Practice Questions 1

---

1. (a) Consider the latent variable model, $y_n^* = \beta' x_n + \varepsilon_n$, $y_n = 1\{y_n^* \geq 0\}$, and $\varepsilon_n \sim N(0,1)$, where (x_n, ε_n) are independent. Derive the probability model for $\Pr(y_n = 1 \mid x_n)$ from the latent variable formulation.

(b) In the model above, it was assumed that $\varepsilon_n \sim N(0,1)$ rather than $\varepsilon_n \sim N(0,\sigma^2)$. Why do we typically assume that $\sigma^2 = 1$ in a probit model?

(c) Derive the log-likelihood function for the probit model.

2. What conditions are necessary for identifying an ordered probit model?

3. (a) Find the maximum of the function $f(x,y) = (\frac{6}{5} - e^{-\frac{1}{4}(x-1)^2})(-1-(y-2x-4)^2)$ using numerical optimization.

(b) Demonstrate that the potential maximum you found is a local maximum.

(c) Can we be sure that this point is a global maximum?

4. Consider the data file mexico.dta which can be accessed using the link, https://sites.google.com/a/stonybrook.edu/mperess/teaching/classdata/mexico.dta. This data file is a post-election study of the 2000 Mexican presidential election, with some aggregate data merged in. We will use this data file to study vote buying by a party machine. The variables in the data set are:

vote_pri - A dummy variable equal to 1 if the respondent voted for the PRI, 0 if the respondent voted for a different party, and missing if the respondent did not vote

gift_pri – A dummy variable equal to 1 if the respondent reported receiving a gift from the PRI in exchange for their vote

approvemucho, approvealgo, disapprovealgo, and disapprovemucho – Dummy variables indicating that the respondent highly approves, somewhat approves, somewhat disapproves, or highly disapproves (respectively) of the incumbent president, who is from the PRI

urban – whether the respondent lives in an urban or rural locality

female – whether the respondent is female

educ – an index indicating the respondent's level of education

employed – whether the respondent is currently employed

dist – the electoral district the respondent resides in. There are about 300 of these in Mexico, and the sample includes 100 such districts. Note that urban does not vary within any electoral district except one, so urban is effectively collinear with dist

pri_97 – The vote share of the PRI in the previous presidential election in the electoral district. All respondents in the same election district should have the same value for pri_97

(i) Estimate a logit model predicting gift_pri using the variables approvemucho, approvealgo, disapprovealgo, disapprovemucho, urban, female, educ, and employed and interpret the coefficients and standard errors.

(ii) Cox and McCubbins (1986) have argued that party machines are more likely to buy voters in their core districts—the idea being that in these districts, the party machine controls the district, so it is cheaper for them to provide these benefits where their machine is strong. An alternative hypothesis is that the party machine will buy swing voters—swing voters should be

those that weakly approve or weakly disapprove of the incumbent PRI president. A second alternative theory is that the party machine will buy voters in swing districts, since swing districts will have the most persuadable voters. Test the Cox and McCubbins theory by adding pri_97 and it's square term (pri_97_sq) to the logit model estimated in (i). Note that since Mexico is not a two party system, a swing district is on necessarily a district where the PRI had a 50% vote share in 1997. Comment on the results.

(c) Calculate some measure of the substantive effect of pri_97 on the probability of observing a value of 1 for the DV (you don't have to include standard errors for the effect size).

(d) Would it make sense to use robust standard errors in the models estimated in (a) or (b)?

(e) What are the disadvantages and advantages of using a logit model vs. using a linear probability model in (a) and (b)?

(f) What are the disadvantages and advantages of using a logit model vs. using a probit model in (a) and (b)?

5. Consider the first column in the table below. The dependent variable is a 5 point scale where respondents were asked whether in the future, they were extremely likely, very likely, moderately likely, a little likely, and not at all likely to join in a protest march, rally, or demonstration. Risk acceptance is a scale that ranges from 0 to 1 with 0 being risk averse and 1 being risk accepting. Strength of partisanship and ideology are the standard 7 point scales from the ANES. Education and Income are 5 and 6 points scales, respectively.

(a) Interpret the variables in the model, as best you can, given the information given.

(b) The main independent variable of interest in this study is the risk acceptance scale. Can we say anything about the magnitude of the effect of risk acceptance (at least, relative to some of the other variables?)

TABLE 3 ***Risk Acceptance*** **and Future Political Participation**

	Rally	Local Meeting	E-petition	Paper Petition	Pol/Soc Donation	Pol/Soc Meeting	Recruit for Meeting	Hand Out Info	Religious Donation
Risk	1.01***	0.62*	0.74**	0.81**	0.20	0.78**	1.22***	0.85**	−0.77**
Acceptance	0.38	0.32	0.36	0.36	0.39	0.38	0.40	0.42	0.38
Female	−0.14	0.13	0.07	0.12	−0.02	0.03	0.04	−0.16	0.34***
	0.11	0.09	0.10	0.10	0.10	0.11	0.11	0.12	0.09
Age	−0.27	0.37*	−0.05	0.46*	0.42	0.44	0.06	0.23	0.33
	0.27	0.21	0.23	0.25	0.26	0.27	0.29	0.30	0.23
Strength of	0.18	0.16	0.16	0.28**	0.22	0.13	0.23	0.32**	0.51***
Partisanship	0.15	0.13	0.13	0.13	0.14	0.14	0.15	0.16	0.14
Ideology	−0.60***	0.10	−0.24*	−0.15	−0.33**	−0.01	−0.05	−0.14	0.95***
	0.17	0.14	0.14	0.15	0.14	0.15	0.16	0.16	0.17
Education	0.41*	0.65***	0.34*	0.31*	0.76***	0.65***	0.17	0.38*	0.19
	0.22	0.18	0.18	0.18	0.21	0.19	0.19	0.21	0.20
Income	−0.48**	−0.03	−0.33**	−0.24*	0.26*	−0.18	−0.10	−0.31**	0.33**
	0.19	0.15	0.16	0.14	0.16	0.15	0.15	0.15	0.16
Black	0.11	0.11	−0.13	−0.05	−0.11	0.06	0.12	0.16	−0.02
	0.17	0.17	0.19	0.19	0.17	0.19	0.17	0.18	0.17
Hispanic	0.85***	0.13	0.27*	0.25*	0.17	0.26	0.33	0.38	−0.17
	0.20	0.16	0.16	0.16	0.20	0.18	0.21	0.25	0.16
τ_1	0.46	0.36	−0.04	−0.14	0.31	0.56	0.95	0.94	0.32
	0.33	0.27	0.30	0.30	0.34	0.36	0.39	0.41	0.29
τ_2	1.03	1.20	0.55	0.72	0.99	1.29	1.70	1.56	0.75
	0.34	0.27	0.30	0.31	0.35	0.37	0.40	0.42	0.29
τ_3	1.89	1.83	1.10	1.36	1.59	2.00	2.31	2.06	1.08
	0.32	0.27	0.31	0.32	0.35	0.39	0.42	0.44	0.30
τ_4	2.47	2.51	1.65	2.01	2.23	2.65	2.73	2.57	1.58
	0.34	0.28	0.32	0.33	0.38	0.42	0.46	0.50	0.30
N	1021	1022	1022	1022	1022	1022	1022	1022	1022

Table entry is the ordered probit regression coefficient with linearized standard error below. Weighted analysis.
Dependent variable is scaled from 0 (not at all likely) to 1 (extremely likely).
***$p < 0.01$; **$p < 0.05$; *$p < 0.10$.