1 Summary

- **Multinomial data:** dependent variable can attain \(m\) possible outcomes \((y_i \in \{0, 1, \ldots, m - 1\})\).

- **Ordered and unordered variables:** variables with or without a natural ordering.
  ordered: e.g. education level, job category; unordered: e.g. means of transport

- **Ordered response model:** a model where the categorical outcome \(y_i\) is related to the latent variable 
  \[y_i^* = x_i' \beta + e_i, \quad e_i \sim IID(0, 1)\] 
  by means of \(m - 1\) unknown threshold values \(\tau_1 < \cdots < \tau_{m - 1}\) as follows
  \[
y_i = \begin{cases} 
  0 & \text{if } -\infty < y_i^* \leq \tau_1, \\
  j & \text{if } \tau_j < y_i^* \leq \tau_{j + 1}, \quad j = 1, \ldots, m - 2, \\
  m - 1 & \text{if } \tau_{m - 1} < y_i^* < \infty
  \end{cases}
  \]
  \((k + m - 2\) parameters, no constant term in \(\beta\) which has \(k - 1\) elements).

- **Multinomial logit:**
  \[
p_{ij} = \frac{\exp(x_i' \beta_j)}{\sum_{h=1}^{m} \exp(x_i' \beta_h)} = \frac{\exp(x_i' \beta_j)}{1 + \sum_{h=2}^{m} \exp(x_i' \beta_h)}
  \]
  \(\Rightarrow\) individual-specific data.

- **Conditional logit:**
  \[
p_{ij} = \frac{\exp(x_i' \beta)}{\sum_{h=1}^{m} \exp(x_i' \beta)}
  \]
  \(\Rightarrow\) alternative-specific data.

- **Marginal effects of explanatory variables:** (in multinomial logit model) all the parameters \(\beta_1, \ldots, \beta_{m - 1}\) 
  together determine the marginal effect of \(x_i\) on the probability to choose the \(j\)th alternative. So the
  sign of the parameter \(\beta_l^{(j)}\) cannot always be interpreted directly as the sign of the effect of the \(x_l\) on the
  probability to choose the \(j\)th alternative.
• **Odds ratio:** the relative odds to choose between the alternatives $j$ and $h$, given by (in multinomial logit):

$$\frac{P(y_i = j|x_i)}{P(y_i = h|x_i)} = \exp(x_i'(\beta_j - \beta_h)).$$

Then: $(\beta_j^{(j)} - \beta_j^{(h)}) > 0$ indicates a positive effect of $x_i$ on $P(y_i = j|x_i)$ relative to $P(y_i = h|x_i)$.

• **Utilities Model:** A model where the observed dependent variable is assumed to be a function of utilities experienced from alternative choices, $U_i^{(j)}$, $j = 0, 1, \ldots, m$. The observed choice depends on the difference in the utilities.

[interpretation of binary logit/probit model alternative to the latent variables model]

• **Multinomial logit:** 3 categories case (for the $j$th variable):

![Diagram](image)

• **Standard extreme value distribution:**

$$G(x) = \exp(-\exp(-x)), \quad (CDF)$$

$$p(x) = \exp(-\exp(-x) - x), \quad (PDF)$$

The difference between two independent variables with (standard) extreme value distribution has (standard) logistic distribution

[used in defining the binary logit model in terms of utilities]

## 2 Extra Topics

**From the last week!**

Check Tutorial Problems No. 1.

## 3 Lecture Problems

Ex. 3: ordered probit model versus binary probit model

*Show that the ordered probit model (with two explanatory variables $x_{i1}$ and $x_{i2}$) with $m = 2$ alternatives is the binary probit model with constant term $\beta_0 = -\tau_1$, by showing that $P(y_i = 1|x_i)$ is the same in both models.*

In ordered probit model in case of 2 categories $y_i \in \{0, 1\}$ and two explanatory variables $x_{i1}$ and $x_{i2}$ we consider a latent variable $y_i^*$:

$$y_i^* = \beta_1 x_{i1} + \beta_2 x_{i2} + e_i,$$
where $e_i \sim N(0, 1)$, i.i.d. We observe the choice $y_i$:

$$y_i = \begin{cases} 
0 & \text{if } -\infty < y_i^* \leq \tau_1, \\
1 & \text{if } \tau_1 < y_i^* \leq \infty,
\end{cases}$$

with threshold value $\tau_1$.

We have:

$$\mathbb{P}(y_i = 1|x_i) = \mathbb{P}(y_i^* > \tau_1|x_i)$$

$$= \mathbb{P}(e_i + \beta_1 x_{i1} + \beta_2 x_{i2} > \tau_1|\tau_1)$$

$$= \mathbb{P}(e_i > \tau_1 - \beta_1 x_{i1} - \beta_2 x_{i2}|\tau_1)$$

$$(\ast) \mathbb{P}(e_i < -\tau_1 + \beta_1 x_{i1} + \beta_2 x_{i2}|\tau_1)$$

$$(\ast\ast) \mathbb{P}(e_i \leq -\tau_1 + \beta_1 x_{i1} + \beta_2 x_{i2}|\tau_1)$$

$$(\ast\ast\ast) \mathbb{P}(e_i \leq -\tau_1 + \beta_1 x_{i1} + \beta_2 x_{i2})$$

$$= \Phi(-\tau_1 + \beta_1 x_{i1} + \beta_2 x_{i2})$$

where we used that the standard normal distribution of $e_i$ is ($\ast$) symmetric around 0, ($\ast\ast$) continuous and ($\ast\ast\ast$) independent of $x_i$, and where $\Phi(.)$ is the cumulative distribution function (CDF) of the standard normal distribution.

Further, since $y_i = 0$ or $y_i = 1$ we have

$$\mathbb{P}(y_i = 0|x_i) + \mathbb{P}(y_i = 1|x_i) = 1,$$

so that

$$\mathbb{P}(y_i = 0|x_i) = 1 - \mathbb{P}(y_i = 1|x_i)$$

$$= 1 - \Phi(-\tau_1 + \beta_1 x_{i1} + \beta_2 x_{i2})$$

$$= \Phi(\tau_1 - \beta_1 x_{i1} - \beta_2 x_{i2}).$$

In the binary probit model we have

$$\mathbb{P}(y_i = 1|x_i) = \Phi(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}),$$

$$\mathbb{P}(y_i = 0|x_i) = 1 - \mathbb{P}(y_i = 1|x_i)$$

$$= 1 - \Phi(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$$

$$= \Phi(-\beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}).$$

So, indeed $\mathbb{P}(y_i = 1|x_i)$ is the same in the binary probit model and in the ordered probit model with $m = 2$ alternatives (with $-\tau_1 = \beta_0$).

Therefore: the ordered probit model reduces to the binary probit model if we have only $m = 2$ alternatives. I.e. they have the same Bernoulli distribution for $y_i$ (conditionally upon $x_i$).

Note: in a similar way it holds that the ordered logit model reduces to the binary logit model if we have only $m = 2$ alternatives.

**Ex. 4: ordered logit model – importance of the ordering**

The EViews file bank_employees_exercise13.wf1 contains the data, where also two variables have been added:

- **admin0_manager1_cust2** (where 0 = administrative, 1 = management, 2 = custodial), where the ordering is done based on average value of male, per category;
- **admin0_cust1_manager2** (where 0 = administrative, 1 = custodial, 2 = management), where the ordering is done based on average value of salary per category.
Estimate two ordered logit models using these series as dependent variable (and education and male as explanatory variables). Compare the AIC, SC and prediction quality with the model where the categories are ordered with education (with dependent variable ORDERED_JOB_CATEGORY, which is used on the slides). Can you explain the differences?

Notice that ORDERED_JOB_CATEGORY could be called ‘cust0_admin1 Manage2’ (where 0 = custodial, 1 = administrative, 2 = management).

We have:

<table>
<thead>
<tr>
<th>dependent variable</th>
<th>AIC</th>
<th>SC</th>
<th>percentage correctly predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORDERED_JOB_CATEGORY</td>
<td>0.829509</td>
<td>0.864624</td>
<td>85.232%</td>
</tr>
<tr>
<td>admin0_manage1_cust2</td>
<td>1.151120</td>
<td>1.186236</td>
<td>77.215%</td>
</tr>
<tr>
<td>admin0_cust1_manage2</td>
<td>1.051928</td>
<td>1.087044</td>
<td>84.810%</td>
</tr>
</tbody>
</table>

Note: The model with (0 = custodial, 1 = administrative, 2 = management) is the best: the lowest (best) AIC and SC, and the highest (best) percentage correctly predicted.

**Reason:** education is the most important explanatory variable (more important than male), so it is best to order the categories with education. A higher education increases the probability of going from category 0=custodial to 1=administrative, and it increases the probability of going from category 1=administrative to 2=management.

Note: The model with (where 0 = administrative, 1 = management, 2 = custodial) is the worst: the highest (worst) AIC and SC, and the lowest (worst) percentage correctly predicted.

**Reason:** male is a relatively unimportant explanatory variable (less important than education), so it is not good to order the categories with male. Here the estimated coefficient of education is ‘damaged’, because education increases the probability of going from category 0=administrative to 1=management, but it decreases the probability of going from category 1=management to 2=custodial.

Note: The model with (0 = administrative, 1 = custodial, 2 = management) is also bad: the AIC, SC and percentage correctly predicted are bad (close to the worst model and much worse than the best model).

**Reason:** Here the estimated coefficient of education is again ‘damaged’, because education decreases the probability of going from category 0=administrative to 1=custodial, but it increases the probability of going from category 1=management to 2=custodial.

Note: Beforehand we could not say whether the model with (0 = administrative, 1 = management, 2 = custodial) or the model with (0 = administrative, 1 = custodial, 2 = management) would be the worst. Both of these models have a poor ordering of the categories (when looking at the effect of education on the probabilities of being in the categories).

### 4 Problem on binary, ordered & multinomial logit models

Consider the binary logit model where

\[ y_i^* = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \]

where the e_i (i = 1, 2, ..., n) are i.i.d. errors that have the (standard) logistic distribution with cumulative distribution function (CDF) given by

\[ G(a) = P(e_i \leq a) = \frac{1}{1 + \exp(-a)} = \frac{\exp(a)}{1 + \exp(a)}. \]

and where the e_i (i = 1, 2, ..., n) are independent of x_j1 and x_j2 (j = 1, 2, ..., n). Further,

\[ y_i = \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{if } y_i^* \leq 0. \end{cases} \]
(a) Derive the probability \( P(y_i = 1|x_{i1}, x_{i2}) \) and the probability \( P(y_i = 0|x_{i1}, x_{i2}) \).

\[
P(y_i = 1|x_i) = P(y_i^* > 0|x_i) = P(x_i^* \beta + e_i > 0|x_i) = P(e_i > -x_i^* \beta|x_i)
\]

\[
\begin{align*}
&\overset{(\ast)}{=} P(e_i < x_i^* \beta|x_i) \\
&\overset{(\ast\ast)}{=} P(e_i \leq x_i^* \beta|x_i) \\
&\overset{(\ast\ast\ast)}{=} P(e_i \leq x_i^* \beta) = G(x_i^* \beta) \\
&= \frac{1}{1 + \exp(-x_i^* \beta)} \\
&= \frac{\exp(x_i^* \beta)}{1 + \exp(x_i^* \beta)},
\end{align*}
\]

where we used that the standard logistic distribution of the error term \( e_i \) is (\ast) symmetric around 0, (\ast\ast) continuous and (\ast\ast\ast) independent of \( x_i \).

Further, \( y_i \) is either 0 or 1, so that

\[
P(y_i = 0|x_{i1}, x_{i2}) + P(y_i = 1|x_{i1}, x_{i2}) = 1,
\]

so we have:

\[
P(y_i = 0|x_{i1}, x_{i2}) = 1 - G(x_i^* \beta) = \frac{1}{1 + \exp(x_i^* \beta)}.
\]

(b) Derive the loglikelihood in this model.

The likelihood per observation \( i \) is the probability function of \( y_i \), conditionally upon \( x_i \):

\[
p(y_i|x_i) = [G(x_i^* \beta)]^{y_i} [1 - G(x_i^* \beta)]^{1-y_i} = \begin{cases} G(x_i^* \beta) & \text{if } y_i = 1, \\ 1 - G(x_i^* \beta) & \text{if } y_i = 0. \end{cases}
\]

The likelihood is the joint probability function of the \( y_i \) \( (i = 1, 2, \ldots, n) \), conditionally upon the \( x_i \) \( (i = 1, 2, \ldots, n) \):

\[
L(\beta) = p(y_1, \ldots, y_n|x_1, \ldots, x_n)
\]

\[
\overset{(\ast)}{=} \prod_{i=1}^n p(y_i|x_i)
\]

\[
= \prod_{i=1}^n [G(x_i^* \beta)]^{y_i} [1 - G(x_i^* \beta)]^{1-y_i},
\]

where in (\ast) we used the assumption that the \( y_i \) are independent (conditionally upon the \( x_i \)). In other words, we assume that the \( e_i \) are independent. The loglikelihood is simply the (natural) logarithm of the likelihood:

\[
\ln L(\beta) = \ln p(y_1, \ldots, y_n|x_1, \ldots, x_n)
\]

\[
= \sum_{i=1}^n \{ y_i \ln[G(x_i^* \beta)] + (1 - y_i) \ln[1 - G(x_i^* \beta)] \}.
\]

(c) Suppose that we analyse data on a presidential election, where there are two candidates, say C and T. We observe \( n = 1000 \) observations. We have:

\[
y_i = \begin{cases} 1 & \text{if person } i \text{ votes for candidate } C, \\ 0 & \text{if person } i \text{ votes for candidate } T, \end{cases}
\]

\[
x_{2i} = \text{number of years of education of person } i, x_{2i} \in [12, 20],
\]

5
and

\[ x_{3i} = \begin{cases} 
1 & \text{if person } i \text{ is a female,} \\
0 & \text{if person } i \text{ is a male.}
\end{cases} \]

Figure 1 contains ML estimation output and graphs of the estimated probability \( \hat{P}(y_i = 1|x_{1i}, x_{2i}) \).

Explain why the estimates of \( \beta_1 \) and \( \beta_2 \) match with the graphs of the estimated probability \( \hat{P}(y_i = 1|x_{1i}, x_{2i}) \).

The estimated coefficients \( \hat{\beta}_1 \) (at education \( x_{1i} \)) and \( \hat{\beta}_2 \) (at female \( x_{2i} \)) are significantly positive, which matches with the fact that \( \hat{P}(y_i = 1|x_{1i}, x_{2i}) \) is increasing with education and is higher for females than for males.

(The graph for males is the graph for females shifted \( 0.91/0.17 = 5.35 \) to the right.)

(d) Now suppose there are three candidates, say C, T and B. We have:

\[ y_i = \begin{cases} 
0 & \text{if person } i \text{ votes for candidate C,} \\
1 & \text{if person } i \text{ votes for candidate T,} \\
2 & \text{if person } i \text{ votes for candidate B.}
\end{cases} \]

Figures 2 and 3 contain ML estimation output and graphs of the estimated probabilities \( \hat{P}(y_i = 0|x_{1i}, x_{2i}) \), \( \hat{P}(y_i = 1|x_{1i}, x_{2i}) \) and \( \hat{P}(y_i = 2|x_{1i}, x_{2i}) \) in the multinomial logit model (with reference category 0). Explain why the estimates of the coefficients match with the graphs of the estimated probabilities \( \hat{P}(y_i = 0|x_{1i}, x_{2i}) \), \( \hat{P}(y_i = 1|x_{1i}, x_{2i}) \) and \( \hat{P}(y_i = 2|x_{1i}, x_{2i}) \).

Figure 2: Multinomial logit model: estimation output.

Figure 3: Multinomial logit model: graphs of the estimated probabilities.
Figure 3: Multinomial logit model: graphs of the estimated probabilities $\hat{P}(y_i = 0|x_{i1}, x_{i2})$, $\hat{P}(y_i = 1|x_{i1}, x_{i2})$ and $\hat{P}(y_i = 2|x_{i1}, x_{i2})$.

In this multinomial logit model we have probabilities:

$$P(y_i = 0|x_i) = \frac{1}{1 + \exp(\beta_0^{(1)} x_{i1} + \beta_1^{(1)} x_{i2}) + \exp(\beta_0^{(2)} + \beta_1^{(2)} x_{i1} + \beta_2^{(2)} x_{i2})},$$

$$P(y_i = 1|x_i) = \frac{\exp(\beta_0^{(1)} + \beta_1^{(1)} x_{i1} + \beta_2^{(1)} x_{i2})}{1 + \exp(\beta_0^{(1)} + \beta_1^{(1)} x_{i1} + \beta_2^{(1)} x_{i2}) + \exp(\beta_0^{(2)} + \beta_1^{(2)} x_{i1} + \beta_2^{(2)} x_{i2})},$$

$$P(y_i = 2|x_i) = \frac{\exp(\beta_0^{(2)} + \beta_1^{(2)} x_{i1} + \beta_2^{(2)} x_{i2})}{1 + \exp(\beta_0^{(1)} + \beta_1^{(1)} x_{i1} + \beta_2^{(1)} x_{i2}) + \exp(\beta_0^{(2)} + \beta_1^{(2)} x_{i1} + \beta_2^{(2)} x_{i2})}.$$

Note: we have **odds ratio**

$$\frac{P(y_i = 1|x_i)}{P(y_i = 0|x_i)} = \exp(\beta_0^{(1)} + \beta_1^{(1)} x_{i1} + \beta_2^{(1)} x_{i2}),$$

so that

$$\ln \left( \frac{P(y_i = 1|x_i)}{P(y_i = 0|x_i)} \right) = \beta_0^{(1)} + \beta_1^{(1)} x_{i1} + \beta_2^{(1)} x_{i2}.$$ 

Looking at the effect of **education**:

- Reference category 0 (voting C) has coefficient 0 (by definition).
- Category 1 (voting T) has significantly negative estimated coefficient $C(2) = -0.12$: an increase in education decreases $\hat{P}(y_i = 1|x_{i1}, x_{i2})$.

- Category 2 (voting B) has significantly positive estimated coefficient $C(5) = 0.82$: an increase in education increases $\hat{P}(y_i = 2|x_{i1}, x_{i2})$.

Hence, an increase in education increases $\hat{P}(y_i = 2|x_{i1}, x_{i2})$ (B) and decreases $\hat{P}(y_i = 1|x_{i1}, x_{i2})$ (T).

Looking at the effect of **female** ($x_{i2} = 1$ for female, $x_{i2} = 0$ for male):

- Reference category 0 (voting C) has coefficient 0 (by definition).
- Category 1 (voting T) has significantly negative estimated coefficient $C(3) = -0.86$:

  $$\frac{\hat{P}(y_i = 1|x_{i1}, x_{i2} = 1)}{\hat{P}(y_i = 0|x_{i1}, x_{i2} = 0)} < 1.$$ 

- Category 2 (voting B) has **insignificant** estimated coefficient $C(6) = 0.19$: we can not reject that $\frac{\hat{P}(y_i = 2|x_{i1}, x_{i2} = 1)}{\hat{P}(y_i = 0|x_{i1}, x_{i2} = 0)} = 1.$
Hence,

- $\hat{P}(y_i = 1|x_{i1}, x_{i2})$ (T) is lower for females than for males.
- $\hat{P}(y_i = 0|x_{i1}, x_{i2})$ (C) and $\hat{P}(y_i = 2|x_{i1}, x_{i2})$ (B) are higher for females than for males.

(e) Could an ordered logit model be appropriate in this case? Motivate your answer.

No: the alternatives cannot be ordered in such a way that the explanatory variables ‘push’ someone from the first to the second alternative and from the second to the third alternative.

- education ‘pushes’ from T to C and from C to B;
- female ‘pushes’ from T to C but not (significantly) from C to B.

However, if we ignore the fact that the positive estimated effect of female on

$$\frac{\hat{P}(y_i = 2|x_{i1}, x_{i2} = 1)}{\hat{P}(y_i = 0|x_{i1}, x_{i2} = 0)}$$

is not significant, then yes: we can order the alternatives T, C, B, where both the variables education and female ‘push’ persons from T to C and from C to B. In that case the ordered logit model could be appropriate.

5 Computer Exercises

W17/C2

Use the data in loanapp.wf1 for this exercise; see also Computer Exercise C8 in Chapter 7.

(i) Estimate a probit model of approve on white. Find the estimated probability of loan approval for both whites and nonwhites. How do these compare with the linear probability estimates?

As there is only one explanatory variable that takes on just two values, there are only two different predicted values: the estimated probabilities of loan approval for white and nonwhite applicants. Rounded to three decimal places these are:

$$\hat{P}(\text{approve} = 1|\text{white} = 0) = \Phi(\beta_0 + \beta_1 \cdot 0) = \Phi(0.547) = 0.708,$$
$$\hat{P}(\text{approve} = 1|\text{white} = 1) = \Phi(\beta_0 + \beta_1 \cdot 1) = \Phi(0.547 + 0.784) = 0.908,$$

for nonwhites and whites, respectively. Without rounding errors, these are identical to the fitted values from the linear probability model. This must always be the case when the independent variables in a binary response model are mutually exclusive and exhaustive binary variables. Then, the predicted probabilities, whether we use the LPM, probit, or logit models, are simply the cell frequencies.

(In other words, 0.708 is the proportion of loans approved for nonwhites and 0.908 is the proportion approved for whites.)

---

1 From the previous week!

(ii) Now, add the variables hrat, obrat, loanprc, unemc, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2, and vr to the probit model. Is there statistically significant evidence of discrimination against nonwhites?

With the set of controls added, the probit estimate on white becomes about 0.520 with the standard error of around 0.097. Therefore, there is still very strong evidence of discrimination against nonwhites. [We can divide this by 2.5 to make it roughly comparable to the LPM estimate in part (iii) of Computer Exercise C7.8: 0.520/2.5 ≈ 0.208, compared with 0.129 in the LPM.]

(iii) Estimate the model from part (ii) by logit. Compare the coefficient on white to the probit estimate.

When we use logit instead of probit, the coefficient on white becomes 0.938 with the standard error of 0.173. [Recall that to make probit and logit estimates roughly comparable, we can multiply the logit estimates by 0.625. The scaled logit coefficient becomes: 0.625 · 0.938 ≈ 0.586, which is reasonably close to the probit estimate of 0.520. A better comparison would be to compare the predicted probabilities by setting the other controls at interesting values, such as their average values in the sample.]
Use equation

\[ n^{-1} \sum_{i=1}^{n} \left\{ G[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_{k-1} x_{ik-1} + \hat{\beta}_k (c_k + 1)] - G[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_{k-1} x_{ik-1} + \hat{\beta}_k c_k] \right\} \]  \tag{17.17} 

\[ (iv) \]

to estimate the sizes of the discrimination effects for probit and logit.

Note that (17.17) is the average partial effect for a discrete explanatory variable. Unfortunately, it seems there is no build-in function for this measure in EViews, so we need to calculate it ourselves using the estimation results from the “augmented” probit and logit models. Figure 4 presents a code to carry out such computations.

We consider all the variables but white. Instead, for each individual we consider two counterfactual scenarios: as if he or she was white and otherwise (new generated variables white1 and white0), which we use to create two groups (variables white1 and variables white0). Then, we use the coefficients from two estimations (coef.probit and coef.logit) to sum all the variables multiplied by their respective coefficient.

This gives us the arguments inside \(G(\cdot)\) in (17.17). To evaluate \(G(\cdot)\) we need to apply the appropriate function for each model. For probit, it is \(\Phi(z)\), the cdf of the standard normal distribution; for logit, it is \(\frac{1}{1+\exp(-z)}\). Finally, we subtract the vector with \(G(\cdot)\) applied to the sum under the “nonwhites scenario” from that under the “whites scenario” and average out. The obtained values are \(APE_{probit} = 0.1042\) and \(APE_{logit} = 0.1009\), hence quite similar.
Figure 4: EViews code for computing APE for probit and logit models, where we are interested in the effect of being white or not on loan approval.

```plaintext
vector(16) coef_probit
coeff_probit = est.probit @ coefs
group variables_white1 white1 marital_loanprob unem male married dep sch consign chist pubrec mortatl mortatl2

vector(16) coef_logit
coeff_logit = est.logit @ coefs
group variables_white1 white1 marital_loanprob unem male married dep sch consign chist pubrec mortatl mortatl2

counterfactual scenarios
g: while white1 = 1
g: while white1 = 0

all variables under counterfactual scenarios
g: while white1 = 1 marital_loanprob unem male married dep sch consign chist pubrec mortatl mortatl2
g: while white1 = 0 marital_loanprob unem male married dep sch consign chist pubrec mortatl mortatl2

sum inside the G functions
series sum_white0_probit
series sum_white1_probit
series sum_white0_logit
series sum_white1_logit

start summing with the intercept (detail)
s: while white1 = 1
sum_white0_probit = coef_probit(1)
sum_white1_probit = sum_white0_probit + temp
sum_white0_logit = coef_logit(1)
sum_white1_logit = sum_white0_logit + temp

for n = 2 to 15

series temp = coef_probit(1) * variables_white1(n-1)
s: while white1 = 1
sum_white0_probit = sum_white0_probit + temp
sum_white1_probit = sum_white1_probit + temp

series temp = coef_logit(1) * variables_white1(n-1)
s: while white1 = 0
sum_white0_logit = sum_white0_logit + temp
sum_white1_logit = sum_white1_logit + temp

end

for probit: compute G as the cdf of the standard normal distribution
series G_white0_probit = @normsdist(sum_white0_probit)
series G_white1_probit = @normsdist(sum_white1_probit)
series diff_probit = G_white1_probit - G_white0_probit
scalar ape_probit = @mean(diff_probit)

for logit: compute G as the logistic function
series G_white0_logit = 1/(1 + @exp(-sum_white0_logit))
series G_white1_logit = 1/(1 + @exp(-sum_white1_logit))
series diff_logit = G_white1_logit - G_white0_logit
scalar ape_logit = @mean(diff_logit)
```
