

# ECON 710B - Problem Set 11

Alex von Hafften\*

4/20/2021

## Exercise 25.1

Emily estimates a probit regression setting her dependent variable to equal  $Y = 1$  for a purchase and  $Y = 0$  for no purchase. Using the same data and regressors, Jacob estimates a probit regression setting the dependent variable to equal  $Y = 1$  if there is no purchase and  $Y = 0$  for a purchase. What is the difference in their estimated slope coefficients?

If Emily estimates  $\hat{\beta}$ , then Jacob's estimates are  $-\hat{\beta}$ .

Let  $Y_J = 1 - Y_E$  and  $e_E = -e_J$ :

$$\begin{aligned} Y_E &= \Phi(X'\beta) + e_E \\ 1 - Y_J &= \Phi(X'\beta) + e_E \\ Y_J &= 1 - \Phi(X'\beta) - e_E \\ &= 1 - P(Z < X'\beta) - e_E \\ &= P(Z > X'\beta) - e_E \\ &= P(Z < -X'\beta) - e_E \\ &= \Phi(-X'\beta) + e_J \end{aligned}$$

Where  $Z \sim N(0, 1)$ . Notice that  $E[e_E|X] = 0 \implies E[e_J|X] = 0$ .

---

\*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Sarah Bass, Emily Case, Danny Edgel, and Katherine Kwok.

### Exercise 25.3

Show (25.1) and (25.2). From Hansen (p. 806):

The variables satisfy the regression framework:  $Y = P(X) + e$  with  $E[e|X] = 0$ . The error  $e$  is not “classical.” It has the two-point conditional distribution:

$$e = \begin{cases} 1 - P(X), & \text{with probability } P(X) \\ -P(X), & \text{with probability } 1 - P(X) \end{cases} \quad (25.1)$$

It is also highly heteroskedastic with conditional variance:

$$\text{var}(e|X) = P(X)(1 - P(X)) \quad (25.2)$$

For (25.1), notice that

$$\begin{aligned} P(Y = 1|X) &= E[Y|X] = E[P(X) + e|X] = P(X) \\ P(Y = 0|X) &= 1 - P(Y = 1|X) = 1 - P(X) \end{aligned}$$

If  $Y = 1$ , then  $1 = P(X) + e \implies e = 1 - P(X)$ . If  $Y = 0$ , then  $0 = P(X) + e \implies e = -P(X)$ .

For (25.2),

$$\begin{aligned} \text{var}(e|X) &= E[e^2|X] \\ &= [1 - P(X)]^2 P(X) + [-P(X)]^2 [1 - P(X)] \\ &= P(X) - 2P(X)^2 + P(X)^3 + P(X)^2 - P(X)^3 \\ &= P(X)(1 - P(X)) \end{aligned}$$

### Exercise 25.9

Find the first-order condition for the logit MLE  $\hat{\beta}^{logit}$ .

From Hansen (p. 810), we know that the log-likelihood function for the logit model is:

$$\ell_n(\beta) = \sum_{i=1}^n \log \Lambda(Z_i' \beta) = \sum_{i=1}^n \log[(1 + \exp(-Z_i' \beta))^{-1}] = - \sum_{i=1}^n \log[1 + \exp(-Z_i' \beta)]$$

where  $Z_i = \begin{cases} X_i & \text{if } Y_i = 1 \\ -X_i & \text{if } Y_i = 0 \end{cases}$ . The FOC with respect to  $\beta$  is:

$$\begin{aligned} 0 &= - \sum_{i=1}^n [1 + \exp(-Z_i' \beta)]^{-1} \exp(-Z_i' \beta) (-Z_i) \\ \implies 0 &= \sum_{i=1}^n \Lambda(Z_i' \beta) \exp(-Z_i' \beta) Z_i \end{aligned}$$

### Exercise 25.12

Show how to use NLLS to estimate a probit model.

The probit model is  $Y = \Phi(X' \beta) + e = m(X, \beta) + e$  with  $E[e|X] = 0$  and  $m(X, \beta) = \Phi(X' \beta)$ . Thus, the sample average of squared errors is

$$S_n(\beta) = \frac{1}{n} \sum_{i=1}^n (Y_i - m(X_i, \beta))^2$$

Thus,  $\hat{\beta}_{nlls} = \arg \min_{\beta \in \mathbb{R}^k} S_n(\beta)$ .

## Exercise 25.14

Take the heteroskedastic nonparametric binary choice model:

$$\begin{aligned} Y^* &= m(X) + e \\ e|X &\sim N(0, \sigma^2(X)) \\ Y &= Y^* \mathbb{1}\{Y^* > 0\} \end{aligned}$$

The observables are  $\{Y_i, X_i : i = 1, \dots, n\}$ . The functions  $m(x)$  and  $\sigma^2(x)$  are nonparametric.

- (a) Find a formula for the response probability.

$$\begin{aligned} P(Y > 0|X) &= P(Y^* \mathbb{1}\{Y^* > 0\} > 0|X) \\ &= P(Y^* > 0|X) \\ &= P(m(X) + e > 0|X) \\ &= P(e > -m(X)|X) \\ &= 1 - P(e < -m(X)|X) \\ &= 1 - F(-m(X)) \end{aligned}$$

where  $F$  is the CDF of a random variable distributed  $N(0, \sigma^2(X))$ .

- (b) Are  $m(x)$  and  $\sigma^2(x)$  both identified? Explain.

No, if  $(m(x), \sigma^2(x))$  is a solution, then  $(a * m(x), a^2 * \sigma^2(x))$  is also a solution. So,  $m(x)$  and  $\sigma^2(x)$  are not identified.

- (c) Find a normalization which achieves identification.

Normalize  $\sigma^2(X) = 1$ , then we can identify  $\tilde{m}(X)$ , which is a transformed version of  $m(X)$  for the normalized values of  $e$ .

- (d) Given your answer to part (c), does it make sense to “allow for heteroskedasticity” in the binary choice model? Explain?

No, because  $m(X)$  and  $\sigma^2(X)$  cannot both be identified, we need to normalize  $e$  to be homoskedastic.

## Exeriser 25.15

Use the `cps09mar` dataset and the subset of men. Set  $Y = 1$  if the individual is a member of a labor union (`union=1`) and  $Y = 0$  otherwise. Estimate a probit model as a linear function of `age`, `education`, and indicators for Black individuals and for Hispanic individuals. Report the coefficient estimates and standard errors. Interpret the results.

```
cps09mar <- read_delim(file = "cps09mar.txt",
  delim = "\t",
  col_names = c("age", "female", "hisp", "education", "earnings", "hours",
    "week", "union", "uncov", "region", "race", "marital"),
  col_types = cols())

sample_25_15 <- cps09mar %>%
  filter(female == 0) %>%
  mutate(black = as.numeric(race == 2))

probit <- glm(union ~ age + education + black + hisp,
  family = binomial(link = "probit"),
  data = sample_25_15)

sum(sample_25_15$union)/nrow(sample_25_15)

## [1] 0.02254633

coeftest(probit, vcov. = vcovHC, type = "HC1")

##
## z test of coefficients:
##
##               Estimate Std. Error  z value  Pr(>|z|)
## (Intercept) -1.9538182   0.0977904 -19.9796 < 2.2e-16 ***
## age          0.0079254   0.0013358   5.9330 2.974e-09 ***
## education    -0.0255046   0.0055826  -4.5686 4.910e-06 ***
## black        -0.0540830   0.0603602  -0.8960  0.3703
## hisp         -0.2977449   0.0598795  -4.9724 6.613e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

margins_summary(probit) %>% kable(digits = 4)
```

factor	AME	SE	z	p	lower	upper
age	0.0004	0.0001	5.5108	0.0000	0.0003	0.0006
black	-0.0029	0.0032	-0.9010	0.3676	-0.0091	0.0034
education	-0.0014	0.0003	-4.0706	0.0000	-0.0020	-0.0007
hisp	-0.0158	0.0031	-5.1742	0.0000	-0.0218	-0.0098

First, note that about 2.2 percent of the sample are in a union.

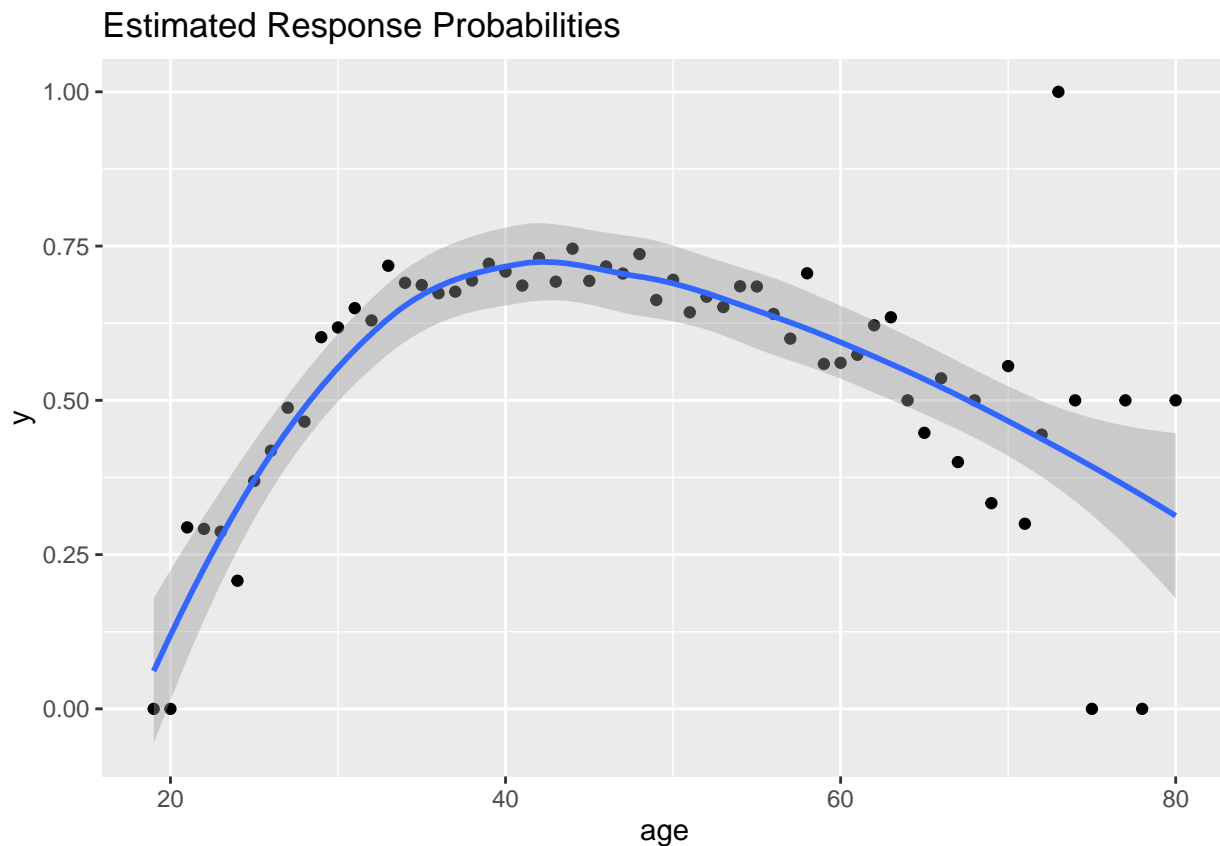
Second, looking at the regression table, we see that the coefficients for **age**, **education**, and **hisp** are statistically significant. The coefficient for **black** is not statistically different than zero and the intercept is not very meaning because probit models are only identified up to scale. While the magnitude of the significant coefficients is difficult to interpret, we can consider at the sign. The coefficient on **age** is positive suggesting that the older a man in the sample the more likely he is in a union. The coefficient on **education** is negative suggesting that more education makes someone less likely to be in a union. The coefficient on **hisp** is also negative suggesting that it is less likely for men in the sample who identified as Hispanic to be in a union.

Third, we can look at the average marginal effects table to get a sense of the economic magnitude of these effects. Note that the unconditional probability of being a union member is 2.2 percent, so these effects are generally large. The average marginal effect of **age** means that if you're 10 years old you're about 0.4 percentage points more likely to be in a union. The average marginal effect of **education** is larger, but similarly small. One more year of education reduces the probability of being in a union by 0.1 percentage points. The marginal effect of **hisp** is even larger. Identify as Hispanic reduces the likelihood of being in a union by about 1.5 percentage points. For context, out of Hispanic men in the sample, about 1.3 percent are in a union.

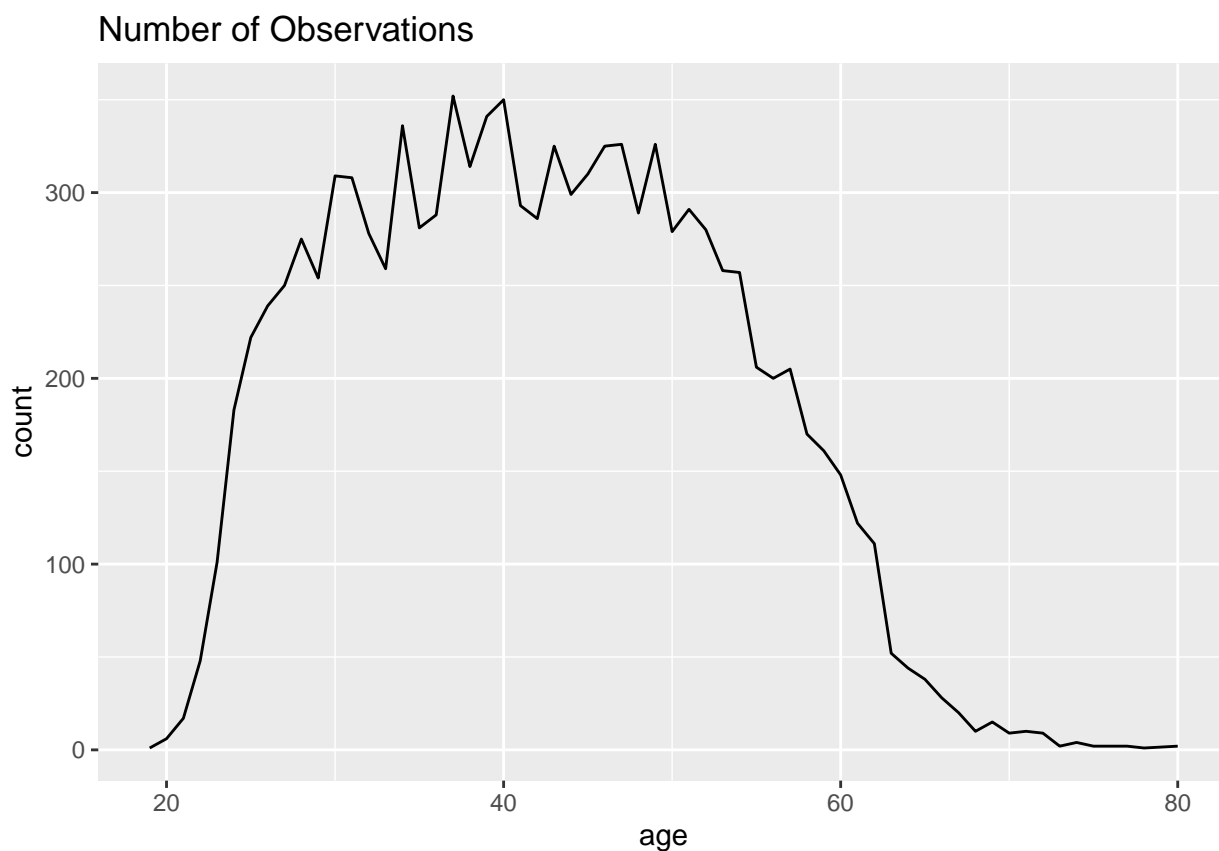
## Exerciser 25.17

Use the `cps09mar` dataset and the subset of women with a college degree. Set  $Y = 1$  if marital equals 1, 2, or 3, and set  $Y = 0$  otherwise. Estimate a binary choice model for  $Y$  as a possibly nonlinear function of age. Describe the motivation for the model you use. Plot the estimated response probability. How do the estimates compare with those for men from Figure 25.1?

```
sample_25_17 <- cps09mar %>%  
  filter(female == 1, education >= 14) %>%  
  mutate(y = as.numeric(marital %in% 1:3))  
  
erp <- sample_25_17 %>% group_by(age) %>% summarise(y = mean(y), count = n())  
  
erp %>%  
  ggplot(aes(y = y, x = age)) +  
  geom_point() +  
  geom_smooth(method = 'loess', formula = y ~ x) +  
  ggtitle("Estimated Response Probabilities")
```



```
erp %>%  
  ggplot(aes(y = count, x = age)) +  
  geom_line() +  
  ggtitle("Number of Observations")
```



```
series_1 <- glm(y ~ age, family = binomial(link = "probit"), data = sample_25_17)
series_2 <- glm(y ~ age + I(age^2), family = binomial(link = "probit"),
               data = sample_25_17)
series_3 <- glm(y ~ age + I(age^2) + I(age^3), family = binomial(link = "probit"),
               data = sample_25_17)
series_4 <- glm(y ~ age + I(age^2) + I(age^3) + I(age^4),
               family = binomial(link = "probit"), data = sample_25_17)

se_1 <- sqrt(diag(vcovHC(series_1, type = "HC1")))
se_2 <- sqrt(diag(vcovHC(series_2, type = "HC1")))
se_3 <- sqrt(diag(vcovHC(series_3, type = "HC1")))
se_4 <- sqrt(diag(vcovHC(series_4, type = "HC1")))

stargazer(series_1, series_2, series_3, series_4,
          se = list(se_1, se_2, se_3, se_4),
          omit.stat = c("f", "aic", "ll"),
          header = FALSE, title = "Series Probit Estimates", float = FALSE)
```



<i>Dependent variable:</i>				
	y			
	(1)	(2)	(3)	(4)
age	0.010*** (0.001)	0.168*** (0.009)	0.449*** (0.040)	0.971*** (0.181)
I(age^2)		-0.002*** (0.0001)	-0.008*** (0.001)	-0.027*** (0.006)
I(age^3)			0.00005*** (0.00001)	0.0003*** (0.0001)
I(age^4)				-0.00000*** (0.00000)
Constant	-0.066 (0.052)	-3.240*** (0.179)	-7.020*** (0.554)	-12.348*** (1.879)
Observations	10,727	10,727	10,727	10,727

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

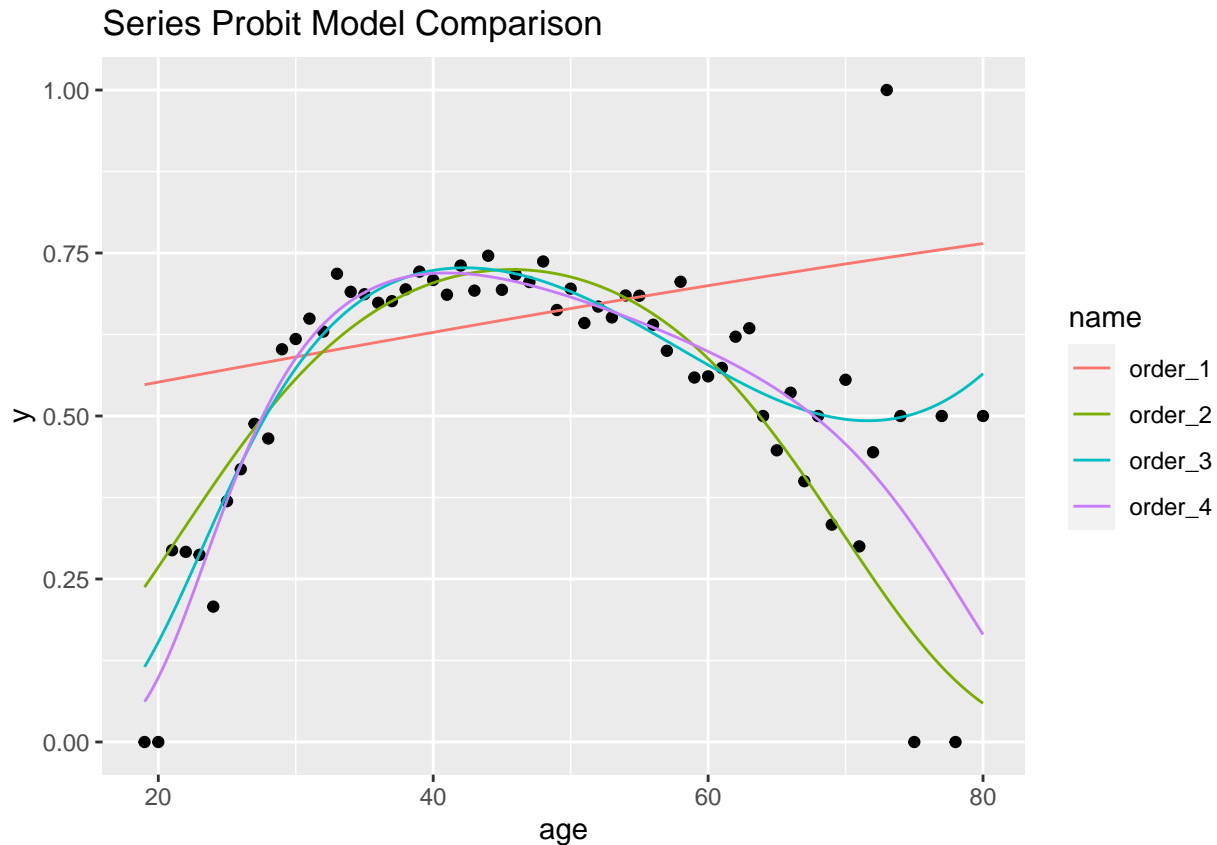
```
bind_rows(margins_summary(series_1),
          margins_summary(series_2),
          margins_summary(series_3),
          margins_summary(series_4)) %>%
mutate(order = 1:4) %>%
kable(digits = 4,
      caption = "Average Marginal Effects")
```

Table 2: Average Marginal Effects

factor	AME	SE	z	p	lower	upper	order
age	0.0037	0.0004	8.7016	0.000	0.0028	0.0045	1
age	0.0052	0.0004	12.1136	0.000	0.0044	0.0061	2
age	0.0070	0.0005	15.0577	0.000	0.0061	0.0080	3
age	0.0078	0.0217	0.3611	0.718	-0.0346	0.0503	4

```
prediction <- tibble(age = seq(min(sample_25_17$age),max(sample_25_17$age), 0.01)) %>%
  mutate(order_1 = predict(series_1, list(age = age), type = "response"),
         order_2 = predict(series_2, list(age = age), type = "response"),
         order_3 = predict(series_3, list(age = age), type = "response"),
         order_4 = predict(series_4, list(age = age), type = "response")) %>%
  pivot_longer(-age)

prediction %>%
  ggplot() +
  geom_point(data = erp, aes(y = y, x = age)) +
  geom_line(aes(y = value, x = age, color = name)) +
  ggtitle("Series Probit Model Comparison")
```



First, I plotted the estimated response probabilities, which is the proportion of married people for each age. I added a local polynomial regression line to get a sense for the shape of the distribution. I think there are two regions to consider. Between 20 and 40, the estimated response probabilities increase steeply and then at 40 the estimated response probabilities decrease at a gentler slope. The variance dramatically increases at around 65, which is consistent with a dramatic decrease in the number of observations. Thus, I think we shouldn't try to find a regression that works well for the right tail of the age distribution.

Second, I estimate series probit models in order to better approximate the nonlinear function of age. I use orders 1, 2, 3, and 4.

Third, since the coefficients of the series probit models are difficult to interpret, I estimate the average marginal effects of age for each model. The size of the average marginal effect is increasing with the polynomial order.

Fourth, I plot the fitted response probabilities from each regression. Clearly, the series probit with order 1 is just a normal probit and as I guess from looking at the estimates response probabilities, it doesn't fit the data well. The series regressions with orders 2, 3, and 4 fit pretty well. As I mentioned above, the number of observations is very low for ages over 65. The order 2 series approximation is slightly too high at the low and high ends and slightly off in the middle. For the relevant window, orders 3 and 4 are very close. My preferred model is the approximation with order 3 because of you do not get much more with a fourth order approximation. In addition, the change in the marginal effect estimate is significantly different going from order 1 to order 2 and order 2 to order 3, but it is insignificant when you add the fourth term. The third order approximation begins to increase in the late-70s, which I think is unfortunate because artifact of the approximation and not a feature of the data.

The data for women is very different from the data for men in figure 25.1. Slightly later peak (40 versus 35) and a much more dramatic decrease after 40. The estimates for men also increased substantial in the late 70s and early 80s which is not apparent for women, but this could similarly be an issue with small sample sizes.

### Exercise 26.1

For the multinomial logit model (26.2), show that  $0 \leq P_j(x) \leq 1$  and  $\sum_{j=1}^J P_j(x) = 1$ .

$$P_j(x) = \frac{\exp(x' \beta_j)}{\sum_{\ell=1}^J \exp(x' \beta_\ell)} \quad (26.2)$$

Notice that  $\exp(x)$  is nonnegative for all  $x \in \mathbb{R}$ , so  $P_j(x) \geq 0$ . Furthermore,

$$P_j(x) = \frac{\exp(x' \beta_j)}{\sum_{\ell=1}^J \exp(x' \beta_\ell)} = \frac{\exp(x' \beta_j)}{\exp(x' \beta_j) + \sum_{\ell=1; \ell \neq j}^J \exp(x' \beta_\ell)} \leq 1$$

Furthermore:

$$\sum_{j=1}^J P_j(x) = \sum_{j=1}^J \frac{\exp(x' \beta_j)}{\sum_{\ell=1}^J \exp(x' \beta_\ell)} = \left( \frac{1}{\sum_{\ell=1}^J \exp(x' \beta_\ell)} \right) \sum_{j=1}^J \exp(x' \beta_j) = 1$$

### Exercise 26.3

For the multinomial logit model (26.2) show that the marginal effects equal (26.4).

$$\delta_j(x) = \frac{\partial}{\partial x} P_j(x) = P_j(x) \left( \beta_j - \sum_{\ell=1}^J \beta_\ell P_\ell(x) \right) \quad (26.4)$$

$$\begin{aligned} \delta_j(x) &= \frac{\partial}{\partial x} P_j(x) \\ &= \frac{\partial}{\partial x} \frac{\exp(x' \beta_j)}{\sum_{\ell=1}^J \exp(x' \beta_\ell)} \\ &= \frac{\beta_j \exp(x' \beta_j)}{\sum_{\ell=1}^J \exp(x' \beta_\ell)} + \frac{\exp(x' \beta_j) \sum_{\ell=1}^J \beta_\ell \exp(x' \beta_\ell)}{[\sum_{w=1}^J \exp(x' \beta_w)]^2} \\ &= \frac{\exp(x' \beta_j)}{\sum_{\ell=1}^J \exp(x' \beta_\ell)} \left( \beta_j - \sum_{\ell=1}^J \beta_\ell \frac{\exp(x' \beta_\ell)}{\sum_{w=1}^J \exp(x' \beta_w)} \right) \\ &= P_j(x) \left( \beta_j - \sum_{\ell=1}^J \beta_\ell P_\ell(x) \right) \end{aligned}$$

### Exerciser 26.7

In the conditional logit model, find an estimator for  $AME_{jj}$ .

From Hansen (p. 828):

$$AME_{jj} = E[\delta_{jj}(W, X)] = E[\gamma P_j(W, X)(1 - P_j(E, X))]$$

Replacing with estimated coefficients and sample moments:

$$\hat{AME}_{jj} = \hat{\gamma} \frac{1}{n} \sum_{i=1}^n \hat{P}_j(W_i, X_i)(1 - \hat{P}_j(W_i, X_i))$$

Where  $\hat{P}_j(W_i, X_i) = \frac{\exp(W_i' \hat{\beta}_j + x_i' \gamma)}{\sum_{\ell=1}^J \exp(W_i' \hat{\beta}_\ell + x_i' \gamma)}$ .

### Exerciser 26.8

Show (26.11).

$$\frac{P_j(W, X|\theta)}{P_\ell(W, X|\theta)} = \frac{\exp(W' \beta_j + X' \gamma)}{\exp(W' \beta_\ell + X' \gamma)}$$

$$\frac{P_j(W, X|\theta)}{P_\ell(W, X|\theta)} = \frac{\frac{\exp(w' \beta_j + x' \gamma)}{\sum_{s=1}^J \exp(w' \beta_s + x' \gamma)}}{\frac{\exp(w' \beta_\ell + x' \gamma)}{\sum_{s=1}^J \exp(w' \beta_s + x' \gamma)}} = \frac{\exp(W' \beta_j + X' \gamma)}{\exp(W' \beta_\ell + X' \gamma)}$$