Causal AI Framework for Unit Selection in Optimizing Electric Vehicle Procurement

Anonymous submission

Abstract

Electric vehicles (EVs) are generally considered more environmental sustainable than internal combustion engine vehicles (ICEVs). Government and policy makers may want to incentivize multi-vehicle households that, if purchsed a new EV, would use their EV to replace a large portion of their ICEV mileage. It is hence important to analyze how EV procurement affects annual EV mileage for different households. Given that many relevant data, especially experimental data are often unavailable in the real-world, we need causal analysis tools to answer this question. Additionally, we aim to compare the expected EV mileage of different combination of vehicles a household owns. It is impossible to observe both combinations since only one might happen, which makes causal inference challenging. In this paper, we construct a causal AI framework utilizing counterfactual reasoning methods to solve this problem.

Introduction

The transportation industry contributes to over a quarter of total greenhouse gas (GHG) emissions in the United States, and light-duty vehicles alone are responsible for more than half of these emissions. There is a widespread consensus that the adoption of electrified vehicles will be a significant factor in future initiatives to achieve carbon neutrality (Jenn 2020; Burnham et al. 2021). When targeting individual choices and when the interventions have a corresponding cost associated with them it is important to take into account the possibility for heterogeneous treatment effects. The benefits from intervening on some groups, or on some individuals, might be smaller or larger than the benefits of intervening on other groups or individuals. Understanding the heterogeneity of driving patterns across individuals, households and groups is important when trying to maximize the desired outputs. A recent paper (Nunes, Woodley, and Rossetti 2022) compares the benefits from from targeting different types of households. The main difference between households the authors considered in their model was in the number of current vehicles in a household. The results indicated that the advantages of acquiring an EV could drastically vary depending on the current vehicle mix.

Nowadays many households own more than one vehicle. In particular, many people choose to purchase an EV as a complementary vehicle, not driving it much while primarily relying on their ICEV (Burlig et al. 2021). This could be due to various reasons including their personal preference towards their ICEVs, insufficient EV mileage ranges, and charging inconvenience. As a result, the carbon emission benefit is not as large as households who drive their EVs as primary vehicles. In the interest of budget, policy makers may want to target EV purchase incentives on those who, upon purchasing new EVs, would use their EVs to replace a large portion of their ICEV driving mileage. To solve this optimization problem, we need to answer the question, "what is the expected difference in EV mileage among households convinced to purchase a new EV versus not convinced?" This gives policy makers a useful criterion for prioritizing incentives.

Note that this question, at the individual household level, is counterfactual. We can never observe or test both actions, one of them cannot occur. There are significant caveats with not treating this at the individual level (Mueller and Pearl 2022). Li and Pearl detail the sometimes severely suboptimal decision making that results from a traditional analysis (Li and Pearl 2019).

In this paper, we focus on multi-vehicle households, and develop a causal AI framework to estimate the counterfactual effects of adding an additional EV to a household on the increment of their EV driving mileage.

Preliminaries

Causal Inference

A causal model is composed of a causal directed acyclic graph (DAG) $G(V, E)$ and a set of structural equations. V are nodes representing model variables and E are edges representing causal relations between two nodes. Directed edges encode the direction of causality, i.e., if a variable A is in the structural equation that determines another variable B, an edge is drawn from A to B.

In this paper, we follow the notation in (Pearl 2009), and use uppercase letters to denote variables, and lowercase letters (combined with symbols and numbers) to denote the values a variable can take on. For example, the values of a binary variable A can be denoted as a and a', and the values of a non-binary variable B can be denoted as $b^1, b^2, \ldots$. A variable C with the value d of another variable D as a subscript represents the event of C with the intervention $D = d$. This is denoted as either C_d or $C_{D=d}$, and they are inter-

changeable.

Causal AI Framework for EV Driving Analysis

Causal Model

We are interested in learning about what happens to a household's total miles on EV if they purchase an additional EV. We first build a graphical representation of the causal relationships using a causal directed acyclic graph.

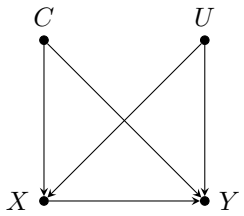

Figure 1: Causal DAG

Since our analysis are not based on real-world data and for illustration purposes, we limit our focus by assuming the variables are categorical and can only take on specific values. Our approach can be easily generalized for larger sets of values. For the same reason, we model one observed confounder and one unobserved confounder while this framework applies to more confounders of either type, too.

In this model, variable X represents the numbers and types of cars a household owns. We focus our discussion on cases where a household has 1) one EV and one ICEV or 2) two EVs and one ICEV. Cases 1 and 2 are represented by $X = x$ and $X = x'$, respectively. Variable Y represents the annual miles driven on all the EVs for a household, which lies in one of the five ranges, $y^1 = [0, 5000)$, $y^2 = [5000, 10000)$, $y^3 = [10000, 15000)$, $y^4 = [15000, 20000)$, $y^5 = [20000, \infty)$. Variables C and U are two confounders causing both X and Y. C is observed (assumed available in the data), which represents the annual total travel needs for a household using all available vehicles. $C = c$ or $C = c'$ denotes a household needs to travel more or less than 15000 miles per year. U is unobserved (assumed not available in the data), which represents whether the typical trip types of a household favors EVs. $U = u$ indicates that the household mainly drives trips that favors EVs (e.g., shorter trips, trips with easy charging options, etc.), and $U = u'$ indicates that the household mainly drives trips that favors ICEVs (e.g., longer trips, trips to cold places, etc.). Note that, although U is unobserved for single households, we might have an estimate of what percentage of the population has $U = u$ or $U = u'$. Hence, we assume that the prior $P(U)$ is given.

Problem Setting

We are interested in assessing for a household where we observe they drove one EV and one ICEV that drove certain EV miles for the past year, what the benefit is on annual EV miles if an additional EV is added. This is a counterfactual query because when we observe a household with certain combination of vehicles (e.g., one EV and one ICEV), we do not simutaneously observe them with a different combination of vehicles (e.g., two EVs and one ICEV). In addition, we are given observational data from the past, and we are specifically interested what happens in the current timestamp if they buy a new EV vs. not buying a new EV. To put this into a counterfactual expression (Pearl 2009; Li and Pearl 2022b), we have

$$P(Y_{X=x} = y^a, Y_{X=x'} = y^b \mid XP = x', YP = y^p), \quad (1)$$

where XP and YP are the variables X and Y at the previous timestamp where the observational data are given. y^a, y^b, y^p are values of Y (or YP) and $1 < a, b, p < 5$. Y and YP are different variables but takes on the same set of values. Note that this expression is the non-binary probability of necessity and sufficiency (PNS(2)) (Li and Pearl 2022b) of X on Y. Here, our goal is to estimate, for a household with one EV and one ICEV that drove y^p EV miles last year, what the probability is that they would drive y^b this year if not added a new EV, and would drive y^a if added a new EV. We want to estimate this probability for all a, b, and p.

Estimation

Without additional assumptions no counterfactual query can be point estimated, even with both observational and experimental data (Tian and Pearl 2000). However, it may be possible to bound the query using observational and/or experimental data (Tian and Pearl 2000; Li and Pearl 2022b; Mueller, Li, and Pearl 2022; Zhang, Tian, and Bareinboim 2022; Dawid, Musio, and Murtas 2017; Li and Pearl 2022c). The listed bounding methods in existing work that are applied to different settings, such as binary, continuous, monotonic, etc. We will adapt methods from (Li and Pearl 2022b) to bound the (1) since the equation is the non-binary probability of causation. We will apply Theorem 8 in (Li and Pearl 2022b) to obtain the query (1) (referred to as Li-Pearl's PNS bounds).

In addition, experimental data are usually unavailable for this problem, since it is costly to conduct an experiment to provide households with EVs. Fortunately, we can use observational data to deduce bounds on experimental data, which are based on the Theorem 4 in (Li and Pearl 2022a) (referred to as Li-Pearl's causal effect bounds.)

Another challenge of counterfactual estimation in this case is that the observational data are from the past, while we are trying to infer the behaviors for the future. To make use of the available observational data, there need to be assumptions on how the past observational data predicts the future observational state. Causal inference frameworks, when applied to real-world problems, usually implicitly assume that what we observe in the past continues to apply for the future. To this end, we will discuss two assumptions of similar purposes, and is up to the practitioner to choose which assumption is more plausible for their setting.

Scenario 1: Constant A simple assumption is to assume the observations from the past has not changed as of the time the study is being conducted. Formally, this means for each

household, $X = XP, Y = YP$. Hence, (1) can be simplified as follows.

$$\begin{aligned}&P(Y_{X=x} = y^a, Y_{X=x'} = y^b \mid XP = x', YP = y^p)\\ =&P(Y_{X=x} = y^a, Y_{X=x'} = y^b \mid X = x', Y = y^p)\\ =&P(Y_{X=x} = y^a \mid X = x', Y = y^p) \qquad (2)\end{aligned}$$

This becomes the non-binary probability of necessity (Li and Pearl 2022b) of X on Y. Under this assumption, (2) can be bounded using the Theorem 7 in (Li and Pearl 2022b) (referred to as Li-Pearl's probability of necessity bounds.)

Scenario 2: Variant A relaxed assumption is to permit change in observations each year, but assume the changes follow the same pattern. Under this assumption, in addition to the observational data from the previous year $P(XP, YP)$, we additionally need the observational data $P(XPP, YPP)$ from the year before the previous year. Formally, this assumption translates to

$$\begin{aligned}&P(XP = x, YP = y^p \mid XPP = x', YPP = y^{pp})\\ =&P(X = x, Y = y^p \mid XP = x', YP = y^{pp})\end{aligned}$$

Hence, we have the observational data $P(X, Y \mid XP, YP)$ for (1). We can then use the observational data to bound the experimental data to obtain $P(Y_X \mid XP, YP)$. Given both observational and experimental data, we can use Li-Pearl's PNS bounds to bound (1).

Computing the Benefit

Once we have the bounds of (1) (if assumption 2 holds) or (2) (if assumption 1 holds), there are multiple ways where the results can be used. For example, for each household, we can compute the bound of the expected EV mileage if added an additional EV and the bound of the expected EV mileage if not added an additional EV. So for each household, the difference in the two expectations is the expected EV mileage increment. We can identify which households are expected to have large mileage increments, and which are not. Another way is to find what most likely to happen for each household, which means finding the PNS or PN bound with the highest probability. The researcher can decide which way best fits their needs.

Experiment and Results

Scenario 1: Constant

In this section, we simulated the following example to illustrate our proposed framework under the first assumption.

We generated $P(X, Y, C)$ and $P(U)$ uniformly, as shown in Tables 1 and 2. We then applied Li-Pearl's causal effect bounds to derive the experimental (RCT) data $P(Y_x)$ using the data from Tables 1 and 2. The results are presented in Table 3. Subsequently, we used Li-Pearl's PN bounds to calculate the non-binary Probability of Necessity, with the results displayed in Table 4. Note that we are only presenting the upper bounds because the lower bounds for this randomly generated example are all zero, which provides no additional information.

Table 1: Scenario 1: Simulated observational distribution of the whole population.

	2 EV & 1 ICEV		**1 EV & 1 ICEV**	
	$\geq 15,000$	$< 15,000$	$\geq 15,000$	$< 15,000$
y^1	0.019	0.008	0.043	0.038
y^2	0.014	0.061	0.178	0.045
y^3	0.016	0.007	0.051	0.043
y^4	0.089	0.018	0.017	0.142
y^5	0.021	0.086	0.033	0.071

Table 2: Scenario 1: Prior knowledge about the trip type.

Typical trip type (U)	**Percentage**
More long trips (u)	6.7%
More short trips (u')	93.3%

From Table 3, we have obtained narrow bounds on causal effects. However, according to our reasoning, the causal effects are not the correct queries we need. In Table 4, although there are only 10 entries that are not 1, we still gather useful information for decision-making. For instance, we can determine that the probability of the population having 1 EV and 1 ICEV, with an EV drive level of y_2, and increasing their EV drive to level y_3 with one more EV, is at most 34.5%.

Scenario 2: Variant

In this section, we simulated another example to illustrate our proposed framework under the second assumption. We generated $P(X, Y, C|XP = x', YP = y^1)$ and $P(U|XP = x', YP = y^1) = P(U)$ (since U is the confounder) uniformly, as shown in Tables 5 and 6. We then applied Li-Pearl's causal effect bounds to derive the experimental (RCT) data $P(Y_{X=x}|XP = x', YP = y^1)$ using the data from Tables 5 and 6. The results are presented in Table 7. Subsequently, we used Li-Pearl's PNS bounds to calculate the non-binary Probability of Necessity and Sufficiency, with the results displayed in Table 8. Note that we are only presenting the upper bounds because the lower bounds for this randomly generated example are all zero, which provides no additional information.

From Table 7, we have obtained narrow bounds on causal effects. However, according to our reasoning, the causal effects are not the correct queries we need and should not be directly used to answer our question. In Table 8, for instance, we can determine that the probability of an individual who have 1 EV and 1 ICEV that would have EV drive level y_5 if having 2 EV and 1 ICEV and would have EV drive level y_2 if still have 1 EV and 1 ICEV, is at most 32.6%.

Conclusion

In this paper, we focus on the problem of optimizing electric vehicle procurement for maximizing environmental sustainability. We showed that the question of how much each household benefits from an additional EV is a counterfactual question, which is hard to solve using available obser-

Table 3: Scenario 1: Bounds of the experimental distribution.

	Lower bound	**Upper bound**
$P(y_x^1)$	0.027	0.117
$P(y_x^2)$	0.075	0.283
$P(y_x^3)$	0.023	0.100
$P(y_x^4)$	0.184	0.472
$P(y_x^5)$	0.164	0.438

Table 4: Scenario 1: Upper bounds of non-binary Probability of Necessity.

$P(y_x^a \mid x', x^p)$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$a=1$	1	0.404	0.957	0.566	0.865
$a=2$	1	0.933	1	1	1
$a=3$	0.951	0.345	0.819	0.484	0.740
$a=4$	1	1	1	1	1
$a=5$	1	1	1	1	1

vational or experimental data. To approach this problem, we developed a causal AI framework based on counterfactual reasoning. We showed how to apply this framework using simulated experiment. For both scenarios discussed, we obtained bounds on the query of interest.

Table 5: Scenario 2: Simulated observational distribution of the population $XP = x', YP = y^1$.

	2 EV & 1 ICEV		**1 EV & 1 ICEV**	
	$\geq 15,000$	$< 15,000$	$\geq 15,000$	$< 15,000$
y^1	0.110	0.062	0.066	0.035
y^2	0.066	0.009	0.030	0.072
y^3	0.142	0.047	0.006	0.063
y^4	0.010	0.098	0.029	0.098
y^5	0.004	0.013	0.012	0.028

Table 6: Scenario 2: Prior knowledge about the trip type.

Typical trip type (U)	**Percentage**
More long trips (u)	10.8%
More short trips (u')	89.2%

Table 7: Scenario 2: Bounds of the experimental distribution of the population $XP = x', YP = y^1$.

	Lower bound	**Upper bound**
$P(y_x^1)$	0.173	0.393
$P(y_x^2)$	0.075	0.139
$P(y_x^3)$	0.211	0.395
$P(y_x^4)$	0.108	0.303
$P(y_x^5)$	0.017	0.051
$P(y_{x'}^1)$	0.101	0.476
$P(y_{x'}^2)$	0.102	0.394
$P(y_{x'}^3)$	0.069	0.260
$P(y_{x'}^4)$	0.127	0.521
$P(y_{x'}^5)$	0.040	0.188

Table 8: Scenario 2: Upper bounds of non-binary Probability of Necessity and Sufficiency of the population $XP = x', YP = y^1$.

$P(y_x^a, y_{x'}^b)$	$b=1$	$b=2$	$b=3$	$b=4$	$b=5$
$a=1$	0.596	0.513	0.412	0.615	0.369
$a=2$	0.439	0.356	0.255	0.458	0.212
$a=3$	0.581	0.498	0.397	0.590	0.354
$a=4$	0.570	0.487	0.386	0.589	0.343
$a=5$	0.409	0.326	0.225	0.428	0.182

References

Burlig, F.; Bushnell, J.; Rapson, D.; and Wolfram, C. 2021. Low Energy: Estimating Electric Vehicle Electricity Use. *AEA Papers and Proceedings*, 111: 430–35.

Burnham, A.; Lu, Z.; Wang, M.; and Elgowainy, A. 2021. Regional emissions analysis of light-duty battery electric vehicles. *Atmosphere*, 12(11): 1482.

Dawid, P.; Musio, M.; and Murtas, R. 2017. The Probability of Causation. *Law, Probability and Risk*, (16): 163–179.

Jenn, A. 2020. Emissions benefits of electric vehicles in Uber and Lyft ride-hailing services. *Nature Energy*, 5(7): 520–525.

Li, A.; and Pearl, J. 2019. Unit Selection Based on Counterfactual Logic. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 1793–1799. International Joint Conferences on Artificial Intelligence Organization.

Li, A.; and Pearl, J. 2022a. Bounds on causal effects and application to high dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 5773–5780.

Li, A.; and Pearl, J. 2022b. Probabilities of Causation with Non-binary Treatment and Effect. Technical Report R-516, Department of Computer Science, University of California, Los Angeles, CA.

Li, A.; and Pearl, J. 2022c. Unit selection with causal diagram. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 5765–5772.

Mueller; and Pearl. 2022. Personalized Decision Making – A Conceptual Introduction. Technical Report R-513, Department of Computer Science, University of California, Los Angeles, CA.

Mueller, S.; Li, A.; and Pearl, J. 2022. Causes of effects: Learning individual responses from population data. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, 2712–2718.

Nunes, A.; Woodley, L.; and Rossetti, P. 2022. Re-thinking procurement incentives for electric vehicles to achieve net-zero emissions. *Nature Sustainability*, 5(6): 527–532.

Pearl, J. 2009. *Causality*. Cambridge university press.

Tian, J.; and Pearl, J. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4): 287–313.

Zhang, J.; Tian, J.; and Bareinboim, E. 2022. Partial counterfactual identification from observational and experimental data. In *International Conference on Machine Learning*, 26548–26558. PMLR.