INSTITUTO DE ECONOMÍA
FACULTAD DE CIENCIAS ECONÓMICAS
Y ADMINISTRATIVAS

Lecture : Econometric Theory II
Professor : Tomás Rau TA : Valle and Barrueto
Date : April 1, 2023
Student : Valentina Andrade
email : vandrade@uc.cl

Report 1

Categorical response models

Abstract

The following report contains the assignment nº1 of the Econometric Theory II course direct by Professor Tomás Rau and the assistants Eduardo Barrueto and Nicolás Valle.

The report consists of 4 parts: 3 sections where the proposed problems are developed and 1 section an appendix where some important statistical tools to consider when designing the previous sections are emphasized. The first section aims to review several discrete choice models using data on parental choice of schools in Chile. The second section seeks to replicate the article by McDevitt, 2011, where the decision to change the name of the firms is studied. This exercise improved our knowledge of multinomial models, particularly the difference between predictors that vary between and within individuals. The third section seeks to reproduce and extend the analysis of the article by Dominguez, 2022 , whose main objective is to understand the effect of a transport system reform on the incidence of crime. This exercise allowed us to learn more about counting models.

In the following web link you can find the GitHub repository that allows you to reproduce the whole

Question 1

a. Maximum Likelihood algorithm for logit model

The **R code** demonstrates how to perform maximum likelihood estimation for a logistic regression model using two different methods and how to obtain robust standard errors for the coefficients. The first part of the code uses the built-in R function glm() to fit a logistic regression model, while the second part defines the likelihood function, its first and second derivatives (Hessian matrix), and uses the *Newton-Raphson* algorithm to maximize the likelihood function.

The *Newton-Raphson* algorithm is an iterative method that approximates the zeros of a function, which is the first-order condition for maximizing the log-likelihood function. In this code, the algorithm is run until the estimated values of the parameters converge to a stable solution with a high degree of precision, defined by a tolerance level and maximum number of iterations.

The use of robust standard errors is also demonstrated in the code. Robust standard errors are recommended when there is heteroscedasticity in the data, as failing to use them can lead to invalid inferences on the estimators. Robust standard errors adjust for heteroscedasticity by estimating the covariance matrix of the parameter estimates using methods that are robust to heteroscedasticity. This results in standard errors that are more reliable and valid for statistical inference.

Overall, this code provides a useful demonstration of how to perform maximum likelihood estimation for a logistic regression model using different methods and how to obtain reliable standard errors for the coefficients, which is essential for valid statistical inference.

b. Marginal effects

	m1 municipal b	m11 _ b
cod_nivel	-0.098*	-0.024*
es_mujer=1	-0.257	-0.063
prioritario=1	0.059	0.014
alto_rendimiento=1	0.352	0.086
Constant	0.961*	
Observations	222	222
Pseudo R^2	.0152323	

Table 1: Coefficients and Marginal effects for Binary Logit model

Table 1 represent the estimation of the following logit model

$$muncipal_i = \beta_0 + \beta_1 curso + \beta_2 mujer + \beta_3 prioritario + \beta_4 rendimiento$$

The first column contains the coefficients but in the logistic model, our estimation scale is the log-odds but we would like to interpret our model in the probability scale. In the second column we show the marginal effects of each dependent variable. For binary variables, the interpretation of the marginal effects is the difference in the probability between category 0 vs 1. For continuos variables, the marginal effect interpretation is "with a marginal change in x we can show that the probability of y decrease/increase". For both interpretations, the other independent variables are *at means.*

Based on the marginal effects, we can interpret the impact of each variable on the probability of being in a municipal school or not:

- A one-unit increase in *codnivel* leads to a 0.024 percentage point decrease in the probability of being in a municipal school, holding other variables constant.

- Being female (*esmujer=1*) decreases the probability of being in a municipal school by 0.063 percentage points, holding other variables constant.
- Being a prioritized student (*prioritario=1*) increases the probability of being in a municipal school by 0.014 percentage points, holding other variables constant.
- Being a high-performing student (*altorendimiento=1*) increases the probability of being in a municipal school by 0.086 percentage points, holding other variables constant.

It is important to note that none of these effects are statistically significant at conventional levels (e.g., $p<0.05$). The overall model's pseudo R-squared indicates that the model explains only a small portion of the variance in the probability of being in a municipal school. Therefore, caution should be taken when interpreting the marginal effects, as they may not be reliable indicators of the true population effects.

c. Type of schools

The table 2 provides a frequency distribution of the *type_school* variable, which classifies schools according to their administration type and the inclusion of the PIE program. There are four categories: municipal without PIE (52 schools or 23.42%), municipal with PIE (82 schools or 36.94%), subvencionado without PIE (46 schools or 20.72%), and subvencionado with PIE (42 schools or 18.92%). The total number of schools in the sample is 222.

The main disadvantage of using only the municipal variable is that it does not capture the heterogeneity of schools within the public or private sectors. For example, public schools can have different characteristics depending on their location, resources, and programs, such as PIE. By using the *type_school* variable, we can differentiate between different types of schools within each sector and incorporate this information into the analysis.

The choice of the model depends on the research question and the nature of the dependent variable. If the interest is only to predict whether a school is municipal or not, the binary model using the municipal variable may be appropriate. However, if we want to understand the factors that determine the type of school a student attends, a multinomial model using the *type_school* variable is necessary. The multinomial model allows us to estimate the probabilities of belonging to each category of the dependent variable, taking into account the relative impact of the independent variables on each outcome.

	Freq.	Percent
municipalnoPIE	52	23.42
municipalPIE	82	36.94
subvencionadonoPIE	46	20.72
subvencionadoPIE	42	18.92
Total	222	100

Table 2: Type school by PIE and administration

In this case, the disaggregation is important in analyzing the choice that parents make when selecting a school for their children, as it provides more detail on the options available to them. Using only the municipal variable described in the previous model could be a disadvantage as it does not capture the variety of options available to parents. By considering the type of administration and inclusion program, we can better understand the factors that influence parents' decisions in choosing a particular school for their children.

It is important to note that the use of the *type_school* variable would require a multinomial choice model instead of a binary choice model. This is because parents have more than two options to choose from when selecting a school, and the decision process is likely to be influenced by different factors depending on the type of school. Thus, a multinomial choice model would be better suited to capture the complexity of the decision-making process.

d. Multinomial logit

In the multinomial logit model with individual variables, the marginal effects represent the change in the probability of choosing one type of school over the base category (municipal No PIE) due to a one-unit change in the corresponding independent variable, holding all other variables constant.

Looking at the marginal effects table, we can see that being a woman ($es\_mujer = 1$) increases the probability of choosing subvencionado PIE schools over municipal No PIE schools, but decreases the probability of choosing the other two types of schools. This is consistent with the results of the binary logit model, which showed that being a woman decreases the probability of being in a municipal school.

Being prioritario also has different effects on the probability of choosing different types of schools. In the multinomial logit model, prioritario status decreases the probability of choosing municipal PIE schools and increases the probability of choosing subvencionado no PIE schools, but has no significant effect on the probability of choosing subvencionado PIE schools. This is somewhat consistent with the results of the binary logit model, which showed that prioritario status increases the probability of being in a subvencionado school (which includes both subvencionado PIE and subvencionado no PIE schools), but did not differentiate between the two types of subvencionado schools.

Having *alto_rendimiento* status also has different effects on the probability of choosing different types of schools in the multinomial logit model. *Alto_rendimiento* status increases the probability of choosing subvencionado no PIE schools and has no significant effect on the probability of choosing the other types of schools. This is somewhat consistent with the results of the binary logit model, which showed that having *alto_rendimiento* status increases the probability of not being in a municipal school.

In general, it is important to note that the multinomial logit model captures the effects of the independent variables on the choice among all four types of schools, while the binary logit model only captures the effects on the choice between municipal and non-municipal schools. Therefore, the results from the two models are not directly comparable. However, we can see some similarities and differences in the effects of the independent variables on the probability of choosing municipal schools in the two models. For example, being a woman and having *alto_rendimiento* status both decrease the probability of being in a municipal school in both models, while the effects of prioritario status are somewhat different in the two models.

	m2		
	municipalPIE	subvencionadonoPIE	subvencionadoPIE
	b	b	b
cod_nivel	-0.016	0.096	-0.151**
es_mujer=1	0.382	1.297***	0.101
prioritario=1	-0.011	0.527	0.969**
alto_rendimiento=1	0.106	-0.249	0.505
Constant	0.325	-1.865*	0.153
Observations	222		
Pseudo R^2	.042493		

Table 3: Coefficients Multinomial Logit

	m21			
	cod_nivel	1.es_mujer	1.prioritario	1.alto_rendimiento
	b	b	b	b
1._predict	0.004	-0.098	-0.066	-0.021
2._predict	-0.000	-0.006	-0.110*	0.008
3._predict	0.021	0.159***	0.047	-0.063
4._predict	-0.024***	-0.054	0.129**	0.076
Observations	222			
Pseudo R^2				

Table 4: Marginal Effects to Multinomial Logit

Economic interpretation

Given that the "municipal" schools are public schools, it is interesting to note that being a priority student (i.e. coming from a lower socioeconomic background) does not necessarily increase the probability of choosing a municipal school with a PIE program. Instead, being a priority student seems to increase the probability of choosing a subvencionado school (with or without a PIE program). This suggests that, at least for the parents in this sample, the perceived quality of education in the subvencionado schools may be higher than that in the municipal schools with PIE programs, even if the subvencionado schools require some level of out-of-pocket expense.

It is also worth noting that being female has a positive effect on the probability of choosing a type 3 school (subvencionado no PIE), but a negative effect on the probability of choosing the other school types. This may reflect gender differences in preferences for different types of schools or in the perceived quality of education offered by different types of schools.

Finally, the results from the binary logit model on the probability of choosing a municipal school suggest that, holding all other variables constant, being a priority student or having high academic performance does not significantly affect the probability of choosing a municipal school. This is in contrast to the multinomial logit results, which suggest that being a priority student actually decreases the probability of choosing a municipal school with a PIE program. This may indicate that the decision to choose a municipal school is driven more by factors such as location or access to transportation, rather than by factors related to the student's socioeconomic status or academic performance (ie. individual variables)

e. Conditional logit

In general we start with a brief summary of both methods: conditional logit and multinomial logit are used to model discrete choice behavior. However, they differ in their assumptions and applications.

Conditional logit is used to model choices made by individuals within a group or cluster of individuals who share similar characteristics. It assumes that each individual in the group faces the same set of alternatives and that the choices made by each individual are independent of the choices made by the others in the group. Therefore, conditional logit is appropriate for modeling choices made by individuals within a fixed set of alternatives.

On the other hand, multinomial logit is used to model choices made by individuals across different groups or clusters, where each group has a different set of alternatives. Multinomial logit assumes that the choices made by each individual are independent of the choices made by others, regardless of the group they belong to. Therefore, multinomial logit is appropriate for modeling choices made by individuals across a variety of alternatives.

To summarize, conditional logit is typically used for between-group analysis, while multinomial logit is used for within-group analysis.

Advantage and disvantage

The advantage of including characteristics such as distance, Simce scores and percentage of priority students is that the model will be able to control for differences in quality of education and accessibility to schools. These variables are likely to be important determinants of the choice of schools, as families will be more likely to choose schools that are closer to their homes and have higher Simce scores. As we can notice, in econometrics terms, an important difference with the previous predictors is that this regressors have not differences within individuals. One major disadvantage of using conditional logit for within-group analysis is that it does not account for unobserved individual-level heterogeneity. In other words, it assumes that all individuals within a group are similar in terms of their preferences and characteristics, which may not be true in reality.

For example, in the school choice example, if there are unobserved individual-level characteristics such as family income, parental education, or prior academic achievement that influence school choice but are not included in the model, then the estimated coefficients may be biased and the predicted probabilities of school choice may be inaccurate. This can lead to incorrect policy recommendations or decisions based on the model's results.

Furthermore, conditional logit assumes that all attributes of the alternatives are equally important for all individuals within a group, which may not be the case. Some individuals may prioritize certain attributes more than others when making choices, and this preference heterogeneity cannot be captured by conditional logit.

Therefore, if there is substantial unobserved individual-level heterogeneity or preference heterogeneity within a group, conditional logit may not be the appropriate method to use and other modeling techniques that account for these factors, such as mixed logit, may be more appropriate.

Interpretation

The table 6 shows the marginal effects of the conditional logit model that was estimated using the characteristics of the four available school alternatives. The model estimates the impact of the different characteristics on the choice of each alternative, holding constant the other variables in the model. The coefficients can be interpreted as the change in the probability of choosing a particular alternative for a one-unit change in the respective variable, while holding the other variables constant.

	m3 type_school b
distancia	-0.043*
p_prioritario	3.652***
Puntaje promedio del establecimiento en Lectura	-0.015***
Puntaje promedio del establecimiento en Matemática	0.015***
Observations	888
Pseudo R^2	

Table 5: Coefficients - Conditional Logit

	m31 distancia b	p_prioritario b	SIMCE promLectura b	SIMCE prom Matemática b
1. × school=1	-0.008*	0.677***	-0.003***	0.003***
1. × school=2	0.003**	-0.243**	0.001***	-0.001***
1. × school=3	0.003*	-0.231***	0.001***	-0.001***
1. × school=4	0.002**	-0.205***	0.001***	-0.001***
2. × school=1	0.003**	-0.242**	0.001***	-0.001***
2. × school=2	-0.008**	0.682**	-0.003***	0.003***
2. × school=3	0.003**	-0.230***	0.001***	-0.001***
2. × school=4	0.002**	-0.207**	0.001***	-0.001***
3. × school=1	0.003*	-0.230***	0.001***	-0.001***
3. × school=2	0.003*	-0.231***	0.001***	-0.001***
3. × school=3	-0.008*	0.653***	-0.003***	0.003***
3. × school=4	0.002**	-0.193***	0.001***	-0.001***
4. × school=1	0.002**	-0.204***	0.001***	-0.001***
4. × school=2	0.002**	-0.208**	0.001***	-0.001***
4. × school=3	0.002**	-0.192***	0.001***	-0.001***
4. × school=4	-0.007**	0.604***	-0.003***	0.002***
Observations	888			
Pseudo R^2				

Table 6: Marginal Effect for the probability to choose each alternative when individuals characteristics varies across alternatives

For example, let's take the *subvencionado no PIE* schools and consider only the effects of the distance variable. An increase in the distance of alternative 1 generates an increase in the probability of choosing a *subvencionado no PIE* school by 0.25 percentage points (pp), while an increase in the distance of alternative 1 do not generate changes in the probability of choosing *subvencionado no PIE* schools. Similarly, an increase in the distance of alternative 3 induces a reduction in the probability of choosing *subvencionado no PIE* schools by 0.24pp, while an increase in the distance of alternative 4 induces an reduce in the probability of choosing *subvencionado no PIE* schools by 0.075pp.

The same interpretation can be made for the other independent variables, such as the percentage of priority students and the Simce scores in reading and math. For example, an increase in the percentage of priority students of a *municipal PIE* school by one unit leads to an increase in the probability of choosing *municipal PIE* schools by 2.59pp, while a decrease in the percentage of priority students of a *municipal no PIE* school by one unit leads to an increase in the probability of choosing *municipal no PIE* schools by 2.5pp.

In general, the results suggest that as the distance of a school type increases, the probability of choosing it decreases. As the percentage of priority students of a school type increases, the probability of choosing that type also increases. Finally, as the SIMCE scores of an alternative type increase, the probability of choosing that type also increases.

f. Alternative specific conditional logit

A mixed logit or alternative-specific conditional logit model is a more flexible model that allows for both between-group (or population-level) and within-group (or individual-level) variation in preferences. This model is useful when we want to estimate the impact of unobserved heterogeneity (i.e. differences in preferences between individuals that are not captured by observed variables) on choices.

In the case of your example, we can use a mixed logit to estimate the choice between the 4 types of schools based on their distance, Simce scores, and percentage of priority students (the variables used in the conditional logit), as well as additional observed and unobserved variables that may affect individual preferences. For example, we could include demographic characteristics of the students and their families, such as class, gender, income, and education level, as well as unobserved factors such as unmeasured school characteristics or unobserved preferences for certain types of schools.

	m4 type_school b	2 b	3 b	4 b
distancia	-0.054**			
p_prioritario	2.696*			
Puntaje promedio del establecimiento en Lectura	-0.017***			
Puntaje promedio del establecimiento en Matemática	0.016***			
cod_nivel		-0.023	0.102	-0.179**
es_mujer=1		0.287	1.350***	0.166
prioritario=1		-0.018	0.631	1.049**
alto_rendimiento=1		0.059	-0.373	0.461
Constant		0.533	-1.906*	0.545
Observations	888			
Pseudo R^2				

Table 7: Coefficients Mixed Multinomial Logit

The mixed logit model assumes that the utility (or desirability) of each alternative is a function of the observed and unobserved variables, as well as a random error term that captures the unobserved variation in preferences. Specifically, the utility of alternative j for individual i can be expressed as:

$$U_{ij} = \beta_{0j} + \beta_{1j} \cdot X_{1ij} + \beta_{2j} \cdot X_{2ij} + \beta_{3j} \cdot X_{3ij} + \varepsilon_{ij}$$

where β_{0j} is the alternative-specific intercept for alternative j,$\beta_{1j}, \beta_{2j}, \beta_{3j}$ are the alternative-specific coefficients for the three observed variables (distance, Simce scores, and percentage of priority students), $X_{1ij}, X_{2ij}, X_{3ij}$ are the corresponding values of these variables for individual i and alternative j, and ε_{ij} is the random error term.

The mixed logit model allows the coefficients $\beta_{1j}, \beta_{2j}, \beta_{3j}$ to vary randomly across individuals, reflecting the fact that different people may have different preferences for these variables. Specifically, we assume that the coefficients follow a normal distribution with mean and variance that are estimated from the data. This allows us to estimate the impact of unobserved heterogeneity on choices, and to capture the fact that some individuals may be more or less sensitive to certain variables than others.

	m41 distancia b	p_prioritario b	SIMCELeg b	SIMCEMat b	codnivel b	mujer b	prioritario b	rendimiento b
1. × school=1	-0.009**	0.464*	-0.003***	0.003***	0.005			
1. × school=2	0.005**	-0.240*	0.002***	-0.001***	0.005	-0.025	0.002	-0.005
1.× school=3	0.002**	-0.118*	0.001***	-0.001***	0.005	-0.055***	-0.028	0.016
1. × school=4	0.002**	-0.112*	0.001***	-0.001***	0.005	-0.007	-0.045**	-0.019
2. × school=1	0.005**	-0.234*	0.001***	-0.001***	0.000			
2. × school=2	-0.012**	0.616*	-0.004***	0.004***	-0.000	0.064	-0.004	0.013
2. × school=3	0.004**	-0.184*	0.001***	-0.001***	0.000	-0.085***	-0.044	0.025
2. × school=4	0.004**	-0.177*	0.001***	-0.001***	-0.001	-0.010	-0.071**	-0.030
3. × school=1	0.002**	-0.118*	0.001***	-0.001***	0.023			
3. × school=2	0.004**	-0.191*	0.001***	-0.001***	0.023	-0.020	0.001	-0.004
3. × school=3	-0.008**	0.390*	-0.002***	0.002***	0.022	0.180***	0.092	-0.053
3. × school=4	0.002**	-0.088*	0.001***	-0.001***	0.023	-0.005	-0.036**	-0.015
4. × school=1	0.002**	-0.112*	0.001***	-0.001***	-0.028***			
4. × school=2	0.004**	-0.185*	0.001***	-0.001***	-0.028***	-0.019	0.001	-0.004
4. × school=3	0.002**	-0.087*	0.001***	-0.001***	-0.028***	-0.040***	-0.021	0.012
4. × school=4	-0.008**	0.376*	-0.002***	0.002***	-0.027***	0.023	0.153**	0.065
Observations	888							
Pseudo R^2								

Table 8: Marginal Effects for Mixed Logit - Full Model

In Table 8 there are present the marginal effects of a mixed logit model for different predictors. The effects are expressed in percentage points, meaning that a one-unit increase in the predictor variable leads to a percentage-point increase in the outcome variable.

The table shows the marginal effects of the multinomial mixed logit model for the choice school model by parents. The dependent variable is a categorical variable with four categories, where 1 = "municipal PIE," 2 = "Municipal no PIE," 3 = "subvencionado PIE," and 4 = "subvencionado no PIE." The independent variables are as follows: distance (from home to the school), pprioritario (whether the student is from a prioritized school), average score of the establishment in reading, average score of the establishment in mathematics, codnivel (school level), 1.esmujer (whether the student is female), 1.prioritario (whether the student is prioritized), and 1.altorendimiento (whether the student has a high academic performance).

The marginal effects indicate the percentage change in the probability of selecting a particular category of school when a given independent variable changes by one unit. For example, if the distance variable increases by one unit, the probability of selecting a municipal PIE school decreases by 0.009 percentage points. Similarly, if the student is prioritized, the probability of selecting a municipal PIE school increases by 0.464 percentage points.

1. outcome × typeschool = 1: The probability of selecting a municipal PIE school decreases by 0.009 percentage points when the distance variable increases by one unit. If the student is prioritized, the probability of selecting a municipal PIE school increases by 0.464 percentage points. An increase of one unit in the average score of the establishment in reading decreases the probability of selecting a municipal PIE school by 0.003 percentage points. An increase of one unit in the average score of the establishment in mathematics increases the probability of selecting a municipal PIE school by 0.003 percentage points.

2. outcome × typeschool = 2: An increase of one unit in the distance variable increases the probability of selecting a municipal no PIE school by 0.005 percentage points. If the student is prioritized, the probability of selecting a municipal no PIE school decreases by 0.24 percentage points. An increase of one unit in the average score of the establishment in reading decreases the probability of selecting a municipal no PIE school by 0.002 percentage points. An increase of one unit in the average score of the establishment in mathematics increases the probability of selecting a municipal no PIE school by 0.001 percentage points. If the student is female, the probability of selecting a municipal no PIE school decreases by 0.025 percentage points. If the student has a high academic performance, the probability of selecting a municipal no PIE school decreases by 0.005 percentage points.

3. outcome × typeschool = 3: An increase of one unit in the distance variable increases the probability of selecting a subvencionado PIE school by 0.002 percentage points. If the student is prioritized, the probability of selecting

a subvencionado PIE school decreases by 0.118 percentage points. An increase of one unit in the average score of the establishment in reading decreases the probability of selecting a subvencionado PIE school by 0.001 percentage points. An increase of one unit in the average score of the establishment in mathematics increases the probability of selecting a subvencionado PIE school by 0.001 percentage points. If the student is female, the probability of selecting a subvencionado PIE school decreases by 0.055 percentage points. If the student is prioritized, the probability of selecting a subvencionado PIE school decreases by 0.028 percentage points. If the student has a high academic performance, the probability of selecting a subvencionado PIE school increases by 0

g. Ordered logit

The Ordered Logit model assumes that there is a logical order among the alternatives, which means that each category has a higher or lower preference than the other categories. However, in the case of the dependent variable being a categorical variable with four categories, where 1 = "municipal PIE," 2 = "Municipal no PIE," 3 = "subvencionado PIE," and 4 = "subvencionado no PIE," it is unclear if there is a logical order among these categories.

Moreover, the Ordered Logit model assumes that there is a probability of selecting any alternative, even if it is not the preferred one. However, in the case of choosing a school, parents generally apply to the school they rank first, which means that they will always choose the alternative (applying) that provides them with the highest utility. Therefore, there is no such thing as the probability of applying to a school that is not the preferred one.

In conclusion, the Ordered Logit model may not be appropriate for analyzing the dependent variable in this case, as there is no clear logical order among the categories, and the assumption that there is a probability of selecting any alternative may not hold true for this particular situation. Other models, such as multinomial logistic regression, may be more appropriate for analyzing this type of data.

Question 2

a. Summary "Names and Reputations: An empirical analysis

The article "Names and Reputations: An Empirical Analysis" by Ryan McDevitt examines the impact of reputation on the decision of firms [1] to change their names or exit the market. Using web data and binary response models, the main hypothesis tested is that firms with poor reputations are more likely to change their names or exit the market.

More specifically, using **Multinomial and Binary probit models** the paper found support for the following five hypotheses:

Hypothesis	Prediction	Result
H_1	Firms with more complaints are more likely to change their names or exit the market	Firm with record complaints one standard deviation above was 74 pct more likely to change name or exit
H_2	Firms with longer track records tend to have established reputations, making them less likely to change their names or exit the market, with all other factors being equal	Firms that had been in operation for one standard deviation longer was 6.5 pct less likely yo change name or exit
H_3	Firms in smaller markets will be less likely to change their names, all else equal	Firms outside city were 32.3 percentil less likely to change name or exit.
H_4	Firms in smaller markets will be more responsive to changes in their reputations, all else equal.	Interaction with complains with market size were more likely to change their names. Also, longer record stronger were less likely: the reputation mechanism had more "bite" in small markets
H_5	Firms that have a higher value in continuing their business will have a lower likelihood of exiting the market, all else being equal.	Negative correlation exist between firm's adverting expenditures and change name (lower cost to add names), but effect differ by size market

Table 9: Hypothesis summary, models and results

b. Table 5, APE and PAE

Interpretation

The results, with standard errors in parenthesis [2], are given in Table 10. Column (1) shows the coefficient estimates for the binary probit model, (2) and (3) marginal effects that we will discuss later.

The predicted probability of change name or exit the market can be calculated using coefficients in column (1). For a given firm, the predicted probability of change name or exit is

$$F(complains \cdot 0.162 + firmage \cdot (-0.094) + empsize \cdot 0.004 + adspend \cdot 0.007 + chicago \cdot 0.243 + complains^2 \cdot (-0.003) + firmage^2 \cdot 0.003 + empsize^2 \cdot (-0.00002) + adspend^2 \cdot (-0.0001) + -0.566) \quad (1)$$

where F is the cumulative distribution function of the standard normal. Interpretation of coefficients in probit regression is not as straightforward as linear regression. However, there are limited ways in which we can interpret the **individual regression coefficients**.

In a binary probit model, the coefficients represent the change in the standard normal distribution of the latent variable for a one-unit increase in the corresponding independent variable, holding all other independent variables constant. A positive coefficient means that an increase in the predictor leads to an increase in the predicted probability. A negative coefficient means that an increase in the predictor leads to a decrease in the predicted probability.

[1] plumbling services in Illinois
[2] Author did not use robust standard errors

	Coefficients Change or Exit	APE _	PEA _
Complaints	0.162***	0.043***	0.043***
Years in Opeartion	-0.094***	-0.025***	-0.025***
Employees	0.004	0.001	0.001
Ad Spending	0.007	0.002	0.002
Metro Chicago	0.244***	0.065***	0.064***
Complaints^2	-0.003***	-0.001***	-0.001***
Firm Age^2	0.003***	0.001***	0.001***
Employees^2	-0.000	-0.000	-0.000
Ad Spending^2	-0.000	-0.000	-0.000
Constant	-0.567***		
Observations	2293	2293	2293
Pseudo R^2	.0559677		

Table 10: A Probit Regression Where the Dependent Variable is Equal to 1 if a Firm Changed Its Name or Exited Completely between 2008 and 2009, and 0 Otherwise

1. The coefficient of complaints is .1621935 and statistically significant. An increase in complaints, increase in the predicted probability of changing name or exit, holding constant the other factors. Also, there is small non-linear relation between complaints and the change/exit (-.0031), that is, there is a concave relationship.

2. The coefficient of the firm age is -.0937624 and is statistically significant. An increase in firm age, decrease in the predicted probability of change name or exit, holding constant the other factors. However, firm age has a non-linear relation with change/exit (0.026), that is, there is a convex relationship.

3. Firm's size effect is very small (0.004144) and not statistically significant. There is no evidence of quadratic relation.

4. Advertising expenditure coefficient is .0069765 but not statistically significant. There is no evidence of quadratic relation.

5. The coefficient of Chicago is 0.24 and statistically significant. This means that firms within Chicago city (a larger market) increase the predicted probability of change/exit, holding the other factors constant.

These results are consistent with the paper results and with the hypothesis summary provided in Table 1 (hypotheses 1 to 3).

APE vs PAE

The only way to determine whether a research study uses partial effect at the average (PEA) or an average of partial effects of each individual (APE) to estimate the marginal effects of an independent variable in a probit model is based on the information provided in the paper. Since this is not the case in this paper, the way to identify it was by calculating both ways of obtaining the marginal effects and seeing which one corresponded to the one reported. With this, we were able to identify that the reported effect corresponds to PEA. Also, by default, Stata uses the PEA method to calculate marginal effects for probit models. In order to understand the differences, we explain the definitions and when use one or the other.

As indicated Wooldridge, 2010 in a probit model, the partial effect at the average (PEA) of regressors and the average of partial effects of each individual (APE) can both be used to estimate the effect of a particular independent variable on the probability of the dependent variable taking a certain value (Wooldridge, 2010, p. 575)

The PEA of regressors ($f[E(\bar{x}\hat{\beta})]$) is the change in the predicted probability of the dependent variable taking a certain value for a one-unit increase in the independent variable, evaluated at the means of all other independent variables. In other words, it estimates the effect of the independent variable on the dependent variable at the average values of all other independent variables.

On the other hand, the APE ($E[f(x\beta)]$) of each individual is the average effect of the independent variable on the predicted probability of the dependent variable taking a certain value, calculated for each individual in the sample. In other words, it estimates the average effect of the independent variable for each individual and then averages those individual effects to get an overall estimate.

When to use PEA versus APE depends on the research question and the context of the analysis. PEA is often used when the focus is on the effect of the independent variable at the average values of all other independent variables. APE, on the other hand, is useful when there is heterogeneity in the effect of the independent variable across individuals, and we want to estimate the average effect across individuals, which can be **useful for policy analysis** (Wooldridge, 2010, p. 578)

It's important to note that PEA and APE estimates are not directly comparable because they are estimating different things but is rare to find different signs unless the coefficients are estimated imprecisely. As we can notice in **Table 2** PEA and APE are very similar.

c. Table 6, interpret results and importance of role base alternative. Econometrics necessary to exist this category

In contrast to the previous analysis, we will consider the name change and the exit from the market as two different decisions that firms can make. In that sense the latent response variable, the decision, is no longer binary, but rather a variable with three response categories: 2 = firm exited the market, 1 = firm name change, and 0 = neither changed its name nor exited (base outcome, persistence henceforth). The reason why a **base category** is used is that neither the coefficients nor the constants for each of the possible decisions can be identified (as a matter of rank), and rather what can be identified is the difference between the coefficients. In this sense, for the models that have name change and exit as a response, we can indicate that the regression coefficients refer to a difference with the base category. Hence, Train, 2009, p. 27 in chapter two of his book indicates that both the constants and the coefficients should be interpreted as the difference in the effect on utility when comparing the decision taken with respect to the base category. This normalization and inability to identify and respond to the characteristics of categorical/discrete response models.

Evidently, the model of the base category is omitted because it does not make sense to obtain coefficients that are compared with themselves. In turn, three more models were specified that seek to test hypotheses 4 and 5. Specifically, interactions are added that allow us to see if there are differentiated effects according to the covariates.

	(1)		(2)		(3)	
	Exit	Name change	Exit	Name change	Exit	Name change
Complaints	0.086*	0.288***	-0.012	0.502***	0.439	0.812***
Years in Opeartion	-0.155***	-0.065**	-0.145***	-0.083**	-0.154***	-0.085**
Employees	0.000	0.010	-0.002	0.004	-0.002	0.004
Ad Spending	-0.004	0.025*	-0.001	0.005	-0.050	0.017
Metro Chicago	0.212**	0.489***	0.307*	0.253	0.138	0.354
Complaints^2	-0.002	-0.006***	-0.002	-0.005***	-0.002	-0.005***
Firm Age^2	0.004***	0.002*	0.004***	0.002	0.004***	0.002
Employees^2	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Ad Spending^2	0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Complaints * Metro Chicago			0.093	-0.232**	-0.361	-0.549***
Firm Age * Metro Chicago			-0.010	0.024*	0.007	0.024
Complaints * Ad Spending			0.000	-0.001	-0.116	-0.016**
Firm Age * Ad Spending			-0.000	0.002***	0.004**	0.001
Complaints * Employees			-0.000	0.000	-0.000	0.000
Firm Age * Employees			0.000	0.000	0.000	0.000
Ad Spending * Metro Chicago					0.062*	-0.020
Complaints * Ad Spending * Chicago					0.117	0.015**
Firm Age * Ad Spending * Chicago					-0.005***	0.000
Constant	-0.786***	-2.068***	-0.854***	-1.790***	-0.743***	-1.862***
Observations	2293		2293		2293	
Pseudo R^2						

Table 11: A Multinomial Probit Regression Where the Choices for a Firm are not Changing Its Name and Not Exiting (Base Outcome), Exiting, and Changing Its Name between 2008 and 2009

Interpretation

First, columns one and two show the results for the multinomial logit for exit (1) and change name (2) with the same variables used in Table 11. As can be seen the results in terms of the sign on the predicted probability is quite similar to when we **collapsed** both categories into one. However, as an exception to that we can notice that spending on advertisements has a different effect in the case of success because now its relationship is negative but not statistically significant (both for its linear and quadratic effect).

Second, we can note that the models that include interactions provide support for hypotheses 4 and 5. According to Hypothesis 4, businesses operating outside of metro Chicago, which received a higher number of complaints, were **more inclined to change their names relative persistence**, while those with longer histories were less likely to do so. This finding aligns with the notion that businesses in rural areas depend more on word-of-mouth and repeat customers. Therefore, businesses in smaller markets may benefit more from maintaining their recognition among past customers. However, if a business in a smaller market develops a reputation for subpar performance, it must either conceal its past or exit the market because the entire community recognizes it as a low-value enterprise, or it might not do business with it at all.

Third, it is interesting to note how the sign of the coefficients changes once the interactions are incorporated. In particular, when we only incorporated the interactions of hypothesis four, it occurred that in some cases there was a difference in sign between the model with exit and change name (columns 3 and 4, respectively). For example, the probability of exit relative to persistence increases when complaints increase in Chicago metro (0.09); while the probability of change name relative to persistence decreases when complaints increase in Chicago metro (-0.231)

Now when incorporating the interactions with advertisement expenditure we can notice that the regressors affecting exit change and now have an effect in the same direction on the predicted probabilities. So it is possible to think that in order to decide to exit the market it is much more important to spend on advertise than to change the name, which makes economic sense. However, the coefficients for these new interactions are not statistically significant either.

d. Table 5 and 6 with logit. Compare results. Can we compare logit and probit directly? Correct your results

As indicates Wooldridge, 2010, p. 578 we can not compare directly coefficient across probit and logit models. Althought the magnitudes are different, it is rare to find probit and logit estimates having different sign unless the coefficients are imprecisly. Also, logit and probit **implicitly use different scale factors**. For both we can look at the function $f(\cdot)$ at zero and obtain a rough idea of how β_j may be different. In fact, there is some rules of thumb are adopted: multiply the probit coefficients by 1.6 to make comparable with logit; or multiply logit coefficient by 0.625 ($\approx \frac{\sqrt{3}}{\pi}$). An interesting thing is that APEs (and, to lesser extent, the PEAs) can be compared across probit and logit models.

	Coefficients Change or Exit	APE _	PEA _
Complaints	0.277***	0.042***	0.041***
Years in Operation	-0.161***	-0.024***	-0.024***
Employees	0.007	0.001	0.001
Ad Spending	0.012	0.002	0.002
Metro Chicago	0.436***	0.066***	0.065***
Complaints^2	-0.005***	-0.001***	-0.001***
Firm Age^2	0.005***	0.001***	0.001***
Employees^2	-0.000	-0.000	-0.000
Ad Spending^2	-0.000	-0.000	-0.000
Constant	-0.934***		
Observations	2293	2293	2293
Pseudo R^2	.0550831		

Table 12: A Logit Regression Where the Dependent Variable is Equal to 1 if a Firm Changed Its Name or Exited Completely between 2008 and 2009, and 0 Otherwise

	(1) Exit	Change name	(2) Exit	Change name	(3) Exit	Change name
Complaints	0.094	0.363***	-0.023	0.633***	0.572	1.045***
Years in Opeartion	-0.198***	-0.076	-0.183***	-0.103**	-0.196***	-0.104**
Employees	-0.002	0.014	-0.004	0.007	-0.002	0.006
Ad Spending	-0.007	0.038*	-0.001	0.012	-0.075	0.035
Metro Chicago	0.264*	0.734***	0.402*	0.349	0.164	0.526
Complaints^2	-0.002	-0.007***	-0.002	-0.006***	-0.002	-0.006***
Firm Age^2	0.005***	0.003	0.005***	0.002	0.005***	0.002
Employees^2	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Ad Spending^2	0.000	-0.001	-0.000	-0.001	-0.000	-0.000
Complaints * Metro Chicago			0.109	-0.292**	-0.496	-0.711***
Firm Age * Metro Chicago			-0.016	0.039*	0.008	0.040
Complaints * Ad Spending			0.001	-0.001	-0.166	-0.020**
Firm Age * Ad Spending			-0.001	0.002***	0.005**	0.002
Complaints * Employees			-0.001	0.000	-0.001	-0.000
Firm Age * Employees			0.000	0.000	0.000	0.000
Ad Spending * Metro Chicago					0.093*	-0.034
Complaints * Ad Spending * Chicago					0.167	0.020**
Firm Age * Ad Spending * Chicago					-0.008***	0.000
Constant	-0.904***	-2.856***	-1.016***	-2.405***	-0.857***	-2.544***
Observations	2293		2293		2293	
Pseudo R^2	.0692919		.0770257		.0832716	

Table 13: A Multinomial Logit Regression Where the Choices for a Firm are not Changing Its Name and Not Exiting (Base Outcome), Exiting, and Changing Its Name between 2008 and 2009

First, as we have anticipated, the signs of the coefficients are statistically the same for the probit and logit models, either binomial or multinomial. I emphasize the word statistically because there are a couple of coefficients where the sign changes but they are statistically zero.

Second, we can do the exercise of verify if the scale can make the comparisons comparable. For example, if we look at the models with the binary response and in particular for the probit coefficient of complains we have 0.16 while for logit 0.27. If we multiply the former by 1.6 we get 0.26 which is close to 0.27 (moreover, from its standard error we can report that statistically, it is between 0.23 and 0.30).

e. Table 7 and 8. Interpret results and indicate involved hypothesis

Although a longer panel of observations would improve the accuracy of identifying the effects of reputation on name choices by controlling for persistent, unobserved heterogeneity among firms, the current dataset limitations do not permit this. However, the current dataset does allow for testing several supporting estimations to verify if the observed data patterns are robust to alternative explanations. Two hypotheses derived from theory are worth noting: (i) Hypothesis 1 posits that firms will only drop names with bad records, and (ii) Hypothesis 2 suggests that only firms with bad records will add names. To test Hypothesis 1, a natural regression test would involve a dependent variable equaling one if the company dropped a name and zero otherwise. To test Hypothesis 2, a regression with a dependent variable equaling one if the company added a name and zero otherwise is appropriate. However, because name drops constitute the majority of the name changes between 2008 and 2009, it is possible to test Hypothesis 1 directly but not Hypothesis 2.

Based on the provided results in Table 14, the Probit regression model indicates that the coefficient for the predictor variable "Complaints" is statistically significant at the 0.01 level ($p<0.001$), with a marginal effect of 0.026. This means that as the number of complaints against a firm increases, the probability of the firm dropping a name in 2009 also increases.

	Coefficients Drop name	Marginal Effect _
Complaints	0.212***	0.026***
Years in Opeartion	-0.008	-0.001
Employees	0.008*	0.001*
Ad Spending	0.025**	0.003**
Metro Chicago	0.323***	0.040***
Complaints^2	-0.005***	-0.001***
Firm Age^2	0.000	0.000
Employees^2	-0.000	-0.000
Ad Spending^2	-0.000*	-0.000*
Constant	-1.881***	
Observations	2293	2293
Pseudo R^2	.0863037	

Table 14: A Probit Regression Where the Dependent Variable is Equal to 1 if a Firm Drops a Name in 2009, and 0 Otherwise

Therefore, we can reject the null hypothesis that the number of complaints against a firm has no effect on whether the firm drops a name or not. This result supports Hypothesis 1, which states that firms will only drop names with bad records.

Furthermore, the marginal effect for the predictor variable "Metro Chicago" is also statistically significant at the 0.01 level ($p<0.001$), with a marginal effect of 0.040. This suggests that firms located in the metropolitan area of Chicago are more likely to drop a name than firms located outside this area.

The coefficients for the predictor variables "Years in Operation," "Employees," "Ad Spending," "Complaints2," "Employees2," and "Ad Spending2" are not statistically significant. The constant term is statistically significant at the 0.01 level ($p<0.001$), with a value of -1.881.

In conclusion, the results of the Probit regression model support Hypothesis 1 that firms will only drop names with bad records. The number of complaints against a firm is positively associated with the likelihood of the firm dropping a name in 2009. Additionally, firms located in the metropolitan area of Chicago are more likely to drop a name.

	Coefficients Multiple Names	Marginal Effects _
Complaints	0.337***	0.050***
Years in Operation	0.004	0.001
Employees	0.014**	0.002**
Ad Spending	0.036***	0.005***
Metro Chicago	0.388***	0.057***
Complaints^2	-0.007***	-0.001***
Firm Age^2	-0.000	-0.000
Employees^2	-0.000*	-0.000*
Ad Spending^2	-0.001***	-0.000***
Constant	-1.922***	
Observations	2293	2293
Pseudo R^2	.1543915	

Table 15: A Probit Regression Where the Dependent Variable is Equal to 1 if a Firm Uses More Than One Name in 2008, and 0 Otherwise

Based on Table 15, the Probit regression model indicates that the coefficient for the predictor variable "Complaints" is statistically significant at the 0.01 level ($p<0.001$), with a marginal effect of 0.050. This means that as the number of complaints against a firm increases, the probability of the firm using more than one name in 2008 decreases.

Therefore, we can reject the null hypothesis that the number of complaints against a firm has no effect on whether the firm uses more than one name or not. This result supports Hypothesis 2, which states that only firms with bad records will add names.

Additionally, the coefficients for the predictor variables "Years in Operation," "Employees," and "Ad Spending" are not statistically significant. The coefficient for the predictor variable "Metro Chicago" is statistically significant at the 0.01 level ($p<0.001$), with a marginal effect of 0.057. This suggests that firms located in the metropolitan area of Chicago are more likely to use more than one name in 2008 than firms located outside this area.

The coefficients for the predictor variables "Complaints2," "Firm Age2," "Employees2," and "Ad Spending2" are all statistically significant at the 0.01 level ($p<0.001$), but their marginal effects are very small. This indicates that the squared terms of these predictor variables have little effect on the probability of a firm using more than one name in 2008.

The constant term is statistically significant at the 0.01 level ($p<0.001$), with a value of -1.922.

In conclusion, the results of the Probit regression model support Hypothesis 2 that only firms with bad records will add names. The number of complaints against a firm is negatively associated with the likelihood of the firm using more than one name in 2008. Additionally, firms located in the metropolitan area of Chicago are more likely to use more than one name in 2008.

Question 3

a. The empirical strategy of "Victims Incentives and Criminal Activity: Evidence from Bus Drivers Robberies in Chile"

The study wants to identify the effect that *Transantiago* reform has on driver's crime, by distinguishing between three moments of the policy: first, pre-reform where bus driver's compensation is relative to passengers that day transported and people pay tickets using currency cash; second, the transition when bus drivers compensation are from fixed salaries; and third, Post reform when all the system convert to the mechanism of fare payments from cash to electronic debit cards.

As we can expect, the outcome variable is the crime that represents the number of robberies reported on a bus in week t. Then, it is direct to notice that crimes are an event count, ie, a realization of a nonnegative integer-valued random variable that can be modeled by a count regression model (Cameron & Trivedi, 2013, Chapter 2). Even though this is the main response variable that motivates the study, crime is computed by dividing the actual number of crimes during a week by the average weekly crimes reported in the pre-reform period for each specific crime category, and also normalize (Dominguez, 2022, p. 16). As a consequence of that, the author estimates the model by a linear regression probably forgetting that the *i.i.d* assumption is too strong (and is essential in OLS), and maybe the observations may depend on the rate of occurrence. To be more precise, the paper must be estimated by different models that consider the number of events that may be characterized as the total number of such realizations over the sun unit of time. This leads to regression models of waiting times or duration, a family of event count models (Cameron & Trivedi, 2013)

Going back to the paper, the author proposes three different strategies to estimate the effect of the policy: (1) interrupted-time series of the cash incidents on buses; (2) difference in differences within buses; and (3) triple differences on buses relative to public spaces. As we focus on count models (and after in estimation presented in Table 2), I will focus in explain strategy (1) with some aspects that consider the related estimation.

The authors estimate the following model

$$crime_t = \alpha + \beta_1 Transition_t + \beta_2 Post_t + \omega_{m(t)} + robbsp_t + \varepsilon_t \qquad (2)$$

- Where $crime_t$ represents the rate of occurrence of robberies during the reference week and normalized.
- $Transition$ and $Post$ represent the temporal structure of the reform. The first variable is a dummy which indicates that it was in the transition period if it is equal to 1 and if it is equal to zero in the pre-reform. Similarly, $Post$ is equal to 1 for the time after the implementation of the regulation and 0 pre-reform. This specification is based on the evidence of a regime shift in the number of cash incidents (Figure 2 in the paper).
- $robbsp_t$ represents robberies in streets and public spaces in the same crime category (cash- or noncash-related incidents).
- $\omega_{m(t)}$ is the month-fixed effect, which is probably included to control for seasonality.

Assumption 1

In order to isolate possible spurious effects, we regress the dependent variable on crimes involving cash and other non-cash. The underlying assumption is that the pattern of non-cash incidents has not changed at any time during the reform. A better way to test this assumption is to use a difference-in-differences model (testing this is called "spurious relationships" in the literature). In addition, one can robust this analysis to the parallel trends assumption that in the absence of the policy, both forms of crime should follow similar trends. Both of the above exercises are tested by the author but are not specific to the strategy reported in Table 2.

Assumption 2

Another important assumption is that on average the proportional change between robberies involving money and those not involving money is stable over the periods of analysis, otherwise, there could be a change in the overall trend in the proportion of incidents. This assumption is not tested in Table 2 (it is in subsequent exercises). However, related to this trend assumption, we seek to ensure that there are no seasonal effects and therefore control for the effect of the month.

Assumptions 1 and 2 are closely connected: both refer to the possibility of factors that are confounding the estimate of cash thefts from drivers. The first refers to the level and the other to the change in the dependent variable. To address both, not only the comparison with non-cash related driver robberies but also with robberies in public spaces (*robbsp*) is considered. The latter is not incorporated as a distinct dependent variable, but as a control in the main model.

b. Reproduce Table 2 and interpret

	(1)	(2)	(3)	(4)	(5)	(6)
tra	0.184**	0.127*	0.103	1.555***	1.596***	1.603***
post	0.126*	0.071	0.048	-0.699***	-0.701***	-0.617***
Observations	313	313	313	313	313	313
R^2	.0179319	.2918532	.2953007	.7422251	.7848292	.7926497

Table 16: Linear model, OLS estimates: crimes to drivers. Table reproduce (Dominguez, 2022)

Table 16 summarize estimation of equation (2), but for our purpose we only report *transition* and *post* coefficients. Despite the author title the model as *"Interrupted time series analysis"*, the paper did multiple **linear regression** by removing and putting the covariates from the original equation, as different specifications of the model. The dependent variable is the weekly number of reported incidents divided by the average number of weekly incidents during the pre-reform period (crime, normalized). The paper distingish between non-cash crime (model 1,2 and 3) and cash-crime (model 4, 5 and 6). The order of the models are same as *Table 2* in the paper.

Coefficients in Table 16 are the "percentage change in crime for each period, relative to the level during the pre-period in the same category" (Dominguez, 2022, p.16). As we can notice, when we include month fixed effect and robberies in public space (*model 3*) the reform do not have any statistical effect on non-cash robberies. On the other hand, Table 16 shows that the reform effect is important. Being more precise, when we include all controls, *transition* period had 160% more incidents than pre-period reform (ie, a positive and statistically significant effect). Also, *post* period represent the purpose of the policy: holding fix the other factors, incidents were 61% lower than in the pre-reform period. Our results are exactly the same with Dominguez, 2022 paper's.

Even the authors did not discuss about the fit of the models, the results shows that cash crime models ($R^2 \approx 0.8$) have better fit than non-cash crime models ($R^2 \approx 0.3$).

c. Poisson and Negative Binomial model

In the following tables, we have two different response variables that are estimate by a Poisson regression, and a negative binomial regression. From the 1st to the 3th the outcome variable is non-cash crime; from 4th to 6th the outcome variable is cash crime. Recall that the dependent variable is a count variable, and Poisson regression models the expected count as a function of the predictor variables. We can interpret the regression as follows: for one unit change in the predictor variable, the difference in expected count, is expected to change by the respective regression coefficients, given the other predictor variables in the model are held constant. Our model is not in log form, but sometimes regressions can be interpreted in terms of incidence rate ratio (IRR), comparing the difference of two log

- For all models, the difference in the expected count of crimes is expected to be higher in transition period compared to pre-reform period, while holding the other variables constant in the model. However, in the model in the third column, when we include robberies in the public space,the transition period has not statistically difference with previous period.
- If we see the post estimate, we can notice that there are difference between the models with response variable as non-cash crime, and cash crime. For example, for cash crime, models the difference in the expected count of cash crimes is expected to be lower in force reform. Compare it to pre reform., While holding the other variables constant. No, if we see 1st to 3rd column the difference in the expected count of non-cash crimes is expected to be higher in post reform compare it to pre-reform previous, ceteris paribus. Obviously, only the cash models show significant evidence.
- In this investigation, the other variables are not important, but the author use them as control factors. For this reason, we do not report all day coefficients.

	(1)	(2)	(3)	(4)	(5)	(6)
tra	0.169**	0.118*	0.097	0.938***	0.981***	0.971***
post	0.119*	0.069	0.050	-1.200***	-1.200***	-1.104***
Observations	313	313	313	313	313	313
Pseudo R^2	.0051599	.0810616	.0819886	.6048185	.6445587	.6512777

Table 17: Poisson model, MV estimates: crimes to drivers. Based on Dominguez, 2022

Until now, we assume Poisson regression is the appropriate model. When we say that, we assume that the dependent variable is not over disperses and does not have an excessive number of zeros. But, in a context of overdispersion Negative Binomial is more appropiate than Poisson; in a context of underdispersion Zero inflated models are the alternative to Poisson when we want to model count data (Cameron & Trivedi, 2001). However, we can reject the null hypothesis of equidispersion with confidence 99% and then there is no statisticall evidence of overdispersion or underdispersion.

Nevertheless, we estimate a Negative Binomial regression (Table 17). As we can notice, there is no difference between Poisson and Negative Binomial results. This is because there is no statistical evidence of overdispersion, and as can be seen in the estimation the alpha parameter (which adjusts the variance) is close to zero. The interpretation of the coefficients of the Negative Binomial regressions is the same as that of the Poisson regressions: when the independent variable is continuous it is interpreted as *"for a one unit change in the predictor variable, the difference in the expected counts of the response variable is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant."*; when the variable is discrete it is interpreted as *"the difference of the expected counts expected to grow or decrease in beta units"*.

	(1)	(2)	(3)	(4)	(5)	(6)
tra	0.169**	0.122**	0.101	0.938***	0.954***	0.953***
post	0.119*	0.072	0.053	-1.200***	-1.210***	-1.111***
Observations	313	313	313	313	313	313
Pseudo R^2	.0031777	.0579194	.0586571	.232068	.2637425	.2733904

Table 18: Negative binomial model, MV estimates: crimes to drivers. Based on Dominguez, 2022

Both questions c and d ask to compare the results with the model reported in the paper (to theoretical comparison see Table 19 below). In order to maintain order and to avoid repeating the answer, it will be left the answer in item d.

Criteria	LLSR	Poisson
Response	Normal	Counts
Variance	Equal for each level of X	Equal to the mean for each level of X (by definition, heterocedastic)
Model fitting	OLS	Maximum Likelihood
Compare models	R^2, F-tests; AIC/BIC	pseudo R^2,Drop in deviance test, AIC/BIC
Interpret coefficients	β_1 = change in y for unit change in X	e^{β_1} = percent change in y for unit change in X
Visual evaluation	plot X vs Y and add linear line	Find $log(\bar{y})$ for subgroups and plot vs X

Table 19: Compare Linear Least Squares vs Poisson regression criteria

d. Differences in results and count model fit

Differences

In *Table 20* we report the results of the specification of model 6 with linear, Poisson and Negative Binomial estimation. We choose this specification because it is the one of interest: it includes all the controls indicated in the identification strategy and the dependent variable is cash robberies. In any case, we checked that the following analysis can be generalized to all models.

	Poisson (6)	Negbin (6)	LLSR (6)
tra	0.971***	0.953***	1.603***
post	-1.104***	-1.111***	-0.617***
Observations	313	313	313
R^2			.7926497
Pseudo R^2	.6512777	.2733904	

Table 20: Comparision between Poisson, Negative binomial and Linear model for crimes to drivers model with all covariates. Based on Dominguez, 2022

As indicated in the answer above, it can be seen that there are no differences in the relevant coefficients for the Poisson and Binomial models. However, there are differences in the effect size between these count models versus the linear regression model. On the one hand, we can see that for the linear model the average difference in the number of robberies with cash is 0.617 lower after the reform compared to before the reform. On the other hand, for the count models this difference in expected counts is higher (1.104).

These results are consistent with the specialized literature (Cameron & Trivedi, 2005, 2013). Although the effect sizes of the coefficients between the different regressions are different (the meaning is different), **the direction and statistical significance should not change**. If so, we should think that the estimation algorithm is probably not converging for the count models (in subdispersion contexts this is possible).

Count models as a good approximation

An obvious question is whether special methods are required to handle count data, or whether the standard Gaussian linear regression may suffice. Common regression estimators and models, such as the ordinary least squares in the linear regression models (LLSR), ignore the **restricted support** for the dependent variable. This produces significant deficiencies unless the mean of the count data is high in which case normal approximation and related regressions methods may be satisfactory.

Also, regressions such as Poisson models seek to preserve and exploit as much as possible the non-negative integer-valued aspect of the outcome, for example, allowing non-linear relations, and respecting the discreetness of the count variable.

If we think in our microeconomic problem (crime to drivers), the count data models reflect modeling discrete aspects of individual economic behavior (crime) such as other types of discrete and limited dependent variables models but with important properties. For example, a key property of the Poisson and negative binomial distribution is the additivity (countable additivity theorem Kingman, 1993) that makes it easy to define their multinomial model and different distributional characterization: the law of rare events, counting process, waiting time, and repetitions events (Cameron & Trivedi, 2001). In this case, this property could be very useful because two reasons:

1. We can asses the distribution of the count process in an interval of time (crimes by week), with a coherent interpretation of the probability of the occurrence of events (crimes)

2. Separate the counting process into two possible events (cash or no cash crime) and then estimate a multinomial model for count data. This is a different specification that the author did in his paper and also a different required exercise that the problem set suggests in part 2c. But if we remember the potential of multinomial plus the count data assumptions we do not need to assume independence of alternatives, an important issue if we consider that cash and no cash crime are restricted by the possibility of doing the other. Crime will always occur, but if the policy constrains the crime of money theft it will probably be transferred to another type of (non-cash) crime. That shows the evidence in this article when the trend of non-cash crime is not affected by the policy.

Also, there exist some problems with some assumptions in this models. For example, Poisson's assumption of equidispersion ($E(u|x) = Var(u|x)$) has a similar qualitative consequence to the failure of the assumption of homoskedasticity in the linear regression model , but in magnitude can be much larger Wooldridge, 2010. In presence of **overdispersion** we can use negative binomial models that corrects this issue (Greene, 2003; Wooldridge, 2010)

However, when we test for a key dispersion, we can not reject the null hypothesis of a cute dispersion on a 1% significance level. In addition, another evidence for **equidispersion**, is that the results obtain it by a negative binomial model is the exactly the same as that Poisson model and also the α parameter is near to zero.

Sometimes the problem is, the **underdispersion** (conditional mean exceeds the conditional variance). In this context Poisson inference of estimators are wrong and no negative binomial can be estimate (the algorithm don't converge because theoretically $\alpha < 0$. But in the models $\alpha > 0$ but near to zero.

References

Cameron, A. C., & Trivedi, P. K. (2001). Essentials of count data regression [Publisher: Blackwell Oxford]. *A companion to theoretical econometrics*, *331*.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge university press.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

McDevitt, R. C. (2011). Names and reputations: An empirical analysis [Publisher: American Economic Association]. *American Economic Journal: Microeconomics*, *3*(3), 193–209.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.

Dominguez, P. (2022). Victim incentives and criminal activity: Evidence from bus driver robberies in chile [Publisher: MIT Press]. *The Review of Economics and Statistics*, *104*(5), 946–961.

Appendix

Mathematical appendix

In statistics, the cumulative distribution function (cdf) give to us $P(X < a)$ and in order to obtain the probability we need to integrate the density $f(z)$ from $-\infty$ to a $\int_{-\infty}^{a} f(z)dz$. When we assume **standard normal cdf** the model becomes a **Probit model**

$$P(y = 1|X = x) = \int_{-\infty}^{\beta_0+\beta_1 x} \frac{1}{2\pi} e^{-\frac{z^2}{2}} dz$$

When u distributes **standard logistic** then the model is a **Logistic model**

$$P(y|X = x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}$$

Both equations are also called *response functions*

Coefficients

Logistic regression is a statistical method for predicting a binary outcome based on one or more independent variables. Logit coefficients are values in the logistic regression equation that predict the dependent variable from the independent variable in log-odds units. The prediction equation is given by:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

where p is the probability of $y = 1$. These coefficients provide information on the relationship between independent variables and the dependent variable on the logit scale, indicating the amount of increase in the **predicted log odds** of $y = 1$ caused by an increase in the predictor, while holding all other predictors constant.

However, interpreting these coefficients can be challenging, as they **are in log-odds units and do not have any useful interpretation in the estimation scale other than the sign**. Moreover, β_1 represents the effect of age on the log-odds of the outcome, **not on the probability**, which is often what we are interested in.

Probit regression is an alternative method to logistic regression, but it still has similar problems. The estimated parameters in the estimation scale are not easily interpretable, and the effect of each independent variable on the outcome is represented as shifts in the standard cumulative normal, which is also of little help.

Marginal effects are a way of presenting results as differences in probabilities, which is more informative than odds ratios and relative risks. In the probability scale, all effects are non-linear, and marginal effects for continuous variables apply to a small change in x when effects are non-linear. Marginal effects use model prediction for interpretation, which helps to better interpret the model in the scale that makes more sense.

To convert the estimation scale to the probability scale, we can solve for p

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1$$

$$p = \frac{e^{\beta_0+\beta_1 x_1}}{1 + e^{\beta_0+\beta_1 x_1}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1)}}$$

However, the effect of X_1 depends on the value of X_1 and the values of all other covariates. Thus, we need to approximate the analytical derivative numerically using the definition of derivative and compute the average effect of X_1 on p. Our goal is to take numerical derivatives of functions for which analytical derivatives are more complicated.

$$\frac{\partial p}{\partial x_1} = \frac{\beta_1 e^{\beta_0+\beta_1 x_1}}{(1 + e^{-(\beta_0+\beta_1 x_1)})^2}$$