

# Siblings paper

## Exploded logit

Javiera Gazmuri

November 12, 2025

## 1 Model

### 1.1 Utility

$$u_{ijk} = \omega_y u_{yij} + (1 - \omega_y) u_{oik} + \gamma \mathbb{1}[j = k] + \varepsilon_{ijk} \quad (1)$$

where  $j$  is the school-assignment of the younger children  $y$  in family  $i$ , and  $k$  is the school-assignment of the older children  $o$ .  $\omega_y$  is the weight that the family places on the utility of the younger children.

$$u_{yij} = \beta_y \text{dist}_{yij} + \delta_y \text{quality}_j + \varepsilon_{yij} \quad (2)$$

$$u_{oik} = \beta_o \text{dist}_{oik} + \delta_o \text{quality}_k + \varepsilon_{oik} \quad (3)$$

where  $\beta_y$ ,  $\beta_o$ ,  $\delta_y$  and  $\delta_o$  are the parameters that govern preferences of the younger and older siblings respectively, and  $\varepsilon_y$  and  $\varepsilon_o$  are idiosyncratic i.i.d. preference shocks.

I'm ignoring the distance between schools for now.

Assumptions:

- When families report marginal applications, it is like they were only applying to one child.
- $\varepsilon_{ijk}$  are i.i.d. type-I extreme value.

### 1.2 Probabilities

Families with more than one common-school applied were asked between,

- a) Worst-joint (WJ)
- b) Best-older-solo (BOS) and best-younger-solo (BYS)

Families with only one common-school applied were asked between,

- a) "Best"-joint (BJ)
- b) BOS & BYS

Let's define,

$$V_{ijk} \equiv \omega_y u_{yij} + (1 - \omega_y) u_{oik} + \gamma \mathbb{1}[j = k]$$

Hence,

$$\begin{aligned} V_i^J &\equiv V_{ijj} && \text{(Assigned together)} \\ V_i^S &\equiv V_{ilk} && \text{(Split assignment)} \end{aligned}$$

The probability that family  $i$  prefers being assigned together is,

$$P_i \equiv \Pr [V_i^J + \varepsilon_{ijj} > V_i^S + \varepsilon_{ilk}] = \frac{\exp(V_i^J)}{\exp(V_i^J) + \exp(V_i^S)}$$

### 1.3 Log-likelihood

Let  $y_i \in \{0, 1\}$  be the observed answer (1 is prefer joint, 0 is prefer split),

$$L_i(\theta) = P_i^{y_i} (1 - P_i)^{1-y_i}, \quad \theta = (\beta_y, \beta_o, \delta_y, \delta_o, \omega_y, \gamma)$$

Assuming independent families  $i = 1, \dots, N$

$$\begin{aligned} \mathcal{L}(\theta) &= \prod_{i=1}^N L_i(\theta) \\ \ell(\theta) &= \sum_{i=1}^N y_i \log P_i + (1 - y_i) \log(1 - P_i) \end{aligned}$$

## 2 Data

From survey responses, we have 4 types of schools:

1. Best-joint (BJ)
2. Worst-joint (WJ)
3. Best-older-solo (BOS)
4. Best-younger-solo (BYS)

From application data, we can say:

- survey showed 10 options max for each, so some lists were trimmed (0,5% sample approx).
- BOS and BYS: 6% selected schools were not found in application data.

Characteristics of schools:

- Specific for each sibling, even for BJ/WJ (since lat-long not always the same between siblings and between campus, quality depends on grade).
- 3% - 6% schools without info of school quality.
- 6% we don't have information of distance for BOS/BYS

## 3 Results

$N = 9,231$  (from original 12,917).