



Discrete Response: The Logit Model

Advanced Microeconometrics

Anders Munk-Nielsen
2022



Plan for lectures: Helicopter

Part I: Linear methods. ✓

Part II: High-dimensional methods. ✓

Part III: M-estimation, theory ✓

Part IV: M-estimation, examples ←

Where are we in the course?

Part	Topic	Parameterization non-linear	Estimation non-linear	Dimension $\dim(x)$	Numerical optimization	M-estimation (Part III)	Outcome (y_i)	Panel (c_i)
I	OLS	÷	÷	low	÷	✓	\mathbb{R}	✓
II	LASSO	÷	✓	high	✓	÷	\mathbb{R}	÷
IV	Probit	✓	✓	low	✓	✓	$\{0, 1\}$	÷
	Logit	✓	✓	low	✓	✓	$\{1, 2, \dots, J\}$	÷
	Tobit	✓	✓	low	✓	✓	$[0; \infty)$	÷
	Sample selection	✓	✓	low	✓	✓	\mathbb{R} and $\{0, 1\}$	÷
	Simulated Likelihood	✓	✓	low	✓	✓	Any	✓
	Quantile Regression	÷	✓	(low)	✓	✓	\mathbb{R}	÷
	Non-parametric	✓	(✓)	∞	÷	÷	\mathbb{R}	÷

1. Overview
2. Random Utility Model
3. Criterion Function
 - 3.1. Identification
4. Features of Interest
5. Specific Issues
 - 5.1. The IIA Restriction
 - 5.2. Market Share Data

Big Big Picture: Modeling Approaches

- **Simplest:** with $q_{jt} = \#$ sold cars and \mathbf{x}_{jt} = car attributes (price, horsepower, km/l, etc.)

$$q_{jt} = \mathbf{x}_{jt}\beta + \text{error}_{jt}$$

- **Con:** Competitor prices / selection does not affect sales.
E.g. zero cross-price elasticity. NB! Of course it matters what my competitor is doing with their prices to demand of my car
- **Big OLS:** Include prices of *all* cars in explaining j 's sales

$$q_{jt} = \mathbf{x}_{jt}\beta + \sum_{k \neq j} \mathbf{x}_{kt}\gamma_{jk} + \text{error}_{jt}.$$

'Almost ideal demand system' (AIDS)

- **Cons:**
 - many parameters (e.g. J^2 cross-price coefficients!!), J -cars?
 - **Too flexible?** Might violate basic logic (Slutsky symmetry, adding up, etc.)
if you don't have much variation in \mathbf{x} 's, then it might be that you impose some basic logical restrictions on how demand works -> slutsky symmetry
- **Logit:**

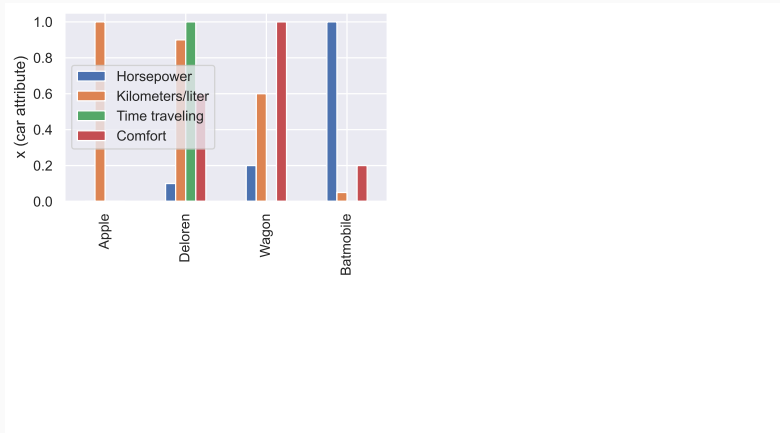
$$\frac{q_{jt}}{\sum_{k=1}^J q_{kt}} = \frac{\exp(\mathbf{x}_{jt}\beta)}{\sum_{k=1}^J \exp(\mathbf{x}_{kt}\beta)}.$$

Essentially, it maps a market share of a j car

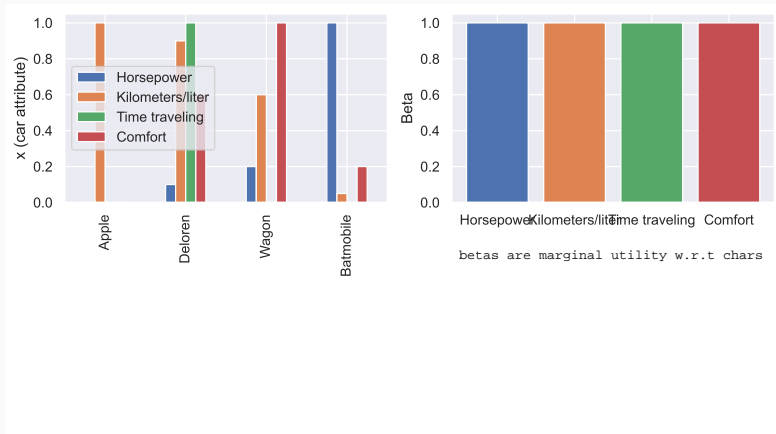
- **Pros:** Few parameters ($\dim \beta = K$), cross-product “market stealing” (substitution)
- **Cons:** Restricted substitution patterns (the “IIA property”)

- **Choices:** four different cars.
- **Preferences:** $u^{\text{bat}} \succ u^{\text{Delorean}} \succ u^{\text{wagon}} \succ u^{\text{apple}}$.
- **Data:**
 - Characteristics of cars, x_j (e.g. price, horsepower, flux-capacitor dummy, ...)
 - Characteristics of decision maker, x_i (e.g. income, work distance).
 - The final decision (which e.g. says $u^{\text{batmobile}} \succ u^j$ for all other j)
- **Model:** A *utility function*, $u : (\mathbf{x}_{ij}, \theta) \mapsto \mathbb{R}$.
 - **Goal:** Choose θ st. $u(\mathbf{x}_{i\text{bat}}, \theta) > u(\mathbf{x}_{ij}, \theta)$ for $j \neq \text{bat}$.

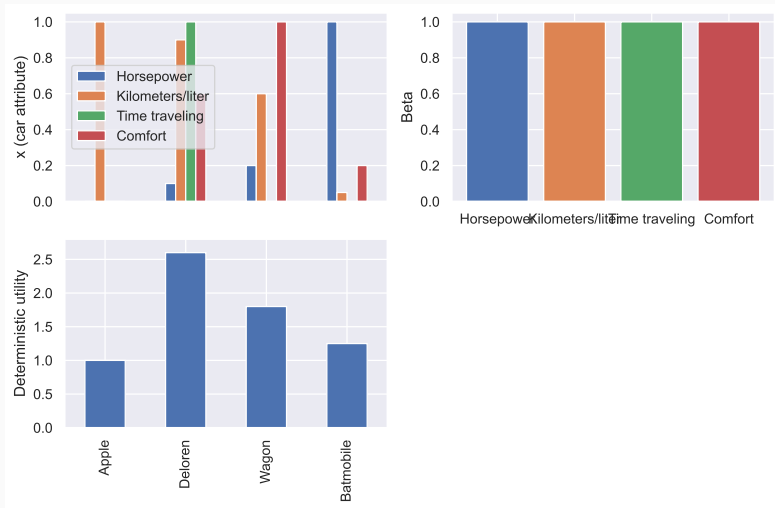
Logit intuition



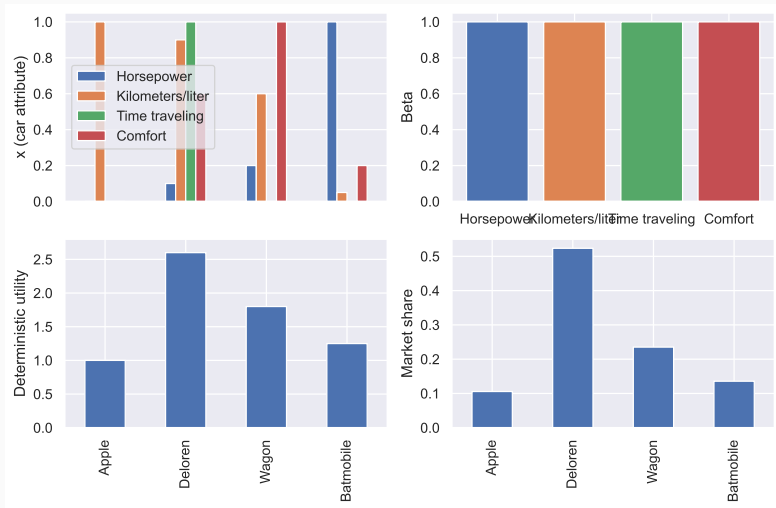
Logit intuition



Logit intuition

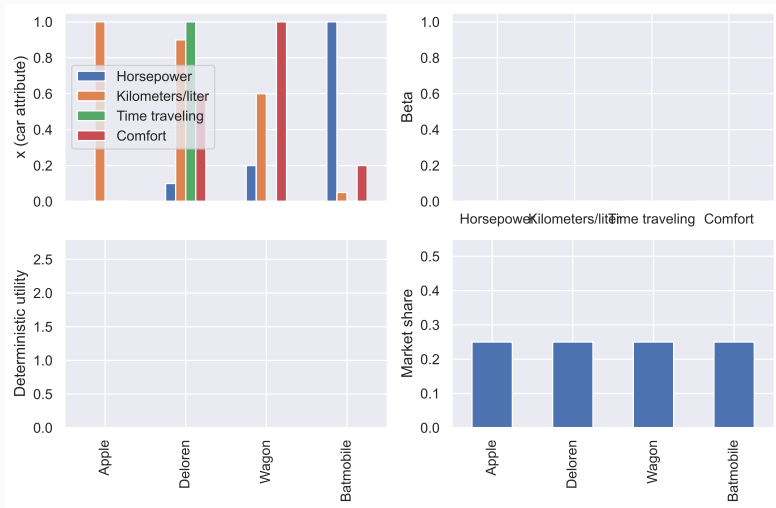


Logit intuition



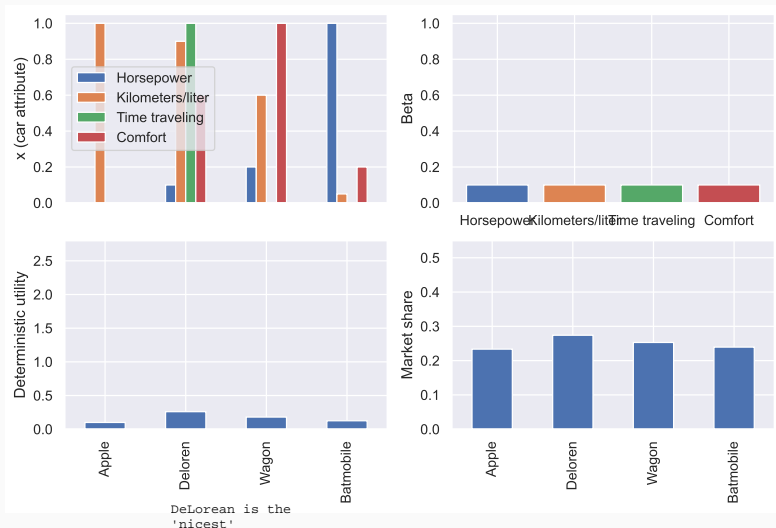
Logit intuition: $\beta = 0$

what if the decision makers do not care about any of the characteristics?



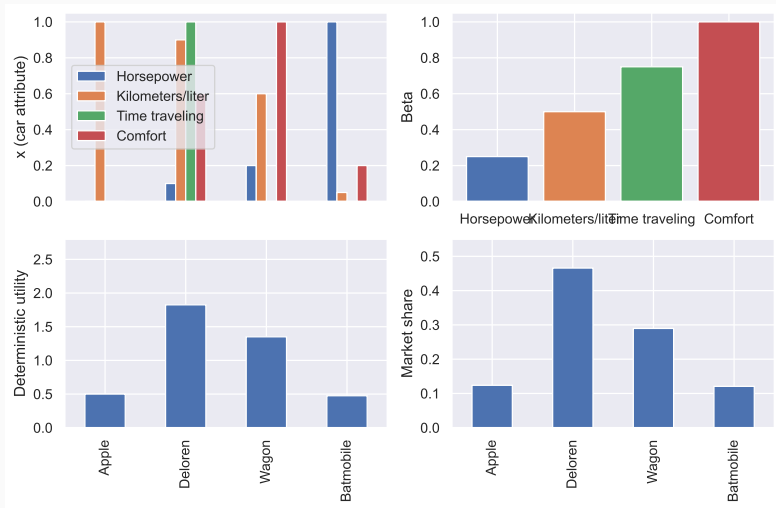
No utility from any characteristics \Rightarrow identical market shares.

Logit intuition: β “small”



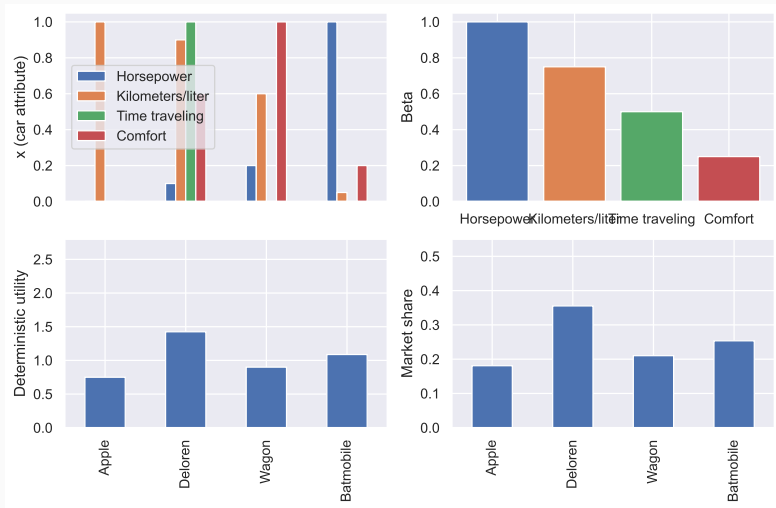
Here, $\beta = (\frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10})$

Logit intuition



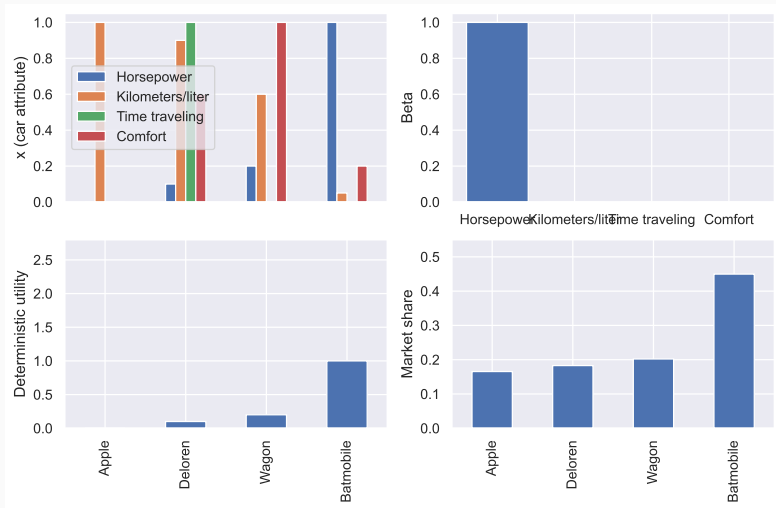
Consumers prefer comfort and time traveling capabilities

Logit intuition



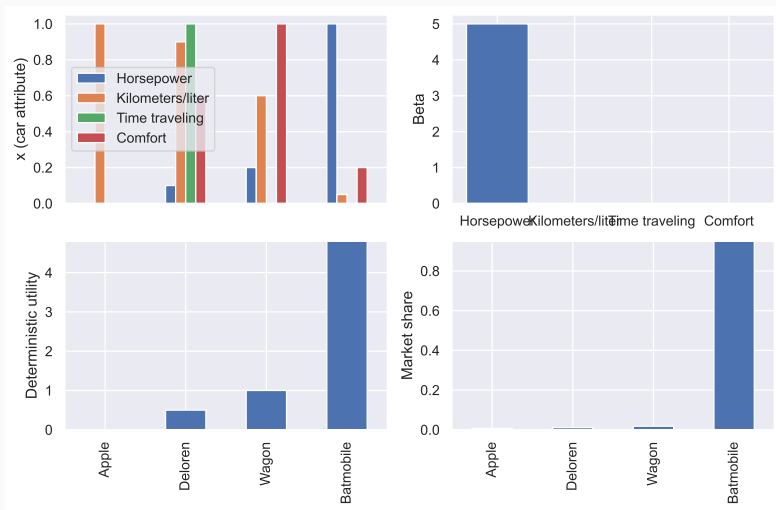
Consumers prefer horsepower and fuel economy.

Logit intuition: $\beta = (1, 0, 0, 0)$



Consumers only care about horsepower (and idiosyncracities)

Logit intuition: $\beta = (5, 0, 0, 0)$



Consumers only care about horsepower, and by a lot!

Overall Intuition: Identification?

Model, 1st idea

For cars $j = 1, \dots, J$,

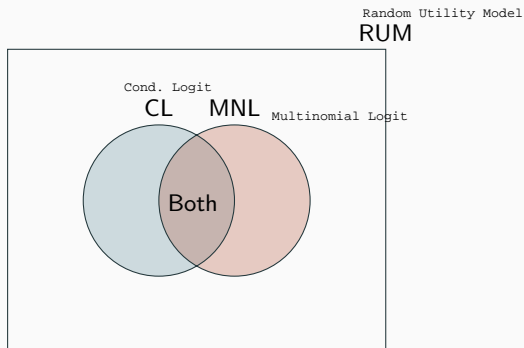
$$u_{ij} = \beta_0 + \beta_1 \text{horsepower} + \beta_2 \text{income} + \text{error}, \quad \text{error} \sim \text{IID}(\mu, \sigma^2).$$

Discuss

What can we hope to identify here?

No way we can identify μ and β_0 -> we can't have two constants

Have to normalize by σ^2



Random Utility Model (RUM)

latent, unobservable
utility

$$u_{ij} = v_{ij} + \varepsilon_{ij},$$

$$y_i = \operatorname{argmax}_{j \in \{1, \dots, J\}} u_{ij}.$$

i : Individual, $i = 1, \dots, N$,

j : Alternative, $j = 1, \dots, J$,

y_i : Chosen alternative by i ,

u_{ij} : Utility from choosing alternative j ,

ε_{ij} : Unobserved error term,

v_{ij} : Observed part (regressors and parameters).

- **Primary** two types of RUM:

1. **Conditional logit:** $v_{ij} = \mathbf{x}_j \beta$, varies over cars J
2. **Multinomial logit:** $v_{ij} = \mathbf{x}_i \beta_j$, varies over decision makers i

Random Utility Model

$$\begin{aligned}u_{ij} &= v_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I}, \\y_i &= \operatorname{argmax}_{j \in \{1, \dots, J\}} u_{ij}.\end{aligned}$$

- **Conditional choice probabilities (CCPs):** to be shown...

$$\Pr(d|v_{ij}) = \frac{\exp(v_{id})}{\sum_{j=1}^J \exp(v_{ij})}.$$

- due to Dan McFadden.

Conditional logit (CL)

- \mathbf{x}_j : characteristics of the *alternatives*.
- **Example**: car choice, $\mathbf{x}_j = (\text{price, horsepower, ...})$.
- **Normalization**: No intercept in \mathbf{x} .
- β_k **interpretation**: marginal utility of characteristic k .

Multinomial logit (MNL)

- \mathbf{x}_i : characteristics of the *individual*.
- **Example**: car choice, $\mathbf{x}_i = (\text{income, work distance, ...})$.
- **Normalization**: $\beta_1 = \mathbf{0}_{K \times 1}$.
- β_{jk} **interpretation**: change in utility of car j (relative to car 1) when (e.g.) income increases (if \mathbf{x}_{ik} is income).

1. Overview
2. Random Utility Model
3. Criterion Function
 - 3.1. Identification
4. Features of Interest
5. Specific Issues
 - 5.1. The IIA Restriction
 - 5.2. Market Share Data

Random Utility Model

$$u_{ij} = v(\mathbf{x}_{ij}, \theta) + \varepsilon_{ij}, \quad \text{IID Extreme Value Type I,}$$

$$y_i = \operatorname{argmax}_{j \in \{1, \dots, J\}} u_{ij}.$$

$v(\cdot, \cdot)$: Known functional form,

\mathbf{x}_{ij} : Data (characteristics of individuals and/or alternatives),

- **Functional form:** CL and MNL are two *linear* forms of v
 - **CL:** $v(\mathbf{x}_{ij}, \beta) = \mathbf{x}_{ij}'\beta$
 - Note: no intercept in \mathbf{x}_{ij} (intercept is normalized)
 - **MNL:** $v(\mathbf{x}_i, \beta_j) = \mathbf{x}_i'\beta_j$.
 - Note: $\theta = (\beta_2, \dots, \beta_J)$ ($\beta_1 := 0$ is normalized)

Multinomial Logit

$$u_{ij} = \mathbf{x}_i \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I}, \quad \beta_1 = \mathbf{0}_{K \times 1},$$
$$y_i = \operatorname{argmax}_{j \in \{1, \dots, J\}} u_{ij}.$$

```
1 from scipy.stats import genextreme # generalized EV distr.
2 def sim_data(theta, N):
3     K, J_1 = theta.shape # theta must be (K, J-1)
4     J = J_1 + 1
5     xx = np.random.normal(size=(N,K-1))
6     oo = np.ones((N,1)) # constant term
7     x = np.hstack([oo,xx]) # full x matrix
8     zz = np.zeros((K,1)) # normalized coefficients
9     beta = np.hstack([zz, theta]) # full (K,J) matrix
10    v = x @ beta # (N,J): observable utility
11    uni = np.random.uniform(size=(N,J)) # uniform draws
12    e = genextreme.ppf(uni, c=0) # generalized extreme value errors
13    u = v + e # full (unobserved) utility
14    y = u.argmax(axis=1) # observed, chosen alternative
15    return v, y
```


Conditional Logit

$$u_{ij} = \mathbf{x}'_{ij}\beta + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I},$$
$$y_i = \operatorname{argmax}_{j \in \{1, \dots, J\}} u_{ij}.$$

```
1 def sim_data(theta, N, J):    \theta = how many characteristics
                                N = no obs
                                J = alternatives
                                NO CONSTANT TERM
2     K = theta.size
3     assert theta.ndim == 1 # theta should be (K,)
4     x = np.random.normal(size=(N,J,K)) # Explanatory variables
5     uni = np.random.uniform(size=(N,J))
6     e = genextreme.ppf(uni, c=0) # generalized extreme value
7     v = x @ theta # (N,J) matrix of "observable utilities"
8     u = v + e # full utility (unobserved)
9     y = u.argmax(axis=1) # observed chosen alternative
10    return y,x
```

1. Overview
2. Random Utility Model
3. Criterion Function
 - 3.1. Identification
4. Features of Interest
5. Specific Issues
 - 5.1. The IIA Restriction
 - 5.2. Market Share Data

- **Goal:** Derive the likelihood function.
- **How** do we begin?
- **Normally:** we always try to isolate the error terms..
- **This time:** there are J of them and only one y_i is observed...

Conditional Logit Model

$$u_{ij} = \mathbf{x}_j \beta_o + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Extreme Value Type I.}$$

$$y_i = \operatorname{argmax}_{j \in \{1, \dots, J\}} u_{ij}.$$

- Choice probability:

$$\begin{aligned} \Pr(y = j | \mathbf{X}) &= \Pr(u_{ij} > u_{ik} \forall k \neq j | \mathbf{X}) \\ &= \Pr(\mathbf{x}_j \beta_o + \varepsilon_{ij} > \mathbf{x}_k \beta_o + \varepsilon_{ik} \forall k \neq j | \mathbf{X}) \\ &= \Pr(\varepsilon_{ik} < \mathbf{x}_j \beta_o - \mathbf{x}_k \beta_o + \varepsilon_{ij} \forall k \neq j | \mathbf{X}) \\ &= \dots \\ &= \frac{\exp(\mathbf{x}_j \beta_o)}{\sum_{k=1}^J \exp(\mathbf{x}_k \beta_o)}. \end{aligned}$$

Note: Nothing relied on the precise form of v_{ij} .

Random Utility Model

$$u_{ij} = v_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Extreme Value Type I.}$$

$$y_i = \operatorname{argmax}_{j \in \{1, \dots, J\}} u_{ij}.$$

- **Choice probability:**

$$\begin{aligned} \Pr(y = j | \mathbf{X}) &= \Pr(u_{ij} > u_{ik} \forall k \neq j | \mathbf{X}) \\ &= \Pr(v_{ij} + \varepsilon_{ij} > v_{ik} + \varepsilon_{ik} \forall k \neq j | \mathbf{X}) \\ &= \Pr[\varepsilon_{ik} < v_{ij} - v_{ik} + \varepsilon_{ij} \forall k \neq j | \mathbf{X}] \\ &= \dots \\ &= \frac{\exp(v_{ij})}{\sum_{k=1}^J \exp(v_{ik})}. \end{aligned}$$

We showed

$$\Pr(y_i = j | \mathbf{X}) = \frac{\exp(v_{ij})}{\sum_{k=1}^J \exp(v_{ik})}.$$

Log Likelihood function

$$\ell_i(\boldsymbol{\theta}) = v_{iy_i} - \log \left[\sum_{k=1}^J \exp(v_{ik}) \right].$$

```
1 def util(theta, x):
2     v = x @ theta # (N,J,K) * (K,) = (N,J)
3     return v
4
5 def loglikelihood(theta, y, x):
6     N,J,K = x.shape
7     v = util(theta, x)
8     v_i = v[np.arange(N), y] # (N,) vector
9     denom = np.exp(v).sum(axis=1) # (N,) vector
10    ll_i = v_i - np.log(denom)
11    return ll_i
```

- **Crucial trick:** *Analytically*, the following holds

$$\frac{\exp(v_{ij})}{\sum_{k=1}^J \exp(v_{ik})} = \frac{\exp(v_{ij} - K_i)}{\sum_{k=1}^J \exp(v_{ik} - K_i)}, \quad \forall K_i \in \mathbb{R}.$$

- **Numerically** however, `exp()` quickly becomes highly imprecise...
 - `np.exp(800) = inf`: overflow
 - `np.exp(-800) = 0.0`: underflow
- **Both** types of error are bad...
 - ... but overflow is much worse.
- **Solution:** *Max-rescaling* our utilities

```
1 def util(theta, x):
2     v = x @ theta
3     vmax = v.max(axis=1, keepdims=True) # (N,1)
4     return v - vmax # subtract columns-wise max
```

Conditional Logit

$$u_{ij} = \mathbf{x}_j \beta + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I},$$
$$y_i = \underset{j \in \{1, \dots, J\}}{\operatorname{argmax}} u_{ij}.$$

Discuss

- Why can't we allow for an intercept (constant) in \mathbf{x}_j ?
- Why can't we estimate $\text{Var}(\varepsilon_{ij})$ here?
- How does this relate to standard micro with *ordinal* vs. *cardinal* utility?

Multinomial Logit

$$u_{ij} = \mathbf{x}_i \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I},$$

$$y_i = \underset{j \in \{1, \dots, J\}}{\operatorname{argmax}} u_{ij},$$

where we normalize $\beta_1 := \mathbf{0}_{K \times 1}$.

Discuss

- Why do we *have* to normalize β for one alternative?
- Why can we allow for an intercept in \mathbf{x}_i and estimate β_{j0} for all j except the normalized alternative?
- What is the utility of the normalized alternative?

1. Overview
2. Random Utility Model
3. Criterion Function
 - 3.1. Identification
- 4. Features of Interest**
5. Specific Issues
 - 5.1. The IIA Restriction
 - 5.2. Market Share Data

- **Generally:** We are rarely interested in β s directly.
- **Note:** The average choice probability is the *market share*.
 - so average partial effects say how market shares change.
- **Car example:** Interested in how car choice is affected by x 'es.
 - i.e. $\frac{\partial}{\partial x_{ijk}} \Pr(j|x)$.
 - Standard errors: via the delta method.
- **Conditional logit:** E.g. the effect raising the price of Volvo V70, holding fixed all other car prices

$$\frac{\partial}{\partial x_{kl}} \Pr(j) = \Pr(j) \left[\mathbf{1}_{\{k=j\}} \beta_l - \Pr(l) \beta_l \right].$$

- **Multinomial logit:** E.g. the effect on $\Pr(V70)$ of raising household income,

$$\frac{\partial}{\partial x_k} \Pr(j) = \Pr(j) \left[\beta_{jk} - \sum_{l=1}^J \Pr(l) \beta_{lk} \right].$$

- **Interpreting β_{jp} , warning:** Partial effect is complicated.
 - The sign is not determined by β_{jp} alone!!!
- **Instead:** Consider the log Odds-ratio (and use $\beta_1 = \mathbf{0}$)

$$\log \frac{\Pr(y = j | \mathbf{x}_i)}{\Pr(y = 1 | \mathbf{x}_i)} = \log \frac{\frac{\exp(\mathbf{x}_i \beta_j)}{\sum_{k=1}^J \exp(\mathbf{x}_i \beta_k)}}{\frac{\exp(\mathbf{x}_i \beta_1)}{\sum_{k=1}^J \exp(\mathbf{x}_i \beta_k)}} = \mathbf{x}_i \beta_j.$$

- **Linear!** This is why some software reports (log) odds ratios.

- **Larger choicesets:** If $J > 1000$, partial effects are not interesting.
- **Instead:** Predicted characteristics,

$$\mathbb{E}(\text{horsepower}) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J \text{Pr}(j) \times \text{horsepower}_j.$$

- **Elasticities?** Computed numerically;
 1. Compute baseline probabilities,
 2. increase the horsepower of car_j by 1%,
 3. Compute new probabilities,
 4. Elasticity is \cong change in probabilities.

Discuss

How does measures like *compensating variation* relate to our model?

Definition: The compensating variation is the monetary transfer that would make an individual indifferent between two regimes.

- **Result:** With $u_{ij} = v_{ij} + \varepsilon_{ij}$, it can be shown that

$$\mathbb{E}[\max_j u_{ij}] = \log \sum_{j=1}^J \exp(v_{ij}) + \gamma,$$

where $\gamma \cong .5772$ is Euler's constant.

- **Suppose** one regressor is price, e.g. x_{j1} .
- **Then** $\frac{1}{\beta_1}$ converts from utils to money.
- **Hence** we can compare “welfare” under $\{v_{ij}\}$ and $\{\tilde{v}_{ij}\}$.

$$CV = \frac{1}{\beta_1} \log \sum_{j=1}^J \exp(v_{ij}) - \frac{1}{\beta_1} \log \sum_{j=1}^J \exp(\tilde{v}_{ij}).$$

1. Overview
2. Random Utility Model
3. Criterion Function
 - 3.1. Identification
4. Features of Interest
5. Specific Issues
 - 5.1. The IIA Restriction
 - 5.2. Market Share Data

Independence of Irrelevant Alternatives (IIA)

- **Model:** Consider the Conditional Logit model,

$$u_{ij}^* = \mathbf{x}_j \beta_o + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Extreme Value Type I.}$$

- **Property:** Let $\Lambda(\mathbf{X}) \equiv \sum_{j=1}^J \exp(\mathbf{x}_j \beta)$.

$$\begin{aligned} \frac{\Pr(y = j)}{\Pr(y = k)} &= \frac{\exp(\mathbf{x}_j \beta) / \Lambda(\mathbf{X})}{\exp(\mathbf{x}_k \beta) / \Lambda(\mathbf{X})} \\ &= \exp[(\mathbf{x}_j - \mathbf{x}_k) \beta]. \end{aligned}$$

This is called the IIA assumption.

- **I.e.** The “odds-ratio” (relative prob.) *does not depend on*
 - x for any $h \neq j, k$,
 - Size of the choicet.
- **Implication:** Restricts substitution patterns...
... two examples clarify.

IIA: Bus Example

- **Assume:** $j = 0$ for (red) bus and $j = 1$ for car.
 - With $\mathbf{x} = (\text{price}, \text{time})$, assume $\mathbf{x}_0 = (1, 2)$, $\mathbf{x}_1 = (2, 1)$.
- **Assume further:** That $\beta_o = (1, 1) \Rightarrow$ then $\Pr(y = 1) = .5$.
- **Since** prob.s sum to 1, $\Pr(y = 0) = .5$.
(*) Therefore $\Pr(y = 1) / \Pr(y = 0) =$.
- **Now:** Add a new (blue) bus; $j = 2$ with $\mathbf{x}_2 = (1, 2)$.
[identical to red bus]
 - $\Rightarrow \Pr(y = 2) =$
- **The problem:** But by IIA and unchanged β_o , (*) must still hold,
 $\Rightarrow \Pr(y = 1) =$.
- **Implication:** The share choosing car would fall to $\frac{1}{3} \dots$
- ... **counterintuitive:** We added an alternative *identical* to an existing.
 - How can that lure people away from cars?

- **Another example:** Suppose $y \in \{1, \dots, J\}$ denotes car type, $\mathbf{x}_j = (\text{price}_j, \text{size}_j)$.
- **Suppose** the price of car J increases.
 - i.e. $x_{J1} := x_{J1} + \Delta$
- **Then** the change in probability (i.e. market share) for *all other* cars, k , are changed **similarly**,

$$\forall k \neq J : \Pr(y = k | \mathbf{X}) = \frac{\exp(\mathbf{x}_k \beta)}{\exp(\beta_1 \cdot \Delta) \exp(\mathbf{x}_J \beta) + \sum_{j \neq J} \exp(\mathbf{x}_j \beta)}.$$

- **Implication:**
 - Car j of course loses some of the market...
 - ... but that loss is divided *similarly* across the remaining cars.
 - ... i.e. not just to the closest substitute.
- **Interpretation:** IIA is a restriction on the possible *substitution patterns*.

- **Model:** Conditional logit.
- **Data:** Aggregate market shares and characteristics of choices.
- **Note:** Expected market share = avg. choice probability (!)

$$s_j \equiv \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y = j\},$$

$$s_j \xrightarrow{N \rightarrow \infty} \Pr(y = j).$$

- **Data:** Time-series information on
 - Characteristics of the available cars at time t , x_{jt} ,
 - Realized market shares during that year, s_{jt} .

Conditional logit in market shares

$$s_{jt} = \Pr(y = j | \mathbf{X}_t) = \frac{\exp(\mathbf{x}_{jt}\beta)}{\sum_{k=1}^J \exp(\mathbf{x}_{kt}\beta)}.$$

- Relative market share of j to l

$$\frac{s_{jt}}{s_{lt}} = \frac{\frac{\exp(\mathbf{x}_{jt}\beta)}{\sum_{k=1}^J \exp(\mathbf{x}_{kt}\beta)}}{\frac{\exp(\mathbf{x}_{lt}\beta)}{\sum_{k=1}^J \exp(\mathbf{x}_{kt}\beta)}} = \frac{\exp(\mathbf{x}_{jt}\beta)}{\exp(\mathbf{x}_{lt}\beta)}.$$

- In logs

$$\log s_{jt} - \log s_{lt} = (\mathbf{x}_{jt} - \mathbf{x}_{lt})\beta.$$

- We showed:

$$\log s_{jt} - \log s_{1t} = (\mathbf{x}_{jt} - \mathbf{x}_{1t})\beta.$$

- Idea:

- $\tilde{y}_{jt} := \log s_{jt} - \log s_{1t},$
- $\tilde{\mathbf{x}}_{jt} := \mathbf{x}_{jt} - \mathbf{x}_{1t}.$

- **Regression:** Obtain $\hat{\beta}$ by regressing \tilde{y}_{jt} on $\tilde{\mathbf{x}}_{jt}$ for all $j \neq 1$ and all t .
- **Applications:** Requires many markets with much exogenous variation in $\mathbf{x}...$
 - but a good first approximation.

6. Appendix

Deriving the likelihood: A few more steps

Choice probability:

$$\begin{aligned}\Pr(y = j|z) &= \Pr(u_{ij} > u_{ik} \forall k \neq j|z) \\&= \Pr(x_j\beta_o + \varepsilon_{ij} > x_k\beta_o + \varepsilon_{ik} \forall k \neq j|z) \\&= \Pr[\varepsilon_{ik} < (x_j - x_k)\beta_o + \varepsilon_{ij} \forall k \neq j|z] \\&= \mathbb{E}_{\varepsilon_{ij}} \{ \Pr[\varepsilon_{ik} < (x_j - x_k)\beta_o + \varepsilon_{ij} \forall k \neq j|z, \varepsilon_{ij}] \} \\&= \mathbb{E}_{\varepsilon_{ij}} \left\{ \prod_{k \neq j} G[(x_j - x_k)\beta_o + \varepsilon_{ij}] \right\} \\&= \int \left\{ \prod_{k \neq j} G[(x_j - x_k)\beta_o + \epsilon] \right\} e^{-\epsilon} e^{-e^{-\epsilon}} d\epsilon \\&= \dots \\&= \frac{\exp(x_j\beta_o)}{\sum_{k=0}^J \exp(x_k\beta_o)}.\end{aligned}$$