy technologies, our findings underscore the need for further evaluation and potential refinement of these programs to ensure they effectively reach and benefit middle-income filers. This missing middle may be overlooked by existing policies, creating an incentive gap that can potentially hinder the adoption of associated sustainable technologies.

A possible explanation for why this missing middle exists is the prevalence of incentive cliff effects, where benefits abruptly disappear above a certain income threshold, often

https://doi.org/10.59720/25-070

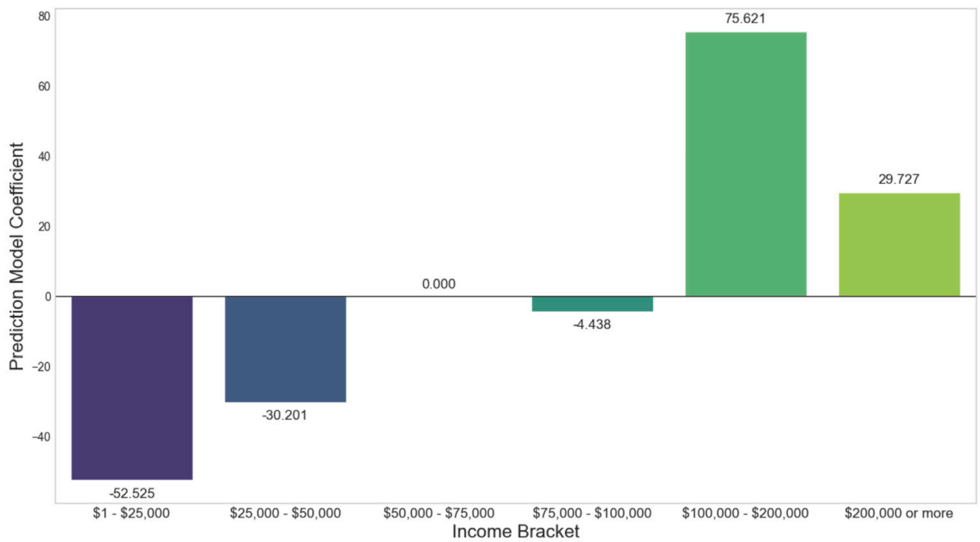

Figure 4: Income brackets and corresponding model coefficients for model predicting magnitude of tax credit taken (A07260) from income bracket. These data were taken from the Statistics of Income Program. The prediction model coefficient, derived from the linear regression model, quantifies how much the predicted tax credit amount changes when a taxpayer belongs to a specific income bracket compared to the baseline of income bracket 3.

set around $75,000 (17). For example, measures in the IRA for pre-owned clean vehicles show a sharp decrease in support right above this income level (17). Similarly, programs like the Low-Income Communities Bonus credit and the Greenhouse Gas Reduction Fund specifically target lower-income communities, increasing access for those groups while leaving middle-income households in a difficult position (17, 18). This structure systematically leaves out the middle-income group: unable to qualify for targeted aid but often lacking the disposable income of higher earners to independently fund projects (18). On the other hand, a filer making less than 80% of the median income of a community will have more costs covered by claiming credits through targeted low-income IRA programs, while a filer in income bracket 5 or 6 would be more likely to have enough disposable income to independently fund such a project (17). In this way, the only group left out of claiming this tax credit is the middle-income group, income bracket 4, and thus explains why households in income bracket 4 were less likely to claim this credit. The nationwide implications of such a cutoff for aid are thus demonstrated by the results found in this study—middle-income groups are systematically less likely to claim the residential energy tax credit and participate in the green transition (19).

The inability to enable the green transition for a significant portion of the US middle class, approximately 29 million filers in 2021, is detrimental as it hinders the pace of the overall green transition and exacerbates existing inequalities in renewable technology adoption (20). To mitigate these issues, it is critical we create policies to counteract the cliff effect. We suggest using a more gradual descent in the level of incentives, instead of a ceiling of $75,000. Future research could model this alternative incentive structure and measure its ability to raise the middle class's likeliness to use incentives available to them and better adopt sustainable technologies.

A key limitation of this study lies in the aggregated nature of the IRS SOI data, which does not allow for analysis of individual household characteristics like homeownership and state-level policy impacts. Given that credit usage varies widely across states, a detailed analysis of the incentive gap at the state level is essential to determine if programs that specifically target middle-income groups can successfully alleviate the decrease we identified (12). Levels of homeownership across income brackets may also have influenced results, as ability to claim the credit is in part dependent on home ownership (18). Future research should use more granular data to explore the localized factors influencing incentive usage and associated sustainability technology adoption within specific communities. Our study only offers a snapshot in time, so a more longitudinal study tracking the impact of specific IRA provisions over several years is necessary to gauge their long-term effectiveness.

Our analysis of residential energy tax credits reveals a non-linear relationship between income and usage of the energy tax credit, with a notable gap in credit usage within income bracket 4. The research is especially important as policymakers move to create incentives with accelerated rigor, as new policies should address gaps in existing programs to better bridge access with the renewable technology associated with these incentives. This research can be used to justify moves to specially target middle-income households when it comes to designing incentives for sustainability.

MATERIALS AND METHODS

The core dataset for this analysis came from the IRS SOI program, specifically the 2021 Individual Income Tax Returns Public Use File (PUF). This file contains anonymized, aggregated data from a statistically representative sample of individual tax returns filed in the United States. At the time of this study, the 2021 dataset was the latest dataset available that provided complete tax returns for public use. The data is organized by zip codes and income bracket. For each zip code, data is split up into six income brackets. A total of 166,221 records were used for this analysis, representing data from all 50 states across the US.

We extracted specific variables relevant to understanding the relationship between income and the ability to use residential energy tax credits. The first is A07260, which represents the dollar amount claimed for the Residential Energy Efficient Property Credit, providing a quantifiable measure of investment in sustainable technologies. N07260 is the number of returns that claim the Residential Energy Efficient Property Credit and measures the number of claims with a non-zero amount of the tax credit claimed. The other variable investigated was AGI_Stub, which categorizes taxpayers into six income brackets based on adjusted gross income (AGI), allowing us to analyze incentive usage patterns across different income brackets: income bracket 1 of $1–25,000, income bracket 2 of $25,000–$50,000; income bracket 3 of $50,000–$75,000, income bracket 4 of $75,000–$100,000, income bracket 5 of $100,000–$200,000, and income bracket 6 of $200,000 and up.

Data preparation was conducted using the Python pandas library (Version 2.2.2). First, the raw dataset was filtered to create a dataframe containing only the income bracket and N07260 columns. To prepare the data for modeling, a binary dependent variable, credtaken, was created from N07260. Records with N07260 = 0 were assigned a value of 0 (credit not taken), and records with N07260 > 0 were assigned a value of 1 (credit taken). The categorical income bracket variable (1 through 6) was converted into six binary dummy variables using one-hot encoding, where each new column represents one of the income brackets.

To analyze the relationship between income bracket and the likelihood of claiming residential energy tax credits, a logistic regression model was employed using the scikit-learn library in Python. The model predicts the log-odds of the credtaken variable being 1 based on the income bracket dummy variables. Logistic regression is specifically designed for modeling binary outcomes, making it the most appropriate model for the data collected in this study. The log-odds from the model can be interpreted based on direction and magnitude. A positive log-odds indicated a higher tax credit amount claimed, while a negative log-odds suggested a reduced amount claimed.

The formula for the logistic regression model is:

$$log\frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \qquad \text{(Equation 1)}$$

Where p is the probability of a household claiming the tax credit (credtaken = 1), X_1 through X_6 were the dummy variables representing each of the six income brackets, and β_1 through β_6 were the coefficients for each income bracket. These coefficients quantified the effect of being in a particular income bracket on the likelihood of claiming the credit (**Figure 3**).

The dataset was split into a 70% training set and a 30% testing set. To address the class imbalance (far more non-claimants than claimants), the class_weight='balanced' parameter was used during model training.

Moreover, to distinguish between likelihood of claiming the credit and the amount of credit claimed, a Least Absolute Shrinkage and Selection Operator (LASSO) linear regression model was also employed. The coefficients presented in this study were generated by fitting the LASSO regression model to the training data. This linear regression model was used in order to reduce overfitting and reduce model complexity arising from the multi-dimensional nature of the IRS SOI data (21). A positive coefficient indicates that being in that income bracket increased the predicted amount of credit claimed relative to the baseline, while a negative coefficient indicates a lower amount of credit claimed (22). The magnitude of the coefficient shows the strength of this relationship.

Received: February 12, 2025
Accepted: June 6, 2025
Published: July 7, 2026

REFERENCES

1. Gielen, Dolf, et al. "The role of renewable energy in the global energy transformation." *Energy Strategy Reviews*, vol. 24, 31 Jan. 2019, pp. 38–50. https://doi.org/10.1016/j.esr.2019.01.006.
2. Gautam, Indu and P.C. Kavidayal. "Green economy: A challenge to inclusive and equitable growth." *Environment Conservation Journal*, vol. 18, no. 1&2, 15 June 2017, pp. 137–142. https://doi.org/10.36953/ECJ.2017.181219.
3. Sovacool, Benjamin K., et al. "Temporality, vulnerability, and energy justice in household low carbon innovations." *Energy Policy*, vol. 128, 23 Jan. 2019, pp. 495–504. https://doi.org/10.1016/j.enpol.2019.01.010.
4. "Transmittal of Fiscal Year (FY) 1998 Public Housing/Section 8 Income Limits." *U.S. Department of Housing and Urban Development.* https://www.huduser.gov/portal/datasets/il/fmr98/sect8.html. Accessed 14 June 2025.
5. Berkouwer, Susanna and Joshua Dean. "Barriers to Energy Efficiency Adoption in Low-Income Communities." *Kleinman Center for Energy Policy*. https://kleinmanenergy.upenn.edu/research/publications/barriers-to-energy-efficiency-adoption-in-low-income-communities/. Accessed 10 October 2025.
6. Curtin, Joseph, et al. "Financial incentives to mobilise local citizens as investors in low-carbon technologies: A systematic literature review." *Renewable and Sustainable Energy Reviews*, vol. 75, 14 April 2017, pp. 534–547. https://doi.org/10.1016/j.rser.2016.11.020.
7. "Residential Energy Tax Credits: Overview and Analysis." *Congress.gov*, Library of Congress, 4 February 2026, https://www.congress.gov/crs-product/R42089.
8. "Residential Clean Energy Credit". *Internal Revenue Service*. www.irs.gov/credits-deductions/residential-clean-energy-credit. Accessed 29 Jan. 2025.
9. Borenstein, Severin and Lucas W. Davis. "The Distributional Effects of U.S. Tax Credits for Heat Pumps, Solar Panels, and Electric Vehicles." *National Bureau of Economic Research*, July 2024. https://doi.org/10.3386/w32688.
10. Bistline, John, et al. "Economic Implications of the Climate Provisions of the Inflation Reduction Act." *Brookings Papers on Economic Activity*, May 2023, pp. 77–182. https://doi.org/10.3386/w31267.
11. "Climate and electrification provisions for disadvantaged communities." *Rewiring America*. https://www.rewiringamerica.org/research/IRA-disadvantaged. Accessed 17 September 2025.
12. "SOI Tax Stats - Individual income tax statistics - ZIP Code data." *Internal Revenue Service*. www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-

data-soi. Accessed 19 July 2024.
13. Ergun, Selim J. and M. Feranda Rivas. "Does higher income lead to more renewable energy consumption? Evidence from emerging-Asian countries." *Heliyon*, vol. 16, no. 9, 15 Jan. 2023, https://doi.org/10.1016/j.heliyon.2023.e13049.
14. McCluskey, Martha. "Framing Middle-Class Insecurity: Tax and the Ideology of Unequal Economic Growth." *Fordham Law Review*, vol. 84, no. 6, 2016. https://ir.lawnet.fordham.edu/flr/vol84/iss6/13.
15. Duflo, Esther, et al. "Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block." *The Quarterly Journal of Economics*, vol. 121, 01 November 2006, pp. 1311–1346. https://doi.org/10.1093/qje/121.4.1311.
16. Winchester, Danielle D. and Sandra J. Huston. "All Financial Advice for the Middle Class is Not Equal." *Journal of Consumer Policy*, vol. 38, 14 May 2015, pp. 247–264. https://doi.org/10.1007/s10603-015-9290-8.
17. "Incentives and assistance programs." *New Jersey Board of Public Utilities*. www.nj.gov/bpu/assistance/incentives/. Accessed 23 Aug. 2024.
18. "Low-income communities bonus credit". *Internal Revenue Service*. www.irs.gov/credits-deductions/low-income-communities-bonus-credit. Accessed 1 Sep. 2024.
19. Maglicic, Martina and Vítor V. Vasconcelos. "Income inequality in the uptake of environmentally friendly products." *iScience*, vol. 28, no. 4, 26 March 2025, https://doi.org/10.1016/j.isci.2025.112277.
20. Kumelys, Ricardo. "Energy Inequality and the Just Transition: How can we make the benefits of clean energy accessible to everyone?." *Berkeley Energy & Resources Collaborative*. https://berc.berkeley.edu/news/energy-inequality-and-just-transition-how-can-we-make-benefits-clean-energy-accessible-everyone. Accessed 20 September 2025.
21. Gauraha, Niharika. "Introduction to the LASSO." *Resonance*, vol. 23, 4 June 2018, pp. 439–464. https://doi.org/10.1007/s12045-018-0635-x.
22. Prasad, Sahana. "Regression." *Advanced Statistical Methods*, 1st ed., Springer, 2024, pp. 1–45. https://doi.org/10.1007/978-981-99-7257-9.

https://doi.org/10.59720/25-070

APPRENDIX

Python code for finding correlation between N07260 (binary variable of having taken Residential energy tax credit or not) and AGI_Stub (income brackets):

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
cred = pd.read_csv("21zpallagi-allparam.csv") # the data set saved locally
cred = cred[['agi_stub', 'N07260']]
x = pd.get_dummies(cred['agi_stub'])
cred = pd.concat([cred, x], axis=1)
cred.columns = cred.columns.astype(str)
def update_credtaken(row):
    if row['N07260'] == 0:
        return 0
    else:
        return 1

# Apply the function to create the 'credtaken' column
cred['credtaken'] = cred.apply(update_credtaken, axis=1)

# Display the updated DataFrame
print(cred.head())
from sklearn.model_selection import train_test_split

X = cred.drop(['credtaken'], axis=1)
y = cred['credtaken']
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, train_size=0.7)

from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.utils.class_weight import compute_class_weight

# Compute class weights
classes = np.array([0, 1])
class_weights = compute_class_weight('balanced', classes=classes, y=y_train)
class_weight_dict = dict(enumerate(class_weights))

# Initialize and train the Logistic Regression model with class weights
model = LogisticRegression(class_weight=class_weight_dict, max_iter=1000)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

model.coef_ # gives the coefficients found in Figure 3
```