Project ID # VAN037

Vehicle Manufacturer's Suggested Retail Price (MSRP) Estimation using Machine Learning

Ayman Moawad, Ehsan Islam, Namdoo Kim, Ram Vijayagopal, Aymeric Rousseau
Argonne National Laboratory

2020 DOE Vehicle Technologies Office
Annual Merit Review

June 2020

PROJECT OVERVIEW

Timeline	Barriers*
• Project start date : 04/01/2019 • Project end date : 03/30/2020 • Percent complete : 100%	• Constant advances in technology. • Cost. • Computational models, design, and simulation methodologies. *from 2011-2015 VTP MYPP
Budget	**Partners**
• FY20 Funding : $150K	• University of Chicago

OBJECTIVES

Update vehicle and component costs to improve Benefit Analysis

- **Background**
  - Argonne has been supporting DOE VTO to estimate the impact of new technologies on energy consumption and **cost.**
  - Component cost estimates outdated (2010).
  - Common cost estimation methods (essentially based on Bill of Materials and teardown methodologies) are lengthy and expensive.

  => New methodology needed to estimate individual technology cost.

- **Methodology**
  - Use a top-down approach: Leverage Machine Learning and Game Theoretical methods to build vehicle cost model and explain the contribution of individual components to the vehicle cost.
  - Extract component cost models at market level (includes direct and indirect costs).
- **Advantage**
  - No need for expensive surveying and teardown data.
  - No need for RPE(1) or ICM(2) adjustment (to mark up direct manufacturing costs to MSRP).
  - Bypass the uncertainty involved in both steps.

(1) RPE: Retail Price Equivalent; (2) ICM: Indirect Cost Multiplier

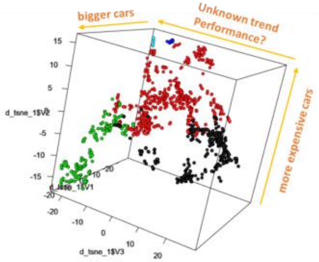
APPROACH

For vehicle MSRP estimation

2 Data Prep. and Analysis

Clean, integrate and feature engineer data

Vehicle "Make-Model Agnostic" Clustering



3 Predictive Model

CatBoost

5 fold cross validated

RMSE	~$950
MAPE	~2.2%
R^2	~0.99
Residuals	Normal

1 Data Collection

Automated web scraping process

VMSbot +

CAR and DRIVER, The Car Connection, MOTORTREND, THE CAR GUIDE, edmunds, cars-data, cars.com, autoblog, QAARS.COM

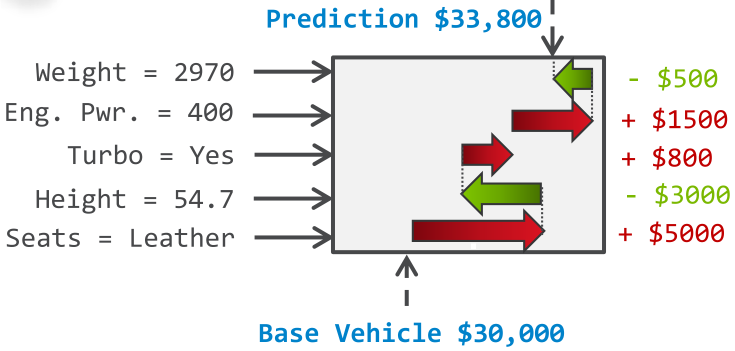
Argonne **V**ehicle **A**ttribute **D**atabase (ArVAD)

60,000 vehicles (MY 1990-2020)

500+ different vehicle specs

Stored in non-relational structure (mongoDB)

4 Surrogate Explainer Model



Additive feature attribution for local explanation i.e. on a per vehicle basis

SHAP

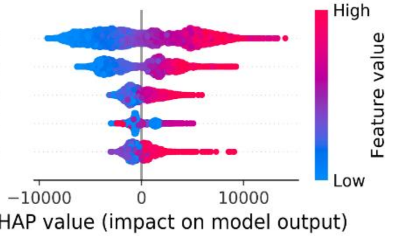
For component level price estimation

5 Global Insights

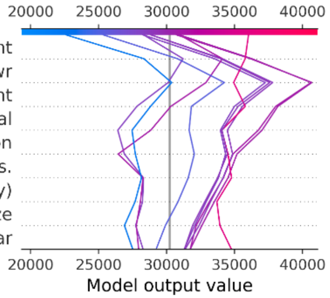
Aggregate local explanations to extract global behavior

(Vehicle & Component level price summaries)

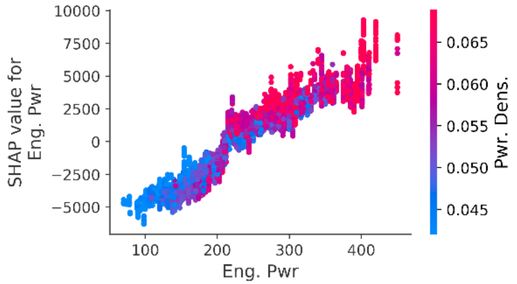
Behavioral Summary



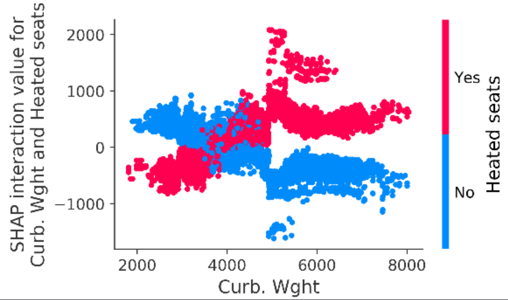
Decision Path



Feature Dependence



Interactional Effects



PROJECT RELEVANCE

- Given the collected data, predicting vehicle price using Machine Learning (ML) is a sensible method.
- We need a new approach to estimate individual technology costs and understand how technology changes affect vehicle costs.

Vehicle

Input specifications

ML Model

Predicts

MSRP[1] Estimate

$46,170

Contributions?

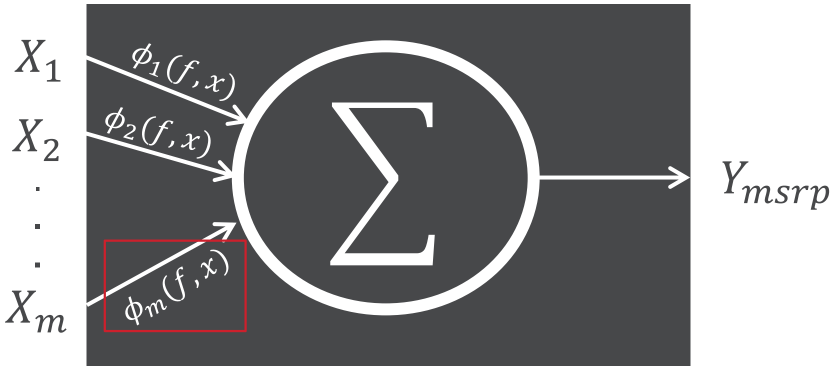
- Vehicle class
- Engine technology
- Fuel
- Transmission technology
- Tires
- Accessories
- …

Can we quantify each component contribution to vehicle price?
Can we extract component level prices?

(1) MSRP: Manufacturer Suggested Retail Price

METHODOLOGY

Additive Feature Attribution



Credit Attributed to component X_m

Several methods leverage this approach

LIME	**Ribeiro et al. 2016**
Shapley Values	**Datta et al. 2016, Lundberg et al. 2019**
Saabas	**Saabas 2014**
DeepLIFT	**Shrikumar et al. 2016**

$46,170

$\boldsymbol{f(x)}$ **: Model prediction**

0 $\quad \phi_0 \quad$ $\mathbb{E}[f(\boldsymbol{X})]$ $\quad$ $\mathbb{E}[f(\boldsymbol{X})|\mathbf{do}(X_1 = x_1)]$ $\quad$ $\mathbb{E}[f(\boldsymbol{X})|\mathbf{do}(x_1, x_2, x_3)]$ $\quad \phi_3$ $\quad$ $\mathbb{E}[f(\boldsymbol{X})|\mathbf{do}(x_1, x_2)]$

x_3: Year = 2010.

$30,000

Average vehicle price, i.e. best price prediction if nothing is known about the vehicle

ϕ_1

x_1: Turbo Engine = TRUE

ϕ_2

x_2: Curb. Weight = 4000 lbs.

Used is Coalitional or Cooperative game theory.

Holds certain fairness properties.

Allows to fairly distribute the contribution of each component from a total.

$$\phi_i(f,x) = \sum_{S \subseteq \mathcal{M} \setminus \{i\}} \frac{|S|!\,(M - |S| - 1)!}{M!}\left[f_x(S \cup \{i\}) - f_x(S)\right]$$

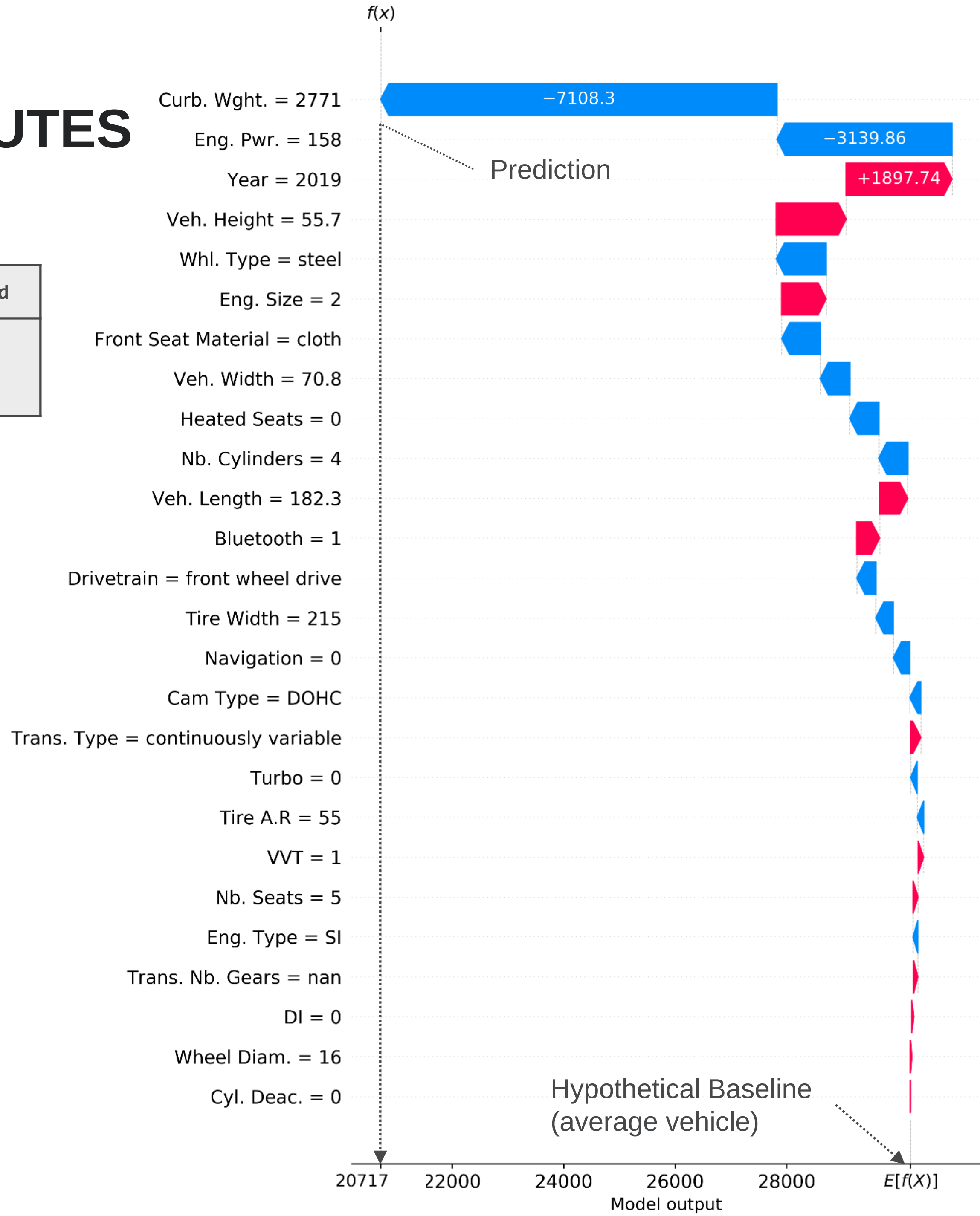
Lloyd Shapley

TECHNICAL ACCOMPLISHMENTS AND PROGRESS

MSRP CAN NOW BE PREDICTED WITH CONTRIBUTIONS OF INDIVIDUAL ATTRIBUTES

Using AVERAGE vehicle within database

year	vehicle	make	model	trim	MSRP	Predicted
2019	Honda Civic	honda	civic	LX 4dr Sedan (2.0L 4cyl CVT)	$20,350	$20,717

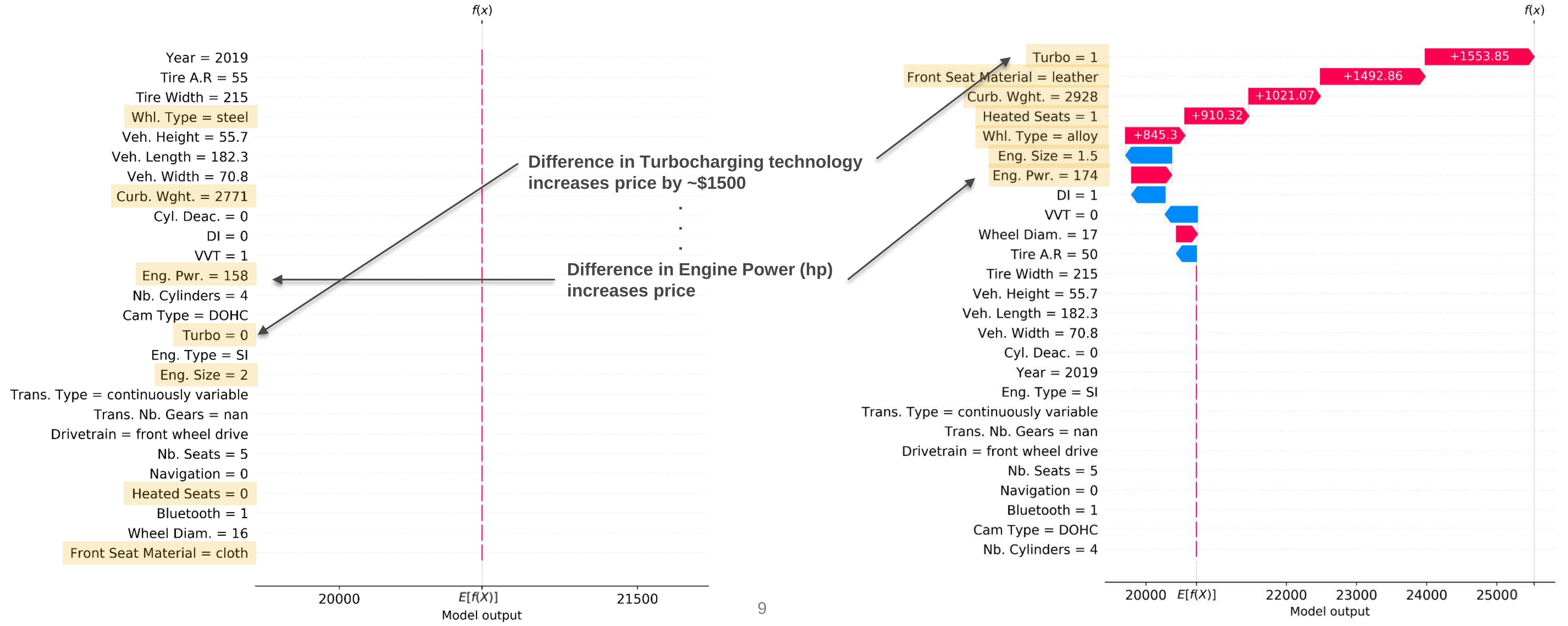


MSRP CAN NOW BE PREDICTED WITH INDIVIDUAL ATTRIBUTES CONTRIBUTIONS

- Using SPECIFIC vehicle within database for one to one comparison. Example: study impact of trim
- Direct trim level comparison allows to better understand and quantify the components involved in the price difference.

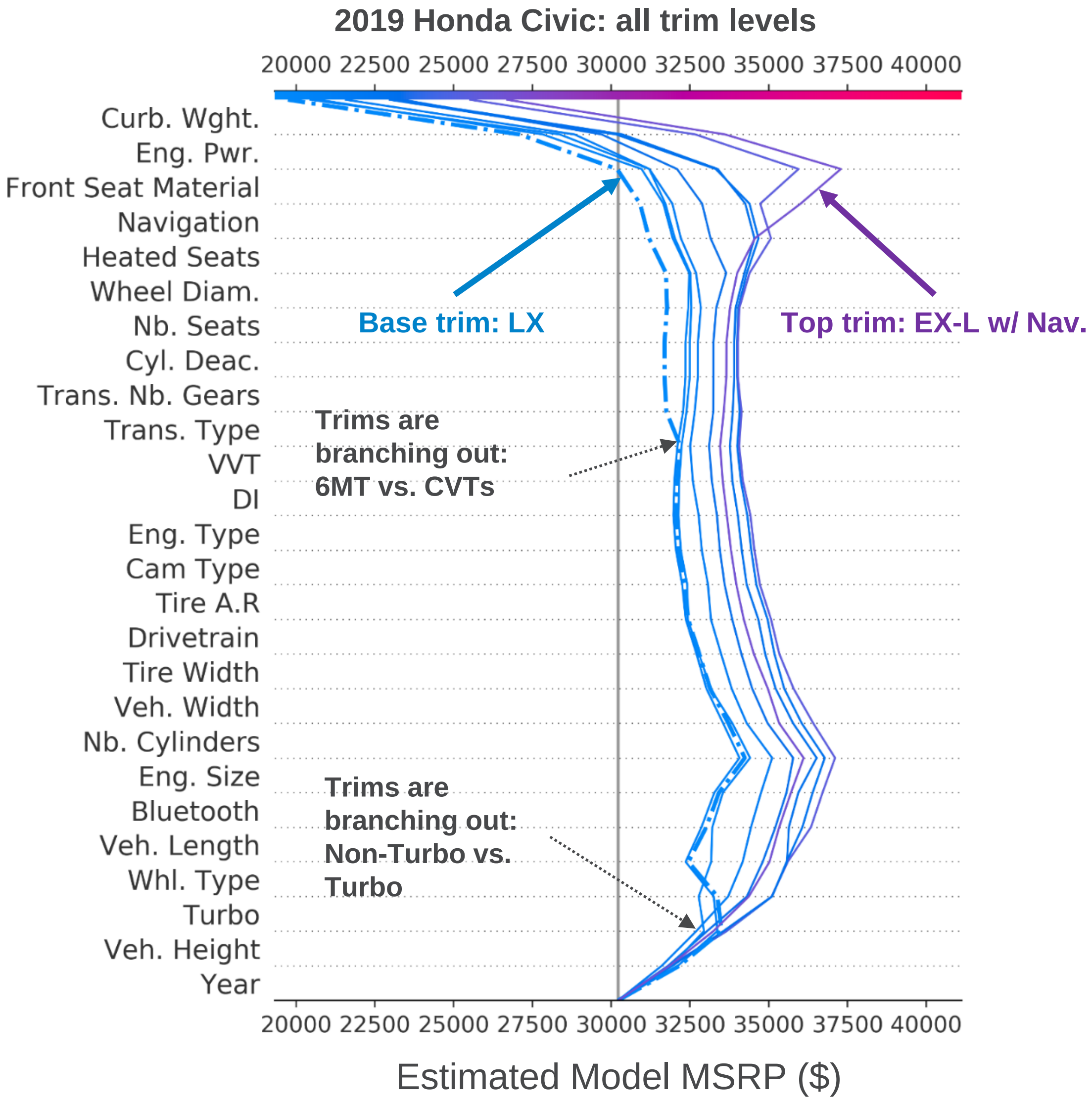
year	vehicle	make	model	trim	MSRP	Predicted
2019	Honda Civic	honda	civic	LX 4dr Sedan	$20,350	$20,717

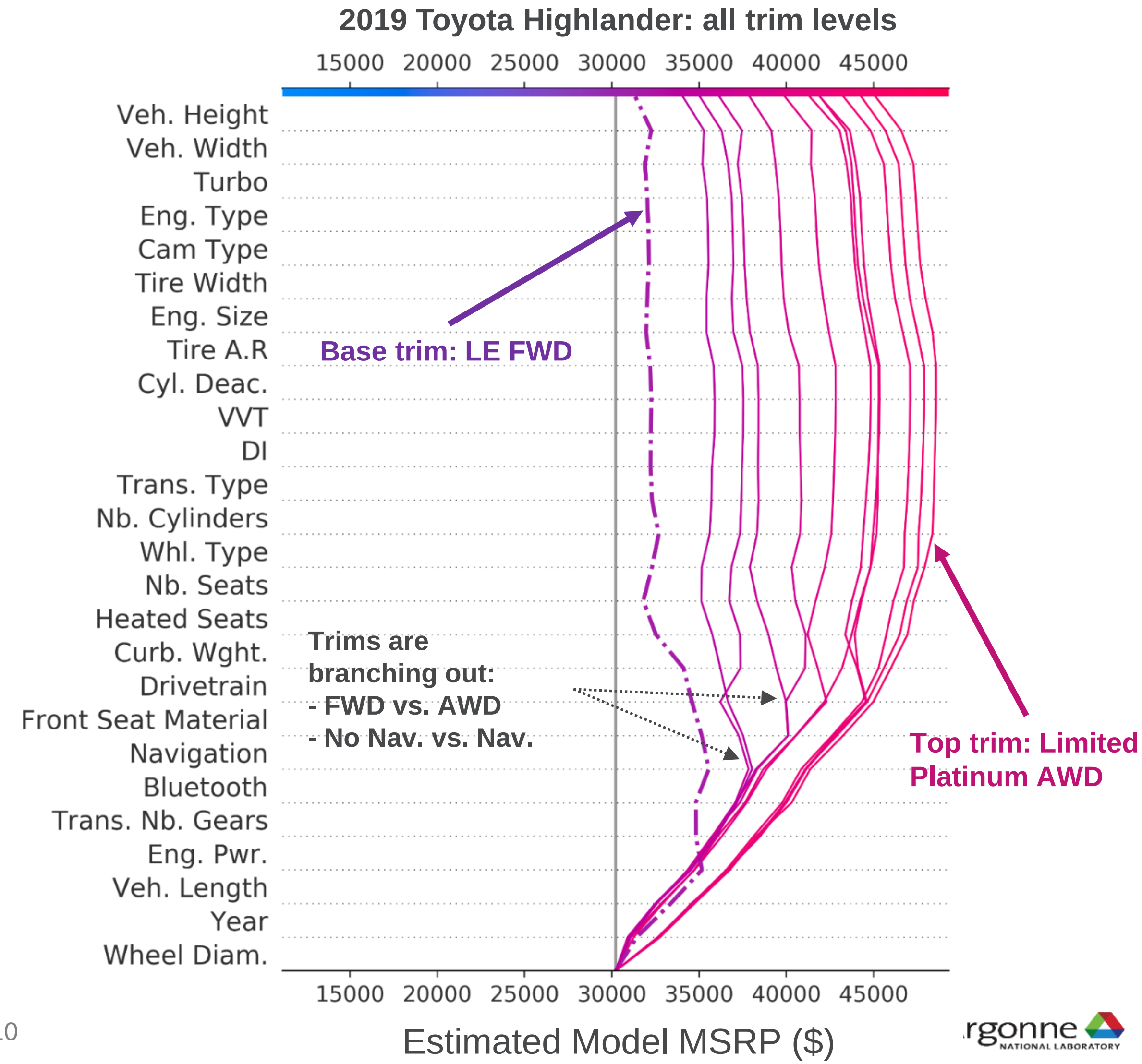
year	vehicle	make	model	trim	MSRP	Predicted
2019	Honda Civic	honda	civic	EX-L 4dr Sedan	$24,700	$25,368



TECHNOLOGIES IMPACTING MSRP MOST CAN BE IDENTIFIED BY COMPARING DIFFERENT ATTRIBUTES FOR A SET OF VEHICLES

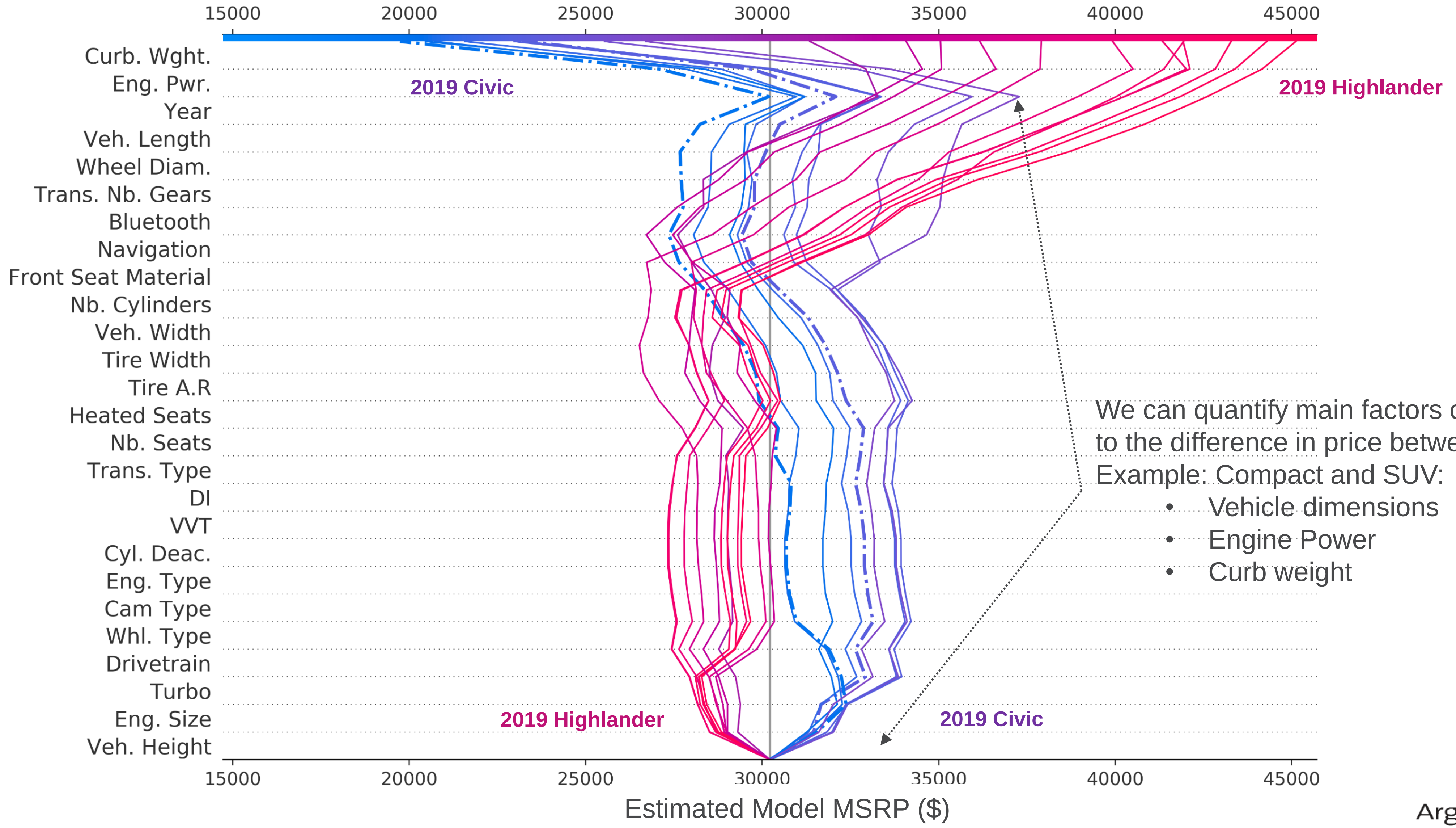
- Vehicles diverge in price as a result of component value differences. Slopes show magnitude of change in price.
- Allows us to better understand the effect of some key vehicle component on pricing





TECHNOLOGIES IMPACTING MSRP MOST CAN BE IDENTIFIED BY COMPARING DIFFERENT ATTRIBUTES FOR A GIVEN VEHICLE

Example of Compact Car vs SUV class

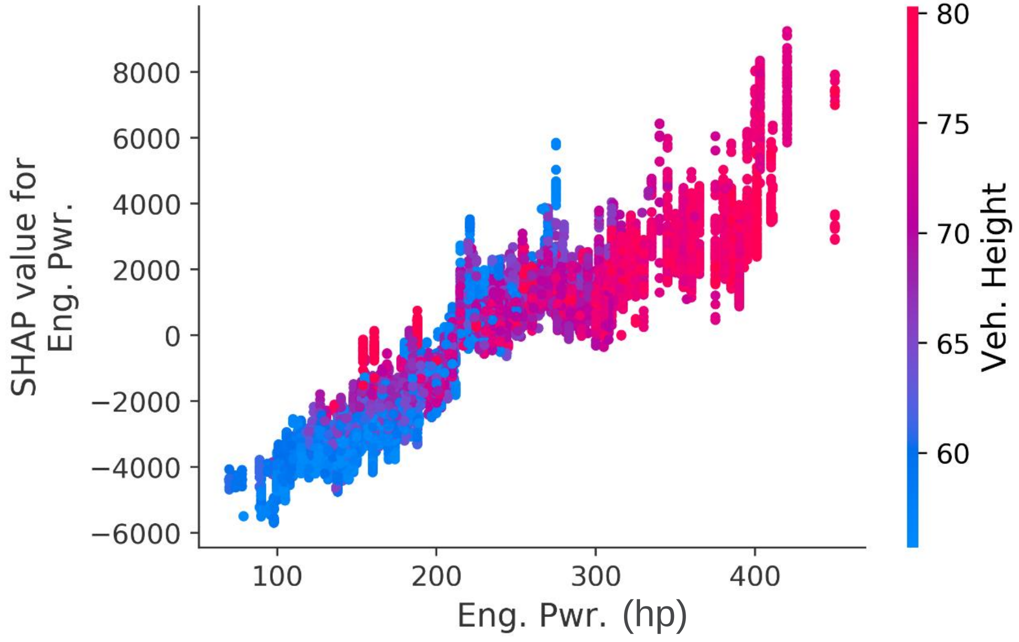


We can quantify main factors contributing to the difference in price between classes.
Example: Compact and SUV:
- Vehicle dimensions
- Engine Power
- Curb weight

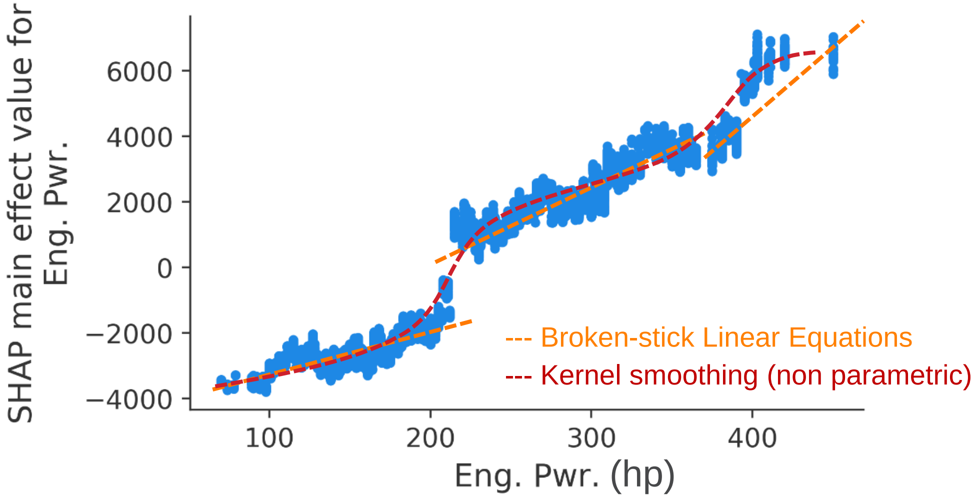
Argonne NATIONAL LABORATORY

IMPACT OF INDIVIDUAL TECHNOLOGY ACROSS ALL VEHICLES

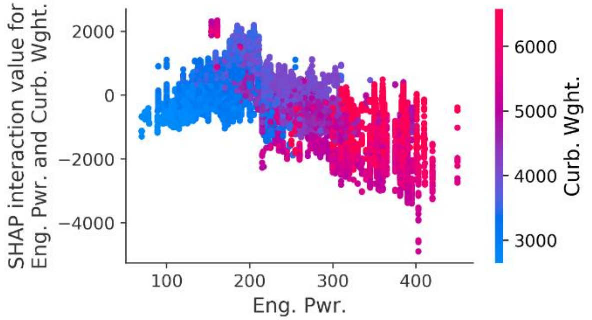
AGGREGATE LOCAL EXPLANATION: ENGINE POWER EXAMPLE



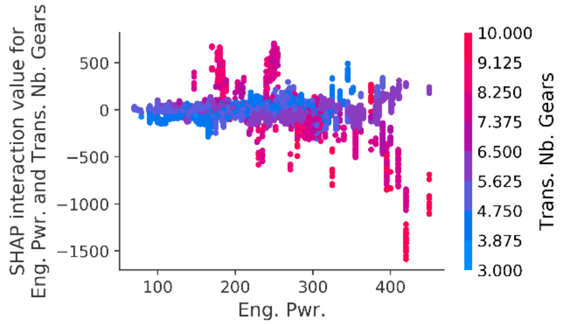
=



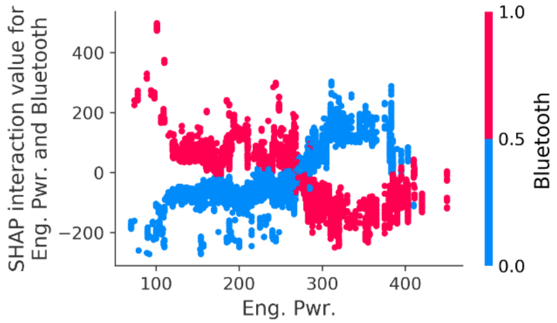
+



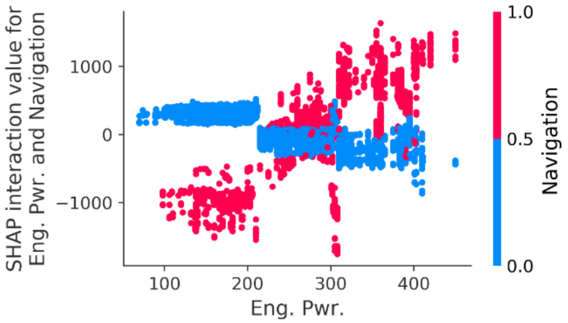
+



+



+



Each point is a vehicle. This form of relationship shows how a feature attribution changes as the feature value varies.

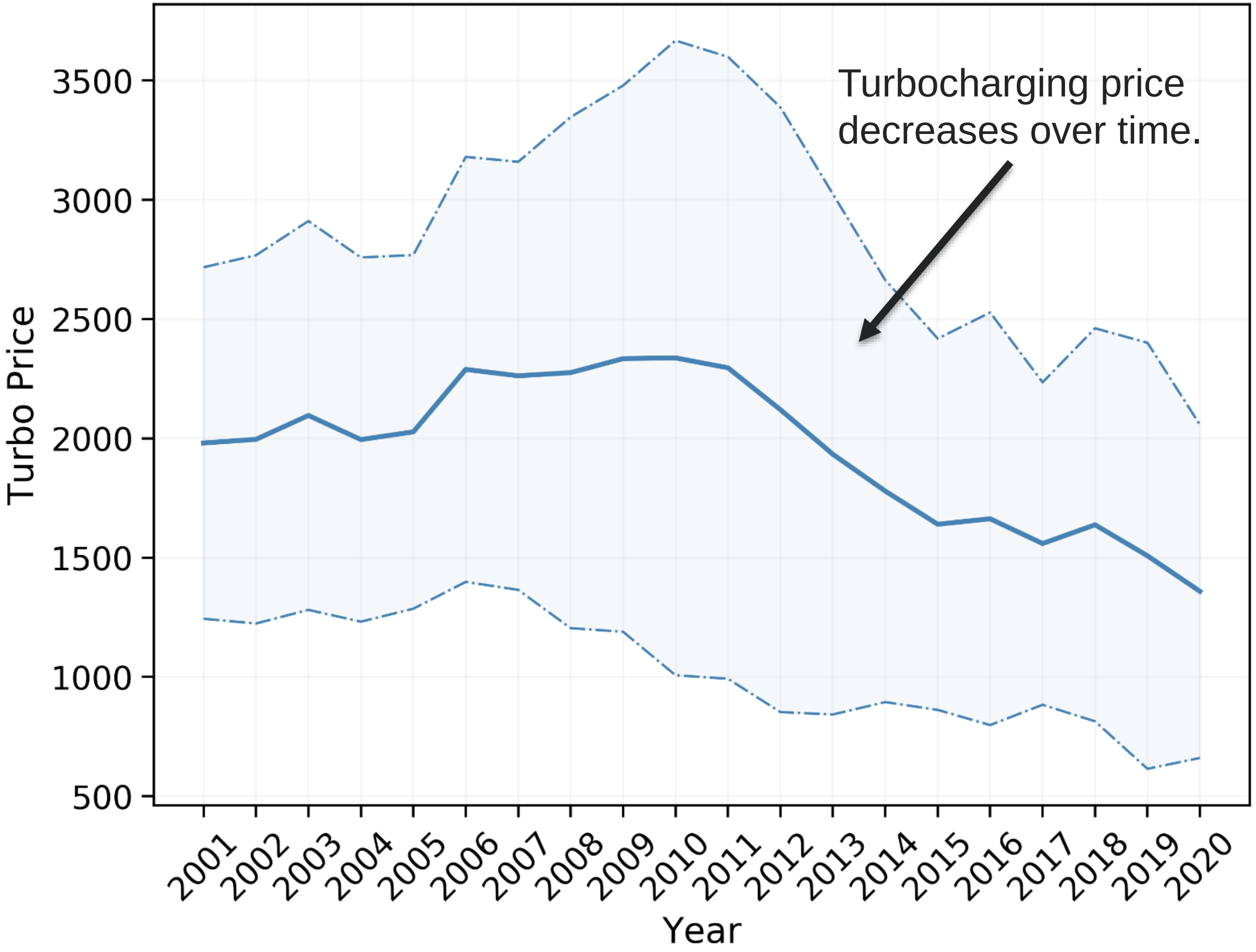
1. We can extract:
   - Marginal effect cost equations (right)
   - Combined effect cost equations (left: includes interactions)
2. **Not restricted to simple linear relationships.**
3. **Not restricted to parametric equations.**

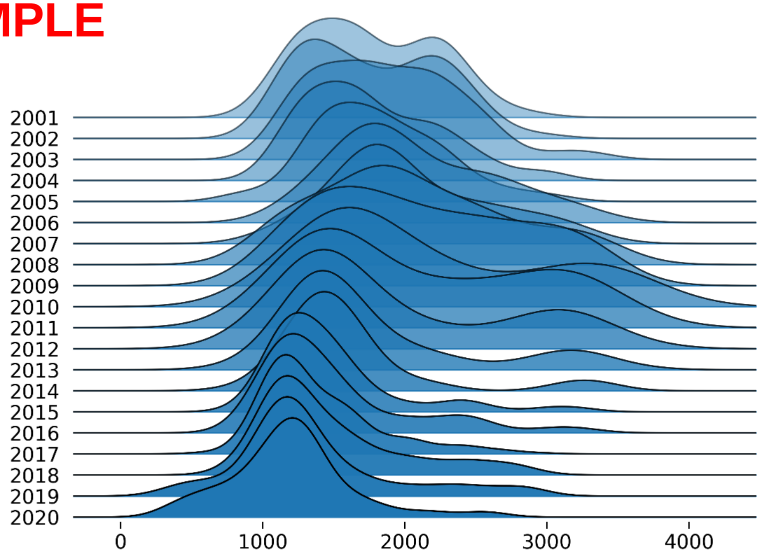
Note: SHAP = Change in price from reference

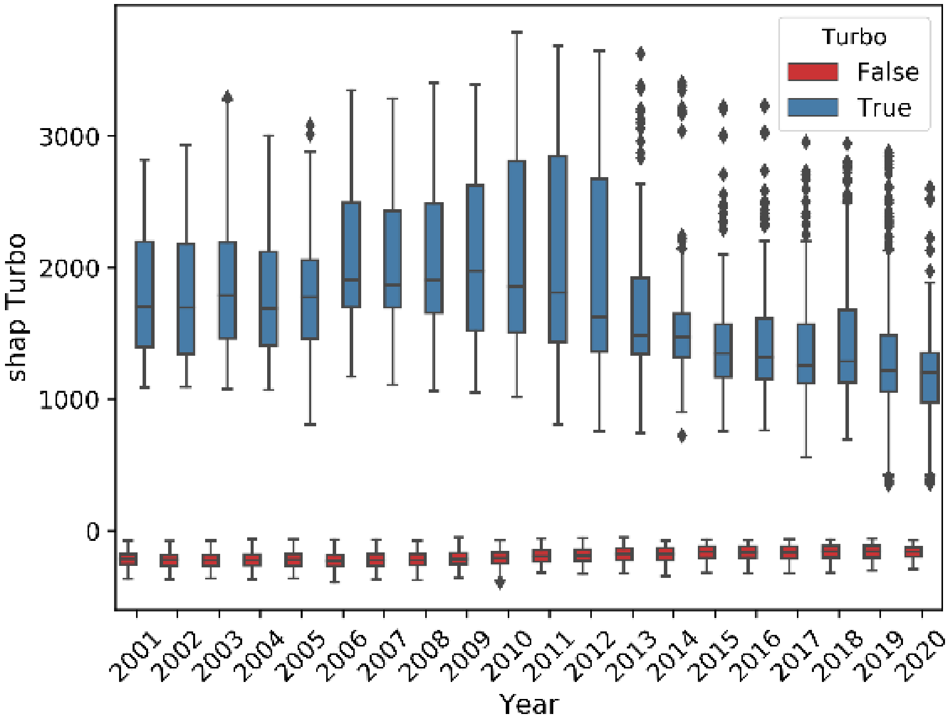
Argonne NATIONAL LABORATORY

INDIVIDUAL TECHNOLOGY PRICE CAN BE ASSESSED

EFFECT OF TIME ON COMPONENT PRICE: TURBOCHARGING EXAMPLE

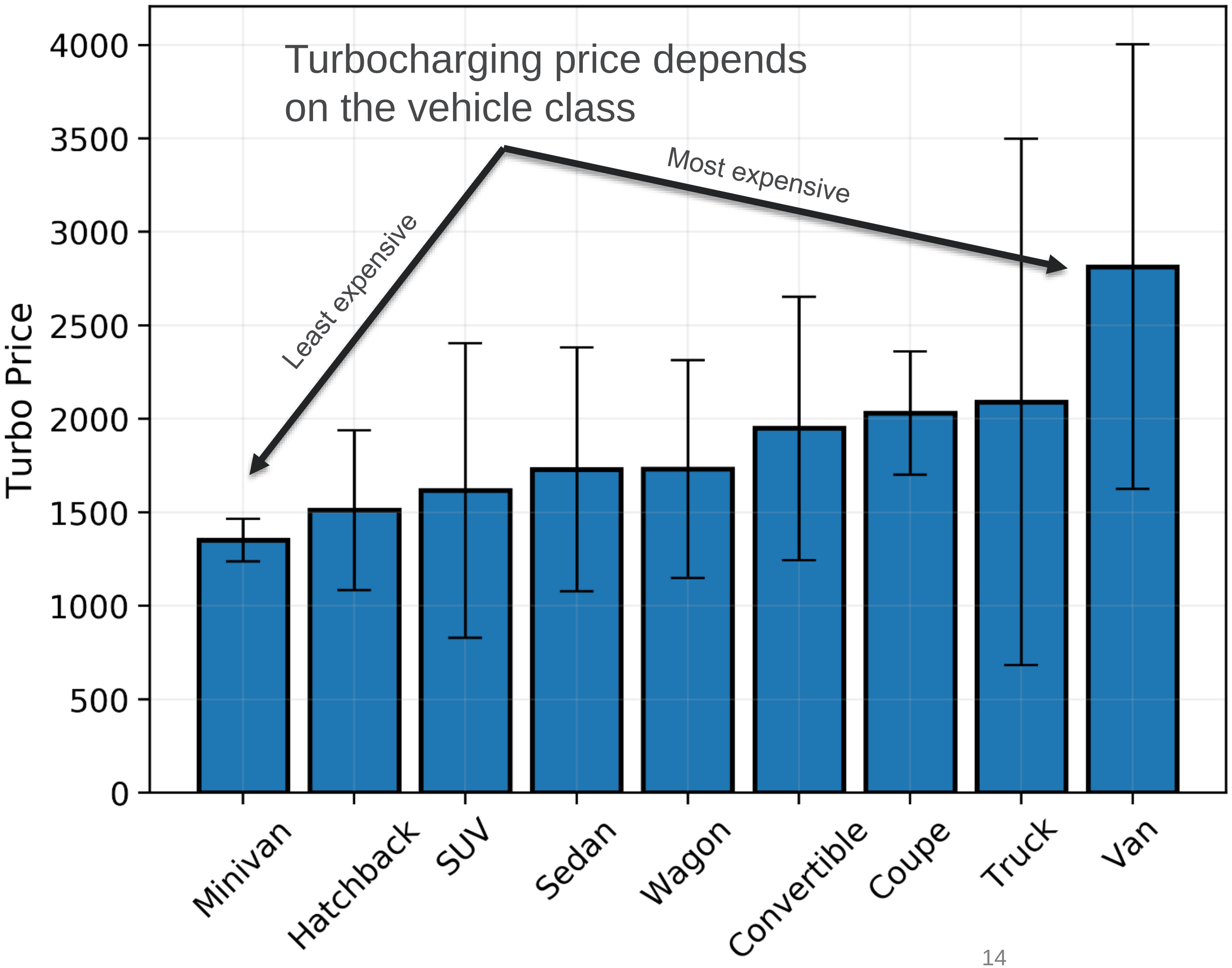


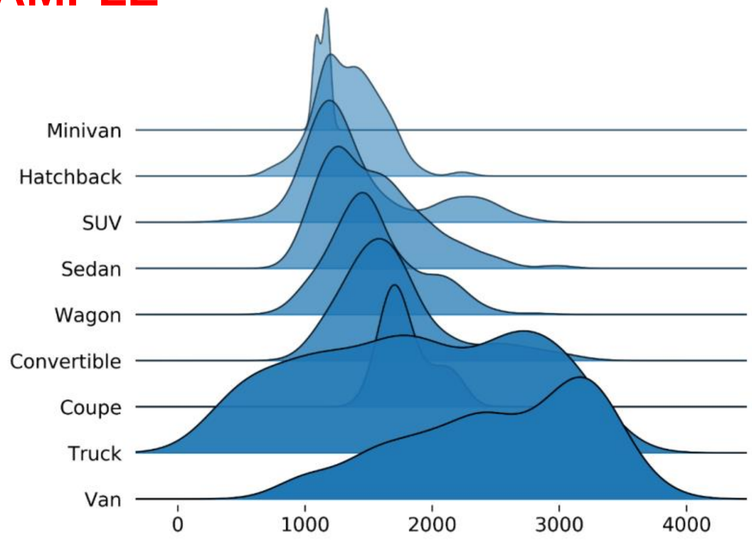


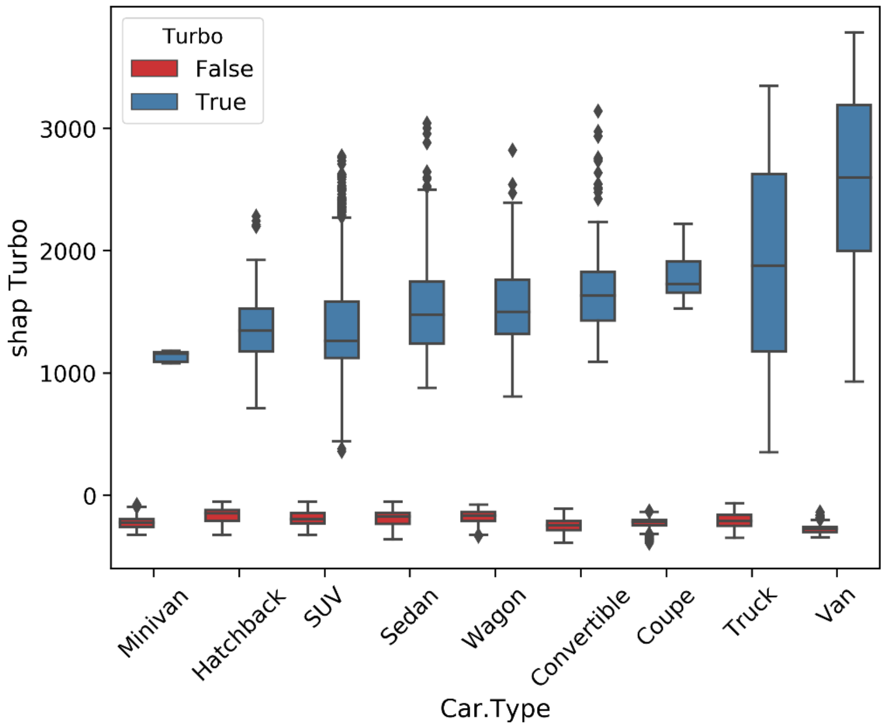


INDIVIDUAL TECHNOLOGY PRICE CAN BE ASSESSED

EFFECT OF CLASS ON COMPONENT PRICE: TURBOCHARGING EXAMPLE







REMAINING CHALLENGES AND BARRIERS

- Very large number of component technologies and attributes => Need to focus on the critical ones.
- Verify / complete / expand database (check all vehicle characteristics, add new model years, new vehicles…).
- Limited number of HEV, PHEV and BEV vehicles.
- Lack of component **pricing** data: need for cost expert validation.
- Need to quantify the uncertainty in estimated attributions (e.g. Confidence Intervals):
  – We have theoretical guarantees for fairness and optimality of split of cost attribution between components, but the uncertainty implicit in the method’s outputs has not been addressed.

POTENTIAL FUTURE RESEARCH

- **Implement methodology into Autonomie/Amber framework** for future VTO related benefits analysis efforts. Since Autonomie relies on manufacturing cost with constant RPE vs. MSRP contribution for ML analysis, tow methods could be considered:
  1. Equation Based
     - Preserve current Autonomie method and derive updated parametric equations or non parametric relationships for each component.
     - Implement independent component prices at the MSRP level (including direct and indirect costs).
  2. Shapley Based Credit/Penalty Component Pricing
     - Use the current predictive model to estimate vehicle price and then generate the (Shapley) attributional values to extract for each component a price contribution
     - A vehicle component price will dependent upon the presence of other components and their feature values. This approach is closest to what has been observed in the data.
     - No need for RPE or ICM adjustment.
- **New analysis**:
  - Study $/mile estimates at the vehicle technology and component levels.
  - Explore tradeoffs between the introduction of more efficient vehicle technologies or more efficient component technologies—and the added price.
  - Connect existing database with sales data to better understand vehicle level, technology level and component level $/mile estimates and the technology's value to the customer.

SUMMARY

- A new vehicle technology database was created with more than 500 individual vehicle attributes for each vehicle over the past 30 years.

- A predictive model with satisfactory accuracy was developed to estimate:
  – Vehicle MSRP
  – Individual component technology price contribution, their evolution over time and across vehicle classes
  – Individual market level component prices

- Potential future work will focus on
  – Integrating the methodology in Autonomie
  – Expanding the analysis use cases

REVIEWER ONLY SLIDES

Publications

Reports submitted to DOE

- A.Moawad, E.Islam, N.Kim, R.Vijayagopal, A.Rousseau, “Vehicle Manufacturer's Suggested Retail Price (MSRP) Estimation using Machine Learning”.

Conferences & Journals

- A.Moawad, E.Islam, N.Kim, R.Vijayagopal, A.Rousseau, W.Wu., “Explainable AI for a No-Teardown Vehicle Component Cost Estimation: A Top Down Approach” *to appear*.