Advanced Econometrics II

EAFIT University

Name: Angel Samuel Suesca Rios (assuescar@eafit.edu.co)

In [1]:
```
import stata_setup
stata_setup.config("/usr/local/stata17/", "mp")
```

```
  ___  ____  ____  ____  ____ ®
 /__    /   ____/   /   ____/      17.0
___/   /   /___/   /   /___/       MP—Parallel Edition

 Statistics and Data Science       Copyright 1985-2021 StataCorp LLC
                                   StataCorp
                                   4905 Lakeway Drive
                                   College Station, Texas 77845 USA
                                   800-STATA-PC        https://www.stata.com
                                   979-696-4600        stata@stata.com

Stata license: Single-user 4-core  perpetual
Serial number: 501706303466
  Licensed to: David Tomas Jacho-Chavez
               Emory University

Notes:
      1. Unicode is supported; see help unicode_advice.
      2. More than 2 billion observations are allowed; see help obs_advice.
      3. Maximum number of variables is set to 5,000 but can be increased;
          see help set_maxvar.
```

Instructions: Please answer *all* questions in the space provided below each cell either in English or Español. Full marks are not only given for correct answers but for elegance, preciseness, and presentation. Please make sure your notebook runs without an error.

Note: If you are seeking to earn my recommendation letter for future doctoral studies, you must (1) Work alone, and (2) score among the highest 2 scores in the class (excluding the scores of people who choose to work in pairs).

E.1

1. For the logistic distribution $\Lambda(x) = (1 + \exp(-x))^{-1}$ verify that

---

(a) [2 points] $\frac{d}{dx}\Lambda(x) = \Lambda(x)(1-\Lambda(x))$.

Answer:

Let's consider $f(g(\cdot)) = g(\cdot)^{-1}$ where $g(\cdot) = 1 + e^{-\cdot}$ then, by chain rule we have that:

$$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}$$

In our case we have then:

$$\frac{d}{dx}\Lambda(x) = -\frac{1}{g^2}\cdot(-1)e^{-x} = \frac{1}{g^2}e^{-x} = \frac{1}{(1+e^{-x})^2}e^{-x}$$

And separing the factors in denominator:

$$\frac{d}{dx}\Lambda(x) = \frac{1}{(1+e^{-x})}\cdot\frac{e^{-x}}{(1+e^{-x})} = \Lambda(x)\left[\frac{e^{-x}}{1+e^{-x}}\right]$$

If we sum and substract 1 in the numerator of the second factor of the right hand side of latest equation, we have:

$$\frac{d}{dx}\Lambda(x) = \Lambda(x)\left[\frac{1+e^{-x}-1}{1+e^{-x}}\right]$$

Then we can distribute the denominator to the 2 first terms and the last term like this:

$$\frac{d}{dx}\Lambda(x) = \Lambda(x)\left[\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right]$$

Finally if we simplified, we have shown the asked expression:

$$\frac{d}{dx}\Lambda(x) = \Lambda(x)[1-\Lambda(x)]$$

(b) [2 points] $h_{\text{logit}}(x) = \frac{d}{dx}\log\Lambda(x) = 1-\Lambda(x)$.

Answer:

Let's consider: $g(x) = \log(\Lambda(x)) = \log\left[(1+e^{-x})^{-1}\right] = -\log(1+e^{-x})$

So we can write the derivates as:

$$\frac{d}{dx}g(x) = \frac{d\left[-\log(1+e^{-x})\right]}{dx} = -\frac{d\log(1+e^{-x})}{dx}$$

Applying chain rule in the last expression we have:

$$\frac{d}{dx}g(x) = -\frac{1}{1+e^{x}} \cdot \frac{d\left(1+e^{-x}\right)}{dx} = \frac{e^{-x}}{1+e^{-x}}$$

And summing and substracting on the last side of the equation, we have:

$$\frac{d}{dx}g(x) = \frac{1+e^{-x}-1}{1+e^{-x}} = \frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}$$

With this we show that:

$$\frac{d}{dx}\log \Lambda(x) = 1 - \Lambda(x)$$

(c) [2 points] $H_{\text{logit}}(x) = -\frac{d^2}{dx^2}\log \Lambda(x) = \Lambda(x)(1-\Lambda(x))$.

Answer:

Since

$$-\frac{d^2}{dx^2}\log(\Lambda(x)) = -\frac{d}{dx}\left(\frac{d}{dx}\log(\Lambda(x))\right) = -\frac{d}{dx}(1-\Lambda(x))$$

For what we shown in the previous exercise (1.b). Then for basic derivative rules:

$$-\frac{d}{dx}(1-\Lambda(x)) = -\left[\frac{d}{dx}[1] - \frac{d}{dx}\Lambda(x)\right] = \frac{d}{dx}\Lambda(x)$$

We now, by exercise (1.a) that the derivative of the logit function with respect to x is $\Lambda(x) \cdot (1-\Lambda(x))$

So we can show that:

$$H_{\text{logit}}(x) = -\frac{d^2}{dx^2}\log \Lambda(x) = \Lambda(x)(1-\Lambda(x))$$

(d) [2 points] $\left|H_{\text{logit}}(x)\right| \leq 1$.

Answer:

We know that $\left|H_{logit}(x)\right| \leq 1 \rightarrow -1 \leq H_{logit}(x) \leq 1$, so we can interpretate this absolute value inequality as two separe inequalities:

The first, the lower bound:

$$H_{logit}(x) \geqslant -1$$
$$\Lambda(x)(1-\Lambda(x)) \geq -1$$
$$\frac{e^{-x}}{\left(1+e^{-x}\right)^2} \geqslant -1$$
$$e^{-x} \geq -\left(1+e^{-x}\right)^2$$
$$e^{-x} \geq -\left(1+2e^{-x}+e^{-2x}\right)$$
$$0 \geq -\left(e^{-2x}+3e^{-x}+1\right)$$
$$1 \geq -\left(e^{-2x}+3e^{-x}\right)$$
$$-1 \leq \left(e^{-x}\right)\left(e^{-x}+3\right)$$

Since $e^{-x} \in (0, \infty)\ \forall x \in \mathbb{R}$, i.e, it is positive in all domine. Thus

$$e^{-x}\left(e^{-x}+3\right) \geqslant 0 \geq -1$$

Then the lower bound inequalities is satisfied for all x.

Let's now concentrate in the upper bound inequality. For $H_{logit}(x) \leqslant 1$, we will try to demostrante that $H_{logit}(x)$ has a unique maximun less than 1. So first, lest find the critic values of the function by finding $H'_{logit}(x)$:

$$H'_{logit}(x) = \frac{d}{dx}[\Lambda(x)(1-\Lambda(x)] = \frac{d}{dx}\left[\frac{e^{-x}}{\left(1+e^{-x}\right)^2}\right]$$

Applying the quotient rule of derivatives, we have:

$$H'_{logit}(x) = \frac{-e^{-x}[1+e^{-x}]^2 + 2e^{-x}\left[1+e^{-x}\right]e^{-x}}{\left(1+e^{-x}\right)^4}$$

And ordering the expression:

$$H'_{logit}(x) = e^{-x}\left[2e^{-x}\left[1+e^{-x}\right]^{-3} - \left[1+e^{-x}\right]^{-2}\right]$$

If we want to find the critic values, we must consider $H'_{logit}(x) = 0$, by factoring the last expression and

$$H'_{logit}(x) = e^{-x}\left[1+e^{-x}\right]^{-2}\left[2e^{-x}\left[1+e^{-x}\right]^{-1} - 1\right] = 0$$

Since $e^{-x} \neq 0\ \forall x \in \mathbb{R}$ and $\left[1+e^{-x}\right]^{-2} \neq 0\ \forall x \in \mathbb{R}$ (Because these are strictic positive functions). Then if we need $H'_{logit}(x)$ be equal to zero, we sould consider $\left[2e^{-x}[1+e^{-x}]^{-1} - 1\right] = 0$. Solving the last expresion for x:

$$
\begin{aligned}
& 2e^{-x}\left[1+e^{-x}\right]^{-1}-1=0 \\
& 2e^{-x}\left[1+e^{-x}\right]^{-1}=1 \\
& 2e^{-x}=1+e^{-x} \\
& e^{-x}=1 \\
& \ln\left(e^{-x}\right)=\ln(1)\rightarrow x=0
\end{aligned}
$$

So we find that $H_{logit}(x)$ has a unique critic value at $x=0$, now we must to show that is a maximun and is less than 1. So we must find $H''_{logit}(x)$ and evaluated at $x=0$, finally show that is less than 0 (maximum criteria).

$$
\begin{aligned}
H''_{logit}(x) &= \frac{d}{dx}\left[\frac{2e^{-2x}}{[1+e^{-x}]^3}-\frac{e^{-x}}{[1+e^{-x}]^2}\right] \\
H''_{logit}(x) &= \frac{6e^{-3x}}{(1+e^{-x})^4}-\frac{6e^{-2x}}{(1+e^{-x})^3}+\frac{e^{-x}}{(1+e^{-x})^2} \\
H''_{logit}(x) &= \frac{e^{-x}}{(1+e^{-x})^2}\left[\frac{6e^{-2x}}{(1+e^{-x})^2}-\frac{6e^{-x}}{(1+e^{-x})}+1\right]
\end{aligned}
$$

Note that we applied twice the quotient rule, and simplified the expression, evaluating in $x=0$ the last expression, we have:

$$
H''_{logit}(0)=\frac{1}{4}\left[\frac{6}{4}-\frac{6}{2}+1\right]=\frac{1}{4}\left[\frac{6-12+4}{4}\right]
$$

$$
\frac{1}{16}[-2]=-\frac{1}{8}<0
$$

So we demostrate that $H_{logit}(x)$ has a unique maximum value at $x=0$, let evaluate $H_{logit}(0)$:

$$
H_{\text{logit}}(0)=\Lambda(0)(1-\Lambda(0))
$$

Consider that $\Lambda(0)=\frac{1}{2}$, then we have:

$$
H_{\text{logit}}(0)=\frac{1}{2}(1-\frac{1}{2})=\frac{1}{2}\frac{1}{2}=\frac{1}{4}
$$

With this, we demostrate that $H_{logit}(x)\leqslant 1$. And finally show:

$$
\left|H_{logit}(x)\right|\leq 1
$$

Assume now that you have a random sample $\{Y_i, X_i'\}_{i=1}^{n}$ taken from the joint distribution of $(Y, X')\in[1,0]\times\mathbb{R}^K$ with $X=[X_1,X_2,\cdots,X_{K-1},1]'$ and fit a *logistic regression* through maximum likelihood methods, i.e., you calculate $\widehat{\beta}_{\text{logit}}$.

(e) [5 points] Prove mathematically that $S_n(\beta) \equiv \frac{d}{d\beta}\mathrm{l}_n = \mathbf{X}'\widehat{\mathbf{u}} = \mathbf{0}$, where $\mathbf{0}$ is the $K \times 1$ vector of zeroes, $\widehat{\mathbf{u}} = [\widehat{u}_1, \widehat{u}_2, \cdots, \widehat{u}_{n-1}, \widehat{u}_n]'$, $\mathbf{X}' = [X_1, X_2, \cdots, X_{n-1}, X_n]$, and $\widehat{u}_i \equiv Y_i - \Lambda(X_i'\widehat{\beta}_{\text{logit}})$.

Answer:

Show that

$$S_n(B) = \frac{d}{d\beta}\mathrm{l}_n(\beta) = X'\hat{U} = 0$$

Where $X' = [X_1, X_2, \ldots, X_{n-1}, X_n]$ and $\hat{U} = \left[\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_{n-1}, \hat{U}_n\right]^9$ where $\hat{U}_i = Y_i - \Lambda\left(X_i\hat{\beta}_{logit}\right)$

Consider the follow expression for the log-ikeliHood of logistic model.

$$\mathrm{l}_n(\beta) = \sum_{i=1}^{n}\left[Y_i \log \Lambda\left(X_i'\beta\right) + (1 - Y_i)\log \Lambda\left(-X_i'\beta\right)\right]$$

If we take the denvative with respect to β :

$$S_n(\beta) = \frac{\partial}{\partial\beta}\mathrm{l}_n(\beta) = \sum_{i=1}^{n}\left[Y_i\left(X_i\right)\left(1 - \Lambda\left(X_i'\beta\right)\right) + (1 - Y_i)\left(-X_i\right)\left(1 - \Lambda\left(-X_i'\beta\right)\right)\right]$$

In the last we consider the shown previously expression $\frac{\partial \log \Lambda(C)}{\partial C} = 1 - \Lambda(C)$ and the chain vile. Considering the symmetric property of logistic function $\Lambda(C) = 1 - \Lambda(-C)$ we have:

$$S_n(\beta) = \sum_{i=1}^{n}\left[Y_iX_i\left[1 - \Lambda\left(X_i'\beta\right)\right] - (1 - Y_i)\left(X_i\right)\left(\Lambda\left(X_i'\beta\right)\right)\right]$$

Simplifying we have that:

$$S_n(\beta) = \sum_{i=1}^{n}\left[Y_iX_i - Y_iX_i\Lambda\left(X_i'\beta\right) - X_i\Lambda\left(X_i'\beta\right) + Y_iX_i\left(\Lambda\left(X_i'\beta\right)\right)\right]$$

$$S_n(\beta) = \sum_{i=1}^{n}\left[Y_iX_i - X_i\Lambda\left(X_i'\beta\right)\right] = \sum_{i=1}^{n}X_i\left[Y_i - \Lambda\left(X_i'\beta\right)\right]$$

We have that a neccesarry condition for $\hat{\beta}_{\text{logit}}$ maximize the likelihood function $\mathrm{l}_n(\beta)$ is that the likelihood score of logit model $S_n(\beta)$ evaluate in $\hat{\beta}_{\text{logit}}$ must be (converge) to 0 . So we have:

$$S_n(\beta) = \sum_{i=1}^{n} X_i \left[Y_i - \Lambda \left(X_i' \hat{\beta}_{logit} \right) \right] = 0$$

Remembering that $\hat{U}_i = Y_i - \Lambda \left(X_i \hat{\beta}_{logit} \right)$ we have:

$$S_n(\beta) = \sum_{i=0}^{n} \left[X_i \hat{U}_i \right] = 0$$

We can express thu in matrix form as:

$$S_n(\beta) = \mathbf{X}' \hat{\mathbf{U}} = 0$$

Where as the beginning of the exercise we have $\mathbf{X}' = [X_1, X_2, X_3, \ldots, X_{n-1}, X_n]$ and $\hat{\mathbf{U}} = \left[\hat{U}_1, \hat{U}_2, \hat{U}_3, \ldots, \hat{U}_{n-1}, \hat{U}_n \right]$

(f) [3 points] Then or otherwise show that the proportion of ones in the sample equals the *average* in-sample fitted probabilities after fitting a *logistic regression*.

Answer:

Considering the expression shown above

$$S_n\left(\beta^{logit}\right) = \sum_{i=1}^{n} \left(X_i \left[Y_i - \Lambda \left(X_i' \hat{\beta}^{logit} \right) \right] \right) = 0$$

If we use the equation for "1" regressor (Constant) implicit in the last expression. (This is not receesarry but is usefull, we can use other justifications) we have:

$$\sum_{i=0}^{n} \left[Y_i - \Lambda \left(X_i' \hat{\beta}^{logit} \right) \right] = 0$$

Operating we have:

$$\sum_{i=0}^{n} Y_i - \sum_{i=0}^{n} \Lambda \left(X_i' \hat{\beta}^{logit} \right) = 0$$
$$\sum_{i=0}^{n} Y_i = \sum_{i=0}^{n} \Lambda \left(X_i' \hat{\beta}^{logit} \right)$$

Dividing by N :

$$N^{-1} \sum_{i=0}^{n} Y_i = N^{-1} \sum_{i=0}^{n} \Lambda \left(X_i' \hat{\beta}^{logit} \right)$$

And since the left than side of the last expression is the sample proportion of ones. And the right-hand side is the in-sampe fitted probability average we have shown that this 2 quantities are equal when logit model is correctly estimate by maximum likelihood.

E.2

2. Use the cps09mar dataset and the subset of men. Set $Y=1$ if marital equals 1,2 , or 3 , and set $Y=0$ otherwise. Estimate a *logistic regression* model for Y as a quadratic function of *age*, a linear function (including a constant term) of *education*, and including indicators for white individuals (equal 1 if race equals 1) and for black individuals (equal 1 if race equals 2).

In [2]:
```
!pwd
```

```
/home/samsuesca/curso/Notebooks
```

In [23]:
```
%%stata
frame change default
use "/home/samsuesca/curso/Notebooks/cps09mar.dta", clear
desc
```

```
. frame change default

. use "/home/samsuesca/curso/Notebooks/cps09mar.dta", clear
(Written by R.              )

. desc

Contains data from /home/samsuesca/curso/Notebooks/cps09mar.dta
 Observations:        50,742                  Written by R.
    Variables:            12                  25 Sep 2013 14:23
----------------------------------------------------------------------------
---
Variable      Storage   Display    Value
    name         type    format    label      Variable label
----------------------------------------------------------------------------
---
age             double  %9.0g                 age
female          double  %9.0g                 female
hisp            double  %9.0g                 Hispanic
education       double  %9.0g                 education
earnings        double  %9.0g                 total annual wage and salary
hours           double  %9.0g                 hours worked per week
week            double  %9.0g                 weeks worked per year
union           double  %9.0g                 union member
uncov           double  %9.0g                 union covered
region          double  %9.0g                 region
race            double  %9.0g                 race
marital         double  %9.0g                 marital status
----------------------------------------------------------------------------
---
Sorted by:

.
```

In [24]:
```
%%stata
*create variables as needed
cap gen Y = inlist(marital,1,2,3)
cap gen white = (race==1)
cap gen black = (race==2)

// check if variables are correctly created
*ta marital Y
*ta race white
*ta race black
```

```
. *create variables as needed
. cap gen Y = inlist(marital,1,2,3)

. cap gen white = (race==1)

. cap gen black = (race==2)

.
. // check if variables are correctly created
. *ta marital Y
. *ta race white
. *ta race black
.
```

In [25]:
```
%%stata
ta education female
```

```
           |        female
 education |         0          1 |     Total
-----------+----------------------+----------
         0 |        44         29 |        73 
         4 |       181         75 |       256 
         6 |       419        183 |       602 
         8 |       282        142 |       424 
         9 |       406        169 |       575 
        10 |       374        191 |       565 
        11 |       865        439 |     1,304 
        12 |     8,354      5,542 |    13,896 
        13 |     5,149      4,105 |     9,254 
        14 |     2,911      2,697 |     5,608 
        16 |     6,441      5,199 |    11,640 
        18 |     2,489      2,181 |     4,670 
        20 |     1,225        650 |     1,875 
-----------+----------------------+----------
     Total |    29,140     21,602 |    50,742 
```

In [26]:
```
%%stata
qui logit Y c.age#c.age education white black if female==0, robust
qui logit Y c.age#c.age white black if female==0, robust
cap drop fitted_prob
predict fitted_prob, pr
```

```
. qui logit Y c.age#c.age education white black if female==0, robust

. qui logit Y c.age#c.age white black if female==0, robust

. cap drop fitted_prob

. predict fitted_prob, pr

.
```

(a) [3 points] Verify numerically your answer to question 1(f) above in this example.

In [27]:
```
%%stata
*Consider the sample proportion of being married:
```

```
qui sum Y if female == 0
di in red "The sample proportion of being married is: `r(mean)'"

qui sum fitted_prob if female == 0
di in red "The average fitted probability is: `r(mean)'"
```

```
. *Consider the sample proportion of being married:
.
. qui sum Y if female == 0

. di in red "The sample proportion of being married is: `r(mean)'"
The sample proportion of being married is: .7179821551132464


.
. qui sum fitted_prob if female == 0

. di in red "The average fitted probability is: `r(mean)'"
The average fitted probability is: .7179821531404021


.
```

Base on this, we can conclude that the sample average of being married is the same as the average fitted probability of being married given by logit model.

(b) [10 points] Plot the fitted probability surfaces $\widehat{\Pr}[Y=1|age, education, race=1]$ and $\widehat{\Pr}[Y=1|age, education, race=2]$ using a *logistic regression* over the entire observed support of age and $education$. **Note:** The data description can be found here.

As is not specify which regressor (and functional forms) must be used, i'll plot the surfaces for the complete sample logit regression. And for the conditions race = 1 and race = 2, i will include the race as a regressor and not black and white.

In [8]:
```
%%stata
*qui logit Y c.age#c.age white black education if female==0, robust
qui logit Y c.age#c.age race education if female==0, robust
qui sum age
local min_age `r(min)'
local max_age `r(max)'
di "min age: `min_age' and max age: `max_age'"
qui sum educ
local min_educ `r(min)'
local max_educ `r(max)'

qui margins, dydx(educ) at(education=(`min_educ'(1)`max_educ') race = 1  age

qui margins, dydx(educ) at(education=(`min_educ'(1)`max_educ') race = 2  age
```

```
. *qui logit Y c.age#c.age white black education if female==0, robust
. qui logit Y c.age#c.age race education if female==0, robust

. qui sum age

. local min_age `r(min)'

. local max_age `r(max)'

. di "min age: `min_age' and max age: `max_age'"
min age: 15 and max age: 85

. qui sum educ

. local min_educ `r(min)'

. local max_educ `r(max)'

.
. qui margins, dydx(educ) at(education=(`min_educ'(1)`max_educ') race = 1  a
ge=
> (`min_age'(1)`max_age')) saving(predictions_white, replace)

.
. qui margins, dydx(educ) at(education=(`min_educ'(1)`max_educ') race = 2  a
ge=
> (`min_age'(1)`max_age')) saving(predictions_black, replace)

.
```

In [9]:

```
%%stata
cap frame create pred
frame change pred
use "/home/samsuesca/curso/Notebooks/predictions_white.dta", clear

rename _at1 age
rename _at3 education
rename _margin pr_inlf
save "/home/samsuesca/curso/Notebooks/predictions_white.dta", replace


use "/home/samsuesca/curso/Notebooks/predictions_black.dta", clear
rename _at1 age
rename _at3 education
rename _margin pr_inlf
save "/home/samsuesca/curso/Notebooks/predictions_black.dta", replace

frame change default
```

```
. cap frame create pred

. frame change pred

. use "/home/samsuesca/curso/Notebooks/predictions_white.dta", clear
(Created by command margins; also see char list)

.
. rename _at1 age

. rename _at3 education

. rename _margin pr_inlf

. save "/home/samsuesca/curso/Notebooks/predictions_white.dta", replace
file /home/samsuesca/curso/Notebooks/predictions_white.dta saved

.
.
. use "/home/samsuesca/curso/Notebooks/predictions_black.dta", clear
(Created by command margins; also see char list)

. rename _at1 age

. rename _at3 education

. rename _margin pr_inlf

. save "/home/samsuesca/curso/Notebooks/predictions_black.dta", replace
file /home/samsuesca/curso/Notebooks/predictions_black.dta saved

.
. frame change default

.
```

In [10]:

```
# Import the necessary Python packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Read (import) the Stata datasets into pandas data frames
data_white = pd.read_stata("/home/samsuesca/curso/Notebooks/predictions_whit
data_black = pd.read_stata("/home/samsuesca/curso/Notebooks/predictions_blac

# Crear una figura con dos subplots 3D
fig = plt.figure(figsize=(14, 7))

# Primer subplot para 'predictions_white'
ax1 = fig.add_subplot(121, projection='3d')
ax1.view_init(elev=25, azim=130)
surf_white = ax1.plot_trisurf(data_white['education'], data_white['age'], da
                              cmap=plt.cm.Spectral_r, alpha=0.7)

# Personalizar el primer subplot
ax1.set_title("Predictions for White", fontsize=14)
```

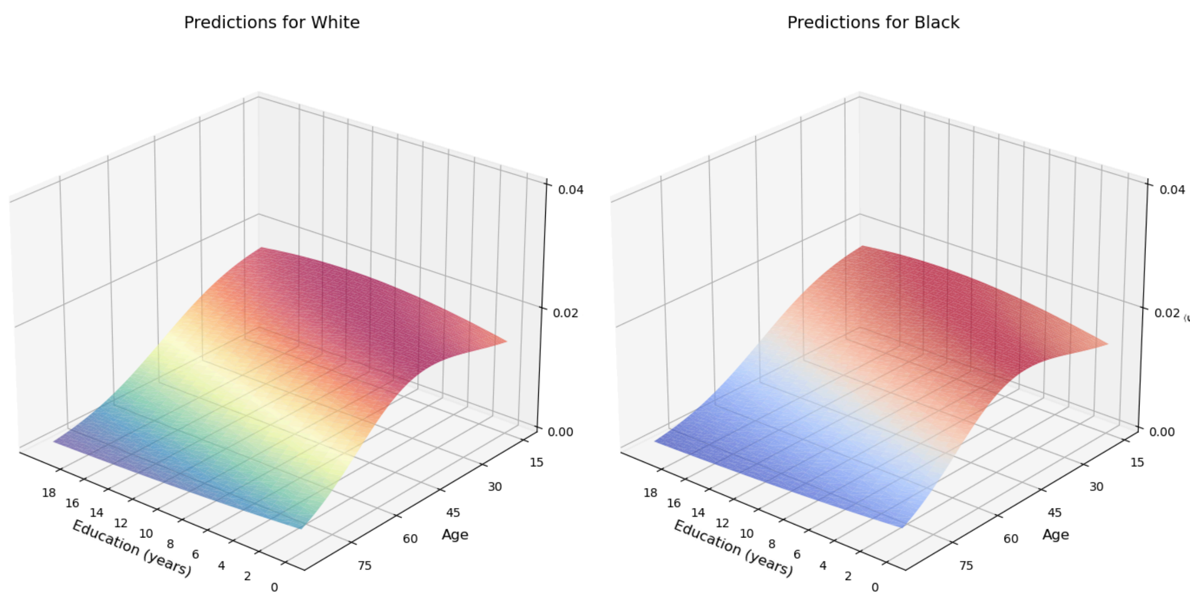
```
ax1.set_xlabel("Education (years)", fontsize=12)
ax1.set_ylabel("Age", fontsize=12)
ax1.set_zlabel(r"$\widehat{\delta}_{educ}$", fontsize=12)
ax1.set_xticks(np.arange(0, 20, step=2))
ax1.set_yticks(np.arange(15, 85, step=15))
ax1.set_zticks(np.arange(0, 0.05, step=0.02))

# Segundo subplot para 'predictions_black'
ax2 = fig.add_subplot(122, projection='3d')
ax2.view_init(elev=25, azim=130)
surf_black = ax2.plot_trisurf(data_black['education'], data_black['age'], da
                              cmap=plt.cm.coolwarm, alpha=0.7)

# Personalizar el segundo subplot
ax2.set_title("Predictions for Black", fontsize=14)
ax2.set_xlabel("Education (years)", fontsize=12)
ax2.set_ylabel("Age", fontsize=12)
ax2.set_zlabel(r"$\widehat{\delta}_{educ}$", fontsize=12)
ax2.set_xticks(np.arange(0, 20, step=2))
ax2.set_yticks(np.arange(15, 85, step=15))
ax2.set_zticks(np.arange(0, 0.05, step=0.02))

# Ajustar espacio entre los subplots
plt.tight_layout()
# Save the graph as a PNG image
plt.savefig("Margins3d_comparison.png")

# Show the plot
plt.show()
```



(c) [3 points] Calculate the average effect of an extra year of education on the probability of being married for a 45-year old white individual. **Note:** You might find useful to read Stata's manual for the margins command.

First consider de model for men with a cuadratic form in age, linear in education and race (then for women, and then a model with full sample just for comparing):

In [11]:
```
%%stata
cap frame change default
qui logit Y c.age#c.age education white black if female==0, robust
margins, dydx(educ) at(age=45 white=1)
```

```
. cap frame change default

. qui logit Y c.age#c.age education white black if female==0, robust

. margins, dydx(educ) at(age=45 white=1)

Average marginal effects                                 Number of obs = 29,1
40
Model VCE: Robust

Expression: Pr(Y), predict()
dy/dx wrt:  education
At: age   = 45
    white =  1

------------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interva
l]
-------------+----------------------------------------------------------------
--
   education |    .012576   .0008639    14.56   0.000     .0108828    .01426
92
------------------------------------------------------------------------------
--

.
```

Doing the same for females:

In [12]:
```
%%stata
qui logit Y c.age#c.age education white black if female==1, robust
margins, dydx(educ) at(age=45 white=1)
```

```
. qui logit Y c.age#c.age education white black if female==1, robust

. margins, dydx(educ) at(age=45 white=1)

Average marginal effects                                  Number of obs = 21,6
02
Model VCE: Robust

Expression: Pr(Y), predict()
dy/dx wrt:  education
At: age   = 45
    white =  1

----------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interva
l]
-------------+--------------------------------------------------------------
--
   education |   .0142234   .0012689    11.21   0.000     .0117365    .01671
03
----------------------------------------------------------------------------
--

.
```

Considering a model with both genders:

In [13]:
```
%%stata
qui logit Y c.age#c.age education white black female, robust
margins, dydx(educ) at(age=45 white=1 female=(0(1)1))
margins, dydx(educ) at(age=45 white=1)
```

```
. qui logit Y c.age#c.age education white black female, robust

. margins, dydx(educ) at(age=45 white=1 female=(0(1)1))

Average marginal effects                                Number of obs = 50,7
42
Model VCE: Robust

Expression: Pr(Y), predict()
dy/dx wrt:  education
1._at: age    = 45
       white  =  1
       female =  0
2._at: age    = 45
       white  =  1
       female =  1

----------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interva
l]
-------------+--------------------------------------------------------------
--
education    |
         _at |
          1  |   .0129969   .0006721    19.34   0.000     .0116795    .01431
43
          2  |   .0158304   .0008182    19.35   0.000     .0142267    .01743
41
----------------------------------------------------------------------------
--

. margins, dydx(educ) at(age=45 white=1)

Average marginal effects                                Number of obs = 50,7
42
Model VCE: Robust

Expression: Pr(Y), predict()
dy/dx wrt:  education
At: age   = 45
    white =  1

----------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interva
l]
-------------+--------------------------------------------------------------
--
   education |   .0141854   .0007327    19.36   0.000     .0127494    .01562
14
----------------------------------------------------------------------------
--
```

```
.
```

Both approaches show similar results. Complete sample regression with gender indicator show that the average marginal effect of a year of education is arround 0.014 for individuals with , and when we consider separating the effects conditioning on gender, we find that the AME is greater for females than for men. Which agrees with the separete estimation approach, in which we estimate a logit model for each subsample of gender.

(d) [2 points] What is the average marginal effect of being white? Calculate the average marginal effect of being black in your sample as well. **Note:** You might find useful to read Stata's manual for the margins command.

For that, consider the latest model estimated in (2.c) for the whole sample

In [14]:
```
%%stata
qui logit Y c.age#c.age education white black if female==0, robust
margins, dydx(white) atmeans
```

```
. qui logit Y c.age#c.age education white black if female==0, robust

. margins, dydx(white) atmeans

Conditional marginal effects                          Number of obs = 29,1
40
Model VCE: Robust

Expression: Pr(Y), predict()
dy/dx wrt:  white
At: age       = 42.04818 (mean)
    education = 13.81949 (mean)
    white     = .8354152 (mean)
    black     = .0828071 (mean)

----------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interva
l]
-------------+--------------------------------------------------------------
--
       white |   .0330199   .0096917     3.41   0.001     .0140245    .05201
52
----------------------------------------------------------------------------
--

.
```

In [15]:
```
%%stata
qui logit Y c.age#c.age education white black if female==0, robust
margins, dydx(black) atmeans
```

```
. qui logit Y c.age#c.age education white black if female==0, robust

. margins, dydx(black) atmeans

Conditional marginal effects                              Number of obs = 29,1
40
Model VCE: Robust

Expression: Pr(Y), predict()
dy/dx wrt:  black
At: age       = 42.04818 (mean)
    education = 13.81949 (mean)
    white     = .8354152 (mean)
    black     = .0828071 (mean)

------------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interva
l]
-------------+----------------------------------------------------------------
--
       black |  -.1312676   .0127193   -10.32   0.000    -.1561969   -.10633
83
------------------------------------------------------------------------------
--

.
```

E.3

3. Take the CHJ2004 dataset. The variables $tinkind$ and $income$ are household transfers received in-kind and household income, respectively. Divide both variables by 1000 to standardize. Create the regressor $Dincome = (income - 1) \times 1\{income > 1\}$. **Note:** The data description can be found here.

In [16]:
```
%%stata
use "/home/samsuesca/curso/Notebooks/CHJ2004", clear
*desc
replace tinkind = tinkind / 1000
replace income = income /1000

gen indicator_income = (income > 1)
*ta indicator_income
*sum income,d
*mdesc income
gen Dincome = (income-1)*indicator_income
*bys indicator_income: sum Dincome
```

```
. use "/home/samsuesca/curso/Notebooks/CHJ2004", clear

. *desc
. replace tinkind = tinkind / 1000
variable tinkind was long now double
(6,481 real changes made)

. replace income = income /1000
variable income was long now double
(8,642 real changes made)

.
. gen indicator_income = (income > 1)

. *ta indicator_income
. *sum income,d
. *mdesc income
. gen Dincome = (income-1)*indicator_income

. *bys indicator_income: sum Dincome
.
```

(a) [2 points] Estimate a linear regression of $tinkind$ on $income$ and $Dincome$. Interpret the results.

In [17]:
```
%%stata
reg tinkind income Dincome
```

```
      Source |       SS           df       MS      Number of obs   =     8,6
84
-------------+----------------------------------   F(2, 8681)      =     43.
97
       Model |  2173.30303         2  1086.65151   Prob > F        =    0.00
00
    Residual |  214520.855     8,681  24.7115372   R-squared       =    0.01
00
-------------+----------------------------------   Adj R-squared   =    0.00
98
       Total |  216694.158     8,683  24.9561393   Root MSE        =    4.97
11

----------------------------------------------------------------------------
--
     tinkind | Coefficient  Std. err.      t    P>|t|     [95% conf. interva
l]
-------------+--------------------------------------------------------------
--
      income |  -1.534968   .6673655    -2.30   0.021    -2.843163   -.22677
38
     Dincome |   1.547418   .6675015     2.32   0.020      .238957     2.855
88
       _cons |   2.702379   .6613564     4.09   0.000     1.405964     3.9987
94
----------------------------------------------------------------------------
--
```

The coefficient on $income$ indicates that for each additional unit increase in income (which is standardized to thousands of dollars), the household's in-kind transfers (tinkind, also standardized) decrease by approximately 1.535 units. This result is statistically significant at the 5% level (p-value = 0.021). In point of view, households with higher incomes tend to receive less in-kind transfers. This might suggest that wealthier households, being more self-sufficient, rely less on external transfers. As income rises by 1000 units (or 1000 dollars in this case), the household's reliance or eligibility for in-kind transfers diminishes.

On the other hand, the coefficient of $Dincome$ 1.547 implies that for households with income greater than 1 i.e., more than 1000, each additional unit of Dincome (or each additional thousand dollars of income beyond 1000) increases in-kind transfers by approximately 1.547 units. This result is statistically significant (p-value = 0.020). This coefficient suggests a more nuanced relationship for higher-income households. While higher income is associated with lower tinkind overall, for households that exceed a certain income threshold (greater than 1000), the relationship between income and transfers becomes positive, implying that these households might receive more in-kind transfers as their income rises beyond the 1000 threshold.

But if we consider the Houlshold where income equals to $1 + e$, where $e > 0$, then the effect on an unit of increase in income estar to be dominated by the to parameters estimate by $income$ and $Dincome$. And the effect is positive but very small. This suggest not linear relationship on income

It is important to know that we are not considering the effect of OVB by censored procces in endogenous variable in this regression.

(b) [2 points] Calculate the percentage of censored observations (the percentage for which $tinkind = 0$). Do you expect censoring bias to be a problem in this example?

In [18]:
```
%%stata
cap gen censored = (tinkind==0)


ta censored
qui sum censored,d
local mean =`r(mean)'*100
di "The percentages of censored observations are `mean' "
```

```
. cap gen censored = (tinkind==0)

.
.
. ta censored

   censored |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      6,481       74.63       74.63
          1 |      2,203       25.37      100.00
------------+-----------------------------------
      Total |      8,684      100.00

. qui sum censored,d

. local mean =`r(mean)'*100

. di "The percentages of censored observations are `mean' "
The percentages of censored observations are 25.36849378166743

.
```

With around 25% of the sample having a $tinkind$ value of zero, though it is not overwhelmingly high, censoring bias could indeed be a concern. When a large number of observations have a value of zero, OLS may mis-estimate the true relationship between the independent variables (income and Dincome) and tinkind. OLS assumes that the relationship between the independent and dependent variables is linear and that the errors are normally distributed. However, with 25% of values of dependent variable at zero, it very possible that error are not normal.

(c) [2 points] Suppose you try and fix the problem by omitting the censored observations. Estimate the regression on the subsample of observations for which $tinkind > 0$, i.e., a truncated regression and interpret the results.

In [19]:
```
%%stata
count if tinkind > 0
reg tinkind income Dincome if tinkind > 0
```

```
. count if tinkind > 0
  6,481

. reg tinkind income Dincome if tinkind > 0

      Source |       SS           df       MS      Number of obs   =     6,4
81
-------------+----------------------------------   F(2, 6478)      =     67.
57
       Model |   4247.7349         2  2123.86745   Prob > F        =    0.00
00
    Residual |  203618.793     6,478  31.4323545   R-squared       =    0.02
04
-------------+----------------------------------   Adj R-squared   =    0.02
01
       Total |  207866.528     6,480  32.0781678   Root MSE        =    5.60
65

----------------------------------------------------------------------------
--
     tinkind | Coefficient  Std. err.      t    P>|t|     [95% conf. interva
l]
-------------+--------------------------------------------------------------
--
      income |  -2.138381   .8657954    -2.47   0.014    -3.835626   -.44113
62
     Dincome |   2.159219   .8659793     2.49   0.013     .4616132    3.8568
24
       _cons |    3.56028   .8578038     4.15   0.000     1.878701    5.2418
59
----------------------------------------------------------------------------
--

.
```

The negative coefficient in $income$ aligns with the intuition that wealthier households are less reliant on in-kind transfers, and is stronger than complete sample OLS. By excluding observations with $tinkind = 0$, we are introducing selection bias. Households with zero transfers might be fundamentally different from those receiving positive transfers. Ignoring these households could lead to biased estimates, as we are not considering the entire population and we are introducing potencial sample bias selection. While this truncated regression helps us focus on the positive transfer group, it does not fully address the censoring problem. A more appropriate approach could be a Tobit model that accounts for both censored and uncensored observations, providing more accurate estimates in the presence of censoring and/or considering a latent variable approach.

(d) [2 points] Estimate the correspoding average marginal effect of $tinkind$ with respect to $income$ for both the Tobit and truncated regression and provide an interpretation.

In [20]:
```
%%stata
qui tobit tinkind income, ll(0) vce(robust)
margins, dydx(income)
```

```
. qui tobit tinkind income, ll(0) vce(robust)

. margins, dydx(income)

Average marginal effects                                Number of obs = 8,6
84
Model VCE: Robust

Expression: Linear prediction, predict()
dy/dx wrt:  income

------------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interva
l]
-------------+----------------------------------------------------------------
--
      income |   .0085848   .0029422     2.92   0.004     .0028174    .01435
22
------------------------------------------------------------------------------
--

.
```

In [21]:
```
%%stata
qui truncreg tinkind income, ll(0) vce(robust) nocons
margins, dydx(income)
```

```
. qui truncreg tinkind income, ll(0) vce(robust) nocons

. margins, dydx(income)

Average marginal effects                                Number of obs = 6,4
81
Model VCE: Robust

Expression: Linear prediction, predict()
dy/dx wrt:  income

------------------------------------------------------------------------------
--
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interva
l]
-------------+----------------------------------------------------------------
--
      income |  -.5767515   .1871308    -3.08   0.002    -.9435211   -.20998
19
------------------------------------------------------------------------------
--

.
```

In the Tobit model, the average marginal effect of income on in-kind transfers (tinkind) is 0.0086. This coefficient reflects the effect of income on both censored observations (where tinkind is zero) and uncensored observations (where tinkind is greater than zero). Thus, this result is interpreted as follows: for each additional thousand dollars of income, the expected value of in-kind transfers increases by 8,6 dollars on average, considering both households that receive in-kind transfers and those that do not. Despite being statistically significant, this effect size is small, suggesting that an increase in income is associated with a slight rise in the amount of in-kind transfers, not necessarily a reduction as we consider earlier. This result highlights that the Tobit model accounts for the entire distribution of tinkind, including those with zero transfers, and captures a latent relationship between income and in-kind transfers.

In contrast, the truncated regression model, which only includes households with positive in-kind transfers, shows a negative marginal effect of -577usd for and increase in 1000usd of income. This indicates that for households that are already receiving in-kind transfers, an increase in income is associated with a decrease in the amount of these transfers. The negative effect is statistically significant, implying that among those who receive positive transfers, higher income is linked to a substantial reduction in the value of in-kind transfers. This interpretation is confined to the subset of households that are not censored and receive in-kind transfers, reflecting how higher income affects the level of support among these households specifically, which agrees with the earlier hipotesis.

(e) [2 points] Estimate the correspoding average marginal effect of $tinkind$ with respect to $income$ for both the Tobit and truncated regression and provide an interpretation.

Same as above?

E.4

4. A latent variable Y^* is generated by

$$\begin{aligned} Y^* &= \beta_0 + X\beta_1 + e \\ e \mid X &\sim \mathrm{N}\left(0, \sigma^2(X)\right) \\ \sigma^2(X) &= \gamma_0 + X^2\gamma_1 \\ Y &= \max\left(Y^*, 0\right) \end{aligned}$$

where X is scalar. Assume $\gamma_0 > 0$ and $\gamma_1 > 0$.

(a) [10 points] Find the log-likelihood function for the conditional distribution of Y given X.

Answer:

The pdf of Y conditional on X is normal as:

$$\phi(Y_i^*) = \frac{1}{\sqrt{2\pi\sigma^2(X_i)}}\exp\left(-\frac{(Y_i^* - \beta_0 - X_i\beta_1)^2}{2\sigma^2(X_i)}\right)$$

The joint likelihood distribution for the sample is then constructed considering that $f^*(Y|X) = \sigma^{-2}(X_i)\phi(\frac{Y_i - \beta_0 - X_i\beta_1}{\sigma^2(X_i)})$ and that $F(0|X) = 1 - \Phi(\beta_0 + X_i\beta_1)$, so:

$$L(\theta|X) = \prod_{i\in\{Y_i>0\}} \frac{1}{\sqrt{2\pi\sigma^2(X_i)}}\exp\left(-\frac{(Y_i - \beta_0 - X_i\beta_1)^2}{2\sigma^2(X_i)}\right) \prod_{i\in\{Y_i=0\}} \Phi\left(-\frac{\beta_0 + X_i\beta_1}{\sigma(X_i)}\right)$$

Then the log-likelihood function for the conditional distribution is (consider symmetric in standar normal cdf):

$$\mathrm{l}_n(\theta) = \log L(\theta|X) = \sum_{i\in\{Y_i>0\}}\left[-\frac{1}{2}\log(2\pi\sigma^2(X_i)) - \frac{(Y_i - \beta_0 - X_i\beta_1)^2}{2\sigma^2(X_i)}\right] + \sum_{i\in\{Y_i=0\}} \mathrm{lo}$$

$$\mathrm{l}_n(\theta) = \log L(\theta|X) = \sum_{i\in\{Y_i>0\}}\left[-\frac{1}{2}\log(2\pi[\gamma_0 + X^2\gamma_1]) - \frac{(Y_i - \beta_0 - X_i\beta_1)^2}{2[\gamma_0 + X^2\gamma_1]}\right] + \sum_{i\in\{Y_i=}$$

(b) [5 points] Are the parameters $\beta_0, \beta_1, \gamma_0, \gamma_1$ identified? Please explain your answer.

Answer:

For the parameters $\beta_0, \beta_1, \gamma_0, \gamma_1$ to be identified in this model, the likelihood function must contain sufficient variation in X and $Y$and the log likelihood function must be globally concave in parameters domine. The identification of β_0 and β_1 relies on the variation in X and Y when output is not censored, as the normal component of the log-likelihood for positive values of Y Y depends on the linear combination of β's and are identified through their influence on the location (mean) of Y. While γ_0 and γ_1 affect the variance of the model which varies with X, this allow us to identify correctly γ_0 and γ_1. We can also see, that as σ appears apart from β's, there are some freedom in conditions to identify, that we can probe checking globally concavity in parameters.

E.5

5. [10 points] Take the model

$$S = 1\left\{X'\gamma + u > 0\right\}$$
$$Y = \begin{cases} X'\beta + e & \text{if } S = 1 \\ \text{missing} & \text{if } S = 0 \end{cases}$$
$$\begin{pmatrix} e \\ u \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} \sigma^2 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix}\right)$$

Show $\mathbb{E}[Y \mid X, S = 1] = X'\beta + \sigma_{21}\lambda\left(X'\gamma\right)$.

Answer:

We can take the conditional expectation as follow:

$$E(Y \mid X, S = 1) = E(Y \mid S = 1, X) = E\left(X'\beta + e \mid X'\gamma + u > 0, x\right)$$

This simple by replacing the assumption equation of the model of the exercise in the require conditional expectation when $S = 1$

Now, we have:

$$E(Y \mid X, S = 1) = X'\beta + E\left(e \mid u > -X'\gamma, X\right)$$

Since e and u are multinomial distribute as proposed the exercise, we can then express the follow equation (also considering that the mean of u is zero and its standar deviation is 1):

$$e = \sigma_{21}u + \epsilon$$

Where ϵ is idenpendent of u.

So we have that:

$$E(Y \mid X, S=1) = X'\beta + E\left(\sigma_{21}u + \epsilon \mid u > -X'\gamma, X\right) = X'\beta + \sigma_{21}E\left(u \mid u > -X'\right.$$

Now we can use the fact that u is normal and applied the theroms 5.7 and 5.8 (especially theorem 5.8.6) of Bruce Hansen Introduction to Econometrics, which establish the truncated and censored moments of normal distribution, i.e , when $a \sim N(m, s)$ we can establish that $E[a \mid a > c] = m + s\lambda(-c^*)$ where $\lambda(\cdot)$ is the inverse mills ratio and $c^* = \frac{c-m}{s}$. Here $c = -x'\gamma$ and $c^* = \frac{0-X'\gamma}{1} = -x'\gamma$ since $m = 0$ and $s = 1$ para u. Then, the second term of the expression above is then $\sigma_{21}\lambda(X'\gamma)$ and we conclude showing that:

$$\mathbb{E}[Y \mid X, S=1] = X'\beta + \sigma_{21}\lambda\left(X'\gamma\right)$$

E.6

6. Suppose $\Pr\{Y = j|X, \varepsilon\} = \exp(-\lambda(X, \varepsilon))\lambda(X, \varepsilon)^j/j!$ where $\lambda(X, \varepsilon) = \exp(\beta_0 + \beta_1 X)$, for $j = 1, 2, \ldots$, $\beta_0 = \gamma_0 + \varepsilon$, with γ_0 and β_1 unknown scalar parameters, and ε is an *unobserved* random variable with $\mathbb{E}[\varepsilon] = 0$ and $\operatorname{var}(\varepsilon) = \sigma^2 > 0$ that is statistically *independent* of the univariate scalar random variable X.

(a) [3 points] Show that $\lambda(X, \varepsilon) = \exp(\gamma_0 + \beta_1 X + \ln \mathbb{E}[\exp(\varepsilon)]) \times W$. Find a mathematical expression for W.

Answer:

Remeber that $\lambda(X, \varepsilon) = e^{\gamma_0 + \beta_1 X} e^{\varepsilon}$

$$\begin{aligned}
&e^{\gamma_0+\beta_1 X} e^{\varepsilon} = e^{\gamma_0+\beta_1 X + \ln \mathbb{E} e^{\varepsilon}} W \\
&e^{\varepsilon} = \mathbb{E}\left(e^{\varepsilon}\right) W \\
&W = \frac{e^{\varepsilon}}{\mathbb{E}\left(e^{\varepsilon}\right)}
\end{aligned}$$

We have then that

$$\lambda(x, e) = e^{\gamma_0 + \beta, X + \ln \mathbb{E}(e^{\varepsilon})} \cdot \frac{e^{\varepsilon}}{\mathbb{E}\left[e^{\varepsilon}\right)}$$

(b) [2 points] Using the mathematical expression for W you found above, show that $\mathbb{E}(W) = 1$.

Answer:

$$\mathbb{E}(w) = \mathbb{E}\left(\frac{e^{\varepsilon}}{\mathbb{E}\left(e^{\varepsilon}\right)}\right) = \frac{1}{\mathbb{E}\left(e^{\varepsilon}\right)}\mathbb{E}\left(e^{\varepsilon}\right) = 1$$

Because $E\left(e^{\varepsilon}\right)$ is not a R.V and can be out of the expectation.

(c) [4 points] Using your answers to questions (a) and (b) above, show that $\mathbb{E}[Y|X] = \exp(\gamma_0 + \beta_1 X + \ln \mathbb{E}[\exp(\varepsilon)])$.

Answer:

To show that $\mathbb{E}[Y|X] = e^{\gamma_0 + \beta_1 X + \ln \mathbb{E}[\exp(\varepsilon)]}$, we proceed as follows:

Recall from part (6.a) that $\lambda(X, \varepsilon)$ is given by:

$$\lambda(X, \varepsilon) = e^{\gamma_0 + \beta_1 X + \ln(\mathbb{E}[\exp(\varepsilon)])} \times W$$

where $W = \frac{\exp(\varepsilon)}{\mathbb{E}[\exp(\varepsilon)]}$. Given that Y follows a Poisson distribution with rate parameter $\lambda(X, \varepsilon)$, the conditional expectation of Y given X and ε is:

$$\mathbb{E}[Y|X, \varepsilon] = \lambda(X, \varepsilon)$$

Substituting the expression for $\lambda(X, \varepsilon)$:

$$\mathbb{E}[Y|X, \varepsilon] = e^{\gamma_0 + \beta_1 X + \ln(\mathbb{E}[\exp(\varepsilon)])} \times W$$

We now want to find $\mathbb{E}[Y|X]$, which involves taking the expectation of $\mathbb{E}[Y|X, \varepsilon]$ over the distribution of ε:

$$\mathbb{E}[Y|X] = \mathbb{E}\left[\mathbb{E}[Y|X, \varepsilon]\right]$$

Substituting the expression for $\mathbb{E}[Y|X, \varepsilon]$:

$$\mathbb{E}[Y|X] = \mathbb{E}\left[e^{\gamma_0 + \beta_1 X + \ln(\mathbb{E}[\exp(\varepsilon)])} \times W\right]$$

Since $e^{\gamma_0 + \beta_1 X + \ln(\mathbb{E}[\exp(\varepsilon)])}$ is independent of ε, it can be factored out of the expectation:

$$\mathbb{E}[Y|X] = e^{\gamma_0 + \beta_1 X + \ln(\mathbb{E}[\exp(\varepsilon)])} \times \mathbb{E}[W]$$

From part (b), we know that $\mathbb{E}[W] = 1$. Therefore:

$$\mathbb{E}[Y|X] = e^{\gamma_0 + \beta_1 X + \ln(\mathbb{E}[\exp(\varepsilon)])} \times 1$$

Thus, we conclude:

$$\mathbb{E}[Y|X] = e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}$$

(d) [6 points] Find a mathematical expression for $\mathbb{E}[Y^2|X]$.

Answer:

To find a mathematical expression for $\mathbb{E}[Y^2|X]$. We must consider the variance definition $V[C] = \mathbb{E}[C^2] - \mathbb{E}[C]^2$ and remember that $\mathbb{E}[Y|X,\varepsilon] = \lambda(X,\varepsilon) = \mathrm{Var}(Y|X,\varepsilon)$. Thus:

$$\mathbb{E}[Y^2|X,\varepsilon] = \mathbb{E}[Y|X,\varepsilon]^2 + \mathrm{Var}(Y|X,\varepsilon) = \lambda(X,\varepsilon)^2 + \lambda(X,\varepsilon)$$

We now then, need to take expectation of the above expression respect to ε:

$$\mathbb{E}[Y^2|X] = \mathbb{E}[\lambda(X,\varepsilon)^2] + \mathbb{E}[\lambda(X,\varepsilon)]$$

Reemplazing be expression calculated earlier:

$$\mathbb{E}[Y^2|X] = \mathbb{E}[(e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])} \times W)^2] + \mathbb{E}[e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])} \times W]$$

Since $e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}$ does not depend of a particular value of ε, we have that:

$$\mathbb{E}[Y^2|X] = e^{2(\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)]))}\mathbb{E}[W^2] + e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}\mathbb{E}[W]$$

We get the expression:

$$\mathbb{E}[Y^2|X] = e^{2(\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)]))}\mathbb{E}[W^2] + e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}$$

(e) [6 points] Find a mathematical expression for $\mathbb{E}^2[Y|X]$.

Answer:

We can use the expression calculated in (6.c) to find $\mathbb{E}^2[Y|X]$:

$$\mathbb{E}^2[Y|X] = (\mathbb{E}[Y|X])^2 = (e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])})^2$$

Where

$$\mathbb{E}[Y|X] = e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}$$

(e) [5 points] Using your answers to questions (d) and (e) above, show that $\mathrm{Var}(Y|X) > \mathbb{E}[Y|X]$.

Answer:

Let's Define the conditional variance as $\text{Var}(Y|X)= \mathbb{E}[Y^2|X]

- \mathbb{E}^2[Y|X]$ we have know, according to previously calculated expression for conditional moments:

$$\text{Var}(Y|X) = e^{2(\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)]))}\mathbb{E}[W^2] + e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])} - (e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])})^2$$

Factorizing $(e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])})^2$ we have:

$$\text{Var}(Y|X) = (e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])})^2[\mathbb{E}[W^2] - 1] + e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}$$

Since $\mathbb{E}[W] = 1$ then we can use $\mathbb{E}[W]^2 = 1$

$$\text{Var}(Y|X) = (e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])})^2[\mathbb{E}[W^2] - \mathbb{E}[W]^2] + e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}$$

And rembering that $\mathbb{E}[Y|X] = e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])}$ we have:

$$\text{Var}(Y|X) = (e^{\gamma_0+\beta_1 X+\ln(\mathbb{E}[\exp(\varepsilon)])})^2[\mathbb{E}[W^2] - \mathbb{E}[W]^2] + \mathbb{E}[Y|X]$$

Since the first term of the right hand side of the last equation is always positive (because it is formed by a cuadractic function multiplying a variance) we can establish that:

$$\text{Var}(Y|X) > \mathbb{E}[Y|X]$$