Answer Keys to Problem Set 2

Microeconometrics I with Joan Llull

IDEA, Fall 2024

TA: Conghan Zheng

October 04, 2024

**This document is not intended to be a complete solution, but rather to provide some examples of interpretations and key points for the answer.*

Logit and Probit

- Estimate model

$$p(\boldsymbol{x}) \equiv \Pr[\text{Same industry} = 1|\boldsymbol{x}] = F(\boldsymbol{x}'\boldsymbol{\beta}),$$

	(1) Logit	(2) Probit
same_ind		
exper	-0.00233	-0.00148
	(0.00445)	(0.00277)
unempl_dur	-0.00278***	-0.00174***
	(0.000659)	(0.000410)
age	0.00767	0.00515
	(0.0340)	(0.0211)
age2	0.000508	0.000311
	(0.000557)	(0.000346)
female	0.133***	0.0822***
	(0.0340)	(0.0212)
ln_wage	0.262***	0.163***
	(0.0300)	(0.0186)
veneto_r~d	-0.199***	-0.124***
	(0.0374)	(0.0232)
_cons	-2.099***	-1.311***
	(0.537)	(0.334)
ll	-10662.5	-10662.6
N	15638	15638

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Odds Ratios

```
Iteration 0:   log likelihood = -10818.448
Iteration 1:   log likelihood = -10662.625
Iteration 2:   log likelihood = -10662.549
Iteration 3:   log likelihood = -10662.549

Logistic regression                                     Number of obs = 15,638
                                                        LR chi2(7)    = 311.80
                                                        Prob > chi2   = 0.0000
Log likelihood = -10662.549                             Pseudo R2     = 0.0144
```

same_ind	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
exper	.9976712	.0044377	-0.52	0.600	.9890113	1.006407
unempl_dur	.9972194	.0006571	-4.23	0.000	.9959325	.9985081
age	1.0077	.0342359	0.23	0.821	.9427838	1.077085
age2	1.000508	.0005574	0.91	0.362	.9994158	1.001601
female	1.141795	.0388167	3.90	0.000	1.068195	1.220466
ln_wage	1.299538	.039032	8.72	0.000	1.225245	1.378336
veneto_resid	.819524	.0306217	-5.33	0.000	.7616516	.8817936
_cons	.1225966	.0658511	-3.91	0.000	.0427826	.3513094

Note: _cons estimates baseline odds.

- Odds Ratio in Logit: e^{β} rather than β
- Example for interpretation: Increasing `exper` by one unit barely change (odds ratio $e^{\beta_j} \approx 0.998$, $p > |z|$ ≈0.600) the decision to stay in the same industry (`same_industry` $= 1$). We can't distinguish its effect from zero ($e^{\beta_j} \to 1 \Rightarrow \beta_j \to 0$). One possible reason for this is that `exper` is general labor market experience, which is highly correlated with another regressor: age.

Marginal Effects

```
Conditional marginal effects                          Number of obs = 15,638
Model VCE: OIM

Expression: Pr(same_ind), predict()
dy/dx wrt:  exper unempl_dur age age2 female ln_wage veneto_resid
At: exper        = 11.50288 (mean)
    unempl_dur   = 15.69095 (mean)
    age          = 30.64529 (mean)
    age2         =  968.246 (mean)
    female       = .4156542 (mean)
    ln_wage      = 6.243519 (mean)
    veneto_resid = .7154368 (mean)
```

	dy/dx	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
exper	-.0005812	.0011089	-0.52	0.600	-.0027547	.0015922
unempl_dur	-.0006942	.0001643	-4.23	0.000	-.0010161	-.0003722
age	.0019122	.0084699	0.23	0.821	-.0146885	.0185128
age2	.0001266	.0001389	0.91	0.362	-.0001457	.0003988
female	.0330579	.0084753	3.90	0.000	.0164466	.0496692
ln_wage	.0653194	.0074877	8.72	0.000	.0506437	.079995
veneto_resid	-.049619	.009315	-5.33	0.000	-.0678761	-.0313619

- The results from the mean marginal effects are consistent with the coefficient estimates. For example, for an individual with average characteristics (30 years old, 42% likely to be female, etc.), as experience increases by one unit, the individual is 0.05 pp less likely to remain in the same industry. This is barely an effect, we can't distinguish it from zero.

Multinomial Logit

```
Multinomial logistic regression                     Number of obs =  15,638
                                                    LR chi2(14)   = 2728.98
                                                    Prob > chi2   =  0.0000
Log likelihood = -14032.645                         Pseudo R2     =  0.0886
```

choice_ind	RRR	Std. err.	z	P>|z|	[95% conf.	interval]
Manufacturing	(base outcome)					
Services						
exper	.9413455	.0048604	-11.71	0.000	.9318673	.9509202
unempl_dur	1.007713	.0007979	9.70	0.000	1.00615	1.009278
age	1.203915	.0458359	4.87	0.000	1.117349	1.297189
age2	.9981284	.0006235	-3.00	0.003	.9969071	.9993511
female	1.650514	.0638165	12.96	0.000	1.530058	1.780453
ln_wage	.7479919	.025463	-8.53	0.000	.6997139	.7996009
veneto_resid	.9728996	.040549	-0.66	0.510	.8965844	1.055711
_cons	.120699	.0727368	-3.51	0.000	.0370462	.393245
Public_sector						
exper	.8622645	.0057761	-22.12	0.000	.8510175	.8736602
unempl_dur	1.019565	.0009598	20.58	0.000	1.017685	1.021448
age	1.743681	.1079043	8.98	0.000	1.544515	1.96853
age2	.9940603	.0009864	-6.00	0.000	.9921288	.9959956
female	6.626412	.3831103	32.71	0.000	5.916511	7.421492
ln_wage	.7437535	.0334216	-6.59	0.000	.6810502	.8122298
veneto_resid	1.021151	.0620432	0.34	0.730	.9065101	1.15029
_cons	.0000283	.0000279	-10.60	0.000	4.07e-06	.0001962

Note: **_cons** estimates baseline relative risk for each outcome.

- Example interpreting the relative risk ratio: an increase of one year of experience reduces the risk of working in the service sector after displacement by a factor of about 0.94 compared to the risk of working in manufacturing (base category), holding the other variables in the model constant.

Marginal Effects for MNL

```
Conditional marginal effects                         Number of obs = 15,638
Model VCE: OIM

Expression: Pr(choice_ind==Services), predict(outcome(2))
dy/dx wrt:  exper unempl_dur age age2 female ln_wage veneto_resid
At: exper        = 11.50288 (mean)
    unempl_dur   = 15.69095 (mean)
    age          = 30.64529 (mean)
    age2         =  968.246 (mean)
    female       = .4156542 (mean)
    ln_wage      = 6.243519 (mean)
    veneto_resid = .7154368 (mean)
```

	dy/dx	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
exper	-.0083323	.0011017	-7.56	0.000	-.0104917	-.0061729
unempl_dur	.0010395	.000167	6.22	0.000	.0007121	.0013668
age	.0218789	.0083329	2.63	0.009	.0055469	.038211
age2	-.0002082	.0001364	-1.53	0.127	-.0004756	.0000591
female	.0447865	.0083423	5.37	0.000	.0284359	.061137
ln_wage	-.0552706	.0072666	-7.61	0.000	-.0695129	-.0410283
veneto_resid	-.0070239	.0090194	-0.78	0.436	-.0247017	.0106538

- You get one table for each sector.
- Examples of interpretation: For individuals with average characteristics, one additional year of experience decreases the probability of working in the service sector by 0.8 pp.

Conditional Logit

Wide-Form Data

- In the wide form data, you have a variable for experience in each sector, for example, `exp_manufacture` for manufacturing, which is not alternative-varying. For conditional logit, you need to transform the data from wide to long.

id	year	age	e_manuf_light	e_manuf_heavy	⋯
1	1999	40	5	0	⋯
2	1999	40	20	0	⋯
3	1999	40	24	0	⋯
4	1999	40	2	7	⋯
5	1999	40	0	23	⋯
⋮	⋮	⋮	⋮		

Conditional Logit

Long-Form Data

- In the long-form data, you have a variable, let's call it e_, for the sector-specific experience.

id	industry	d	e_	⋯
1	manuf_elecon	0	0	⋯
1	manuf_heavey	0	0	⋯
1	manuf_light	0	5	⋯
1	public_adm	0	0	⋯
1	public_hlth	0	0	⋯
1	serv_fin	0	0	⋯
1	serv_other	0	0	⋯
1	serv_sales	1	13	⋯
⋮	⋮	⋮	⋮	

- Now each row refers to a sector, you have a variable d indicating whether that individual chooses that sector or not, and the variable e_ indicates the individual's experience in that sector. The variable e_ is alternative-specific variable.

Conditional Logit

Collinearity

Log likelihood = -21561.652				Wald chi2(43) = 10720.73		
				Prob > chi2 = 0.0000		

d_	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
industry						
e_	.2368502	.0027354	86.59	0.000	.2314889	.2422115
manuf_elecon	(base alternative)					
manuf_heavy						
unempl_dur	-.0148817	.0014249	-10.44	0.000	-.0176745	-.012089
age	-.0754779	.074813	-1.01	0.313	-.2221087	.0711529
age2	.0012089	.0012399	0.97	0.330	-.0012213	.003639
female	2.15553	.1623102	13.28	0.000	1.837408	2.473652
ln_wage	.0362148	.07177	0.50	0.614	-.104452	.1768815
veneto_resid	.5966138	.0767017	7.78	0.000	.4462812	.7469464
_cons	1.411464	1.175378	1.20	0.230	-.8922342	3.715163

- For conditional logit and nested logit, it is better to exclude general experience when including industry-specific experience in a regression.
- In fact, we don't include general cumulative experience not only here but also in the recent labor literature when we distinguish between job types.
- Some of the information in general experience is captured by industry-specific experience, and some is captured by age, education, and other related regressors.

Conditional Logit

Coefficients

				Wald chi2(43)	=	10720.73
Log likelihood = -21561.652				Prob > chi2	=	0.0000

d_	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
industry						
e_	.2368502	.0027354	86.59	0.000	.2314889	.2422115
manuf_elecon	(base alternative)					
manuf_heavy						
unempl_dur	-.0148817	.0014249	-10.44	0.000	-.0176745	-.012089
age	-.0754779	.074813	-1.01	0.313	-.2221087	.0711529
age2	.0012089	.0012399	0.97	0.330	-.0012213	.003639
female	2.15553	.1623102	13.28	0.000	1.837408	2.473652
ln_wage	.0362148	.07177	0.50	0.614	-.104452	.1768815
veneto_resid	.5966138	.0767017	7.78	0.000	.4462812	.7469464
_cons	1.411464	1.175378	1.20	0.230	-.8922342	3.715163

- Example of interpreting the coefficients: The estimate for the alternative-varying regressor e_ is positive ($\beta \approx 0.237$). This implies that an increase in experience in sector j leads to an increase in the probability of choosing sector j, while the probability of choosing the other alternatives decreases. We obtain consistent results with at-means marginal effects.

Conditional Logit

Marginal Effects

Pr(choice = manuf_heavy|1 selected) = .32037399

variable	dp/dx	Std. err.	z	P>|z|	[95% C.I.]		X
e_							
manuf_elecon	-.003246	.000213	-15.22	0.000	-.003664	-.002828	.87006
manuf_heavy	.05157	.000756	68.18	0.000	.050088	.053053	2.6819
manuf_light	-.009298	.000267	-34.78	0.000	-.009822	-.008774	2.2968
public_adm	-.002716	.000162	-16.74	0.000	-.003034	-.002398	.31251
public_hlth	-.006653	.000235	-28.33	0.000	-.007113	-.006193	.67189
serv_fin	-.001704	.000113	-15.12	0.000	-.001924	-.001483	.18468
serv_other	-.020046	.000453	-44.24	0.000	-.020934	-.019157	1.0052
serv_sales	-.007908	.000247	-31.98	0.000	-.008393	-.007424	.83105

- We get consistent results from the average marginal effects: positive own effects, negative cross effects (and this is true for all sectors).
- Example of interpretation: A one-year increase in experience in heavy manufacturing is associated with a 5 pp increase in the probability of choosing that sector after displacement.

Conditional Logit

Subscripts of the Coefficients

- Recall the *additive random utitlity model* we have seen before,

$$U_j = X\beta_j + Z_j\gamma + \varepsilon_j, \;\; j \in \{1, \ldots, J\}, J > 2$$

- The response probability:

$$\begin{aligned} p_j(x, z) &\equiv \mathbb{P}(y = j | X = x, Z = z) \\ &= \mathbb{P}(U_j \geq U_k) \;\; , \forall k \neq j \\ &= \mathbb{P}(\varepsilon_k - \varepsilon_j \leq x(\beta_j - \beta_k) + (z_j - z_k)\gamma) \;\; , \forall k \neq j \end{aligned}$$

- For the individual-specific regressors X, why should I have β_j instead of β?
  - Think about this: Can β be identified if it's the same for all alternatives?
- Think about identification for the subscripts of all coefficients.
  - γ is always identified.
  - For βs, we have $J-1$ differences to solve for J parameters, one of the errors need to be normalized. Only $J-1$ errors of $\{\varepsilon_1, \ldots, \varepsilon_J\}$ are free to vary, and similarly, only $J-1$ of $\{\beta_1, \ldots, \beta_J\}$ are free to vary.

Nested Logit

	(1)	(2)
	CLogit8	NestedL~t
industry		
e_	0.237***	0.278***
	(0.00274)	(0.00477)
manuf_he~y		
unempl_dur	-0.0149***	-0.0226***
	(0.00142)	(0.00213)
age	-0.0755	0.00932
	(0.0748)	(0.0412)
age2	0.00121	-0.000516
	(0.00124)	(0.000716)
female	2.156***	3.454***
	(0.162)	(0.277)
ln_wage	0.0362	0.139
	(0.0718)	(0.0962)
veneto_r~d	0.597***	1.097***
	(0.0767)	(0.118)
_cons	1.411	
	(1.175)	

- Compared to the conditional logit, the coefficient estimates for the nested logit are slightly larger in magnitude. Part of the reason is that we don't have a bottom level intercept, the coefficients absorb it.

Nested Logit

Dissimilarity Parameter (McFadden (1978) and Maddala (1983))

/type		
manufact~u		1.597***
		(0.0502)
services~u		1.157***
		(0.0493)
public_tau		0.673***
		(0.0344)
ll	-21561.7	-21471.2
N	125104	125104

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

- For first-level choice j (one of the 3 sectors) and second-level choices k and l (industries), the dissimilarity parameter is

$$\sigma_j = 1 - \sqrt{1 - Cor(\varepsilon_{jk}, \varepsilon_{jl})}$$

  The name dissimilarity parameter is intuitive as the parameter is a measure of the degree of independence in unobserved utility among the alternatives (industries) in the j-th limb (sector). $\rho_j = 1 - \sigma_j$ is called the *scale parameters* because they scale the regression parameters in the nested logit model.

Log Likelihood Function

Conditional Logit

Suppose we have m alternatives: $j = 1, \ldots, m$. The dependent variable y is defined to take value j if the j-th alternative is chosen:

$$y_{ij} = \mathbb{1}(y = j)$$

The probability that individual i chooses alternative j is

$$p_{ij} = \frac{e^{x'_{ij}\beta + z'_i\gamma_j}}{\sum_{l=1}^{m} e^{x'_{il}\beta + z'_i\gamma_l}}$$

We use the "mixed" formulation of the utilities for teaching purposes. The likelihood function is

$$\mathcal{L} = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln p_{ij}$$

with all the components defined above.

Log Likelihood Function

Nested Logit

Suppose at the top level there are J limbs to choose from. The j-th limb has K_j branches. Denote by y_{ijk} the indicator for individual i choosing the branch k of limb j. The deterministic utility function is

$$\begin{aligned} V_{jk} &= \texttt{exper}'_{jk}\beta + w'\gamma_{jk} \\ &= x'_{jk}\beta + w'\gamma_{jk} \end{aligned}$$

The choice probability for branch k of of limb j is

$$p_j \times p_{k|j} = \frac{e^{\rho_j I_j}}{\sum_{l=1}^{J} e^{\rho_l I_l}} \times \frac{e^{\frac{1}{\rho_j}\left(x'_{jk}\beta + w'\gamma_{jk}\right)}}{\sum_{s=1}^{K_j} e^{\frac{1}{\rho_j}\left(x'_{js}\beta + w'\gamma_{js}\right)}}$$

$$\text{where} \quad I_j = \ln\left(\sum_{k=1}^{K_j} e^{\frac{1}{\rho_j}\left(x'_{jk}\beta + w'\gamma_{jk}\right)}\right)$$

Note that in the specification we used in this exercise, we don't have a regressor that affects the choice of the first level.

Log Likelihood Function

Nested Logit

The FIML estimator maximizes

$$\mathcal{L}(\alpha, \{\beta_j\}_j, \{\rho_j\}_j) = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \ln p_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K_j} y_{ijk} \ln p_{ik|j}$$

In estimation, we might use sequential estimation (LIML) to provide starting values for FIML when its log-likelihood is not globally concave.

Nested Logit

Subscripts of the Coefficients

- Things are similar to that we have for the conditional logit.
- The deterministic part of the utility:

$$V_{jk} = X\beta_j + Z_j\gamma + Y_{jk}\alpha_j$$

- Why α_j? Consider this: if instead, we have α, can we identify ρ_j?

$$p_{jk} = p_j \times p_{k|j} = \underbrace{\frac{e^{X\beta_j + Z_j\gamma + \rho_j I_j}}{\sum_{m=1}^{J} e^{X\beta_m + Z_m\gamma + \rho_m I_m}}}_{\rightarrow\ \widehat{\beta_j}, \widehat{\gamma}, \widehat{\rho_j}} \times \underbrace{\frac{e^{Y_{jk}\alpha_j/\rho_j}}{\sum_{l=1}^{K_j} e^{Y_{jl}\alpha_j/\rho_j}}}_{\rightarrow\ \widehat{\alpha_j/\rho_j}}$$

where ρ_j is the dissimilarity parameter and I_j is the inclusive value or log-sum:

$$I_j = \ln\left(\sum_{k=1}^{K_j} e^{Y_{jk}\alpha_j/\rho_j}\right)$$

Nested Logit

Dissimilarity Parameter

- You can relate the dissimilarity parameter ρ_j to the *elasticity of substitution* you have seen in
  - utility functions (when there are at least two goods, CRRA is an example), and
  - production functions (when there are more than one inputs, capital-skill complementarity is an example).
- The grouping of alternatives is a modeling decision; alternatives with a high degree of substitutability should be placed in the same group.
- Therefore, even in Exercise 2, you don't have a regressor that affects the choice of the first level, the first level still exists.
- If a group has a single alternative, its dissimilarity parameter is not identified, so it should be set to one.
- Some authors interpret the nested model as a nested sequential choice, which may help structure the groupings but is not technically correct. The correct interpretation is degree of substituability, not the timing of decisions (textbook Hansen (2022), Ch26.6).