Project 3 – Limited Dependent Variable Models:

Creating a Mortgage Loan Denial Model

Patrick Cataldo

Economics Department, James Madison University

ECON485: Advanced Econometrics

Professor Joane Doyle

December 18th, 2023

Table of Contents

1 Summary of Munnell

Research Agenda

Munnell's paper critically examines previous studies of credit discrimination in lending and addresses their methodological limitations. Previous studies, including the work of Black et al., King, and Schaefer & Ladd, suggested, but argued against, a focus on selected organizations and omitted some important variables such as costs and operating history.

In response, the Munnell survey aims to provide a comprehensive analysis by including a wide range of lenders and sources in its analysis. Specifically, it focuses on the need to account for all financial factors affecting borrower decisions, therefore addressing the issue of omitted variable bias plaguing previous studies Research uses 1990 data and provides information Another is included to provide a more complete view of borrower decision This approach attempts to provide an accurate understanding of the role of caste in lending decision.

Estimators used in Loan Denial

In Munnell's paper, multiple variables are used to estimate the probability of a loan applicant defaulting, which helps determine if an applicant should be denied a loan.

- <u>Financial Indicators</u>: This variable includes net wealth, liquid assets, and obligation ratios. These indicators are used to measure an applicant's finances and estimate whether they can pay back the loan.
- <u>Loan-to-Value Ratio</u>: This variable is included to measure the loan amount in relation to the estimated value of the property the applicant is seeking to get a loan for. The lower the ratio, the higher the equity stake by the applicant has, and a lower risk for the lender.

- Credit History: The applicants credit history. This variable is measured by looking at the applicant's credit behaviors like paying loans on time, proper credit management, or any red flags that pop up such as defaults or declaring bankruptcy.
- Income Stability: Because mortgage loans are long term payments, an applicant's income must be measured. The applicant's job, income, education, industry, and possibility of future unemployment are all accounted for.
- Regional Employment Conditions: These variable measures area specific and industry specific unemployment rates to measure if an applicant could be at risk of losing income or becoming unemployed.
- Probability of Unemployment: The authors looked at data from the Panel Study of Income Dynamics (PSID) to estimate the likelihood of an applicant becoming unemployment soon. Again, industry specific characteristics are considered to find any factors that might influence job stability.

Variables used to capture the cost of loan defaults

In Munnell's paper, the authors have considered and included multiple relevant variables in hopes to capture and measure the potential cost of default to the lender, should an applicant default on their loan.

- Private Mortgage Insurance (PMI): PMI reduces the potential loss to the lender if an applicant defaults on the loan. The insurance provided covers part of the lender's losses, making it so that loans with PMI less risky on a cost basis for lenders.
- Property and Neighborhood Characteristics: The variables relating to the resale value and stability of a mortgaged property include location, neighborhood characteristics, housing

price trends, local market conditions, and neighborhood demographic factors affect the profitability of selling the property, should a foreclosure occur.

- Housing Price Variability: The variability of housing prices in the location of the property to be mortgaged can impact the potential loss if the applicant defaults on the loan. The higher the variability, the higher risk of loss.
- Loan Specifics: Term and conditions, interest rates, whether the loan is fixed or adjustable rate, and the duration can influence the cost if the applicant defaults.
- Housing Market Indicators: To determine the cost of a default, vacancy rates and rent to-Value ratios are measured to assess the risk associated with the property.

2 Data Analysis

Overall Loan Applications

Total Applications Denied: 402 (14.18%)

Total Applications Accepted: 2433 (85.82%)

Applications Denied and Accepted by Race

Race	Denied	Accepted	Percent Denied	Percent Accepted
White	222	1,961	10.169%	89.831%
Minority	180	472	27.607%	72.393%

Applications Denied and Accepted by Gender

Gender	Denied	Accepted	Percent Denied	Percent Accepted
Male	311	1,938	13.828%	86.172%
Female	91	495	15.529%	84.471%

To test if there is a true difference in loan denial based on gender, a t-test will be used:

$$H_0: Males - Females = 0$$

$$H_a: Males - Females \neq 0$$

$$t = \frac{0.13828 - 0.15529}{\sqrt{0.12173 \times (\frac{1}{2249} + \frac{1}{586})}}$$

$$t = 1.0213$$

A T-value of 1.0213 corresponds to a P-value of 0.3074. Because this is a two-tailed test, and testing at α = 0.05, we cannot reject the null hypothesis that there is no true difference in application denial based on gender. To summarize, there is not enough statistical evidence to support that application denial based on gender doesn't exist.

Table 1. Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Unemployment Probability	2,835	3.82	2.08	1.80	10.60
Loan-to-Value Ratio	2,835	0.75	0.18	0.02	1.95
Debt-to-Income Ratio	2,835	33.27	11.17	0.00	300.00
Credit History	2,835	2.16	1.70	1	6
Credit Status	2,835	0.08	0.27	0	1
Self Employed	2,835	0.12	0.32	0	1
Adjustable Rate	2,835	0.34	0.47	0	1
Gender	2,835	0.79	0.41	0	1
Multi-Family Unit	2,835	0.13	0.34	0	1
Condo Unit	2,835	0.25	0.43	0	1
Single Family Unit	2,835	0.62	0.49	0	1
Race	2,835	0.23	0.42	0	1
Denied	2,835	0.14	0.35	0	1

3 Estimating a Linear Probability Model (OLS)

Table 1. Linear Probability Model Results

	Dependent variable:
	Denied
B_0	-0.291***
	(0.039)
Debt-to-Income Ratio	0.006***
	(0.001)
Race	0.087***
	(0.015)
Credit History	0.038***
	(0.004)
Loan-to-Value Ratio	0.201***
	(0.035)
Credit Status	0.209***
	(0.023)
Unemployment Probability	0.008***
	(0.003)
Self Employed	0.056***
	(0.019)
Condo Unit	-0.066***
	(0.021)
Single Family Unit	-0.072***
	(0.019)
Gender	-0.003
	(0.015)
Adjustable Rate	-0.031**
	(0.013)
Observations	2,835
R^2	0.199
Adjusted R^2	0.195
Residual Std. Error	0.313 (df = 2823)
F Statistic	63.572*** (df = 11; 2823)
Note:	*p<0.1; **p<0.05; ***p<0.01

The variable PI (Debt-To-Income Ratio) is statistically significant at the 99% confidence level. We are 99% sure that the estimated coefficient of this variable is not equal to zero.

The predicted effect on the probability of being denied a loan when the applicant's PI is 10% points higher can be calculated as:

$$Denied = 0.006\ x\ 10$$

$$Denied = 0.60$$

We are estimating an applicant who experiences a 10% increase in their Debt-To-Income ratio to be more likely to be denied a loan.

The predicted effect on the probability of being denied a loan when the applicant has bad credit can be calculated similarly to the equation above, but using the bad credit coefficient estimate:

$$Denied = 0.209\ x\ 1$$

$$Denied = 0.209$$

All else equal, and compared to an applicant with good credit, an applicant with bad credit is estimated to be more likely to be denied a loan. And this effect is statistically significant at a 99% confidence level.

If an applicant is applying for an adjustable-rate loan, the impact on the probability of denial can be calculated as:

$$Denied = -0.031\ x\ 1$$

$$Denied = -0.031$$

Our model estimates that an applicant applying for an adjustable-rate loan is less likely to be denied compared to fixed-rate loans. This effect is statistically significant at a confidence level of 95%. The reason for the negative sign on the coefficient could be due to the nature of the loan. The adjustable-rate loan offers less stability than the fixed-rate loan, which make them less desirable, so banks would probably be more willing to give out these types of loans to individuals, despite them being considered a default risk.

The difference between the predicted probability of denial for the two racial groups: White and Minority, is 0.087. An applicant who is considered a minority (Race = 1) is more likely to be denied a loan compared to an applicant who is white (Race = 0). This effect is statistically significant at a 99% confidence level. Munnell's paper finds similar results with a coefficient estimate equal to 0.07 for Race. However, Munnell's model is less confident in the estimation with a standard deviation of 3.34, compared to our 0.015 standard deviation.

Applicants with Extreme Predicted Probabilities

Lowest Predicted Probability:

Unemployment Probability	Loan-To-Value	Debt-To-Income	Credit History	Credit Status	Self-Employed	Adjustable Rate	Male	Multi - Family Unit	Condo Unit	Single Family Unit	Race	Actual Denied	Estimated Denied
3.100	0.103	16.500	1	0	0	1	1	0	1	0	0	0	-0.208

Predicted Probability: -0.2084

Highest Predicted Probability:

Unemployment Probability	Loan-To-Value	Debt-To-Income	Credit History	Credit Status	Self-Employed	Adjustable Rate	Male	Multi - Family Unit	Condo Unit	Single Family Unit	Race	Actual Denied	Estimated Denied
10.600	0.571	300	1	0	1	0	1	0	0	1	0	1	1.720

Predicted Probability: 1.7203

The characteristics of the person who has the lowest probability of being denied is someone who:

1. Has a low probability of unemployment (3.1)
2. Applies for a loan with a low Loan-To-Value ratio (0.103)
3. Has a low Debt-To-Income ratio (16.5)
4. Has a good credit history with minimal delinquent accounts (1)
5. Has good credit. (0)
6. Is not self-employed (0)
7. Applies for an Adjustable-Rate loan (1)
8. Is a male (1)
9. Applies for a condo unit (1)
10. Is white (0)

The characteristics of the person who has the highest probability of being denied is someone who:

1. Has a high probability of unemployment (10.6)
2. Applies for a loan with a medium Loan-To-Value ratio (0.571)
3. Has a high Debt-To-Income Ratio (300)
4. Has a good credit history with minimal delinquent accounts (1)
5. Has good credit (0)
6. Is self-employed (1)
7. Applies for a Fixed-Rate loan (0)
8. Is a male (1)
9. Applies for a Single-Family unit (1)
10. Is white (0)

Comparing these two applicants to the sample averages in Table 1. Descriptive Statistics, Debt-To-Income Ratio and Unemployment Probability had a big impact on these two applicants and determining their loan denial probabilities. For reference, the average Debt-To-Income ratio is around 33.27 for the sample, and the Unemployment Probability is around 3.82.

4 Quick review of Lim and Ky

After quickly reviewing the Lim and Ky paper, there are some key differences in their dataset and estimated model from Munnell's paper.

1. Lim and Ky's dataset covers 2018 through 2020.
2. The dataset used covers 6.1 million mortgage applications, while Munnell's dataset only covers 2835 mortgage applications.
3. The percentage of mortgage applications that were denied is around 3.8%, while the sample average of denied mortgage applications in Munnell's paper is around 14.8%
4. Lim and Ky estimated their Model using Race/Ethnicity (Asian, Black, Latinx, and All other Races), log loan amount, and log income. The main difference between Munnell's model and theirs in terms of races is the number of races included and the measurement / categorization of the races. Model 1 only includes races, and their second Model controls both State and Time effects, log loan amount, and log income. When controlling for these factors, Lim and Ky increase the accuracy of their estimates.

5 Estimating a Probit Model

Table 3. Linear Probability and Probit Model Results

	Dependent variable:	
	Denied	
	(OLS)	(Probit)
B_0	-0.291***	-3.825***
	(0.039)	(0.260)
Debt-to-Income Ratio	0.006***	0.031***
	(0.001)	(0.004)
Race	0.087***	0.361***
	(0.015)	(0.076)
Credit History	0.038***	0.166***
	(0.004)	(0.018)
Loan-to-Value Ratio	0.201***	1.429***
	(0.035)	(0.225)
Credit Status	0.209***	0.649***
	(0.023)	(0.101)
Unemployment Probability	0.008***	0.045***
	(0.003)	(0.015)
Self Employed	0.056***	0.305***
	(0.019)	(0.100)
Condo Unit	-0.066***	-0.245**
	(0.021)	(0.104)
Single Family Unit	-0.072***	-0.280***
	(0.019)	(0.094)
Gender	-0.003	0.006
	(0.015)	(0.084)
Adjustable Rate	-0.031**	-0.166**
	(0.013)	(0.071)
Observations	2,835	2,835
Log Likelihood		-901.728
Akaike Inf. Crit.		1,827.456
Note:	*p<0.1; **p<0.05; ***p<0.01	

Model Accuracy: 88.01%

From the results of the Probit Model, we don't see any big changes in the significance of estimated coefficients. We do see a sign flip on the gender variable, but that variable remains statistically insignificant and should be taken with a grain of salt.

6 Marginal Effects

Continuous Variables

Variable	Marginal Effect
PI (10 pt increase)	0.0553
CCS (+1 slow pay)	0.0292
LV (10 pt increase)	2.5144

Comparing the marginal effect of a 10% increase in Debt-To-Income ratio to the Linear Probability Model, the results are similar, but this estimate is slightly lower. (Table 2. Linear Probability Model Results).

Above are also the Marginal Effects of adding an additional “slow pay” credit account and a 10% increase to the Loan-To-Value ratio. All marginal effects were calculated and exported using R.

Dummy Variables

Variable	Marginal Effect
Race (Minority to White)	0.0709
Adjustable-Rate Mortgage (0 to 1)	-0.0279
Self-Employed (0 to 1)	0.0612

Comparing the marginal effects of Race on the probability of denial to our linear model, the values are similar (Table 2. Linear Probability Model Results). Given that Race is statistically significant at a 99% confidence level, we can claim that there is statistical evidence supporting racial discrimination in mortgage lending.

The rest of the marginal effects of our dummy variables also remain similar to the estimations found in the Linear Probability Model but provide a better insight to the true probability of loan denial.

7 References

Munnell, Alicia H., et al. “Mortgage Lending in Boston: Interpreting HMDA Data.” The American Economic Review, vol. 86, no. 1, Mar. 1996, pp. 25–53.

Ky, Kim-Eng, and Katherine Lim. “The Role of Race in Mortgage Application Denials.” Federal Reverse Bank of Minneapolis, May 2022, pp. 1–61.

8 Appendix

R Code

```
# -------------------------------------------------------- Code Information ----
# Project:              Project 3 - Limited Dependent Variable Models
#
# Author:               Patrick Cataldo
# Date:                 12/18/2023
# Last Modified:        11/9/2024
#
# Purpose:              The purpose of this code is to analyze the loan dataset
#                       and extract insights, as well as create both linear
#                       and probit probability models to estimate the denial of
#                       a loan.
#
# ----------------------------------------------------------------- Outline ----
# Part 1:               Summary of Munnell                                   N/A
# Part 2:               Analyze the data                               Line  35
# Part 3:               Estimate a Linear Probability Model            Line 102
# Part 4:               Review Lim and Ky Paper                              N/A
# Part 5:               Estimate the Probit Model                      Line 117
# Part 6:               Marginal Effects                               Line 132
# Part 7:               Deliverable Tables and Graphs                  Line 261
#
# ---------------------------------------------- Install necessary packages ----
# Install packages if not already installed
# install.packages("stargazer")
# install.packages("dplyr")
# install.packages("ggplot2")
# install.packages("gridExtra")

# Load Libraries
library(stargazer)
library(dplyr)
library(ggplot2)
library(gridExtra)
```

```
# ------------------------------------------------ Part 2: Analyze the data ----

# Load in data file
hdmaFile <- "C:/Users/School Account/Desktop/JMU Classes/2023-24/Fall Semester
2023/ECON485 - Advanced Econometrics/Projects/Project 3 Overhaul/hmda_subset_F23.csv"
myData <- read.csv(hdmaFile)

# Percentage of apps denied
totalDenied <- sum(myData$denied == 1)
percentDenied <- (totalDenied / nrow(myData)) * 100

# Percent of apps accepted
totalAccepted <- sum(myData$denied == 0)
percentAccepted <- (totalAccepted / nrow(myData)) * 100

# Part 2b: Separate data by RACE
# Minority applicants (race == 1)
minorityData <- subset(myData, race == 1)
minorityTotalDenied <- sum(minorityData$denied == 1)
minorityPercentDenied <- minorityTotalDenied / nrow(minorityData) * 100
minorityTotalAccepted <- sum(minorityData$denied == 0)
minorityPercentAccepted <- minorityTotalAccepted / nrow(minorityData) * 100

# White applicants (race == 0)
whiteData <- subset(myData, race == 0)
whiteTotalDenied <- sum(whiteData$denied == 1)
whitePercentDenied <- whiteTotalDenied / nrow(whiteData) * 100
whiteTotalAccepted <- sum(whiteData$denied == 0)
whitePercentAccepted <- whiteTotalAccepted / nrow(whiteData) * 100

# Create a table of applications denied and accepted by race
denied_by_race <- data.frame(
  Race = c("White", "Minority"),
  Denied = c(whiteTotalDenied, minorityTotalDenied),
  Accepted = c(whiteTotalAccepted, minorityTotalAccepted),
```

```
  Percent_Denied = c(whitePercentDenied, minorityPercentDenied),
  Percent_Accepted = c(whitePercentAccepted, minorityPercentAccepted)
)

# Part 2c: Separate data by MALE
# Male applicants (male == 1)
maleData <- subset(myData, male == 1)
maleTotalDenied <- sum(maleData$denied == 1)
malePercentDenied <- maleTotalDenied / nrow(maleData) * 100
maleTotalAccepted <- sum(maleData$denied == 0)
malePercentAccepted <- maleTotalAccepted / nrow(maleData) * 100

# Female applicants (male == 0)
femaleData <- subset(myData, male == 0)
femaleTotalDenied <- sum(femaleData$denied == 1)
femalePercentDenied <- femaleTotalDenied / nrow(femaleData) * 100
femaleTotalAccepted <- sum(femaleData$denied == 0)
femalePercentAccepted <- femaleTotalAccepted / nrow(femaleData) * 100

# Create a table of applications denied and accepted by gender
denied_by_gender <- data.frame(
  Gender = c("Male", "Female"),
  Denied = c(maleTotalDenied, femaleTotalDenied),
  Accepted = c(maleTotalAccepted, femaleTotalAccepted),
  Percent_Denied = c(malePercentDenied, femalePercentDenied),
  Percent_Accepted = c(malePercentAccepted, femalePercentAccepted)
)

# Get descriptive statistics to view sample averages
descriptiveStatsTable <- stargazer(myData, type = "html",
                                   title = "Descriptive Statistics",
                                   summary = TRUE, digits = 2)

# ---------------------------- Part 3: Estimate a Linear Probability Model ----
olsModel <- lm(denied ~ pi + race + ccs + lv + pbcr + uria + se + condo + sfam +
                 male + iraj, data = myData)
```

```r
# Add predicted values to the dataset
myData$denied_LPM_hat <- fitted(olsModel)

# Identify the individuals with the lowest and highest predicted probabilities
min_prob <- min(myData$denied_LPM_hat)
max_prob <- max(myData$denied_LPM_hat)

# Get the observations with min and max predicted probabilities
min_prob_obs <- myData[which.min(myData$denied_LPM_hat), ]
max_prob_obs <- myData[which.max(myData$denied_LPM_hat), ]

# -------------------------------------- Part 5: Estimate the Probit Model ----
probitModel <- glm(denied ~ pi + race + ccs + lv + pbcr + uria + se + condo +
                     sfam + male + iraj, family = binomial(link = "probit"),
                   data = myData)

# Add predicted probabilities to the dataset
myData$denied_Probit_hat <- predict(probitModel, type = "response")

# Construct Denied_hat values based on a cutoff of 0.5
estimatedDenied <- ifelse(myData$denied_Probit_hat > 0.5, 1, 0)
myData$denied_HAT <- estimatedDenied

# Measure accuracy of the model
accuracy <- mean(myData$denied == myData$denied_HAT) * 100

# ---------------------------------------------- Part 6: Marginal Effects ----

## A: Continuous X Variables

# Calculate the linear predictor (z) for each observation
linearPredictor <- predict(probitModel, type = "link")

# Compute the standard normal density function phi(z) for each observation
phi_z <- dnorm(linearPredictor)
```

```
# Extract coefficients from the probit model
coefficients <- coef(probitModel)

# a.1 Marginal Effect of a 10 Percentage Point Increase in PI
delta_PI <- 10
marginalEffects_PI <- phi_z * coefficients["pi"] * delta_PI
averageMarginalEffect_PI <- mean(marginalEffects_PI)

# a.2 Effect of One More "Slow Pay" on Credit Account (CCS)
delta_CCS <- 1
marginalEffects_CCS <- phi_z * coefficients["ccs"] * delta_CCS
averageMarginalEffect_CCS <- mean(marginalEffects_CCS)

# a.3 Marginal Effect of a 10 Percentage Point Increase in Loan-to-Value Ratio
delta_LV <- 10
marginalEffects_LV <- phi_z * coefficients["lv"] * delta_LV
averageMarginalEffect_LV <- mean(marginalEffects_LV)

## B: Dichotomous X Variables (Dummy Variables)

# Calculate the mean of each variable (excluding the dependent variable)
meanValues <- myData %>%
  select(pi, race, ccs, lv, pbcr, uria, se, condo, sfam, male, iraj) %>%
  summarise_all(mean)
meanValues <- as.list(meanValues)

# b.1 Marginal Effect for RACE
z_race1 <- coefficients["(Intercept)"] +
  coefficients["race"] * 1 +
  coefficients["pi"]   * meanValues$pi +
  coefficients["ccs"]  * meanValues$ccs +
  coefficients["lv"]   * meanValues$lv +
  coefficients["pbcr"] * meanValues$pbcr +
  coefficients["uria"] * meanValues$uria +
  coefficients["se"]   * meanValues$se +
```

```
  coefficients["condo"]* meanValues$condo +
  coefficients["sfam"] * meanValues$sfam +
  coefficients["male"] * meanValues$male +
  coefficients["iraj"] * meanValues$iraj

z_race0 <- coefficients["(Intercept)"] +
  coefficients["race"] * 0 +
  coefficients["pi"]   * meanValues$pi +
  coefficients["ccs"]  * meanValues$ccs +
  coefficients["lv"]   * meanValues$lv +
  coefficients["pbcr"] * meanValues$pbcr +
  coefficients["uria"] * meanValues$uria +
  coefficients["se"]   * meanValues$se +
  coefficients["condo"]* meanValues$condo +
  coefficients["sfam"] * meanValues$sfam +
  coefficients["male"] * meanValues$male +
  coefficients["iraj"] * meanValues$iraj

P_race1 <- pnorm(z_race1)
P_race0 <- pnorm(z_race0)
marginalEffect_RACE <- P_race1 - P_race0

# b.2 Marginal Effect for IRAJ
z_iraj1 <- coefficients["(Intercept)"] +
  coefficients["iraj"] * 1 +
  coefficients["pi"]   * meanValues$pi +
  coefficients["race"] * meanValues$race +
  coefficients["ccs"]  * meanValues$ccs +
  coefficients["lv"]   * meanValues$lv +
  coefficients["pbcr"] * meanValues$pbcr +
  coefficients["uria"] * meanValues$uria +
  coefficients["se"]   * meanValues$se +
  coefficients["condo"]* meanValues$condo +
  coefficients["sfam"] * meanValues$sfam +
  coefficients["male"] * meanValues$male
```

```
z_iraj0 <- coefficients["(Intercept)"] +
  coefficients["iraj"] * 0 +
  coefficients["pi"]   * meanValues$pi +
  coefficients["race"] * meanValues$race +
  coefficients["ccs"]  * meanValues$ccs +
  coefficients["lv"]   * meanValues$lv +
  coefficients["pbcr"] * meanValues$pbcr +
  coefficients["uria"] * meanValues$uria +
  coefficients["se"]   * meanValues$se +
  coefficients["condo"]* meanValues$condo +
  coefficients["sfam"] * meanValues$sfam +
  coefficients["male"] * meanValues$male

P_iraj1 <- pnorm(z_iraj1)
P_iraj0 <- pnorm(z_iraj0)
marginalEffect_IRAJ <- P_iraj1 - P_iraj0

# b.3 Marginal Effect for SE
z_se1 <- coefficients["(Intercept)"] +
  coefficients["se"]   * 1 +
  coefficients["pi"]   * meanValues$pi +
  coefficients["race"] * meanValues$race +
  coefficients["ccs"]  * meanValues$ccs +
  coefficients["lv"]   * meanValues$lv +
  coefficients["pbcr"] * meanValues$pbcr +
  coefficients["uria"] * meanValues$uria +
  coefficients["condo"]* meanValues$condo +
  coefficients["sfam"] * meanValues$sfam +
  coefficients["male"] * meanValues$male +
  coefficients["iraj"] * meanValues$iraj

z_se0 <- coefficients["(Intercept)"] +
  coefficients["se"]   * 0 +
  coefficients["pi"]   * meanValues$pi +
  coefficients["race"] * meanValues$race +
  coefficients["ccs"]  * meanValues$ccs +
```

```
  coefficients["lv"]   * meanValues$lv +
  coefficients["pbcr"] * meanValues$pbcr +
  coefficients["uria"] * meanValues$uria +
  coefficients["condo"]* meanValues$condo +
  coefficients["sfam"] * meanValues$sfam +
  coefficients["male"] * meanValues$male +
  coefficients["iraj"] * meanValues$iraj

P_se1 <- pnorm(z_se1)
P_se0 <- pnorm(z_se0)
marginalEffect_SE <- P_se1 - P_se0

# ------------------------------------------------------ Deliverable Tables ----

# Set the output HTML file path to your specified directory
htmlFile <- "C:/Users/School Account/Desktop/JMU Classes/2023-24/Fall Semester
2023/ECON485 - Advanced Econometrics/Projects/Project 3
Overhaul/Project3_Outputs.html"

# Create the HTML file and write outputs
sink(htmlFile, type = "output", split = TRUE)

cat("<html><head><title>Project 3 Outputs</title></head><body>")
cat("<h1>Project 3 - Limited Dependent Variable Models</h1>")

# Part 2 Outputs
cat("<h2>Part 2: Analyze the Data</h2>")

cat("<h3>Overall Loan Applications</h3>")
cat(sprintf("<p>Total Applications Denied: %d (%.2f%%)</p>", totalDenied,
            percentDenied))
cat(sprintf("<p>Total Applications Accepted: %d (%.2f%%)</p>", totalAccepted,
            percentAccepted))

cat("<h3>Applications Denied and Accepted by Race</h3>")
stargazer(denied_by_race, type = "html", summary = FALSE, rownames = FALSE)
```

```
cat("<h3>Applications Denied and Accepted by Gender</h3>")
stargazer(denied_by_gender, type = "html", summary = FALSE, rownames = FALSE)

cat("<h3>Descriptive Statistics</h3>")
descriptiveStatsTable

# Part 3 Outputs
cat("<h2>Part 3: Estimate a Linear Probability Model</h2>")
stargazer(olsModel, type = "html",
          title = "Linear Probability Model Results",
          dep.var.labels = "Denied",
          covariate.labels = c("B0", "Debt-to-Income Ratio", "Race",
                              "Credit History", "Loan-to-Value Ratio",
                              "Credit Status",
                              "Unemployment Probability", "Self Employed",
                              "Condo Unit", "Single Family Unit",
                              "Gender", "Adjustable Rate"),
          intercept.bottom = FALSE, intercept.top = TRUE,
          digits = 3)

cat("<h3>Individuals with Extreme Predicted Probabilities</h3>")
cat("<p><strong>Lowest Predicted Probability:</strong></p>")
stargazer(min_prob_obs, type = "html", summary = FALSE, rownames = FALSE)
cat(sprintf("<p>Predicted Probability: %.4f</p>", min_prob))

cat("<p><strong>Highest Predicted Probability:</strong></p>")
stargazer(max_prob_obs, type = "html", summary = FALSE, rownames = FALSE)
cat(sprintf("<p>Predicted Probability: %.4f</p>", max_prob))

# Part 5 Outputs
cat("<h2>Part 5: Estimate the Probit Model</h2>")
stargazer(olsModel, probitModel, type = "html",
          title = "Regression Results - OLS and Probit Models",
          dep.var.labels = c("Denied (OLS)", "Denied (Probit)"),
          covariate.labels = c("B0", "Debt-to-Income Ratio", "Race",
```

```
                                "Credit History", "Loan-to-Value Ratio",
                                "Credit Status",
                                "Unemployment Probability", "Self Employed",
                                "Condo Unit", "Single Family Unit",
                                "Gender", "Adjustable Rate"),
          model.names = FALSE, intercept.bottom = FALSE, intercept.top = TRUE,
          digits = 3)

cat(sprintf("<p>Model Accuracy: %.2f%%</p>", accuracy))

# Part 6 Outputs
cat("<h2>Part 6: Marginal Effects</h2>")

# Output Marginal Effects for Continuous Variables
cat("<h3>A: Continuous Variables</h3>")
cat("<table border='1'>")
cat("<tr><th>Variable</th><th>Marginal Effect</th></tr>")
cat(sprintf("<tr><td>PI (10 pt increase)</td><td>%.4f</td></tr>",
            averageMarginalEffect_PI))
cat(sprintf("<tr><td>CCS (+1 slow pay)</td><td>%.4f</td></tr>",
            averageMarginalEffect_CCS))
cat(sprintf("<tr><td>LV (10 pt increase)</td><td>%.4f</td></tr>",
            averageMarginalEffect_LV))
cat("</table>")

# Output Marginal Effects for Dummy Variables
cat("<h3>B: Dummy Variables</h3>")
cat("<table border='1'>")
cat("<tr><th>Variable</th><th>Marginal Effect</th></tr>")
cat(sprintf("<tr><td>Race (Minority to White)</td><td>%.4f</td></tr>",
            marginalEffect_RACE))
cat(sprintf("<tr><td>Adjustable Rate Mortgage (0 to 1)</td><td>%.4f</td></tr>",
            marginalEffect_IRAJ))
cat(sprintf("<tr><td>Self-Employed (0 to 1)</td><td>%.4f</td></tr>",
            marginalEffect_SE))
cat("</table>")
```

```
cat("</body></html>")

# Close the HTML file
sink()

# ---------------------------------------------------------- End of Script ----
```