# AAEC 6312: Homework 2
# Due October 26, 2022

1. Mortgage lenders are interested in determining borrower and loan factors that may lead to delinquency or foreclosure. In the file *lasvegas.xlsx* are 1,000 observations on mortgages for single family homes in Las Vegas, Nevada during 2008. The variable of interest is $DELINQUENT$, an indicator variable = 1 if the borrower missed at least three payments (90þ days late), but 0 otherwise. Explanatory variables: are $LVR$ = the ratio of the loan amount to the value of the property; $REF = 1$ if purpose of the loan was a "refinance" and= 0 if loan was for a purchase;$INSUR = 1$ if mortgage carries mortgage insurance, 0 otherwise; $RATE$ = initial interest rate of the mortgage; $AMOUNT$ = dollar value of mortgage (in \$100,000); $CREDIT$ = credit score, $TERM$ = number of years between disbursement of the loan and the date it is expected to be fully repaid, $ARM = 1$ if mortgage has an adjustable rate, and = 0 if mortgage has a fixed rate.

   (a) Estimate the linear probability (regression) model explaining $DELINQUENT$ as a function of the remaining variables. Use White robust standard errors. Are the signs of the estimated coefficients reasonable?

   (b) Use probit to estimate the model in (a). Are the signs and significance of the estimated coefficients the same as for the linear probability model?

   (c) Compute the predicted value of $DELINQUENT$ for the 500th and 1,000th observations using both the linear probability model and the probit model. Interpret the values.

   (d) Construct a histogram of $CREDIT$. Using both linear probability and probit models, calculate the probability of delinquency for$CREDIT = 500, 600$, and 700 for a loan of \$250,000 ($AMOUNT = 2.5$). For the other variables, loan to value ratio ($LVR$) is 80%, initial interest rate is 8%, indicator variables take the value one, and $TERM = 30$. Discuss similarities and differences among the predicted probabilities from the two models.

   (e) Compute the marginal effect of $CREDIT$ on the probability of delinquency for $CREDIT = 500, 600$, and 700, given that the other explanatory variables take the values in (d). Discuss the interpretation of the marginal effect.

   (f) Construct a histogram of $LVR$. Using both linear probability and probit models, calculate theprobabilityofdelinquencyfor $LVR = 20$ and $LVR = 80$, with $CREDIT = 600$ and other variables set as they are in (d). Compare and contrast the results.

   (g) Compare the percentage of correct predictions from the linear probability model and the probit model, using a predicted probability of 0.5 as the threshold.

   (h) As a loan officer, you wish to provide loans to customers who repay on schedule and are not delinquent. Suppose you have available to you the first 500 observations in the data on which to base your loan decision on the second 500 applications (501–1,000). Is using the probit model, with a threshold of 0.5 for the predicted probability the best decision rule for deciding on loan applications? If not, what is a better rule.

2. Use the data on college choice contained in *nels_small.xlsx*.

   (a) Estimate a multinomial logit model explaining $PSECHOICE$. Use the group who did not attend college as the base group. Use as explanatory variables $GRADES$, $FAMINC$, $FEMALE$, and $BLACK$. Are the estimated coefficients statistically significant?

   (b) Compute the estimated probability that a white male student with median values of $GRADES$ and $FAMINC$ will attend a four-year college.

   (c) Compute the probability ratio that a white male student with median values of $GRADES$ and $FAMINC$ will attend a four-year college rather than not attend any college.

(d) Compute the change in probability of attending a four-year college for a white male student with median $FAMINC$ whose $GRADES$ change from 6.64 (the median value) to 4.905 (top 25th percentile).

(e) From the full data set create a subsample, omitting the group who attended a two-year college. Estimate a logit model explaining student's choice between attending a four-year college and not attending college, using the same explanatory variables in (a). Compute the probability ratio that a white male student with median values of $GRADES$ and $FAMINC$ will attend a four-year college rather than not attend any college. Compare the result to that in (c).

3. Consider a conditional logit model of choice among three brands of soda: Coke, Pepsi, and 7-Up. The data are in the file *cola.xlsx*.

(a) In addition to $PRICE$, the data file contains variables indicating whether the product was "featured" at the time ($FEATURE$) or whether there was a store display ($DISPLAY$). Estimate a conditional logit model explaining choice of soda using $PRICE$, $DISPLAY$, and $FEATURE$ as explanatory variables. Discuss the signs of the estimated coefficients and their significance.

(b) Compute the probability ratio of choosing *Coke* relative to *Pepsi* and $7-Up$ if the price of each is \$1.25 and no display or feature is present.

(c) Compute the probability ratio of choosing *Coke* relative to *Pepsi* and $7-Up$ if the price of each is \$1.25, a display is present for *Coke* but not for the others, and none of the items is featured.

(d) Compute the change in the probability of purchase of each type of soda if the price of *Coke* changes from \$1.25 to \$1.30, with the prices of the *Pepsi* and $7-Up$ remaining at \$1.25. Assume that a display is present for *Coke*, but not for the others, and none of the items is featured.

(e) Add the alternative specific "intercept" terms for *Pepsi* and $7-Up$ to the model in (a). Estimate the conditional logit model. Compute the probability ratio in (c) based upon these new estimates.

(f) Based on the estimates in (e), calculate the effects of the price change in (d) on the choice probability for each brand.

(g) Estimate a nested logit model with cola and non-cola nest, and repeat (d).

4. In the data $nels_small.xlsx$, consider the variable $PSECHOICE$ as an ordered variable with 1 representing the least favored alternative (no college) and 3 denoting the most favored alternative (four-year college).

(a) Use an ordered probit to explain the probability of $PSECHOICE$ as a function of $GRADES$. Calculate the probability that a student will choose no college, a two-year college, and a four-year college if the student's grades are the median value, $GRADES = 6.64$. Recompute these probabilities assuming that $GRADES = 4.905$. Discuss the probability changes. Are they what you anticipated? Explain.

(b) Expand the ordered probit model to include family income ($FAMINC$), family size ($FAMSIZ$), and the indicator variables $BLACK$ and $PARCOLL$. Discuss the estimates and their signs and significance.

(c) Test the joint significance of the variables added in (b) using a likelihood ratio test.

(d) Compute the probability that a black student from a household of four members, including a parent who went to college, and household income of \$52,000, will attend a four-year college if (i) $GRADES = 6.64$ and (ii) $GRADES = 4.905$.

(e) Repeat (d) for a "nonblack" student and discuss the differences in your findings.

5. **Censoring/Truncation**. Greene (2007) analyzed the default behavior and monthly behavior of a large sample of credit card users (13,444). The Data consist of

- Cardhldr = Dummy variable, 1 if application for credit card accepted, 0 if **not**
- Default = 1 if defaulted 0 if not (observed when Cardhldr = 1, 10,499 observations),
- Age = Age in years plus twelfths of a year,
- Adepcnt = 1 + number of dependents,
- Acadmos = months living at current address,
- Majordrg = Number of major derogatory reports,

- Minordrg = Number of minor derogatory reports,
- Ownrent = 1 if owns their home, 0 if rent
- Income = Monthly income (divided by 10,000),
- Selfempl = 1 if self employed, 0 if not,
- Inc_per = Income divided by number of dependents,
- Exp_Inc = Ratio of monthly credit card expenditure to yearly income,
- Spending = Average monthly credit card expenditure (for Cardhldr = 1),
- Logspend = Log of spending.

Using this data, do the following:

1. Estimate the following model

$$logspend = \beta_1 + \beta_2 \ln income + \beta_3 Age + \beta_4 Adepcnt + \beta_5 ownrent + \epsilon$$

   (a) Using OLS, what is the effect of 10% increase in income on credit card expenditure?

   (b) Using Censored regression, what is the effect of 10% increase in income on credit card expenditure?

   (c) Using Heckman Two-Step Estimator, what the is effect of 10% increase in income on credit card expenditure?

2. Create a subsample where only credit cardholders appear and do the following

   (a) Estimate the above model using OLS. What is the difference in credit card spending between home owner and renter?

   (b) Estimate the above model using truncated regression. What is the difference in credit card spending between home owner and renter?

3. Now we are interested in explaining the number of major derogatory reports as function of log income, age, the number of dependents, home ownership status and ratio of monthly credit card expenditure to yearly income.

   (a) Estimate this model using Poisson regression for credit cardholders only. What is the effect of 10% increase in income on the expected value (mean) of the number of major derogatory reports? Is Poisson regression a good specification for the data at hand?

   (b) Estimate this model using negative binomial regression for credit cardholders only. What is the effect of 10% increase in income on the expected value (mean) of the number of major derogatory reports?

   (c) Estimate the two models taking into account the truncation.What is the effect of 10% increase in income on the expected value (mean) of the number of major derogatory reports?