From econometrics to machine learning:
Insights from transportation studies

Benjamin Cottreau
bencottreau@proton.me

Academic and Professional Background

- 2012-2014: Biology, Mathematics
- 2014-2018: Engineering School - Transport Economics
- 2018-2021: Consultant for transportation projects
- 2021-2025: PhD Student in Economics

My main research interest lies in short-term forecasting of economic demand under disturbing events. Applied to transportation, this aims to optimize transport infrastructures and understand "erratic" mobility behaviours.

Outline

A. **Economic theory**

Heavily relies on Ben Akiva (1985) and Bonnel (2002)

A1. The concept of Utility

- **Utility** measures satisfaction. It can be described as a function U with the following (non-exhaustive) hypotheses:

  - Decision makers are *economically rational*.

    Ex. Transitivity - For 3 different quantities of goods q_1, q_2 and q_3 if $U(q_1) > U(q_2)$ and $U(q_2) > U(q_3)$, then $U(q_1) > U(q_3)$

  - Utility increases and marginal utility decreases (*diminishing returns*)

$$\frac{dU}{dq} \geq 0 \quad \text{and} \quad \frac{d^2U}{dq^2} \leq 0 \tag{1}$$

A2. Consumer theory

- The consumer theory focuses on the quantities of goods $q_1, ..., q_n$ with prices $p_1, ..., p_n$.
- Under a constraint of income I, the consumer wants to maximize its utility such that

$$max_{q_i}[U(q_1, ..., q_n)] \tag{2}$$

$$\sum_{i=1}^{n} p_i q_i = I \tag{3}$$

- This is an optimization problem, for which we can find optimal quantities $q_1^*, ..., q_n^*$. **We are interested in finding explanations**, i.e. <u>why</u> would we reach a certain optimum.

A2. Consumer theory

- Resolution in a simple case with 2 goods q_1 and q_2:
  - We assume the utility function has a multiplicative form (Cobb-Douglas)

$$U(q_1, q_2) = \beta_0 q_1^{\beta_1} q_2^{\beta_2} \tag{4}$$

  - Under the constraint of income (2), we can maximize the utility function using the Lagrangian multiplier. We obtain:

$$q_1^* = \frac{\beta_1}{\beta_1 + \beta_2} \cdot \frac{I}{p_1} \qquad q_2^* = \frac{\beta_2}{\beta_1 + \beta_2} \cdot \frac{I}{p_2} \tag{5}$$

  - With enough data (surveys) on prices, quantities and income levels, we can retrieve coefficients β_1 and β_2.

A3. Discrete choice theory

- Two goals:
  - When we isolate economic problems by field (Strotz, 1959), we no longer chose quantities of several different goods, but **we pick up one good among similar competing goods** (discrete choice).
  - We want to get rid of inconsistencies due to **transitivity issues**, that might occur often with discrete choices.

- Therefore, we introduce new utility function with a deterministic component and a random component. For decision maker n and choice i:

$$U_{in} = V_i n + \epsilon_i n \tag{6}$$

A3. Discrete choice theory

- Let's consider a binary choice set $C = \{i, j\}$. The deterministic component is defined as a linear combination of K variables x_k ($x \in \mathbb{R}, k \in [\![1, K]\!], K \in \mathbb{N}$). Utility functions are computed as:

$$U_{in} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{k_{in}} + \epsilon_{in} \quad \text{and} \quad U_{jn} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{k_{jn}} + \epsilon_{jn} \tag{7}$$

- For decision maker n, the probability to make choice i is:

$$\begin{aligned} P_n(i) &= P(U_{in} \geq U_{jn}) \\ &= P(\underbrace{\epsilon_{jn} - \epsilon_{in}}_{\epsilon_n} \leq V_{in} - V_{jn}) \end{aligned} \tag{8}$$

- We assume ϵ_n follows a logistic distribution

$$P_n(i) = \frac{1}{1 + e^{-\sum_{k=1}^{K} \beta_k (x_{k_{in}} - x_{k_{jn}})}} \tag{9}$$

A3. Discrete choice theory

- We find $\beta_1, ..., \beta_K$ through Maximum Likelihood Estimation. Let's define the variable

$$y_{in} = \begin{cases} 1, & \text{if decision maker } n \text{ chose } i \\ 0, & \text{if decision maker } n \text{ chose } j \end{cases} \tag{10}$$

- We compute the likelihood function

$$\mathcal{L}(\beta_1, ..., \beta_K) = \prod_{n=1}^{N} P_n(i)^{y_{in}}[1 - P_n(i)]^{1-y_{in}} \tag{11}$$

- We linearize and seek estimates $\hat{\beta}_1, ..., \hat{\beta}_K$ by maximizing the likelihood function. This can be hard to solve, so we approximate it numerically using the Newton-Raphson algorithm.

B. Transportation problems

B1. Application of discrete choice modeling to transportation problems

- **Mode choice** - few possibilities; uses socio-economic attributes and eventually land-use attributes; purpose of trip; costs...
- **Route choice** - lots of possibilities (routes of a network) ; independence issues; uses access/egress time, in-vehicle time; costs; purpose of trip; congestion...

B2. Limitations

- **Time focus**: This econometric modeling framework is not really suitable for short term analysis (e.g. real time optimization). Research at this level relies on traffic theory.
- **Data**: Stated preferences (survey) vs. Revealed preferences (passive data), especially to gather data on abnornal events.
- **Flexibility**: Econometric tools require to specify the modeling framework beforehand, which might not be ideal in the event of disruptions.

C. Machine Learning

C1. Supervised learning for binary classification problems

- Let's consider the binary choice set $C = \{i, j\}$.
  - We have:

$$y_{in} = \begin{cases} 1, & \text{if decision maker } n \text{ chose } i \\ 0, & \text{if decision maker } n \text{ chose } j \end{cases} \tag{12}$$

  - Given a set of explanatory variables $X = \{X_k\}_{k \in [1,K]}$, ML models can be used to find a mapping h that will estimate new data points based on the observed distribution of the couple (y, X). These estimated values are called $\widehat{y} = h(X)$.
  - The mapping h is found thanks to a loss function $\mathcal{L}(\widehat{y}, y)$, that we want to minimize.
- The main advantage is that we don't have to specify any utility function
- But sadly, we don't have βs! :(

C2. Shapley additive explaination (SHAP)

The SHAP theory is based on game theory, where *"a prediction can be viewed as a coalitional game by considering each [variable] value of an instance as a player in a game"* (Molnar, 2023)

- **Intuition** - Max and Alice are treasure hunters. Together, they find 100 gold coins with the following distribution :
  - Alice finds 60 gold coins alone
  - Max finds 30 gold coins alone
  - Together, they find 10 gold coins
- We compute the following value functions:
  - $v(\{\}) = 0$
  - $v(\{\text{Alice}\}) = 60$
  - $v(\{\text{Max}\}) = 30$
  - $v(\{\text{Alice}, \text{Max}\}) = 100$

C2. Shapley additive explaination (SHAP)

- We compute marginal contributions of each player:

  - **Alice's Marginal Contribution**

    Alice with Max: $v(\{\text{Alice}, \text{Max}\}) - v(\{\text{Max}\}) = 100 - 30 = 70$
    Alice alone: $v(\{\text{Alice}\}) = 60$

  - **Max's Marginal Contribution**

    Max with Alice: $v(\{\text{Alice}, \text{Max}\}) - v(\{\text{Alice}\}) = 100 - 60 = 40$
    Max alone: $v(\{\text{Max}\}) = 30$

- We compute the average marginal contribution ϕ of each player to all possible coalitions:

$$\phi(\text{Alice}) = \frac{70 + 60}{2} = 65 \qquad \phi(\text{Max}) = \frac{40 + 30}{2} = 35 \tag{13}$$

C2. Shapley additive explaination (SHAP)

- SHAP can be used as a post-hoc and model-agnostic technique to retrieve the marginal effects of variables X on output y.
- Tree-SHAP computes exact values when used in combination with tree-based models (Random Forest, Gradient Boosting)
- SHAP deals with interactions between variables
- SHAP has a nice additive property
- for more info on SHAP: Lundberg and Lee (2017) and Molnar (2023)

D. **A Funkier Example**

Based on Cottreau et al. (2025)

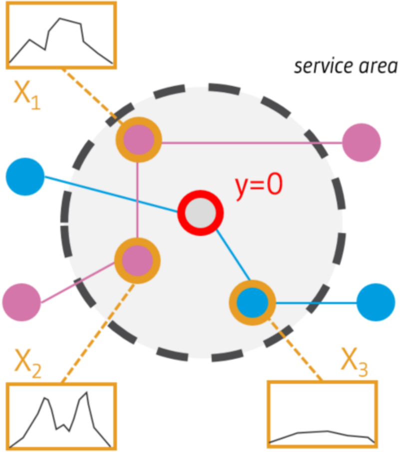
D1. Binary classification setting

- Let's consider a subway station which status is defined by the binary variable y at each time interval t, such that:

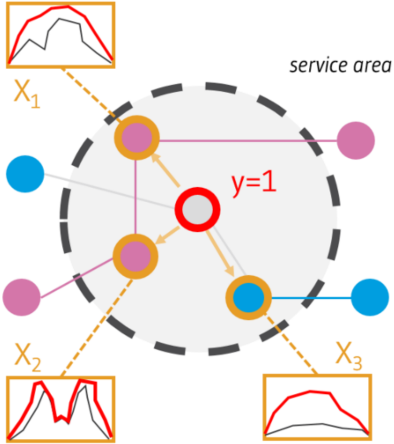
$$y_t = \begin{cases} 1, & \text{if the subway station experiences a disruption at time } t \\ 0, & \text{if not} \end{cases} \tag{14}$$

- Let's define the variable $x_{k,t}$ that captures the demand levels at stop k at each time t. Stop k is selected based on its proximity to the subway station ($<$ 600m)
- We seek a mapping h that will estimate the occurrence of disruptions based on demand levels $x_{k,t}$. These estimated values are called $\widehat{y} = h(x_{k,t})$. **[Random Forest]**
- We also want to know the contribution of each stop k in the estimation process. **[SHAP]**

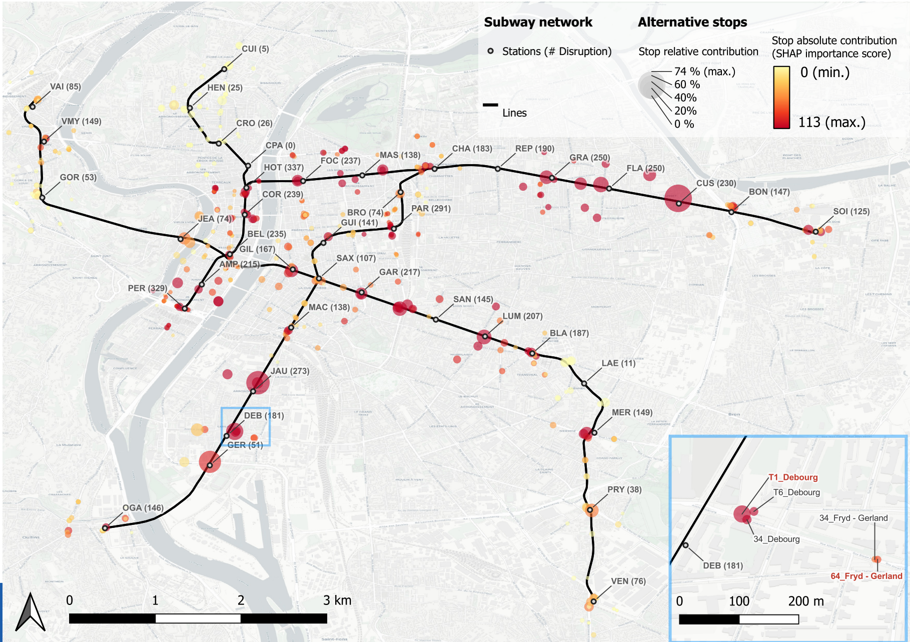
D1. Binary classification setting



D1. Binary classification setting

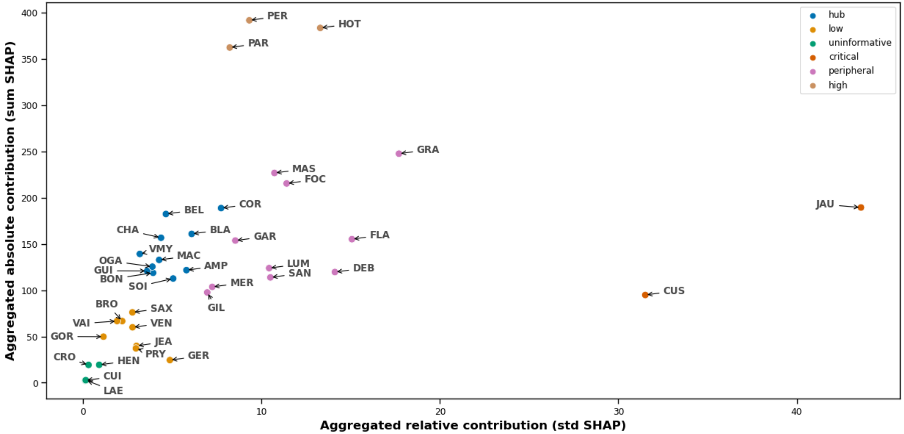


D2. Results



Subway network
Stations (# Disruption)
Lines
Alternative stops
Stop relative contribution
74 % (max.)
60 %
40%
20%
0 %
Stop absolute contribution (SHAP importance score)
0 (min.)
113 (max.)
CUI (5)
VAI (85)
HEN (25)
VMY (149)
CRO (26)
CPA (0)
HOT (337)
FOC (237)
MAS (138)
CHA (183)
REP (190)
GRA (250)
FLA (250)
CUS (230)
BON (147)
SOI (125)
GOR (53)
COR (239)
BRO (74)
GUI (141)
PAR (291)
JEA (74)
BEL (235)
GIL (167)
AMP (215)
SAX (107)
GAR (217)
PER (329)
SAN (145)
MAC (138)
LUM (207)
BLA (187)
LAE (11)
JAU (273)
DEB (181)
GER (51)
MER (149)
PRY (38)
OGA (146)
VEN (76)
0
1
2
3 km
T1_Debourg
T6_Debourg
34_Fryd - Gerland
34_Debourg
DEB (181)
64_Fryd - Gerland
0
100
200 m

D2.Results



PER
HOT
PAR
GRA
MAS
FOC
COR
BEL
JAU
CHA
BLA
GAR
FLA
VMY
MAC
OGA
GUI
BON
AMP
LUM
SAN
DEB
SOI
MER
GIL
CUS
BRO
SAX
VAI
VEN
GOR
JEA
PRY
GER
CRO
HEN
CUI
LAE
hub
low
uninformative
critical
peripheral
high
Aggregated absolute contribution (sum SHAP)
Aggregated relative contribution (std SHAP)
0
50
100
150
200
250
300
350
400
0
10
20
30
40

References

Ben-Akiva, M., Lerman, S. R. (1985). Discrete choice analysis: Theory and application to travel demand. MIT Press.

Bonnel, P. (2002). Prévision de la demande de transport. Economies et finances. Université Lumière - Lyon II. tel-00268919

Cottreau, B., Celbiș, M.G., Manout, O., & Bouzouina, L. (2025). Detection of subway service disruptions and contribution of alternative stops to public transit resilience. Transportation Research Part A: Policy & Practice, Volume 199, 2025, 104550, ISSN 0965-8564, https://doi.org/10.1016/j.tra.2025.104550

Lundberg, S., Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. arXiv. Retrieved 2024-10-21, from https://arxiv.org/abs/1705.07874 (Version Number: 2) doi:10.48550/ARXIV.1705.07874

Molnar, C. (2023). Interpreting machine learning models with SAP: a guide with Python examples and theory on shapley values (First Edition ed.). München, Germany: Chistoph Molnar c/o MUCBOOK, Heidi Seibold.

Strotz, R. (1975). The Empirical Implications of a Utility Tree. Econometrics, 25: 269-280.

Thanks!

Benjamin Cottreau
bencottreau@proton.me