Stefano Sperti
September 24, 2025

# Problem set 1
## Code Appendix

---

### 1. Look at the data and plot the distribution of distance to all schools, and the distribution of distance to the chosen school.

We first examine the distribution of distances between students and all available schools. Next, we consider the distribution of the distance to the chosen school (i.e., the minimum distance). These two histograms allow us to compare the general accessibility of schools with the actual distances driven by the choice decision.

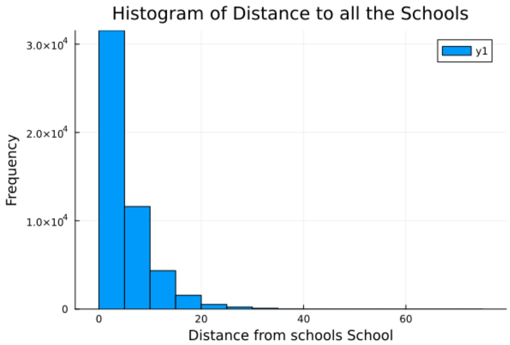


Figure 1: Histogram of distances to all schools.

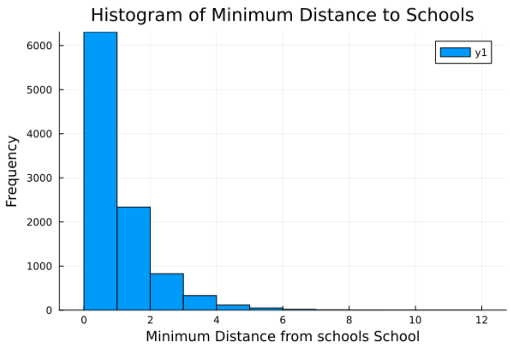


Figure 2: Histogram of distances to the chosen school (minimum distance).

### 2. Write down the market share and log-likelihood for a plain logit model.

Let $y_{ij}$ be an indicator equal to 1 if household $i$ chooses alternative $j$ and 0 otherwise, so for each $i$ we have $\sum_j y_{ij} = 1$.
The observed (empirical) market share of alternative $j$ is

$$\hat{s}_j = \frac{1}{N}\sum_{i=1}^{N} y_{ij}.$$

Let $P_{ij}$ denote the model probability that household $i$ chooses alternative $j$. In the multinomial logit model,

$$P_{ij} = \frac{\exp(\beta' x_{ij})}{\sum_k \exp(\beta' x_{ik})},$$

where $x_{ij}$ is the vector of attributes for alternative $j$ as faced by household $i$, and $\beta$ is the parameter vector.
The likelihood of observing the sample of choices $\{y_{ij}\}$ is

$$L(\beta) \;=\; \prod_{i=1}^{N} \prod_{j} P_{ij}^{y_{ij}}.$$

Taking logs gives the log-likelihood

$$\ell(\beta) \;=\; \sum_{i=1}^{N} \sum_{j} y_{ij} \ln P_{ij}.$$

Substituting the logit probabilities yields

$$\ell(\beta) = \sum_{i=1}^{N} \sum_{j} y_{ij} \left(\beta' x_{ij}\right) - \sum_{i=1}^{N} \ln\Big( \sum_{k} \exp\big(\beta' x_{ik}\big)\Big),$$

where we used $\sum_j y_{ij} = 1$ to collapse the second term. The maximum likelihood estimator $\hat{\beta}$ is the value of $\beta$ that maximizes $\ell(\beta)$.

## 3. Write down the score and the gradient of your log-likelihood.

Starting from the log-likelihood function

$$\ell(\beta) = \sum_{i=1}^{N} \sum_{j} y_{ij} \big(\beta' x_{ij}\big) - \sum_{i=1}^{N} \ln\left( \sum_{k} e^{\beta' x_{ik}} \right),$$

its derivative with respect to $\beta$ is

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{N} \sum_{j} y_{ij} \frac{\partial (\beta' x_{ij})}{\partial \beta} - \sum_{i=1}^{N} \frac{\partial}{\partial \beta} \ln\left( \sum_{k} e^{\beta' x_{ik}} \right).$$

The first term simplifies to

$$\sum_{i=1}^{N} \sum_{j} y_{ij} x_{ij},$$

and the second term to

$$\sum_{i=1}^{N} \sum_{k} P_{ik} x_{ik},$$

where $P_{ik} = \dfrac{e^{\beta' x_{ik}}}{\sum_m e^{\beta' x_{im}}}$.
Thus, the score vector is

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{N} \sum_{j} \left(y_{ij} - P_{ij}\right) x_{ij}.$$

The score is the sum over households of the difference between observed choices $(y_{ij})$ and model-predicted probabilities $(P_{ij})$, weighted by the covariates $x_{ij}$.

## 4. Estimate the plain logit model by maximimum likelihood.

In the following table, we present the estimates of the plain logit model obtained by maximum likelihood.

| Parameter | Estimate |
|---|---|
| $\alpha$ | 0.201097 |
| $\beta_1$ | -0.175195 |
| $\beta_2$ | 0.185489 |
| $\xi_1$ | 0.0 |
| $\xi_2$ | -0.65893 |
| $\xi_3$ | -1.014664 |
| $\xi_4$ | 0.391017 |
| $\xi_5$ | -0.894889 |

Table 1: Parameter estimates for the school full choice model.

## 5. Estimate a restricted model with only $\xi_j$ parameters. Add that to your table.

In the following table, we present the estimates of the restricted model with only $\xi_j$ parameters.

| Parameter | Estimate |
|---|---|
| $\xi_1$ | 0.0 |
| $\xi_2$ | -0.966312 |
| $\xi_3$ | -0.750957 |
| $\xi_4$ | 0.57213 |
| $\xi_5$ | -0.743168 |

Table 2: Parameter estimates for the school restricted choice model.

## 6) Random coefficient on test scores: simulated market share and gradient

Assume parents' taste for test scores is random,

$$\beta_{1i} \sim \mathcal{N}(\beta_1, \sigma_b^2).$$

Then, we can define

$$\beta_1^{(r)} \;=\; \beta_1 + \sigma_b\, z^{(r)}.$$

Conditional on draw $r$ the representative utility for household $i$ and school $k$ is

$$V_{ij}^{(r)} | \beta_1^{(r)} \;=\; \beta_1^{(r)} \cdot \mathrm{test}_j \;+\; \beta_2 \cdot \mathrm{sports}_j \;+\; \xi_j \;-\; \alpha\, d_{ij},$$

and the conditional (given $\beta_1^{(r)}$) choice probability is the usual logit

$$P_{ij}^{(r)} | \beta_1^{(r)} \;=\; \frac{\exp\big(V_{ij}^{(r)}\big)}{\sum_{m=1}^{J} \exp\big(V_{im}^{(r)}\big)}.$$

We approximate the integrated choice probability by averaging over the $R$ draws:

$$\widehat{P}_{ik} \;=\; \frac{1}{R} \sum_{r=1}^{R} P_{ik}^{(r)}.$$

The, The simulated market share of school $j$ is

$$\widehat{S}_j \;=\; \frac{1}{N} \sum_{i=1}^{N} \widehat{P}_{ij} \;=\; \frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} P_{ij}^{(r)}.$$

Let $j(i)$ denote the observed choice of household $i$. The simulated log-likelihood is

$$\widehat{\mathcal{L}}(\theta) \;=\; \sum_{i=1}^{N} \log \widehat{P}_{ij(i)},$$

where $\theta = (\beta_1, \sigma_b, \beta_2, \xi_{1:J}, \alpha)$.
For a generic parameter $\theta \in \{\beta_1, \sigma_b, \beta_2, \xi_1, \dots, \xi_J, \alpha\}$ we have

$$\frac{\partial \widehat{\mathcal{L}}}{\partial \theta} = \sum_{i=1}^{N} \frac{1}{\widehat{P}_{ij(i)}} \, \frac{\partial \widehat{P}_{ij(i)}}{\partial \theta} = \sum_{i=1}^{N} \frac{1}{\widehat{P}_{ij(i)}} \, \frac{1}{R} \sum_{r=1}^{R} \frac{\partial P^{(r)}_{ij(i)}}{\partial \theta}.$$

For computing the $\frac{\partial P^{(r)}_{ij(i)}}{\partial \theta}$ we can use the chain rule and starting analyzing the he derivative of a logit probability with respect to the utilities is

$$\frac{\partial P^{(r)}_{ij}}{\partial V^{(r)}_{ik}} = P^{(r)}_{ij} \big( 1_{\{j=k\}} - P^{(r)}_{ik} \big).$$

Hence

$$\frac{\partial P^{(r)}_{ij}}{\partial \theta} = \sum_{k=1}^{J} P^{(r)}_{ij} \big( 1_{\{j=k\}} - P^{(r)}_{ik} \big) \frac{\partial V^{(r)}_{ik}}{\partial \theta}.$$

The derivatives of $V^{(r)}_{ik}$ for the parameters depends specifically of the functional form of the utility and in this case we get

$$\frac{\partial V^{(r)}_{ik}}{\partial \beta_1} = \text{test}_k, \qquad \frac{\partial V^{(r)}_{ik}}{\partial \sigma_b} = z^{(r)}\, \text{test}_k,$$

$$\frac{\partial V^{(r)}_{ik}}{\partial \beta_2} = \text{sports}_k, \qquad \frac{\partial V^{(r)}_{ik}}{\partial \xi_\ell} = 1_{\{k=\ell\}}, \qquad \frac{\partial V^{(r)}_{ik}}{\partial \alpha} = -d_{ik}.$$

Combining the pieces,

$$\frac{\partial \widehat{\mathcal{L}}}{\partial \theta} = \sum_{i=1}^{N} \frac{1}{\widehat{P}_{ij(i)}} \, \frac{1}{R} \sum_{r=1}^{R} \sum_{k=1}^{J} P^{(r)}_{ij(i)} \big( 1_{\{j(i)=k\}} - P^{(r)}_{ik} \big) \frac{\partial V^{(r)}_{ik}}{\partial \theta},$$

with $\partial V^{(r)}_{ik} / \partial \theta$ given above for each parameter.

### 7. Estimate this expanded model via maximum likelihood: (a) Using 100 Monte Carlo Draws from an apporpriately transformed standard normal. (b) Using a Gauss Hermite quadrature rule.

In the following table, we present the estimates of the expanded model via maximum likelihood using 100 Monte Carlo Draws from an apporpriately transformed standard normal.

| Parameter | Estimate |
|---|---|
| $\alpha$ | 0.201097 |
| $\beta_1$ | -1.437858 |
| $\beta_2$ | -0.307116 |
| $\xi_1$ | 0.0 |
| $\xi_2$ | -0.04998 |
| $\xi_3$ | -1.303247 |
| $\xi_4$ | 1.854057 |
| $\xi_5$ | -1.651327 |
| $\sigma_b$ | 7.251364 |

Table 3: Parameter estimates for the school full simulated choice model using Monte Carlo methods.

In the following table, we present the estimates of the expanded model via maximum likelihood using a Gauss Hermite quadrature rule.

| Parameter | Estimate |
|---|---|
| $\alpha$ | 0.201097 |
| $\beta_1$ | -0.211839 |
| $\beta_2$ | 0.201481 |
| $\xi_1$ | 0.0 |
| $\xi_2$ | -0.632793 |
| $\xi_3$ | -1.030053 |
| $\xi_4$ | 0.364165 |
| $\xi_5$ | -0.897929 |
| $\sigma_b$ | 1.0e-6 |

Table 4: Parameter estimates for the school full simulated choice model using Gaussian Quadrature methods.

### 8. Read Chapter 10 in Train and write down the MSM estimator for the expanded model. What are your "instruments"?

### 8. MSM estimator and instruments

The Method of Simulated Moments (MSM) estimator is obtained by choosing $\hat{\theta}$ to minimize

$$Q_N(\theta) \;=\; g_N(\theta)^\top W_N g_N(\theta),$$

where

$$g_N(\theta) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j}\left[y_{ij} - \widehat{P}_{ij}(\theta)\right] z_{ij},$$

with $y_{ij}$ the observed choice indicator, $\widehat{P}_{ij}(\theta)$ the simulated choice probabilities, and $z_{ij}$ an instrument. The instruments are functions of the exogenous variables that shift utilities but are uncorrelated with the unobserved errors. In this setting, the instruments are the school characteristics that are test scores, sports, distance.

### 9. Calculate the Jacobian of the MSM estimator.

Let's copute the derivative

$$\frac{\partial g_H(\theta)}{\partial \theta} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J} z_{ij}\left(\frac{\partial \hat{P}_{ij}(\theta)}{\partial \theta}\right).$$

The derivatives with respect to $\hat{P}_{ij}$ is computed in section 6.

### 10. Estimate the Parameters of the MSM estimator.

In the following table, we present the estimates of the Parameters of the MSM estimator.

| Parameter | Estimate |
| --- | --- |
| $\alpha$ | 0.197131 |
| $\beta_1$ | -0.21552 |
| $\beta_2$ | 0.213248 |
| $\xi_1$ | 0.0 |
| $\xi_2$ | -0.626077 |
| $\xi_3$ | -1.033541 |
| $\xi_4$ | 0.384117 |
| $\xi_5$ | -0.89063 |
| $\sigma_b$ | 0.985956 |

Table 5: Parameter estimates for the school full simulated choice model using Simulated Method of Moments.

### 11. Bonus: Using your initial MSM estimates as a starting point, explain how to construct an "efficient" MSM estimator, and produce "efficient" estimates.

Starting from the initial MSM estimates, we obtain an efficient estimator by re-estimating the model using as weighting matrix the inverse of the optimal covariance matrix of the sample moments:

$$W_N^* = \hat{\Omega}_N^{-1},$$

where $\hat{\Omega}_N$ is the estimated variance-covariance matrix of the simulated moments. This two-step procedure delivers the efficient MSM estimator and the corresponding efficient estimates.