JOURNAL OF
EMERGING INVESTIGATORS

Article

Analyzing carbon dividends' impact on financial security via ML & metaheuristic search

Amay Babel[1], **Anuj Babel**[1]
[1] Coppell High School, Coppell, Texas

SUMMARY

The question of a carbon tax and dividend has long intrigued policymakers and researchers. If enacted, this policy would levy a tax on carbon emissions, with the resulting revenue being distributed to American citizens through an annual dividend. Although current literature extensively covers the climate impacts of a carbon dividend, its economic effects in the short term remain unclear. We hypothesized that the modeled implementation of a carbon dividend would increase predicted financial security and stability. To analyze the tax's effect on financial security across various income levels, we used the US Federal Reserve Board's Survey of Household Economics and Decisionmaking (SHED) as a representative sample for the population of the United States. We then developed the Financial Security and Resilience Index (FSRI) using the response data and applied machine learning algorithms to predict changes in financial security based on each individual's net income from a carbon dividend. Under an equal dividend scenario, we observed a 0.19% mean increase in financial security. After introducing a genetic algorithm that allocated varying amounts of the carbon dividend's total revenue to different income ranges, we observed an increase of ~0.5% in mean financial security. The varied allocation of funds achieved through metaheuristic search algorithms as a concept can certainly become an effective instrument, helping policymakers better understand possible improvements in policy design. These findings can aid in generating a strong movement in favor of the dividend, as our results indicate that a substantial majority of the population would benefit from the policy.

INTRODUCTION

A carbon tax and dividend is a proposed tax on carbon emissions on a per-metric-ton basis from which the revenue will be returned to the people of the U.S. through an annual dividend (1). The main goal of the carbon tax and dividend is to limit carbon emissions to help mitigate the effects of climate change, such as food supply chain disruptions, disease, air pollution, and increases in natural disasters (2). This policy is expected to incentivize the switch of firms as well as citizens towards green energy in hopes that these forms of energy will become more cost-efficient than their carbon alternatives (3, 4). A gradual switch in energy sources can decrease emissions and facilitate the transformation of the U.S. towards cleaner energy, which can help mitigate the effects of climate change as a whole (3, 4). The carbon dividend is also considered to be revenue-neutral because the government does not retain any funds collected from this policy, unlike the majority of other taxes. The proposition appeals to policymakers and households alike due to its remarkable ability to offset the impact of price increases on households while simultaneously decreasing net carbon emissions. Other types of climate policy usually involve a tradeoff between effectiveness and financial impact that the carbon tax and dividend does not suffer from.

Researchers of this topic have attempted to analyze the economic impact of such a policy (5, 6). However, these papers have focused on quantifying basic effects of the tax and dividend, such as calculating household net incomes from a modeled policy implementation. High-level analyses can sometimes obscure the finer details of nuanced impacts that are essential for truly understanding the effects on households across the nation. One example of this can be seen in analyzing the policy's impacts on financial security. The effects on financial security are of the utmost importance to understand because individuals with different incomes are impacted variously by the same change in household income. For example, a $1,000 loss on an annual income of $20,000 will have a greater negative effect when compared with the same loss on an annual income of $100,000. We see evidence of this pattern already in looking at inflation, as low-income households are more heavily impacted by inflation than high-income households (7). Therefore, comparing metrics that go beyond the surface level uncovers insights not available in broader analyses that are prevalent in today's research base.

The goal of our study was to fortify our understanding of the true impact of a carbon tax and dividend by contextualizing net income calculations in terms of change in financial security. This allows for more developed decision-making as we can better show policymakers the actual impact on individuals across all income ranges. In our scenario, the total annual revenue for the tax side of the policy amounted to $257.4 billion if implemented across the United States. To put this into perspective, under an equal per capita dividend, each person would receive around $753 to spend at their leisure. However, this amount would not represent each individual's exact net income from the policy, since prices for carbon-based products are expected to rise significantly (4). The dividend is designed to more than compensate for these price increases while also reducing carbon emissions. We hypothesized that a carbon dividend would indeed increase overall financial security in a modeled scenario. To test this, we employed random forest and ridge regression models to predict how implementing a carbon dividend might enhance household financial stability (8, 9). We used

https://doi.org/10.59720/24-234

the U.S. Federal Reserve Board's Survey of Household Economics and Decisionmaking (SHED) 2022 as our primary dataset. This dataset was instrumental in our creation of the Financial Security and Resilience Index (FSRI) (9). The "FSRI Score" was a quantitative metric that we utilized to measure an individual's financial security. Ultimately, we analyzed two different dividend distribution scenarios. The first was a simple, equal dividends scenario where every income decile—or tenth of the population—received the same cash dividend. The machine learning models predicted that the mean increase in financial security for this scenario was ~0.19%. The second dividend distribution was more complex, as it involved using a genetic algorithm to allocate different amounts of funds to different income deciles (11). The specific algorithm that we used—a Non-Dominated Sorting Algorithm (NSGA-II)— found a solution that exhibited a mean increase in financial security of ~0.5% (12). This paper has the ability to provide key economic data and analysis that supports a carbon dividend, which can correlate to increases average financial security and radically decrease carbon emissions across the United States.

RESULTS

We sought out to model the implications of a carbon dividend and predict change in mean financial security it would cause. The goal was to uncover potential repercussions of the dividend and evaluate whether the outcome would support its enactment from an economic perspective. To do this, we trained machine learning models on extensive survey data to predict changes in financial security based on changes in net income for individual people.

We used the Survey of Household Economics and Decisionmaking (SHED) 2022 survey dataset as a foundation for our models. This dataset contains extensive financial survey data for 12,000 United States citizens, containing information regarding their finances, employment status, and living standards. First, we examined the per-person income distribution to establish whether the data was representative of the U.S. population based on a total population of 342,000,000 people (**Figure 1**). The data was heavily skewed right, as the vast majority of individuals had an annual income under $100,000. The mean of this dataset ($60,169) as well as the median ($42,500) closely match true U.S. mean and median values (13, 14). Next, the Financial Security and Resilience Index (FSRI) was developed, mainly utilizing individual response data from the SHED survey that were key in determining financial security (**Table 1**). The FSRI drew ideological inspiration from a neighborhood-level index in Orange County, CA, which examined local financial stability on the household level (15). Although this index was built for a different reason and measured financial resilience through alternative means and frameworks, the Orange County index was still a useful tool that aided us in the creation of our index. Income, savings, and investment statistics were also incorporated into the calculation. Each individual's responses to the selected questions were systematically ranked and summed to generate a total FSRI score ranging from 0-1. Higher scores indicate greater financial stability as demonstrated through their response data. For example, those who had saved "Over $1,000,000" for retirement were awarded more index points then someone who has saved "$10,000 to $24,000". Afterward, three indicators of financial

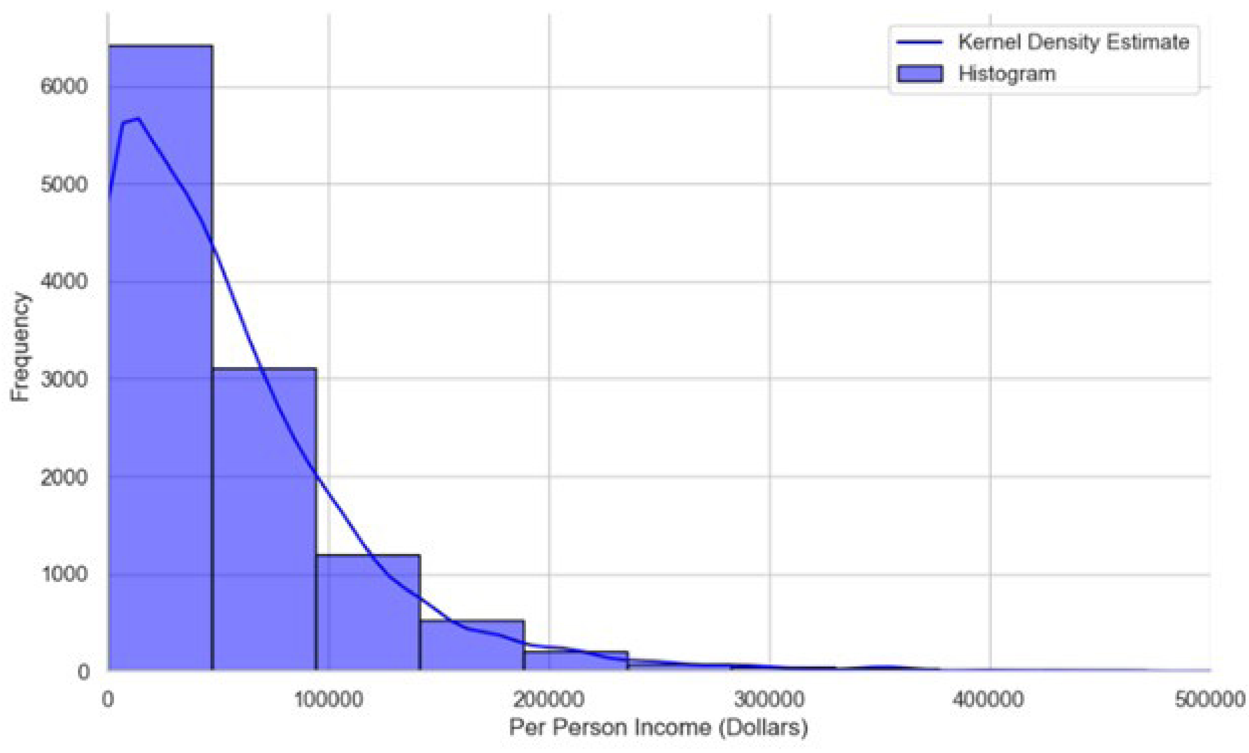

Figure 1: Distribution of per-person income in the SHED 2022 dataset. Graph showing the distribution of per-person income in the dataset (n = 11,215), with a kernel density estimate displaying the frequencies of particular incomes. The graph was limited at an annual income of $500,000; however, there were individuals with incomes greater than that in the dataset. The data is heavily right skewed, with a mean of $60,169 and a median of $42,500. This data came from the SHED 2022 survey (9).

https://doi.org/10.59720/24-234

Variable	Description
B2	Overall, which one of the following best describes how well you are managing financially these days?
D1A	Last month, did you do any work for either pay or profit?
K20	Approximately how much money do you currently have saved for retirement?
I40	Which of the following best describes the income that you received from all sources, before taxes and deductions, in the past 12 months?
I9	In the past 12 months, which one of the following best describes your (and/or your spouse's or partner's) income?
3_month_expense_coverage	Are you able to cover 3 months of expenses in case of emergency?
EF3_h	Suppose that you have an emergency expense that costs ($400/$500). Based on your current financial situation, how would you pay for this expense?
EF5A	Which best describes your ability to pay all of your bills in full this month?
EF7	Based on your current financial situation, what is the largest emergency expense that you could handle right now using only your savings?
ppeduc5	Education (5 Categories).
ppemploy	Employment status.
ppfs0596	What is the approximate total amount of your household's savings and investments?
affect_of_price_inc	Ability to withstand price increases.

Table 1: Survey data questions that were used in FSRI score calculations. The table displays all of the primary survey data categories that were used in the calculation of FSRI score. They were chosen based on their ability to represent financial stability. Subjective questions as well as income, savings, and investment statistics were included to account for financial stability as well as statistical security when it came to an individual's finances. These questions were selected from the SHED 2022 survey (9).

security were created as input data for model training to increase our ability to predict FSRI score. These three indicators were distance from the poverty line, distance from the mean income, and a weighted sum. The weighted sum measured certain aspects of financial security, calculated using a financial security gradient. The other two indicators were based on self-reported income statistics that participants provided in response to the SHED survey. Though income in and of itself was a strong predictor of FSRI score, its accuracy was not perfect (R^2 of ~0.6 after we trained it on a random forest model). For this reason, the indicators were created so that FSRI could be more accurately predicted.

Our machine learning pipeline to predict FSRI score based off of income begins with the training a random forest model (we will call this model the "indicator model") to predict the three financial indicators using per-person income. Then, a second pair of models (random forest and ridge regression models, which we will refer to as the "final models") use these predicted indicators to make final predictions of the FSRI score, a measure of financial stability. Essentially, the final models use the predictions of the indicator model as inputs to estimate an individual's financial security score. This allows us to effectively take an individual's income, run it through the pipeline, and end up with a predicted FSRI score value (**Figure 2**).

When comparing the pair of final models, we found that the random forest model exhibited better accuracy with a higher R^2 (0.95 for the random forest model and 0.79 for the ridge model) and a lower mean absolute error (MAE) (0.035 for the random forest model and 0.073 for the ridge model). However, we decided to combine the models' predictions together because the ridge regression was better able to detect granularity in the data, which allowed for more nuanced analysis (**Table 2**). The random forest model contributed 75% of the total weight, while the ridge model accounted for the remaining 25%. The combined model had an R^2 of 0.94, an MAE of 0.038, and a mean squared error (MSE) of 0.0023. In context, this means that the model was able to explain 94% of the variability in the data and on average was only off by about 0.038 FSRI score points per prediction. Given that the FSRI is on a scale of 0–1, these results provided a high degree of confidence.

After training the models, we added inputs of the specific details pertaining to the carbon dividend to obtain the results for both scenarios. This included the costs to consumers, which we adapted from Fremstad and Paul, as well as the

https://doi.org/10.59720/24-234

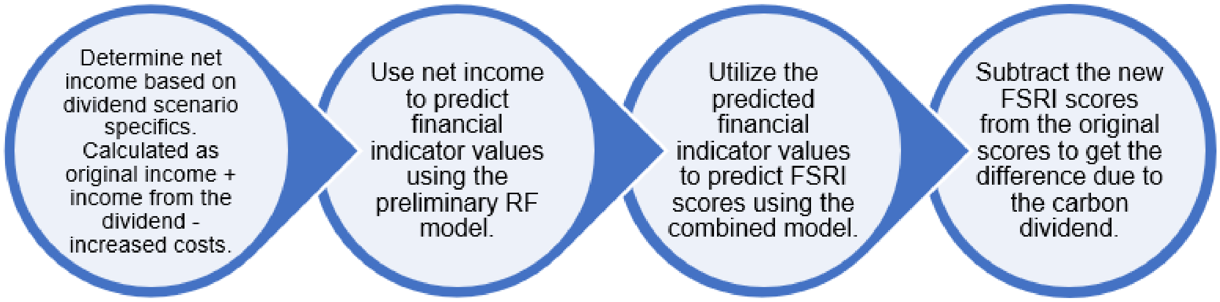

Figure 2: Machine learning pipeline for converting income into financial security. Flowchart showing all of the steps that we took to analyze each scenario (both equal dividends and the genetic algorithm). The pipeline was constructed by maximizing model R2 and minimizing mean absolute error aimed at increasing our ability to correctly predict changes in FSRI score. The extra step of converting income into indicators allowed us to predict financial security.

tax revenue figures (5). For each individual, net income from the tax was calculated, which allowed us to run the income changes through the pipeline to output the net change in FSRI values. There were two scenarios that we considered, the equal dividends scenario and the genetic algorithm scenario. The purpose of the models in the equal-dividends scenario was quite straightforward, as they allowed us to simply calculate the difference in financial stability before and after the tax, which allowed us to measure the benefit or harm that each person experienced due to the policy in terms of their FSRI score. The use of the models with the genetic algorithm was slightly more complicated. The models acted as evaluators of solution "fitness," measuring the benefit of each fund allocation (referred to as a solution) by assessing its impact on individual financial stability. They provided feedback to the algorithm on whether a particular solution was "strong" or "weak," enabling the algorithm to iteratively discard weaker solutions and retain stronger ones for an improved result. In the equal dividends scenario, we observed a 0.19% average increase in FSRI score, a 0.41% increase for the bottom half of the income distribution, and a 0.04% decrease for the top half (**Table 3**). In the genetic algorithm scenario, we saw a 0.50% mean increase in FSRI score, a 0.98% increase for the bottom half of the income distribution, and a 0.01% decrease for the top half of the distribution (**Table 3**).

Both scenarios demonstrated that, on average, individuals experienced increased financial security (based off of FSRI score) under the carbon tax and dividend than without the tax; however, the genetic algorithm's allocation of funds made individuals 0.31% more financially secure on average than the equal dividends scenario (0.19% mean increase for the equal dividends and 0.50% mean increase for the genetic algorithm). We also saw increased benefit for the bottom 50th percentile of the income range, as the equal dividends scenario displayed a mean positive increase of 0.41% in financial security (in this case, mean is the average net FSRI score resulting from the tax across all individuals), and the genetic algorithm solution showed a mean increase of almost 1%. The top 50th percentile of the income range was not majorly affected, although they still experienced fractional decreases in financial security in both scenarios (-0.04% and -0.01% in equal dividends and the genetic algorithm, respectively).

DISCUSSION

We hypothesized that an economically modeled carbon dividend would lead to increases in predicted financial security, providing policymakers with key data to better understand its impacts from a financial point of view. A combined model that consisted of a random forest regression (75% of the weightage) and a ridge regression (25% of the weightage) was used to predict the impact of a given dividend allocation on FSRI score. We chose this weighting because the random forest model was more accurate, as it generally is a more advanced method of modeling non-linear data, while the ridge model better captured how more subtle changes in income could impact FSRI, justifying the inclusion of both (16). This is not surprising because ridge regressions are parametric methods (they use an equation) while random forests are not (17). The combined model had an R^2 of 0.94, an MAE of 0.038, and a mean squared error (MSE) of 0.0023, suggesting the model's reliability to accurately predict the impacts of a carbon tax and dividend on financial security of individuals in the U.S.

Our findings strongly supported the hypothesis that a carbon tax and dividend with equal dividends does indeed result in increases in predicted financial security of individuals as measured through FSRI. We observed a mean 0.19% increase in financial security across all income ranges, with an increase of 0.41% for the bottom half of the income range, and just a minor 0.04% decrease for the top half. These results identified the possible financial benefits of the carbon dividend, supporting its movement forward as a strong option to combat climate change and create economic benefit. Although these percentages may seem small, the greatest benefit is that this policy is completely revenue neutral, which makes it an efficient method to increase financial security for individuals. A 0.41% increase in financial security for the bottom half of the population would certainly make a substantial difference, helping stabilize those households. These figures are also encouraging in the fact that they can generate a wide base of support, as a majority of the public will not be financially worse off. It could further bolster the notion of a double dividend—creating benefits for both the environment and our country's citizens (18).

The genetic algorithm demonstrated higher gains in

Model	R^2	Mean Absolute Error	Mean Squared Error
Random Forest	0.95	0.035	0.002
Ridge	0.79	0.073	0.009
Combined (75% RF, 25% Ridge)	0.94	0.038	0.0023

Table 2: Machine learning model statistics. RF = random forest regression model. The table displays the R^2, mean absolute error, and the mean squared error of the primary machine learning models that were trained to predict changes in FSRI score. Though the statistics of the combined model were slightly lower (R^2 dropped by 0.01) as compared to the RF model, the inclusion of the ridge regression model allowed us to capture finer changes in FSRI score. This allowed for more nuanced analysis, warranting the fractional sacrifice in model accuracy.

financial security as compared to the equal dividends scenario, as the algorithm displayed a 0.5% mean increase in financial security—0.31% higher than the equal dividends scenario. However, the algorithm's dividend allocation is likely not politically feasible due to the varying fund distribution across income levels. Rather, this algorithm's utility lies in its ability to push the boundaries of policy design. For another example, we can look at renewable energy investment and development. Genetic algorithms can explore large search spaces to determine the optimal combination of renewable energy sources (such as solar, wind, and hydropower) to maximize multiple objectives, such as cost efficiency and energy output. Other researchers have also used genetic algorithms to design renewable energy systems and optimize renewable energy communities (19, 20). A genetic algorithm was used in our research to show its capacity to solve complex problems, as well as to propagate its use in other policy domains. Therefore, it should be stated that an equal dividend scenario would be more realistically feasible if policymakers attempt to enact a carbon tax and dividend.

The primary constraint in this analysis was the fact that the carbon dividend as a policy has not been implemented yet. Therefore, we had to structure our hypothesis around predicted changes in financial security, not actual ones. Also, some components, such as cost increases for the population, had to be modeled. Though they were modeled empirically, they were estimates nonetheless. Extending the genetic algorithm's runtime was another action that could have been implemented to improve overall results. We ran ours for 50 hours on an NVDA V100 GPU, which completed 500 generations. Instead of 500, we could run up to 1000 or even 2000 generations, which would likely reveal a solution with a higher fitness value. However, that much computation can quickly get expensive, making it difficult for this to be feasible.

Our results can impact potential policy decisions by providing support for a carbon tax and dividend from a financial security perspective. These findings can aid in building public and political support for the carbon tax and dividend method by challenging pre-existing notions that a carbon pricing scheme will negatively impact individuals financially.

MATERIALS AND METHODS

Carbon Dividend Revenue and Cost Increase Calculations

In our analysis, like many others', it was assumed that firms will pass 100% of the burden of the tax on to the consumer through price increases, which was a key assumption that went into our revenue calculations. It's also important to note that our analysis focused on the short-term effects of the tax and dividend, as the long-term effects are more difficult to gauge in terms of judging the effect of the policy on total taxable emissions, making revenue calculations burdensome. Our revenue calculations required two components: the total emissions to be taxed as well as a tax rate per metric ton of carbon. Starting with total taxable emissions, we first used the Congressional Budget Office's estimates to quantify the amount of carbon that would be taxed under the jurisdiction of a carbon dividend, which amounted to 3.9 billion metric tons (21). This includes taxation on commodities such as energy and gasoline, thereby raising consumer prices for these goods (21). The tax rate was set at $66 per metric ton of carbon emissions equivalent (tCO2e). To arrive at this number, we used the Interagency Working Group's (IWC) social cost of carbon at a 3% discount rate for 2025 (set at $56), which quantifies the average cost to society produced by each metric ton of CO_2 (22). This initial $56 dollar rate set by the IWC was in 2020 dollars, so after adjusting for inflation using the Bureau of Labor Statistics' inflation calculator, we ended up with a final rate of $66/tCO2e to be used in our analysis (23). To calculate the total revenue, we multiplied the tax rate ($66/tCO2e) by the amount of emissions being taxed (3.9 billion tCO2e). This yielded a total tax revenue of 257.4 billion dollars to be distributed to citizens through a dividend. For cost increase calculations, we adapted code from Fremstad and Paul, which calculates the average cost increases per income decile using input-output tables and other economic modelling techniques (5). In their code, we changed the tax rates (from their rate of $50/tCO2e to our rate of $66/tCO2e), income terms (mean and median incomes), and population weights (342,000,000 in our scenario) to match recent developments in economic data so that we would have access to the most accurate results.

The Creation of the FSRI

The SHED 2022 dataset provided a representative sample of the U.S. population necessary for the creation of the FSRI (10). Data cleaning of this survey was conducted in Python version 3.12.4, primarily using the Pandas and NumPy packages (24, 25). The Seaborn and Matplotlib packages were used for graphing and visualization (26, 27). We reduced the dataset to the survey questions and categories that were required to build the FSRI. These categories were selected based on their ability to represent an individual's financial security (**Table 1**). The exact categories we chose (from

https://doi.org/10.59720/24-234

Difference in FSRI (Income Ranges)	Genetic Algorithm	Equal Dividends	Difference (GA - Equal Dividends)
Mean Overall	0.50%	0.19%	0.31%
Mean Bottom Half	0.98%	0.41%	0.57%
Mean Top Half	-0.01%	-0.04%	0.03%

Table 3: Changes FSRI score by allocation scenario. GA = Genetic Algorithm, FSRI = Financial Security and Resilience Index. The table displays the final changes in financial stability under each scenario, as well as showing the difference between the two. The genetic algorithm exceeds the equal dividends in every category, but it is important to note that the top half of the income decile experience fractional decreases in stability under both scenarios.

the survey data) can be condensed down to the following: personal financial sentiment, employment, retirement savings, income in the past 12 months, ability to cover certain types of emergency expenses, ability to pay bills, education, household savings and investments, and ability to withstand price increases for normal goods. We categorized missing values (where a survey respondent chose not to answer any particular question) in any categories as "not sure" (meaning that the respondent would not receive any FSRI points for that specific response) instead of inputting them to preserve data purity. Two categories, "3_month_expense coverage" and "affect_of_price_inc", were derived from combining multiple original survey questions, which we hoped would better capture their impact on financial security. The variable "3_ month_expense_coverage" merges two survey items: one asking whether the respondent has an emergency fund and another asking whether they could cover living expenses for three months if they lost their primary income. If a respondent answered "yes" to either question, they received credit in this category. The "affect_of_price_inc" category gauges how respondents cope with inflation and rising prices. Based on answers to several related questions about their ability to absorb cost increases, respondents were classified as "not affected", "somewhat affected", "very affected", or "extremely affected" and were awarded FSRI points accordingly.

Each individual, based on their responses to these questions, was assigned an FSRI rating on a scale of 0–1. Each question contributed to an individual's score, with 11 out of the 13 questions having equal importance. However, annual household income was double weighted to emphasize its importance in financial stability. To allow for this weightage, we reduced the importance of income variability to cap the maximum score any individual could achieve at 1 full point. The points were dependent on how strong an individual's response was to each particular question. For any given question, the strongest response received 0.075 FSRI point, the weakest received 0 points, and the rest of the responses were evenly spaced between. For example, consider the following survey question: Based on your current financial situation, what is the largest emergency expense that you could handle right now using only your savings? A response of "$2,000 or more", the strongest response, would have awarded that individual 0.075 FSRI points for that question alone; On the other end of the spectrum, a response of "Under $100", the weakest response, would not have added anything to that individual's overall score. These calculations were done for the responses to all of the selected questions to create an FSRI index score for every single individual in the dataset. This rating was the primary metric that we used to measure changes in financial security. The survey originally had 11,775 participants, but after excluding people who chose not to disclose their income, we narrowed the field of individuals down to 11,215. Because the survey reported income on a household basis, we needed to convert the data to reflect individual incomes. This was done using the square root scale. The square root scale was employed to acknowledge economies of scale in consumer spending and expenditures, which aimed to create more representative income statistics (28). The equation used to scale household income can be seen below.

$$\text{Scaled Household Income} = \frac{\text{Household Income}}{\sqrt{\text{Household Size}}}$$

To confirm that the data was indeed representative of the population, we plotted the distribution of our income variable. Since the median and mean income of the dataset matched closely with real U.S. economic data, we were satisfied to continue moving forward.

Machine Learning Model Training

Our machine learning model training utilized NVIDIA RAPIDS—a GPU-accelerated machine learning library—to provide us with the random forest and ridge regression models that we fitted. The entirety of the training of models and modeling of the tax was done in Python version 3.12 (29). To assess the impact of a carbon dividend on financial security, we first developed indicators that provided input data for training our machine learning models. These indicators consisted of the following: distance from the poverty line, distance from the mean income, and weighted sum. Weighted sum consisted of a multi-class classification model trained on similar input data as what makes up the FSRI score calculations. This model classified individuals via a financial security gradient, deciding how financially secure an individual was. We used XGBoost, a gradient boosting algorithm, due to its accuracy as well as fast training times (30). This model allowed us to compute the predicted probabilities—where each entry corresponds to the probability of an individual belonging to a certain class—for each individual. These probabilities were then weighted and summed together to create the weighted sum indicator. The other two indicators, distance from the poverty line and distance from the mean income were calculated based off of reported income from the survey data. These indicators were somewhat arbitrary but were required to better predict changes in FSRI score.

https://doi.org/10.59720/24-234

Next, we trained the preliminary random forest model (the indicator model), using per-person income to predict weighted sum. A python function was created, taking income as its input parameter, calculating the first two indicators and predicting weighted sum, then returning a list of all three. This allowed us to integrate income into the pipeline, as changes in income could now compute changes in the indicators. Then, we trained a second random forest model as well as a ridge regression model (called the final models) to predict FSRI score based on income. The final models were combined with a weightage of 75%/25% for the random forest model and the ridge model, respectively. This particular weighting was selected because it struck an effective balance between maintaining accuracy and capturing detailed granularity during inference.

Finally, we modelled the equal dividends implementation. First, we calculated every individual's net income after a simulated carbon tax had been implemented, the values of which were then passed through the machine learning pipeline. This produced the predicted post-tax FSRI scores for this specific scenario. Once this was completed, we took the difference between the post-tax predicted FSRI score values and the original predicted FSRI score values to find the overall welfare gain from the policy (**Table 3**).

The Genetic Algorithm - Setup

Genetic algorithms are evolutionary based metaheuristic search algorithms that solve complex problems by iteratively improving solutions within a population (11). We employed PyGAD, an open-source Python package used for constructing and running genetic algorithms (24). The algorithm required a fitness function to evaluate solution quality, as well as to compare that solution with others that the algorithm was considering. The fitness function was a completely separate python function from the machine learning pipeline; however, the pipeline was used in the calculation of solution fitness. The fitness function returned values from 0–1 depending on how well any particular solution increases financial stability to guide the search process. Before we defined the fitness function, basic constraints were set up, such as the gene space, which defined the possible dividend allocations for each income decile (ranging from 80% to 120% of what each decile would receive under an equal dividend scenario). This ensured that any one decile wouldn't suffer substantial disadvantages when it came to fund allocation. Other variables, such as the total revenue, the amount of people in each decile, and the cost increases, were also defined.

The Genetic Algorithm - Implementation

The actual fitness function was very similar to the function that calculated the welfare gain from an equal dividend's scenario. It took the net income from the policy scenario and went through the same machine learning pipeline (formed out of the combined model) to obtain the differences in FSRI score for each individual. The fitness values computed included the mean overall gain as well as the mean gain for both the bottom and top income deciles. This classified the problem as multi-objective; therefore, we implemented a non-dominated sorting genetic algorithm (NSGA-II) which was specifically configured for maximizing all three objectives to create the most overall benefit (12). These algorithms are typically populated with an initial solution; in this case, we selected the equal dividends scenario to start off as it was the main focus of this policy analysis. Then, the algorithm iterated through solutions, maintaining a population of the ones that were of the highest fitness. These strongest solutions were chosen through tournaments of selection after each generation. The selected solutions then mated with each other, which created new offspring solutions that contained properties from both parent solutions. These mating processes (which consist of crossover and mutation), like the selection tournaments, also took place at the end of every generation, allowing the algorithm to keep exploring more and more solutions. We selected the scattered crossover method, where fund allocation figures are exchanged between parent solutions to create diverse child solutions. As is typical, we set the probability of crossover between parents at 50% (31). Mutation, on the other hand, consists of random changes to the solution. Mutation occurs after the parents have crossed over, introducing new properties to the offspring solutions rather than just exchanging pre-existing ones. In our case, we swapped deciles' fund allocation with another value from the gene space at a low probability (10% chance of a decile being selected for mutation). This means that as a result of mutation, deciles could either lose or gain revenue funds based on the randomly selected value from the gene space. This random mutation-maintained solution diversity within the overall population. Our scenario required a custom mutation function to ensure that these funding swaps didn't imbalance the use of the carbon dividends' total revenue. In our function, if one decile got a funding increase due to mutation, another decile would be randomly selected for a proportional decrease to compensate (and vice versa if the decile selected for mutation received a funding decrease instead of an increase). Once these necessary variables and functions were defined, we used PyGAD's documentation to create a genetic algorithm instance and run the algorithm. Within this instance, we defined key parameters, such as the number of generations (set to 500) as well as a random state, so that the results are entirely reproducible. The algorithm had a runtime of ~50 hours on an NVIDIA V100 GPU to complete it's 500 generations.

ACKNOWLEDGMENTS

We would like to thank Anders Fremstad for his help and guidance, as well as for sharing his methods to calculate the price increases due to a carbon dividend which we were then able to adapt to our own needs.

Received: July 29, 2024
Accepted: February 04, 2024
Published: June 14, 2025

REFERENCES

1. "The Basics of Carbon Fee and Dividend." *Citizen's Climate Lobby*. https://www.citizensclimatelobby.org/basics-carbon-fee-dividend/. Accessed 14 May 2024.
2. Nunez, Christina. "Carbon dioxide levels are at a record high. Here's what you need to know." *National Geographic.* 13 May 2019, https://www.nationalgeographic.com/environment/article/greenhouse-gases. Accessed 17 May 2024.
3. Baker, James A., III, *et al*. "THE CONSERVATIVE CASE FOR CARBON DIVIDENDS." *Climate Leadership Council*.

https://doi.org/10.59720/24-234

Feb. 2017, https://www.clcouncil.org/media/2017/03/The-Conservative-Case-for-Carbon-Dividends.pdf. Accessed 25 Dec. 2023.
4. Kaufman, Noah. "What You Need to Know About a Federal Carbon Tax in the United States." *Center on Global Energy Policy.* 2 Apr. 2019, https://www.energypolicy.columbia.edu/publications/what-you-need-to-know-about-a-federal-carbon-tax-in-the-united-states/. Accessed 20 May 2024.
5. Fremstad, Anders and Paul, Mark. "The Impact of a Carbon Tax on Inequality." *Ecological Economics*, vol. 163, Sept. 2019, pp. 88-97. https://doi.org/10.1016/j.ecolecon.2019.04.016.
6. Nystrom, Scott and Luckow, Patrick. "The Economic, Climate, Fiscal, Power, and Demographic Impact of a National Fee-and-Dividend Carbon Tax". June, 2014. https://static.prod01.ue1.p.pcomm.net/cclobby/content/resources/remi/The-Economic-Climate-Fiscal-Power-and-Demographic-Impact-of-a-National-Fee-and-Dividend-Carbon-Tax-5.25.18.pdf
7. Charalampakis, Evangelos, *et al*. "The impact of the recent rise in inflation on low-income households." *European Central Bank Economic Bulletin*. July 2022, https://www.ecb.europa.eu/press/economic-bulletin/focus/2022/html/ecb.ebbox202207_04~a89ec1a6fe.en.html. Accessed 2 June 2024.
8. Breiman, Leo. "Random Forests." *Machine Learning*, vol. 45, Oct. 2001, pp. 5-32. https://doi.org/10.1023/A:1010933404324.
9. Hoerl, Arthur E., and Kennard, Robert W. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". *Technometrics,* vol 42, no. 1, Feb. 1970, pp. 80-86, https://doi.org/10.2307/1271436.
10. "Survey of Household Economics and Decisionmaking." *Federal Reserve.* 22 May 2023, https://www.federalreserve.gov/consumerscommunities/shed_data.htm. Accessed 8 Oct. 2023.
11. Kanade, Vijay. "What Are Genetic Algorithms? Working, Applications, and Examples." *Spiceworks.* Sept. 6 2023, https://www.spiceworks.com/tech/artificial-intelligence/articles/what-are-genetic-algorithms/. Accessed 3 Nov. 2023.
12. Deb, Kalyanmoy, *et al*. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *Transactions on Evolutionary Computation,* vol. 6, no. 2, Apr. 2002, pp. 182-197, https://doi.org/10.1109/4235.996017.
13. "Real Mean Personal Income in the United States." *Federal Reserve Economic Data.* Sept. 12 2023, fred.stlouisfed.org/series/MAPAINUSA672N. Accessed 9 Oct. 2023.
14. "Real Median Personal Income in the United States." *Federal Reserve Economic Data.* Sept. 12 2023, fred.stlouisfed.org/series/MEPAINUSA672N. Accessed 9 Oct. 2023.
15. "Family Financial Stability Index." *United Way.* May 2022, www.unitedwayoc.org/wp-content/uploads/2022/06/FFSI-Summary-and-2020-FdFSI-OC-Results_FINAL.pdf.
16. Datta, Abhirup and Arkajyoti Saha. "Random forests for binary geospatial data." *arXiv*, Feb. 2023, https://doi.org/10.48550/arXiv.2302.13828.
17. Taboga, Macro. "Ridge Regression." *StatLect.* 2021, https://www.statlect.com/fundamentals-of-statistics/ridge-regression. Accessed 17 Jan 2024.
18. Jaeger, William K. "Double Dividend." *Encyclopedia of Energy, Natural Resource, and Environmental Economics*, vol 1, 2 Apr. 2013, pp. 37-40, https://doi.org/10.1016/B978-0-12-375067-9.00073-5.
19. Ismail, Mahmoud S., *et al.* "Genetic algorithm based optimization on modeling and design of hybrid renewable energy systems." *Energy Conversion and Management*, vol. 85, Sept. 2014, pp. 120-130, https://doi.org/10.1016/j.enconman.2014.05.064.
20. Lazzari, Florencia, *et al*. "Optimizing planning and operation of renewable energy communities with genetic algorithms." *Applied Energy,* vol. 338, 15 May 2023, https://doi.org/10.1016/j.apenergy.2023.120906.
21. "Impose a Tax on Emissions of Greenhouse Gases." *Congressional Budget Office.* 7 Dec. 2022, https://www.cbo.gov/budget-options/58638. Accessed 23 Aug. 2023.
22. "Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide Interim Estimates under Executive Order 13990." *Interagency Working Group.* Feb. 2021, https://www.whitehouse.gov/wp-content/uploads/2021/02/TechnicalSupportDocument_SocialCostofCarbonMethaneNitrousOxide.pdf. Accessed 17 Aug. 2023.
23. "CPI Inflation Calculator." *Bureau of Labor Statistics.* https://data.bls.gov/cgi-bin/cpicalc.pl. Accessed 2 Aug 2023.
24. Harris, Charles R., *et al*. "Array programming with NumPy." *Nature,* vol. 585, 16 Sept. 2020, pp. 357-362, https://doi.org/10.1038/s41586-020-2649-2.
25. McKinney, Wes, *et al*. "Data structures for statistical computing in python." *Proceedings of the 9th Python in Science Conference*, vol. 445, 2010, pp. 51–56, https://doi.org/10.5281/zenodo.3509134.
26. Waskom, Micheal L. "seaborn: statistical data visualization." *Journal of Open Source Software,* 6 Apr. 2021, https://doi.org/10.21105/joss.03021.
27. Hunter, John. "Matplotlib: A 2D Graphics Environment.", *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, May 2007, https://doi.org/10.1109/MCSE.2007.55.
28. "Appendix B: Adjusting Household Income for Household Size*." Pew Research Center*, Oct. 3 2011, https://www.pewresearch.org/social-trends/2011/10/03/appendix-b-adjusting-household-income-for-household-size/. Accessed 26 Feb. 2024.
29. van Rossum, Guido. "Python Reference Manual". *CWI*, 1995, https://ir.cwi.nl/pub/5008.
30. Chen, Tianqi, et al. "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* Aug. 2016, pp. 785-794, https://doi.org/10.1145/2939672.2939785.
31. Gad, Ahmed F. "PyGAD: An Intuitive Genetic Algorithm Python Library." *Computer Science > Neural and Evolutionary Computing,* 11 June 2021, https://doi.org/10.48550/arXiv.2106.06158.

https://doi.org/10.59720/24-234

Appendix

The following is a link to our public GitHub repository where all of our code and data are stored: https://github.com/MostAardvark224/The-effect-of-carbon-dividends-on-financial-security-using-machine-learning-and-metaheuristic-search.