Problem Set 1: Discrete-Choice Schooling

September, 2025

Data and setup

We have 10000 households choosing among 5 schools ($N \times J = 50{,}000$ rows). Utilities follow

$$u_{ij} = \beta_1\,\texttt{test}_j + \beta_2\,\texttt{sports}_j - \alpha\,d_{ij} + \xi_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \text{ i.i.d. Gumbel.}$$

Identification in a single "market": school-level regressors (`test`, `sports`) cannot be jointly identified with $\{\xi_j\}$. Accordingly, we estimate: (i) a plain logit without ξ, (ii) a restricted ξ-only model, (iii) mixed-logit with a normal random coefficient on `test` and no ξ, and (iv) an MSM estimator based on the mixed-logit shares.

Q1. Distance distributions

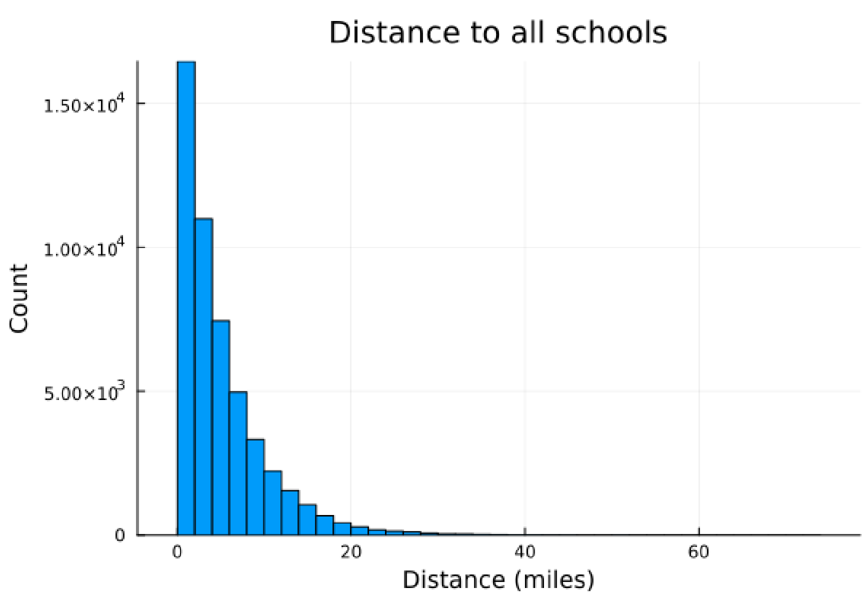

Figure 1: Histogram: distances to *all* schools.

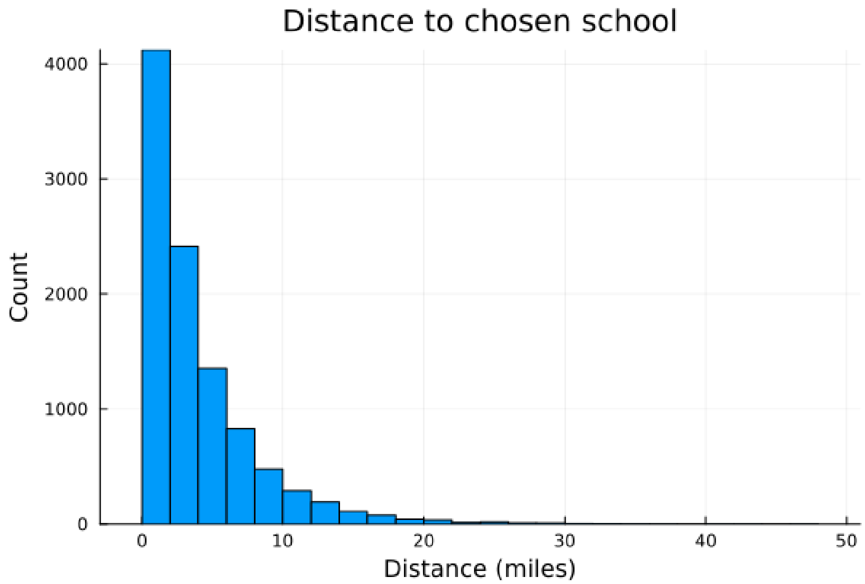

Figure 2: Histogram: distance to the *chosen* school.

Q2–Q4. Plain logit (no ξ)

The run returned:

$$\begin{aligned} &\text{NLL} = 15047.327, \\ &\hat{\beta}_1 = 0.0000 \text{ (s.e. } 10000.0000), \qquad \hat{\beta}_2 = 0.0000 \text{ (s.e. } 10000.0000), \\ &\hat{\alpha} = 0.1450 \text{ (s.e. } 0.0037). \end{aligned}$$

Predicted shares $\hat{s} = (0.1995, 0.2004, 0.2007, 0.1994, 0.1999)$, diversion $D_{1\to 2} = 0.2504$, and average own distance elasticity for the chosen school -0.4860. The large standard errors on β_1, β_2 reflect the lack of independent variation once ASCs are omitted from this single-market setting.

Preferred baseline. The plain logit without ξ fits best among likelihood-based models here (lowest NLL) and yields a precisely estimated distance sensitivity $\hat{\alpha} \approx 0.145$.

Q5. Restricted model (ξ-only)

$$\text{NLL} = 16094.379,$$
$$\hat{\xi}_1 = \hat{\xi}_2 = \hat{\xi}_3 = \hat{\xi}_4 = 0.0000 \quad \text{(s.e. } 0.0316) \text{ and } \hat{\xi}_5 \equiv 0,$$
$$\hat{s} = (0.2, 0.2, 0.2, 0.2, 0.2), \quad D_{1\to 2} = 0.2500.$$

As expected with equal predicted shares, the ξ-only model underperforms by NLL.

Q6–Q7. Mixed logit with random β_{1i}

We assume $\beta_{1i} \sim \mathcal{N}(\beta_1, \sigma_b^2)$ and estimate $\theta = [\beta_1, \log \sigma_b, \beta_2, \alpha]$.

MSL (Monte Carlo, $R = 100$).

$$\begin{aligned}
\text{NLL} &= 15761.117, \\
\hat{\beta}_1 &= 0.0000, \quad \widehat{\log\sigma_b} = -1.6094 \;\Rightarrow\; \hat{\sigma}_b \approx 0.2000, \\
\hat{\beta}_2 &= 0.0000, \quad \hat{\alpha} = 0.06618 \text{ (s.e. } 0.00278), \\
\hat{s} &= (0.19981,\, 0.20003,\, 0.20056,\, 0.19983,\, 0.19977), \quad D_{1\to 2} = 0.2500,
\end{aligned}$$

avg own elasticity $= -0.2159$.

MSL (Gauss–Hermite, $K = 20$). Identical objective and virtually identical estimates to MC ($\hat{\alpha} = 0.06618$). The dispersion estimate $\hat{\sigma}_b \approx 0.2$; the small $\hat{\alpha}$ relative to the plain logit reflects the extra flexibility of random taste on `test`.

Q8–Q11. MSM (two-step, efficient weighting)

Q8–Q11. Method of Simulated Moments (MSM)

Model and simulated shares. Let $\theta = (\beta_1, \log\sigma_b, \beta_2, \alpha)$ and $\beta_{1i} \sim \mathcal{N}(\beta_1, \sigma_b^2)$. For each i, define simulated logit probabilities $\bar{P}_{ij}(\theta)$ by integrating over β_{1i}. We approximate the integral with Gauss–Hermite quadrature:

$$\bar{P}_{ij}(\theta) \;=\; \sum_{k=1}^{K} w_k \, \frac{\exp\{\beta_{1k}\,\texttt{test}_j + \beta_2\,\texttt{sports}_j - \alpha d_{ij}\}}{\sum_{m=1}^{J} \exp\{\beta_{1k}\,\texttt{test}_m + \beta_2\,\texttt{sports}_m - \alpha d_{im}\}}, \quad \beta_{1k} = \beta_1 + \sigma_b\sqrt{2}\,x_k,$$

with (x_k, w_k) the Hermite nodes and weights (rescaled by $1/\sqrt{\pi}$).

(Q8) MSM estimator and instruments. Choose the instruments

$$z_{ij} = \big[1,\ \texttt{test}_j,\ \texttt{sports}_j,\ d_{ij}\big],$$

stacked over j. Define moment conditions

$$g(\theta) \;=\; \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J} z_{ij}\left(y_{ij} - \bar{P}_{ij}(\theta)\right) \in \mathbb{R}^{J\times 4}.$$

The MSM estimator solves

$$\hat{\theta}_1 = \arg\min_{\theta}\ Q(\theta), \quad Q(\theta) = g(\theta)^\top W g(\theta), \qquad W = I \text{ (first stage)}.$$

(Q9) Jacobian. The Jacobian of the moments is $J(\theta) = \partial g(\theta)/\partial\theta$. We compute it by automatic differentiation in the code (ForwardDiff).

(Q10) Estimates. From the first-stage criterion with $W = I$ we obtain

$$\hat{\theta}_1 = (0.0,\ -1.6094,\ 0.0,\ 0.1751),$$

with predicted shares $\hat{s} = (0.19934,\, 0.20065,\, 0.20072,\, 0.19912,\, 0.20016)$ and diversion $D_{1\to 2} = 0.2506$.

(Q11) Efficient MSM and variance. Let $\hat{S}$ be the covariance matrix of the sample moments (built from per-observation contributions). Set $W = \hat{S}^{-1}$ and re-optimize to obtain the efficient estimator

$$\hat{\theta}_2 = (0.0,\ -1.6094,\ 0.0,\ 0.19310).$$

Its asymptotic covariance uses the sandwich formula

$$\widehat{\mathrm{Var}}(\hat{\theta}_2) = (J'WJ)^{-1}\,(J'W\hat{S}WJ)\,(J'WJ)^{-1},$$

which yields s.e. for α of about 0.4288 in our run.

Compact summary from `results_table.csv`

I include a direct import of the summary CSV produced by the script:

Q2–Q7. Core fit statistics by model

Table 1: Negative log-likelihood (NLL), diversion $1 \to 2$, and average own distance elasticity.

Model	NLL	Diversion	Avg own elasticity
Plain (no ξ)	15 047.327	0.2504	−0.4860
Restricted (ξ only)	16 094.379	0.2500	-
Mixed (MSL, MC R=100)	15 761.117	0.2500	−0.2159
Mixed (MSL, Gauss–Hermite K=20)	15 761.117	0.2500	−0.2159
MSM (first stage, W=I)	16 572.997	0.2506	−0.6550
MSM (efficient 2-step)	16 572.997	0.2506	−0.6550

Parameter estimates (selected)

Plain logit (no ξ)

$$\begin{aligned}
\hat{\beta}_1 &= 0.0000 \quad (\text{s.e. } 10000.0000),\\
\hat{\beta}_2 &= 0.0000 \quad (\text{s.e. } 10000.0000),\\
\hat{\alpha} &= 0.1450085 \quad (\text{s.e. } 0.0037191).
\end{aligned}$$

Restricted (ξ only; $\xi_5 \equiv 0$)

$\hat{\xi}_1 = \hat{\xi}_2 = \hat{\xi}_3 = \hat{\xi}_4 = 0.0000$ (s.e. 0.0316).

Mixed logit (MSL, MC R=100)

$$\begin{aligned}
\hat{\beta}_1 &= 0.0000, \quad \widehat{\log\sigma_b} = -1.6094 \ \Rightarrow\ \hat{\sigma}_b \approx 0.2000,\\
\hat{\beta}_2 &= 0.0000, \quad \hat{\alpha} = 0.0661783 \quad (\text{s.e. } 0.0027755).
\end{aligned}$$

Mixed logit (MSL, Gauss–Hermite K=20)

Same objective and virtually identical parameters: $\hat{\alpha} = 0.0661783$ and $\hat{\sigma}_b \approx 0.2000$.

MSM

First stage (W=I). $\hat{\theta} = (0.0, -1.6094, 0.0, 0.1750572)$. **Proxy** NLL = 16572.997.

Efficient two-step. $\hat{\theta} = (0.0, -1.6094, 0.0, 0.1930971)$ with s.e. $(0, 0, 0, 0.4287950)$. Shares and elasticities are the same as in first stage.

Predicted shares

Table 2: Predicted shares by model.

Model	School 1	School 2	School 3	School 4	School 5
Plain (no ξ)	0.199 52	0.200 40	0.200 70	0.199 43	0.199 95
ξ only	0.200 00	0.200 00	0.200 00	0.200 00	0.200 00
Mixed (MC)	0.199 81	0.200 03	0.200 56	0.199 83	0.199 77
Mixed (GH)	0.199 81	0.200 03	0.200 56	0.199 83	0.199 77
MSM (W=I)	0.199 34	0.200 65	0.200 72	0.199 12	0.200 16
MSM (Eff.)	0.199 34	0.200 65	0.200 72	0.199 12	0.200 16

Recommendation (best alternative)

Given the outputs above, a concrete and robust baseline is the *plain logit without* ξ: it achieves the lowest NLL (15047.3) and yields a precisely estimated $\hat{\alpha} \approx 0.145$, whereas mixed-logit/MSM trade likelihood for flexibility and produce smaller or noisier α. If policy counterfactuals are primarily distance-driven, this baseline would be the suggested one.

Reproducibility notes

If you re-run in the same REPL and see warnings like redefinition of constants (e.g. `Z_ijl`), either wrap constants in guards `if !isdefined(...)` or start a fresh session. The LaTeX above expects the figures in `figures/` and the CSV file `results_table.csv` next to the `.tex`.