Answers to Empirical IO I: Problem Set 1

Q1. Look at the data and plot the distribution of distance to all schools, and the distribution of distance to the chosen school.

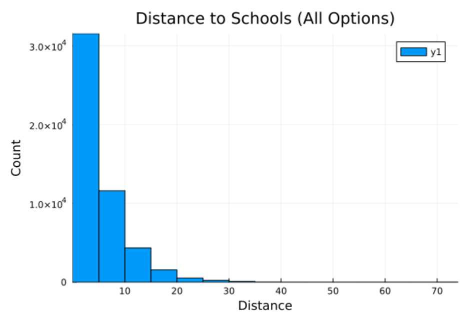

Figure 1: Distance to Schools (All Options)

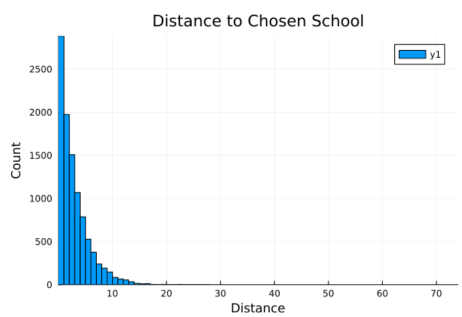

Figure 2: Distance to Chosen School

Clearly, chosen schools tend to be closer.

Q2. Write down the market share and log-likelihood for a plain logit model.

Market share for a school j comes from aggregating up the individual choice probabilities. We don't observe school quality and the inclusion of a school fixed effect would make the school-level observables collinear. The observed part of utility is:

$$V_{ij} := \beta_1 \, \text{tests}_j + \beta_2 \, \text{sports}_j - \alpha \, d_{ij}$$

The choice probability that household i chooses school j is

$$P_{ij}(\theta) \;=\; \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}.$$

So the implied market share for school j is jsut the average over all households:

$$s_j(\theta) \;=\; \frac{1}{H} \sum_{i=1}^{H} P_{ij}(\theta).$$

Let $y_{ij} \in \{0, 1\}$ indicate whether household i chose school j. Given the choice probabilities above, the likelihood of the observed choice for household i can be written as

$$L_i(\theta) \;=\; \prod_{j=1}^{J} P_{ij}(\theta)^{\,y_{ij}}.$$

With independence across households, the likelihood can be written out multiplicatively as:

$$L(\theta) \;=\; \prod_{i=1}^{H} L_i(\theta) \;=\; \prod_{i=1}^{H} \prod_{j=1}^{J} P_{ij}(\theta)^{\,y_{ij}}.$$

With the log-likelihood therefore simplifying to

$$\ell(\theta) \;=\; \sum_{i=1}^{H} \sum_{j=1}^{J} y_{ij} \; \log P_{ij}(\theta).$$

Q3. Write down the score and the gradient of your log-likelihood.

The score and the gradient refer to the same thing: the first derivatives with respect to the model parameters (?)

Let $x_{ij} = \big(\text{tests}_j, \; \text{sports}_j, \; -d_{ij}\big)$. Using the log-likelihood above, the score is then

$$\nabla_\theta \ell(\theta) = \left(\sum_{i=1}^{H} \sum_{j=1}^{J} x_{ij} \, (y_{ij} - P_{ij}) \right)$$

Which is a 3 dim vector of the derivatives wrt $(\beta_1, \beta_2, \alpha)$.

Or written out component-wise:

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{H} \sum_{j=1}^{J} \text{tests}_j \, (y_{ij} - P_{ij}), \qquad \frac{\partial \ell}{\partial \beta_2} = \sum_{i=1}^{H} \sum_{j=1}^{J} \text{sports}_j \, (y_{ij} - P_{ij}),$$

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{H} \sum_{j=1}^{J} (-d_{ij}) \, (y_{ij} - P_{ij}),$$

Q4. Estimate the plain logit model by maximimum likelihood.

Model	NLL	β_1	β_2	α	D_{12}	Avg. own dist. elas.
Plain Logit	12982.80	-3.44	0.52	0.20	0.15	-0.43

Table 1: Estimation results for logit models.

Q5. Estimate a restricted model with only ξ_j parameters. Add that to your table.

Model	NLL	β_1	β_2	α	D_{12}	Avg. own dist. elas.
Plain Logit	12982.80	-3.44	0.52	0.20	0.15	-0.43
Only-ξ	14260.39	–	–	–	–	–

Table 2: Estimation results for logit models.

Q6. Now allow for parents to have different preferences for test scores$_j$ so that $\beta_{1i} \sim \mathcal{N}(\beta_1, \sigma_b)$. Write down the (simulated) market share and gradient expressions.

Each household i has a random coefficient β_{1i} such that

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

The observable part of utility becomes

$$V_{ij} = \beta_{1i} \, \text{tests}_j + \beta_2 \, \text{sports}_j - \alpha \, d_{ij}.$$

For a given household i and a given draw r from the normal, the choice probability is

$$P_{ij}^{(r)}(\theta) \;=\; \frac{\exp\Big(\beta_{1i}^{(r)} \, \text{tests}_j + \beta_2 \, \text{sports}_j - \alpha \, d_{ij}\Big)}{\sum_{k=1}^{J} \exp\Big(\beta_{1i}^{(r)} \, \text{tests}_k + \beta_2 \, \text{sports}_k - \alpha \, d_{ik}\Big)}.$$

We can obtain the simulated market share by averaging over R draws for each household:

$$\tilde{s}_j(\theta) \;=\; \frac{1}{H}\sum_{i=1}^{H}\frac{1}{R}\sum_{r=1}^{R}P_{ij}^{(r)}(\theta).$$

For the log-likelihood

$$\ell(\theta) = \sum_{i=1}^{H}\sum_{j=1}^{J} y_{ij}\log P_{ij}(\theta),$$

with P_{ij} replaced by its simulated counterpart, the (simulated) score components are

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{H}\frac{1}{p_i(\theta)}\cdot\frac{1}{R}\sum_{r=1}^{R}P_{i\,j(i)}^{(r)}\Big(\text{tests}_{j(i)} - \sum_{k=1}^{J}P_{ik}^{(r)}\,\text{tests}_k\Big),$$

$$\frac{\partial \ell}{\partial \beta_2} = \sum_{i=1}^{H}\frac{1}{p_i(\theta)}\cdot\frac{1}{R}\sum_{r=1}^{R}P_{i\,j(i)}^{(r)}\Big(\text{sports}_{j(i)} - \sum_{k=1}^{J}P_{ik}^{(r)}\,\text{sports}_k\Big),$$

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{H}\frac{1}{p_i(\theta)}\cdot\frac{1}{R}\sum_{r=1}^{R}P_{i\,j(i)}^{(r)}\Big(-d_{i\,j(i)} + \sum_{k=1}^{J}P_{ik}^{(r)}\,d_{ik}\Big),$$

where

$$p_i(\theta) \;=\; \frac{1}{R}\sum_{r=1}^{R}P_{i\,j(i)}^{(r)}(\theta).$$

There is also the additional parameter σ_b governing the dispersion of β_{1i}:

$$\frac{\partial \ell}{\partial \sigma_b} = \sum_{i=1}^{H}\frac{1}{p_i(\theta)}\cdot\frac{1}{R}\sum_{r=1}^{R}P_{i\,j(i)}^{(r)}\,z_{ir}\Big(\text{tests}_{j(i)} - \sum_{k=1}^{J}P_{ik}^{(r)}\,\text{tests}_k\Big).$$

Because $\beta_{1i}^{(r)} = \beta_1 + \sigma_b z_{ir}$, the chain rule gives

$$\frac{\partial \beta_{1i}^{(r)}}{\partial \sigma_b} = z_{ir},$$

which is why a factor z_{ir} multiplies the same test-score score component as for β_1.

> Q7. Estimate this expanded model via maximum likelihood: (a) Using 100 Monte Carlo Draws from an appropriately transformed standard normal. (b) Using a Gauss Hermite quadrature rule.

I had some trouble getting these to run. When I coded up the analytical gradient, (even with $R = 1$) my optimisation (BFGS) was running indefinitely.

When I instead numerically differentiated, $\hat{\sigma_b}$ was going towards zero and I was effectively getting the plain logit results. These are what I've tabulated below but I don't think should be behaving like this:

Model	NLL	β_1	β_2	α	D_{12}	Avg. own dist. elas.
Plain Logit	12982.80	-3.44	0.52	0.20	0.15	-0.43
Only-ξ	14260.39	–	–	–	–	–
Expanded Logit (MC, R=100)	12982.80	-3.44	0.52	0.20	0.15	-0.43
Expanded Logit (GH, M=20)	12982.80	-3.44	0.52	0.20	0.15	-0.43

Table 3: Estimation results for logit models.

Q8. Read Chapter 10 in Train and write down the MSM estimator for the expanded model. What are your "instruments"?

The MSM estimator takes the residuals from the simulated model of choice probabilities and projects them onto a set of (exogenous) variables. Specifically, for each household i and school j we observe $y_{ij} \in \{0, 1\}$, and we compute simulated probabilities $\hat{P}_{ij}(\theta)$ given parameters $\theta = (\beta_1, \beta_2, \alpha, \sigma_b)$.

To construct the MSM estimator, we take the simulated residuals

$$y_{ij} - \hat{P}_{ij}(\theta)$$

and require that, on average, they are uncorrelated with the chosen instruments z_{ij}. Formally, this means imposing the moment conditions

$$\frac{1}{H}\sum_{i=1}^{H}\sum_{j=1}^{J}\big(y_{ij} - \hat{P}_{ij}(\theta)\big)\, z_{ij} \;=\; 0.$$

Since in practice the sample moments will not be exactly zero, the MSM estimator chooses θ to make them as close to zero as possible. That is,

$$\hat{\theta}_{MSM} = \arg\min_{\theta} \left\| \frac{1}{H}\sum_{i=1}^{H}\sum_{j=1}^{J}\big(y_{ij} - \hat{P}_{ij}(\theta)\big)\, z_{ij} \right\|^2,$$

where the norm reflects the weighting we put on different instruments.

The instruments z_{ij} need to be variables that exogenosuly shift the observed choices. We might plausibly just use our observables to instrument themselves.

So in practice, the MSM estimator here would be computed by driving the average residuals interacted with distance, tests, and sports to zero.

Q9. Calculate the Jacobian of the MSM estimator.

As above, we have the moment conditions:

$$g_H(\theta) \;=\; \frac{1}{H}\sum_{i=1}^{H}\sum_{j=1}^{J}\big(y_{ij} - \hat{P}_{ij}(\theta)\big)\, z_{ij}.$$

The Jacobian is the matrix of first derivatives, which we can calculate by the chain rule:

$$G(\theta) \;:=\; \frac{\partial g_H(\theta)}{\partial \theta^\top} \;=\; -\frac{1}{H}\sum_{i=1}^{H}\sum_{j=1}^{J} z_{ij}\,\Big(\frac{\partial \hat{P}_{ij}(\theta)}{\partial \theta^\top}\Big).$$

$\hat{P}_{ij}(\theta)$ is a simulation (I'll use GH) average over draws $r = 1, \ldots, R$:

$$\hat{P}_{ij}(\theta) \;=\; \frac{1}{R}\sum_{r=1}^{R} P_{ij}^{(r)}(\theta), \qquad P_{ij}^{(r)} \;=\; \frac{\exp(V_{ij}^{(r)})}{\sum_{k=1}^{J}\exp(V_{ik}^{(r)})},$$

so

$$\frac{\partial \hat{P}_{ij}}{\partial \theta} \;=\; \frac{1}{R}\sum_{r=1}^{R}\frac{\partial P_{ij}^{(r)}}{\partial \theta}.$$

For the logit probability

$$P_{ij}^{(r)}(\theta) = \frac{\exp(V_{ij}^{(r)})}{\sum_{k=1}^{J}\exp(V_{ik}^{(r)})},$$

the derivative with respect to a parameter θ is

$$\frac{\partial P_{ij}^{(r)}}{\partial \theta} = P_{ij}^{(r)}\Big(x_{ij,\theta}^{(r)} - \sum_{k=1}^{J} P_{ik}^{(r)}\, x_{ik,\theta}^{(r)}\Big),$$

where $x_{ij,\theta}^{(r)} := \partial V_{ij}^{(r)}/\partial \theta$.

In our model

$$V_{ij}^{(r)} = \beta_{1i}^{(r)}\,\text{tests}_j + \beta_2\,\text{sports}_j - \alpha\, d_{ij}, \qquad \beta_{1i}^{(r)} = \beta_1 + \sigma_b z_{ir}.$$

Hence the relevant derivatives are

$$x_{ij,\beta_1}^{(r)} = \text{tests}_j, \qquad x_{ij,\sigma_b}^{(r)} = z_{ir}\,\text{tests}_j, \qquad x_{ij,\beta_2}^{(r)} = \text{sports}_j, \qquad x_{ij,\alpha}^{(r)} = -d_{ij},$$

Substituting these into the expression above gives the derivatives of the probabilities with respect to each parameter, which are then used in the Jacobian $G(\theta)$.

Q10. Estimate the Parameters of the MSM estimator.

Model	NLL	β_1	β_2	α	D_{12}	Avg. own dist. elas.
Plain Logit	12982.80	-3.44	0.52	0.20	0.15	-0.43
Only-ξ	14260.39	–	–	–	–	–
MSM (GH, M=20)	–	-3.48	0.53	0.20	0.15	-0.43

Table 4: Estimation results for logit models.

> Q11. Bonus: Using your initial MSM estimates as a starting point, explain how to construct an "efficient" MSM estimator, and produce "efficient" estimates.

Efficiency comes from using the right weighting matrix W on the moment conditions. We can construct the efficient MSM by:
Step 1 (initial): estimate $\hat{\theta}^{(1)}$ with one–step MSM using $W = I$, as in the last question. This gives us consistent but not necessarily efficient estimates, since it treats each moment condition as equally informative.

Step 2 (weighting): at $\hat{\theta}^{(1)}$, form per-household moment vectors

$$m_i(\theta) = \sum_{j=1}^{J} \big(y_{ij} - \hat{P}_{ij}(\theta)\big)\, z_{ij}, \qquad g_H(\theta) = \frac{1}{H}\sum_{i=1}^{H} m_i(\theta).$$

Here $m_i(\theta)$ measures the discrepancy between observed choices and predicted probabilities, weighted by the instruments z_{ij}. The covariance of these discrepancies tells us which moments are measured more precisely. Following Train Chapter 10, we can estimate this covariance as

$$\widehat{S} = \frac{1}{H}\sum_{i=1}^{H} \Big(m_i(\hat{\theta}^{(1)}) - g_H(\hat{\theta}^{(1)})\Big)\Big(m_i(\hat{\theta}^{(1)}) - g_H(\hat{\theta}^{(1)})\Big)^{\top},$$

and set $W = \widehat{S}^{-1}$ to give higher weight to the more informative moments.

Step 3 (re-estimate): solve

$$\hat{\theta}^{(2)} = \arg\min_{\theta} g_H(\theta)^{\top} W\, g_H(\theta).$$

This re-estimation with the optimal weight matrix yields the efficient MSM estimator. See estimates in cumulative trade table below:

Model	NLL	β_1	β_2	α	D_{12}	Avg. own dist. elas.
Plain Logit	12982.80	-3.44	0.52	0.20	0.15	-0.43
Only-ξ	14260.39	–	–	–	–	–
MSM (GH, M=20)	–	-3.48	0.53	0.20	0.15	-0.43
Efficient MSM (GH, M=20))	–	-3.48	0.53	0.20	0.15	-0.43

Table 5: Estimation results for logit models.