Pontificia Universidad Católica de Chile
Facultad de Economía & Administración
Instituto de Economía

Problem Set # 1

Jose Carlo Bermúdez

jcbermudez@uc.cl

Econometric Theory II

Professor: Tomás Rau

April 12th, 2024

Summary

This report includes my solution to the problem set # 1 of the Econometric Theory II graduate course. The first section of the report dives into the theoretical fundamentals behind discrete choice non-linear models. The second part of the document includes an empirical analysis of educational trajectories in Chile for the cohort of undergraduate students of 2018. This section aims to strengthen not only our knowledge of empirical applications of non-linear econometrics tools but to improve our skills as researchers as well. The replication package for the empirical analysis is available at my GitHub account.

Contents

1 Theoretical Section

1.1 Question 1 (Part A)

Suppose that an economic model suggests that latent dependent variable y_i^* satisfies the following classical linear model:

$$y_i^* = x_i'\beta_0 + \epsilon_i$$

but y_i^* it is not observable across its entire range. On the contrary, we can observe a random sample of size N from y_i and x_i, where:

$$y_i \equiv \tau_i(y_i^*) = \begin{cases} 0 & if \quad y_i^* \leq 0 \\ y_i^* & if \quad 0 < y_i^* \leq L_i \\ L_i & if \quad L_i < y_i^* \leq U_i \\ y_i^* & if \quad U_i < y_i^* \end{cases}$$

This means that latent variable y_i^* is observed unless it is lower than zero or lies in the interval (L_i, U_i). Threshold variables L_i, U_i with $U_i > L_i > 0$ are assumed to be observable $\forall i$, and both are linear functions from regressors of x_i. Assume ϵ_i to be normally distributed with media zero and unknown variance σ_0^2.

Log-Likelihood for Unknown Parameters

In this subsection, I derive the average maximum likelihood function (MLF) for the set of unknown parameters. Before that, I shall state that we are facing a special case of Tobit models, also known as "corner solution models" (Wooldridge, 2010), because it combines elements from a multinomial unordered probit (we have more than two categories but also we are assuming a normal distribution for residuals) and a censored model (since the observable is equal to the latent variable in two of the four brackets where y_i is defined). This fact is important not only when it comes to the estimation of the MLF –because it should express both types of conditional probabilities, for uncensored $\mathbb{P}(y_i = 0^\vee y_i = L_i|x_i)$ and censored densities $\mathbb{P}(y_i = y_i^*|x_i)$ as well–, but because the interpretation for empirical analysis might be different regarding another discrete choice specifications. Next, I describe every step until obtaining the MLF.

1. Let β_0, σ_0 be the unknown parameters.

2. Let y_i^* be a latent variable with homokedastic conditional distribution $y_i^*|x_i \sim \mathcal{N}(x_i'\beta_0, \sigma_0^2)$.

   Proof: By statement, we know that $\epsilon_i \sim \mathcal{N}(0, \sigma_0^2)$. Then:

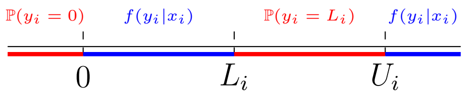
$$\begin{aligned} \mathbb{E}(y_i^*|x_i) &= \mathbb{E}(x_i'\beta_0 + \epsilon_i|x_i) \\ &= \mathbb{E}(x_i'\beta_0|x_i) + \cancelto{0}{\mathbb{E}(\epsilon_i|x_i)} \\ \therefore \mathbb{E}(y_i^*|x_i) &= x_i'\beta_0 \end{aligned}$$

   and

$$\begin{aligned} \mathbb{V}(y_i^*|x_i) &= \mathbb{E}[(y_i^*)^2|x_i] - \mathbb{E}^2[y_i^*|x_i] \\ &= \mathbb{E}[(x_i'\beta_0 + \epsilon_i)^2|x_i] - (x_i'\beta_0)^2 \\ &= \cancel{(x_i'\beta_0)^2} + 2x_i'\beta_0\cancelto{0}{\mathbb{E}(\epsilon_i|x_i)} + \mathbb{E}(\epsilon_i^2|x_i) - \cancel{(x_i'\beta_0)^2} \\ &= \mathbb{E}(\epsilon_i^2|x_i) \\ &= \sigma_0^2 + \cancelto{0}{\mathbb{E}^2(\epsilon_i|x_i)} \\ \therefore \mathbb{V}(y_i^*|x_i) &= \sigma_0^2 \end{aligned}$$

■

3. Notice that we have two cases: when y_i is continuous and discrete, respectively. For the continuous case, probabilities are expressed in terms of the density function (pdf). In contrast, when y_i turns discrete, then the probabilities are expressed in terms of the cumulative function (cdf). The following figure illustrates how the MLF is characterized across each interval.



Following the figure above, I define probabilities for those brackets where the latent variable is unobservable, given $\epsilon_i \sim \mathcal{N}(0, \sigma_0^2)$.

$$\begin{aligned}
\mathbb{P}(y_i = 0|x_i) \quad & \equiv \mathbb{P}(y_i^* \leq 0|x_i) \\
& = \mathbb{P}(-\infty < y_i^* \leq 0|x_i) \\
& = \mathbb{P}(y_i^* \leq 0|x_i) - \mathbb{P}(y_i^* < -\infty|x_i) \\
& = \mathbb{P}\left(x_i'\beta_0 + \epsilon_i \leq 0|x_i\right) - \mathbb{P}\left(x_i'\beta_0 + \epsilon_i < -\infty|x_i\right) \\
& = \mathbb{P}\left(\tfrac{\epsilon_i}{\sigma_0} \leq \tfrac{0 - x_i'\beta_0}{\sigma_0}|x_i\right) - \mathbb{P}\left(\tfrac{\epsilon_i}{\sigma_0} < \tfrac{-\infty - x_i'\beta_0}{\sigma_0}|x_i\right) \\
& = \Phi\left(\tfrac{-x_i'\beta_0}{\sigma_0}\right) - \cancelto{0}{\Phi\left(\tfrac{-\infty - x_i'\beta_0}{\sigma_0}\right)} \\
\therefore \mathbb{P}(y_i = 0|x_i) \quad & = 1 - \Phi\left(\tfrac{x_i'\beta_0}{\sigma_0}\right)
\end{aligned} \tag{1}$$

$$\begin{aligned}
\mathbb{P}(y_i = L_i|x_i) \quad & \equiv \mathbb{P}(L_i < y_i^* \leq U_i|x_i) \\
& = \mathbb{P}(y_i^* \leq U_i|x_i) - \mathbb{P}(y_i^* < L_i|x_i) \\
& = \mathbb{P}(x_i'\beta_0 + \epsilon_i \leq U_i|x_i) - \mathbb{P}(x_i'\beta_0 + \epsilon_i < L_i|x_i) \\
& = \mathbb{P}\left(\tfrac{\epsilon_i}{\sigma_0} \leq \tfrac{U_i - x_i'\beta_0}{\sigma_0}|x_i\right) - \mathbb{P}\left(\tfrac{\epsilon_i}{\sigma_0} < \tfrac{L_i - x_i'\beta_0}{\sigma_0}|x_i\right) \\
\therefore \mathbb{P}(y_i = L_i|x_i) \quad & = \Phi\left(\tfrac{U_i - x_i'\beta_0}{\sigma_0}\right) - \Phi\left(\tfrac{L_i - x_i'\beta_0}{\sigma_0}\right)
\end{aligned} \tag{2}$$

For the continuous section, when the latent variable is observable we have the density function from a normal distribution, given $y_i^*|x_i \sim \mathcal{N}(x_i'\beta_0, \sigma_0^2)$. Thus:

$$\begin{aligned}
f(y_i|x_i) \quad & = \tfrac{1}{\sigma_0\sqrt{2\pi}} exp\left\{\tfrac{-(y_i - x_i'\beta_0)^2}{2\sigma_0^2}\right\} \\
& = \tfrac{1}{\sigma_0} \cdot \underbrace{\frac{1}{\sqrt{2\pi}} exp\left\{-\frac{1}{2} \cdot \left(\frac{y_i - x_i'\beta_0}{\sigma_0}\right)^2\right\}}_{\phi} \\
\therefore f(y_i|x_i) \quad & \equiv \tfrac{1}{\sigma_0}\phi\left(\tfrac{y_i - x_i'\beta_0}{\sigma_0}\right)
\end{aligned} \tag{3}$$

4. Let $\mathbb{1}(.)$ be the indicator function. Since probabilities are defined in terms of random samples of y_i and x_i then the MLF can be expressed as a product of probabilities. Following a general multinomial Bernoulli for discrete choice models, the MLF is given by:

$$\mathcal{L}(\beta_0, \sigma_0) = \prod_{i=1}^{N}\left\{\mathbb{P}(y_i = 0|x_i)^{\mathbb{1}\{y_i^* \in [-\infty, 0]\}} f(y_i|x_i)^{\mathbb{1}\{y_i^* \in (0, L_i]\}} \mathbb{P}(y_i = 0|x_i)^{\mathbb{1}\{y_i^* \in (L_i, U_i]\}} f(y_i|x_i)^{\mathbb{1}\{y_i^* \in (U_i, +\infty]\}}\right\} \tag{4}$$

Now, replacing expressions 1, 2, and 3 in the expression above, we have:

$$\mathcal{L}(\beta_0, \sigma_0) = \prod_{i=1}^{N} \left[1 - \Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\right]^{\mathbb{1}[y_i^* \leq 0]} \cdot \prod_{i=1}^{N} \left[\frac{1}{\sigma_0}\phi\left(\frac{y_i - x_i'\beta_0}{\sigma_0}\right)\right]^{\mathbb{1}[0<y_i^* \leq L_i]}$$
$$\cdot \prod_{i=1}^{N} \left[\Phi\left(\frac{U_i - x_i'\beta_0}{\sigma_0}\right) - \Phi\left(\frac{L_i - x_i'\beta_0}{\sigma_0}\right)\right]^{\mathbb{1}[L_i<y_i^* \leq U_i]} \cdot \prod_{i=1}^{N} \left[\frac{1}{\sigma_0}\phi\left(\frac{y_i - x_i'\beta_0}{\sigma_0}\right)\right]^{\mathbb{1}[U_i<y_i^*]} \tag{5}$$

5. Let $\ell(\beta_0, \sigma_0) := ln\mathcal{L}(\beta_0, \sigma_0)$ be the log-likelihood of the MLF expressed in Equation 5. Then, based on the dominance assumption[1] and a lemma for the expected log-likelihood inequality[2], Fisher (1922, 1925) proposes maximizing the sample analog of $\mathbb{E}[\ell(\beta_0, \sigma_0)]$. This gives the average MLF requested by the problem:

$$\ell(\beta_0, \sigma_0) = \frac{1}{N}\sum_{i=1}^{N}\Bigg\{\mathbb{1}[y_i^* \leq 0] \cdot ln\left[1 - \Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\right] + \mathbb{1}[0 < y_i^* \leq L_i] \cdot ln\left[\frac{1}{\sigma_0}\phi\left(\frac{y_i - x_i'\beta_0}{\sigma_0}\right)\right] +$$
$$\mathbb{1}[L_i < y_i^* \leq U_i] \cdot ln\left[\Phi\left(\frac{U_i - x_i'\beta_0}{\sigma_0}\right) - \Phi\left(\frac{L_i - x_i'\beta_0}{\sigma_0}\right)\right] + \mathbb{1}[U_i < y_i^*] \cdot ln\left[\frac{1}{\sigma_0}\phi\left(\frac{y_i - x_i'\beta_0}{\sigma_0}\right)\right]\Bigg\} \tag{6}$$

6. The first order condition for β_0 is derived from Equation 6:

$$[\beta_0] : \sum_{i=1}^{N}\Bigg\{\frac{\mathbb{1}[y_i^* \leq 0]}{\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) - 1} \cdot \frac{x_i}{\sigma_0}\phi\left(\frac{x_i'\hat{\beta}_0}{\sigma_0}\right) - \frac{\mathbb{1}[0 < y_i^* \leq L_i]}{\phi\left(\frac{y_i - x_i'\hat{\beta}_0}{\sigma_0}\right)} \cdot x_i\phi'\left(\frac{y_i - x_i'\hat{\beta}_0}{\sigma_0}\right) +$$
$$\frac{\mathbb{1}[L_i < y_i^* \leq U_i]}{\Phi\left(\frac{U_i - x_i'\hat{\beta}_0}{\sigma_0}\right) - \Phi\left(\frac{L_i - x_i'\hat{\beta}_0}{\sigma_0}\right)} \cdot \frac{x_i}{\sigma_0}\left[\phi\left(\frac{L_i - x_i'\hat{\beta}_0}{\sigma_0}\right) - \phi\left(\frac{U_i - x_i'\hat{\beta}_0}{\sigma_0}\right)\right] -$$
$$\frac{\mathbb{1}[U_i < y_i^*]}{\phi\left(\frac{y_i - x_i'\hat{\beta}_0}{\sigma_0}\right)} \cdot x_i\phi'\left(\frac{y_i - x_i'\hat{\beta}_0}{\sigma_0}\right)\Bigg\} = 0$$

Censored Estimation: Discussion

We know that opposite to standard linear regression models –say a linear probability model estimated using OLS– where partial effects are constant $\left(\frac{\partial \mathbb{E}(y_i|x_i)}{\partial x_j} = \hat{\beta}_j\right)$ across any realization of the covariates because y_i is observable in all its domain, for censored regression models or corner solution models, we have $\frac{\partial \mathbb{E}(y_i|x_i)}{\partial x_j} = \Phi(\frac{x_i\beta}{\sigma})\hat{\beta}_j$. That is, $\Phi(\cdot)$ is the probability of observing a positive response given x_i. This censorship in the model has several implications in terms of the estimation of β_0 and interpretation of marginal effects.

On one hand, regarding the estimation, we must be aware that in Tobit models β_0 is not identified, because it is part of a non-linear function in the inverse Mill ratio when we analyze $\mathbb{E}(y_i|x, y > 0)$. Moreover, even when estimation is made using maximum likelihood and including all the available data –opposite, for instance, to a Heckprobit specification–, the fact that our parameter of interest is normalized

[1]This assumption implies that $\mathbb{E}[sup_{\theta\in\Theta}|\ell(\theta)]$ exists, where θ and Θ are the subset of sample estimator and the set of population estimators, respectively. Then, this assumption guarantees that $\mathbb{E}[\ell(\theta)]$ exists $\forall\theta \in \Theta$ (Ruud, 2000).

[2]The lemma for the expected log-likelihood inequality states that if $\ell(\theta)$ is the conditional log-likelihood and we meet the dominance assumption, then $\mathbb{E}[\ell(\theta)] \leq \mathbb{E}[\ell(\theta_0)]$, with θ_0 the population parameter, so the log-likelihood reaches a maximum if $\theta = \theta_0$ (Ruud, 2000).

with respect to the variance of the distribution, implies that there are different combinations for unknown parameters β_0, σ_0 that meet the required functional form of the distribution.

When it comes to interpretation, we are often interested in the average partial effects (APE), where the expected value of the observable y_i is averaged over the population distribution of ϵ_i, and then derivatives or differences concerning elements of x_i are obtained. However, (Wooldridge, 2010, p. 524) argues that when $\Phi(\cdot) \rightarrow 1$ then it is unlikely we observe $y_i = 0|x_i$ and the adjustment factor becomes irrelevant. If the estimated probability of a positive response is close to one at the sample means of the covariates, the adjustment factor can be ignored. In most interesting Tobit applications, $\Phi(\cdot)$ is notably less than unity. Then, an important limitation of the standard Tobit model is that a single mechanism determines the choice between $y = 0$ versus $y > 0$. This is why we often get the same sign for $\frac{\partial \mathbb{P}(y_i>0|x_i)}{\partial x_j}$ and $\mathbb{E}(y_i|x, y > 0)$.

1.2 Question 2 (Part B)

Now, let's suppose that y_i^* is never observed, but is only observed in the range in which it falls. More specifically, the dependent variable y_i is now defined by:

$$y_i \equiv \tau_i(y_i^*) = \begin{cases} 0 & if \quad y_i^* \leq 0 \\ 1 & if \quad 0 < y_i^* \leq L_i \\ 2 & if \quad L_i < y_i^* \leq U_i \\ 3 & if \quad U_i < y_i^* \end{cases}$$

Again, assume that ϵ_i is normally distributed with mean zero and unknown variance σ_0^2.

Log-Likelihood of Unknown Parameters

In this case, we are facing an ordered multinomial probit because we have more than two categories ordinarily defined, and also the error term ϵ_i has a normal distribution. As in the previous problem, I describe the step-by-step until the final derivation of the MML.

1. Let β_0, σ_0^2 be the unknown parameters.
2. Let y_i^* to be a latent variable with homokedastic conditional distribution $y_i^*|x_i \sim \mathcal{N}(x_i'\beta_0, \sigma_0^2)$. This has already been proved in question 1.
3. As $\epsilon_i \sim (0, \sigma_0^2)$, probabilities are expressed in terms of the normal distribution's cdf as follows.

$$\begin{aligned} \mathbb{P}(y_i = 0|x_i) \quad & \equiv \mathbb{P}(y_i^* \leq 0|x_i) \\ & = \mathbb{P}(-\infty < y_i^* \leq 0|x_i) \\ & = \mathbb{P}(y_i^* \leq 0|x_i) - \mathbb{P}(y_i^* < -\infty|x_i) \\ & = \mathbb{P}(x_i'\beta + \epsilon_i \leq 0|x_i) - \mathbb{P}(x_i'\beta + \epsilon_i < -\infty|x_i) \\ & = \mathbb{P}(\epsilon_i \leq 0 - x_i'\beta|x_i) - \mathbb{P}(\epsilon_i < -\infty - x_i'\beta_0|x_i) \\ & = \Phi(-x_i'\beta_0) - \Phi(\cancelto{0}{-\infty - x_i'\beta_0}) \\ \therefore \mathbb{P}(y_i = 0|x_i) \quad & = 1 - \Phi(x_i'\beta_0) \end{aligned} \tag{7}$$

$$
\begin{aligned}
\mathbb{P}(y_i = 1|x_i) \quad &\equiv \mathbb{P}(0 < y_i^* \leq L_i|x_i) \\
&= \mathbb{P}(y_i^* \leq L_i|x_i) - \mathbb{P}(y_i^* < 0|x_i) \\
&= \mathbb{P}(x_i'\beta_0 + \epsilon_i \leq L_i|x_i) - \mathbb{P}(x_i'\beta_0 + \epsilon_i < 0|x_i) \\
&= \mathbb{P}(\epsilon_i \leq L_i - x_i'\beta_0|x_i) - \mathbb{P}(\epsilon_i < 0 - x_i'\beta_0|x_i) \\
&= \Phi(L_i - x_i'\beta_0) - \Phi(-x_i'\beta_0) \\
\therefore \mathbb{P}(y_i = 1|x_i) \quad &= \Phi(L_i - x_i'\beta_0) - 1 + \Phi(x_i'\beta_0)
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
\mathbb{P}(y_i = 2|x_i) \quad &\equiv \mathbb{P}(L_i < y_i^* \leq U_i|x_i) \\
&= \mathbb{P}(y_i^* \leq U_i|x_i) - \mathbb{P}(y_i^* < L_i|x_i) \\
&= \mathbb{P}(x_i'\beta_0 + \epsilon_i \leq U_i|x_i) - \mathbb{P}(x_i'\beta_0 + \epsilon_i < L_i|x_i) \\
&= \mathbb{P}(\epsilon_i \leq U_i - x_i'\beta_0|x_i) - \mathbb{P}(\epsilon_i < L_i - x_i'\beta_0|x_i) \\
\therefore \mathbb{P}(y_i = 2|x_i) \quad &= \Phi(U_i - x_i'\beta_0) - \Phi(L_i - x_i'\beta_0)
\end{aligned} \tag{9}
$$

$$
\begin{aligned}
\mathbb{P}(y_i = 3|x_i) \quad &\equiv \mathbb{P}(U_i < y_i^*|x_i) \\
&= \mathbb{P}(U_i < y_i^* \leq +\infty|x_i) \\
&= \mathbb{P}(y_i^* \leq +\infty|x_i) - \mathbb{P}(y_i^* < U_i|x_i) \\
&= \mathbb{P}(x_i'\beta_0 + \epsilon_i \leq +\infty|x_i) - \mathbb{P}(x_i'\beta_0 + \epsilon_i < U_i|x_i) \\
&= \mathbb{P}(\epsilon_i < +\infty - x_i'\beta_0|x_i) - \mathbb{P}(\epsilon_i < U_i - x_i'\beta_0|x_i) \\
&= \Phi(+\infty - x_i'\beta_0) - \Phi(U_i - x_i'\beta_0) \\
\therefore \mathbb{P}(y_i = 3|x_i) \quad &= 1 - \Phi(U_i - x_i'\beta_0)
\end{aligned} \tag{10}
$$

4. Let $\mathbb{1}(.)$ be the indicator function. Since probabilities are defined in terms of random samples of y_i and x_i then the MLF can be expressed as a product of probabilities. Following a general multinomial Bernoulli for discrete choice models, the MLF is given by:

$$
\mathcal{L}(\beta_0, \sigma_0) = \prod_{i=1}^{N} \prod_{j=0}^{3} \mathbb{P}(y_i = j|x_i)^{\mathbb{1}\{y_i = j\}} \tag{11}
$$

Now, replacing equations from (7) to (10), we have:

$$
\begin{aligned}
\mathcal{L}(\beta_0, \sigma_0) = \prod_{i=1}^{N} [1 - \Phi(x_i'\beta_0)]^{\mathbb{1}\{y_i=0\}} \cdot \prod_{i=1}^{N} [\Phi(L_i - x_i'\beta_0) - 1 + \Phi(x_i'\beta_0)]^{\mathbb{1}\{y_i=1\}} \cdot \\
\prod_{i=1}^{N} [\Phi(U_i - x_i'\beta_0) - \Phi(L_i - x_i'\beta_0)]^{\mathbb{1}\{y_i=2\}} \cdot \prod_{i=1}^{N} [1 - \Phi(U_i - x_i'\beta_0)]^{\mathbb{1}\{y_i=3\}}
\end{aligned} \tag{12}
$$

5. Let $\ell(\beta_0, \sigma_0) := ln\mathcal{L}(\beta_0, \sigma_0)$ be the log-likelihood of the MLF expressed in Equation 12. Then, the average MLF requested by the problem takes the following functional form:

$$
\begin{aligned}
\ell(\beta_0, \sigma_0) = \frac{1}{N} \sum_{i=1}^{N} \Big\{ \mathbb{1}[y_i = 0] \cdot ln\,[1 - \Phi(x_i'\beta_0)] + \mathbb{1}[y_i = 1] \cdot ln\,[\Phi(L_i - x_i'\beta_0) - 1 + \Phi(x_i'\beta_0)] + \\
\mathbb{1}[y_i = 2] \cdot ln\,[\Phi(U_i - x_i'\beta_0) - \Phi(L_i - x_i'\beta_0)] + \mathbb{1}[y_i = 3] \cdot ln\,[1 - \Phi(U_i - x_i'\beta_0)] \Big\}
\end{aligned} \tag{13}
$$

6. Finally, the first-order condition for β_0 can be retrieved from Equation 13 as follows:

$$[\beta_0] : \sum_{i=1}^{N} \Bigg\{ \mathbb{1}[y_i = 0] \cdot \frac{x_i \phi(x_i'\hat{\beta}_0)}{\Phi(x_i'\hat{\beta}_0) - 1} + \mathbb{1}[y_i = 1] \cdot \frac{x_i[\phi(x_i'\hat{\beta}_0) - \phi(L_i - x_i'\hat{\beta}_0)]}{\Phi(L_i - x_i'\hat{\beta}_0) - 1 + \Phi(x_i'\hat{\beta}_0)} +$$
$$\mathbb{1}[y_i = 2] \cdot \frac{x_i[\phi(L_i - x_i'\hat{\beta}_0) - \phi(U_i - x_i'\hat{\beta}_0)]}{\Phi(U_i - x_i'\hat{\beta}_0) - \Phi(L_i - x_i'\hat{\beta}_0)} + \mathbb{1}[y_i = 3] \cdot \frac{x_i \phi(U_i - x_i'\hat{\beta}_0)}{1 - \Phi(U_i - x_i'\hat{\beta}_0)} \Bigg\} = 0$$

Normalization/Identification of β_0 and σ^2

In the case of the multinomial ordered probit, we need one parameter to be normalized for identification purposes. Normalization is necessary when there are additional restrictions or when the independent variables perfectly predict some of the categories of the dependent variable. Following (Colin Cameron & Trivedi, 2005, p. 517), let us consider a generalization of our multinomial probit model $y_j^* = x\beta_{0,j} + \varepsilon_j$ with $j = 1, ..., m$, and errors are multivariate normally distributed $\varepsilon \sim \mathcal{N}(0, \mathbf{\Sigma})$, then different multinomial models arise from different specifications of the covariance matrix $\mathbf{\Sigma}$. Some of the off-diagonal entries are specified to be nonzero, to permit correlation across the errors, though some restrictrictions need to be placed on $\mathbf{\Sigma}$. Note that if the errors are uncorrelated the multinomial probit still yields no closed-form solution for the probabilities and it is easier to assume instead that the errors are extreme value and use multinomial logit models. Hence, restrictions on $\mathbf{\Sigma}$ are needed to ensure identification. One way to achieve this identification is to normalize $\varepsilon_1 = 0$.

1.3 Question 3 (Part C)

Let's consider the model of question 2 (part B), where the dependent variable y_i only indicates the range in which the latent dependent variable y_i^* falls in, classifying it in one out of four categories: 0, 1, 2, or 3. Opposite to the case above, under this scenario thresholds defining the ranges (L_i, U_i) are now **unknown**.

Uncertainty from Unknown Parameters

If U_i and L_i are unknown, then we are adding more uncertainty to our model estimation. However, this should not necessarily be interpreted as a limitation, because we can always recover threshold parameters by maximizing the log-likelihood. In terms of the identification of the parameters of the model, having unknown threshold outcomes implies a scale ambiguity issue due to the arbitrary scaling of the threshold parameters. This means that multiplying all threshold parameters by a constant factor does not change the model's predictions. Moreover, the threshold parameters introduce constraints on the cumulative probabilities associated with each category. These constraints must be satisfied to ensure that the model's predictions align with the observed data. Failure to satisfy these constraints can lead to rank identification problems, where the model cannot uniquely determine the ordering of the categories.

Estimation Strategies

To the best of my knowledge, the most feasible estimation is by maximizing the log-likelihood, as in Equation 13, but now represented by $\ell(\beta_0, \sigma_0, U_i, L_i)$. The maximum likelihood estimation method is desirable because provides consistent and asymptotically efficient estimates under certain regularity conditions, nonetheless, additional methods for estimations include simulation-based methodologies such as Markov Chain Monte Carlo (MCMC) or Generalized Method of Moments (GMM).

2 Empirical Section

2.1 Question 1

This question aims to study the factors that are relevant to predicting student permanence at the undergraduate level in Chile for 2018-2019. For that purpose, I rely on administrative records from the Ministry of Education (MINEDUC)[3]. Specifically, I built a dataset in cross-section format including enrollment, applications, and awarding of scholarships. In this sense, a relevant data source was admission tests, but these records are unavailable for the period under analysis.

Data cleaning. Before any estimations, I made minor adjustments to the data. The most relevant is that I restrict my sample only to students who enrolled in at least one program/degree at the undergraduate level in 2018. Also, I dropped from my sample students with missing identification (RUN, which stands for acronyms in Spanish of *Rol Único Nacional*), which barely accounted for $\approx 0.08\%$ of the sample. Table A1 of the Appendix illustrates that most of these missing RUN correspond to foreign students, so maybe it is the case that they did not have their RUN at hand at the moment of enrollment, since the process for getting an identification can last around two-three months. Due to the presence of duplicated students -$\approx 1\%$- of the final sample, I keep a core dataset for 2018 including duplicated students but also create secondary datasets that deal with duplicated cases to precisely identify students staying in the system or changing from a degree program, respectively. This allows me to flexibly go back and forth throughout estimations without concern about duplicates in each question. These datasets are included in the replication package.

Persistence of Undergraduate Students in 2019

In this section, I analyze the undergrad dropouts in 2019. The retention rate of students in higher education, especially the first year, is one of the most used indicators internationally to evaluate the internal efficiency of tertiary institutions, considering that the highest dropout of students occurs in that period. I start by identifying the cohort of undergrads in 2018. This is the total amount of students who enrolled at the undergraduate level for the very first time in 2018. After cleaning duplicated cases, I remain with a cross-section dataset of 338,336 unique RUN, so this is what I will refer to as my "cohort 2018". As a sanity check, this number is quite similar to the one reported by MINEDUC for 2018[4]. In addition, it's important to argue the type of indicator that is going to be analyzed. MINEDUC defines three different indicators to measure school dropout:

- 1^{st} **year retention rate.** It is calculated as the ratio between the number of students who enter a degree or program as first-year students in a given year, and the number of those same students who remain in the same institution the following year, expressed in percentage points.
- 1^{st} **year persistence rate at the same institution.** It corresponds to the ratio between the number of students who enter an undergraduate program as first-year students in a given year, and the number of those students who remain enrolled in the same institution the following year, expressed in percentage points.
- 1^{st} **year persistence rate.** Is defined as the ratio between the number of students who enrolled in a program as first-year students in a given year, and the number of those same students who remain enrolled in an institution the following year, expressed in percentage points. In this case, those students who appear linked to a tertiary education institution, independent of the cohort and institution of origin, are considered persistent.

[3] Data retrieved from: `https://datosabiertos.mineduc.cl`.

[4] In its *Informe Matrícula 2018 en Educación Superior en Chile*, MINEDUC states that 337,936 students enrolled for the very first time in at least one undergrad program in Chile. Notice there is a difference of 400 students between official records and my estimates, but is negligible anyway. For reference, see table titled "*Evolución Matrícula 1er año de Pregrado por tipo de institución período 2009 - 2018*" in page 5 of the report.

I focus on the 1^{st} year retention rate. My numerator is the cohort of students who enrolled in any undergraduate degree in 2018. The denominator is those same students who enrolled in 2019. Results are reported in Table 1. I get that 74.4% of cohort 2018 persisted in the tertiary system of education in 2019. This is a reasonable number compared to the retention rate reported by the *Servicio de Información de Educación Superior* of MINEDUC in its *Informe de Retención Cohortes 2014-2018* who states that dropout for 1st-year undergrad students was 25% –see Table 1, page #2 for reference.

Table 1: Dropout in 2019

Condition	Count	Fraction (%)
Not continue	86,736	25.6
Continues	251,600	74.4
Total	338,336	100.0

Note: This table reports the retention rate for the cohort of 2018. Missing and duplicates *RUN* were dropped from the sample.

Binary Response Model: Estimation & Marginal Effects

In this section, I run a logit model to estimate the effects of a set of covariates on the probability of persistence at the undergraduate level. As argued by Cabrera (1994), logistic regressions are an appropriate statistical method for the study of variables that influence qualitative, dichotomous outcomes, such as persistence. For this reason, I build on previous literature studying retention in the education system that has also applied this type of non-linear model (Hu & John, 2001; Paulsen & John, 2002).

The Model. Let's first consider a student i who, conditional on a set of individual observable characteristics x_i, perceives some utility level $y_i^* : \mathbb{R}^n \to \mathbb{R}$ for enrolling at any university to start an undergraduate degree. Notice that y_i^* is a latent variable. Hence, econometricians do not directly observe the utility realization of the student but observe a binary decision: the student drops school if perceived utility levels are null, and continues to enroll otherwise. This problem can be summed in the following binary choice model:

$$\begin{cases} y_i^* = x_i\beta + \varepsilon_i \\ \\ y_i = 1 \text{ if } y_i^* > 0, \quad 0 \text{ otherwise} \end{cases} \tag{14}$$

From Equation 14 is easy to realize that:

$$\mathbb{P}(y_i = 1|x_i) = F(x_i\beta) \quad \forall i = 1, ..., n$$

So our logit model to be estimated is defined according to the assumption that ε has a standard logistic distribution. Thus, we can express our model as a response function of the form:

$$\mathbb{P}(persists|x_i) = F(x_i\beta) \equiv \Lambda(x_i\beta) = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}} \tag{15}$$

Equation 15 is our model of interest. The estimation is made using the maximum likelihood method and the average log-likelihood function of my model is defined as follows:

$$\ell(\beta; y_i|x_i) = \frac{1}{n}\sum_{i=1}^{n}\Big\{y_i \cdot ln[\Lambda(x_i\beta)] + (1 - y_i) \cdot ln[1 - \Lambda(x_i\beta)]\Big\}$$

Our solution is given by $\hat{\beta}_{ML} = \text{argmax}_\beta\, \ell(\beta; y_i|x_i)$.

Covariates. To make an adequate empirical analysis, I rely upon previous studies analyzing undergraduate dropouts in Chile (MINEDUC, 2012; Barrios, 2013; Bordón et al., 2015; Ayala Reyes & Atencio Abarca, 2018). I include a sub-set of sociodemographics (age, gender, and region where the university is located), education-related (type of institution, degree level, and area of study), and economic-related (students' income quintile[5], whether they receive any type of financial support[6]) outcomes.

Marginal Effects. Results for the estimations of $\hat{\beta}$ are reported in columns (1) and (2) of Table 2, and standard errors are estimated with a robust covariance matrix *à la* White. However, these coefficients are at the scale of log odds, so we cannot directly infer anything from raw estimates. To analyze $\hat{\beta}$ in terms of probabilities, I compute marginal effects from Equation 15 as follows:

$$\frac{\partial \mathbb{P}(persists|x_i)}{\partial x_j} = \Lambda(x_i\hat{\beta}) \cdot [1 - \Lambda(x_i\hat{\beta})] \cdot \beta_j$$

Results from marginal effects are reported in columns (3) and (4) of Table 2. I evaluate marginal effects at means and standard errors for marginal effects are computed using the delta method. For comparison purposes of marginal effects, columns (5) and (6) report estimates from a linear probability model using OLS. In all cases, columns entitled with even numbers correspond to models including the students' income quantile, while columns with odd numbers omit this covariate. Next, I briefly interpret my results, with the consideration that all of my independent variables are categorical outcomes, so marginal effects are interpreted as the difference in the probability between category 0 and 1 (when binary), or between some categories with respect to a baseline category (when more than one choice).

Sociodemographics

- **Gender.** Marginal effects for a dummy taking the value of 1 if the student is female and 0 otherwise, suggests that being a woman increases the probability of persisting in the undergraduate degree by $\approx$ 0.03 percentage points (p.p.) with respect to males, *ceteris paribus*[7]. Results are quite similar between my logit estimates and the linear probability model by OLS. Moreover, estimates are consistent with previous literature for Chile. For instance, MINEDUC (2012) finds a marginal effect of 0.02 p.p., Bordón et al. (2015) find a marginal effect of 0.01 p.p., and Barrios (2013) find a marginal effect of 0.03 p.p.

- **Age.** Consistent with other studies for Chile, I also find that the likelihood of staying in the education system decreases with age (MINEDUC, 2012; Bordón et al., 2015), *ceteris paribus*. In my five categories, the sign of coefficients turns out negative and statistically significant. For instance, being older than 40 years old decreases the probability of persistence at any undergraduate level with respect to the baseline category of students ranging between 15 to 19 years old.

- **Location of the university.** Marginal effects suggest that being enrolled in any university/program in the Metropolitan Region (RM) increases the probability of persisting in the education system by $\approx 0.003 - 0.01$ p.p., *ceteris paribus*. This is with respect to students enrolled in educational institutions outside of the Metropolitan Region. Results are quite consistent between my logit and OLS estimates. An insightful fact was made by Barrios (2013), who find that studying in a region

[5]Income quintile is obtained from FUAS filling (*Formulario Único de Acreditación Socioeconómica*) of 2018 and 2019. In some cases, students who were not granted any financial support in 2018 re-apply in 2019, but they tend to miss-report the perceived family income, so their quintile might change. In those cases, I keep the income quintile for 2018.

[6]This variable was built using scholarship allocation records. I restrict the dummy to take the value of 1 if the student was granted any type of financial benefit (scholarship or credit) and belongs to the type of student *"Estudiante que ingresa por primera vez a educación superior y obtiene un beneficio"* since I am interested in studying how certainty in financial aid affects educational persistence decisions, just like some literature has done before (Hu & John, 2001).

[7]This term was firstly adopted by Alfred Marshall to state "holding all other variables constant". Sometimes I will refer to this expression for simplicity.

different from the region of origin increases the probability of dropping out of college. However, I was not able to replicate this particular estimation.

Education-Related

- **Type of institution.** Results indicate that attending a Professional Institute or a University increases the probability of persistence in the educational system by ≈ 0.03, and 0.6 p.p., respectively. This is with regard to attending a Technical Formation Center, the baseline category, and holding constant the remaining variables.

- **Type of Degree.** Classification of the specific level to which the studies of the career or program belong is also relevant. Results indicate that pursuing a technical degree decreases the probability of persisting in the education system by ≈ 0.05 p.p.. This is with respect to *carreras profesionales*, the baseline category, and holding all other variables constant.

- **Area of study.** A very common phenomenon studied in the education literature is how career/degree decisions affect educational and even gender gaps. In this sense, my results suggest that studying arts, basic sciences, social sciences, humanities, and technology decreases the probability of persistence in the educational system. On the other hand, studying education increases the probability of staying at the undergraduate level by ≈ 0.01 p.p. These results are measured with respect to studying business and management, the baseline category, and holding other variables constant. Results reported here are quite consistent with previous estimates for Chile (Bordón et al., 2015).

Economic-Related

- **Beneficiary.** Being granted any type of financial support (scholarship or credit) increases the probability of persistence at the undergraduate level by ≈ 0.1 p.p. *ceteris paribus*. This is the most relevant covariate predicting persistence in education and is also consistent with the education literature for Chile (MINEDUC, 2012; Barrios, 2013; Bordón et al., 2015; Ayala Reyes & Atencio Abarca, 2018).

- **Income level.** Having income levels in the fifth quantile of the distribution increases the probability of persistence in the undergraduate level by $\approx 0.03 - 0.05$ p.p. with respect to the first quintile, the baseline category, and holding constant remaining outcomes. Even though previous studies for Chile have not included the student's income quantile but the income level instead, results are qualitatively similar since higher levels of income increase the probability of continuing in the education system. It is worth noting that for quintiles between 2 to 4 signs of the coefficients turn out negative and statistically significant. It is worth highlighting that the income quantile as a covariate implies a reduction in the number of observations because there is a non-negligible amount of students with missing values. Notwithstanding this caveat, I decided to include it because, as the results suggest, income is a relevant source of heterogeneity when predicting educational decisions.

Table 2: Estimates from Logit model

	$\mathbb{P}$(Persist=1)		Margins		OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Sociodemographics						
Gender						
Mujer	0.162***	0.161***	0.030***	0.028***	0.030***	0.028***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Age						
20 a 24 años	-0.290***	-0.211***	-0.052***	-0.036***	-0.049***	-0.035***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
25 a 29 años	-0.477***	-0.307***	-0.090***	-0.054***	-0.089***	-0.054***
	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
30 a 34 años	-0.477***	-0.284***	-0.090***	-0.050***	-0.091***	-0.050***
	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
35 a 39 años	-0.460***	-0.221***	-0.086***	-0.038***	-0.087***	-0.037***
	(0.02)	(0.03)	(0.00)	(0.01)	(0.00)	(0.01)
40 y más años	-0.454***	-0.225***	-0.085***	-0.039***	-0.086***	-0.038***
	(0.02)	(0.03)	(0.00)	(0.01)	(0.00)	(0.01)
Location						
En la RM	0.024***	0.014	0.004***	0.002	0.004***	0.002
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Panel B: Education-Related						
Type of Institution						
Institutos Profesionales	0.084***	0.092***	0.016***	0.016***	0.014***	0.017***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Universidades	0.154***	0.127***	0.029***	0.022***	0.026***	0.021***
	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Type of Degree						
Carreras Técnicas	-0.306***	-0.256***	-0.058***	-0.045***	-0.058***	-0.045***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Area of Study						
Agropecuaria	0.050*	0.023	0.009*	0.004	0.008*	0.004
	(0.03)	(0.03)	(0.00)	(0.01)	(0.00)	(0.01)
Arte y Arquitectura	-0.135***	-0.179***	-0.025***	-0.032***	-0.024***	-0.031***
	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Ciencias Básicas	-0.362***	-0.417***	-0.072***	-0.079***	-0.064***	-0.071***
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Ciencias Sociales	-0.097***	0.004	-0.018***	0.001	-0.017***	0.001
	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Derecho	-0.041	-0.009	-0.007	-0.001	-0.007	-0.002
	(0.03)	(0.03)	(0.00)	(0.01)	(0.00)	(0.00)
Educación	0.037***	0.081***	0.007***	0.013***	0.007**	0.012***
	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Humanidades	-0.373***	-0.331***	-0.074***	-0.061***	-0.072***	-0.058***
	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)
Salud	0.029**	0.073***	0.005**	0.012***	0.004	0.011***
	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Tecnología	-0.143***	-0.153***	-0.027***	-0.027***	-0.027***	-0.028***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Panel C: Economic-Related						
Financial Aid						
Beneficiario	0.563***	0.681***	0.104***	0.123***	0.103***	0.127***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Income Level						
Quintil 2		-0.360***		-0.061***		-0.056***
		(0.03)		(0.00)		(0.00)
Quintil 3		-0.230***		-0.038***		-0.034***
		(0.03)		(0.00)		(0.00)
Quintil 4		-0.182***		-0.030***		-0.027***
		(0.03)		(0.00)		(0.00)
Quintil 5		0.194***		0.028***		0.050***
		(0.03)		(0.00)		(0.00)
Constant	1.036***	1.114***			0.731***	0.734***
	(0.02)	(0.04)			(0.00)	(0.01)
N	338,256	265,359	338,256	265,359	338,256	265,359
R-Squared					0.04	0.03
Pseudo R-Squared	0.03	0.03				
Log-likelihood	-186,021.84	-138,923.46			-193,101.88	-142,452.78

Note: Standard errors are displayed in parenthesis. For columns (1) and (2) standard errors are computed using the White method. For columns (3) and (4) standard errors are computed using the delta method. For columns (5) and (6) standard errors are robust using the White method as well. Significance levels are denoted by * 0.10 ** 0.05 *** 0.01, respectively.

2.2 Question 2

In this question, we are interested in exploring the suggestion made by a MINEDUC advisor who states that educational trajectories are more complex than just persisting in the superior educational system. It is proposed that students can persist, change, or drop out of school. For this question, I tackle the problem by using the **"RUN-Código Único"** pair as my unique identifier. In my interpretation, this approach should give me a parsimonious identification of the status of each student because "codigo único" uniquely identifies the program degree code, considering institution, headquarters, program/career, journey, and version. Considering the presence of duplicated students, I merge 2018 and 2019 and create an auxiliary variable '*max merge*' that takes the maximum value of the merge between 2018 and 2019 for each student (whether is duplicated or not). I define trajectories following the next criteria:

- **Stays in the same degree.** Any (duplicated) student is labeled as "stayer" if had (more than two) any degree in 2018 and also enrolled (in at least one of them) in the same degree in 2019, so '*max merge*' takes the value of 3. That is, the student did merge at the "RUN-Código Único" level and still enrolled in 2019.

- **Changes degree.** Any (duplicated) student is labeled as if "changes degree" in those cases where enrolled (in more than two) in any degree in 2018, has a '*max merge*'$=1$ but appears to be enrolled in the educational system in 2019. That is, the student did not merge at the "RUN-Código Único" level but still enrolled in 2019, so I interpret this mismatch as the student changed her degree program in 2019.

- **Dropout.** Any (duplicated) student is labeled as if "dropout" in those cases where enrolled (in more than two degrees) in any degree in 2018, has a '*max merge*'$=1$ but does not appear enrolled in 2019 anymore. That is, the student did not merge at the "RUN-Código Único" level and did not enroll in 2019 either.

Persistence, Changes, & Drop Outs in 2019

Table 3 reports trajectories for the 2018 cohort. I get that $\approx 67\%$ of cohort 2018 persisted in the same degree during 2019, only a small fraction ($\approx 6\%$) changed the degree and the remaining ($\approx 26\%$) dropped out of school. For this number, I was not able to find any official report from MINEDUC, so I do not have any benchmark to infer whether it is a high or low estimate. To the best of my knowledge, the only report documenting switching patterns at the undergraduate level in Chile is Valenzuela & Kuzmanic (2023), who states that more than 50% of students under analysis changed their degree program before graduation. However, their study focuses on cohort 2011 and followed a total of 90,535 students from CFT, IP, and universities, which represent 34% of the total secondary education graduates in 2010. So their results do not directly mirror those reported here. For the sake of the remaining analysis, I assume that this number is correct.

Table 3: Trajectories in 2019

Condition	Count	Fraction (%)
Stays in same degree	232,547	68.7
Changes degree	19,053	5.6
Dropout	86,736	25.6
Total	338,336	100.0

Note: This table reports the trajectories for the cohort of 2018. Students with missing *RUN* were dropped from the sample, and duplicated students were treated according to the criteria described above.

Multinomial Choice Model: Estimation & Marginal Effects

In this sub-section, I analyze students' characteristics behind the decision to stay, change degrees, or leave university, using the decision to stay as my baseline category. For this goal, and even though is more costly computationally speaking, I estimate a multinomial probit model, so the Independence of Irrelevant Alternatives (IIA from now on) assumption is not a threat to my estimations. Since the problem set requests to argue about the IIA assumption and its implications, I will go back to this discussion further down a few lines. From now let us focus on the estimation of the multinomial probit model.

The Model. Following Wooldridge (2010), let's consider a general definition for unobserved utility perceived by any individual i in comparison to up to k different alternatives, represented by:

$$y_j^* = V_j + \varepsilon_i \quad j = 1, ..., J$$

where $V_j = x_i\beta_j$, and x_i is a vector with dimension $(k \times 1)$. For instance, the probability of the i-th alternative to be chosen is given by:

$$\mathbb{P}(y_j^* > y_{j+1}^*, ..., y_j^* > y_J^*) = \mathbb{P}(\varepsilon_{j+1} - \varepsilon_j < V_j - V_{j+1}, ..., \varepsilon_J - \varepsilon_j < V_j - V_J)$$

For our specific case, we have $J = 3$ (stays in the same degree, changes degree, and dropout). Hence, the average log-likelihood function to be estimated is given by:

$$\ell(\beta; y_i|x_i) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{3}\left\{\mathbb{1}(y_j^* = j) \cdot ln[\mathbb{P}_{ij}(y_j^* > y_{j+1}^*, ..., y_j^* > y_J^*)]\right\} \quad \text{s.t.} \sum_{j=1}^{3}\mathbb{P}_{ij} = 1 \tag{16}$$

Notice that Equation 16 implies solving two integrals because we need to normalize one parameter, as a matter of rank. So the model is precisely identified as the difference between coefficients from $j + 1, .., J$ categories with respect to the j-th category.

Marginal effects and interpretation. As is the case for any discrete choice model, estimates from $\hat{\beta}$ cannot be directly interpreted because they are not at a probability scale. Moreover, we can obtain marginal effects:

$$\frac{\partial \mathbb{P}(y_j^* = j)}{\partial x_j} = \frac{\partial \mathbb{P}(y_j^* > y_{j+1}^*, ..., y_j^* > y_J^*)}{\partial x_j}$$

Wooldridge (2010) flags that even when multinomial probits are very attractive their computational complexity not only makes it difficult to obtain the partial effects on the response probabilities but also makes maximum likelihood infeasible for more than five alternatives. In my case $J = 3$, so I might be optimistic and pray for our computer not to blow away :) Interpretation of marginal effect is also different in multinomial probits because we no longer have a binary variable as our dependent outcome, but a categorical variable instead. In this case, coefficients are interpreted as the difference in the effect on utility when comparing any j-th decision with respect to the baseline, *ceteris paribus*.

Table 4 displays the results for the multinomial probit model. I estimate the model including the same covariates as in question 1 above, with the twist that now I reduce the dimensionality of three variables (age, area of study, and income quintile) from more than two categories to a binary format, so estimations are easier to run on the computer. There are several takeaways from the results. In the first place, even though results are not directly comparable, most of the predicted probabilities turn out qualitatively similar to those reported across even columns in the collapsed version of the logit model in Table 2. Secondly, the sociodemographic traits of the student and attending a professional institute are not relevant for predicting the probability of changing the degree program. In the third place, is worth noting that some predictors change their signs according to the educational decision. For instance, studying at a university affects positively the likelihood of staying in the same degree program, and decreases the probability of changing from degree programs or even dropping out of school, with respect to the baseline category, holding all other variables constant. The same intuition holds when analyzing covariates that are economic-related.

Table 4: ESTIMATES FROM MULTINOMIAL PROBIT MODEL

	Estimates for $\hat{\beta}$		Marginal Effects		
	(1)		(2)	(3)	(4)
	Cambiarse	Desertar	Quedarse _	Cambiarse _	Desertar _
Panel A: Sociodemographics					
Gender					
Mujer	-0.054***	-0.142***	0.031***	-0.001	-0.030***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Location					
En la RM	-0.011	-0.012	0.003	-0.001	-0.002
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Panel B: Education-Related					
Type of Institution					
Institutos Profesionales	-0.022	-0.064***	0.014***	-0.000	-0.014***
	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
Universidades	-0.627***	-0.216***	0.078***	-0.047***	-0.031***
	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
Type of Degree					
Carreras Técnicas	-0.105***	0.195***	-0.032***	-0.014***	0.046***
	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
Area of Study					
Tecnología	0.241***	0.148***	-0.042***	0.016***	0.026***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Panel C: Economic-Related					
Financial Aid					
Beneficiario	-0.292***	-0.610***	0.135***	-0.008***	-0.127***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Income Level					
Quintil 5	-0.179***	-0.418***	0.091***	-0.004***	-0.087***
	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
Constant	-1.394***	-0.454***			
	(0.02)	(0.02)			
N	265,362		265,362	265,362	265,362
Log-likelihood	-191,154.57				

Note: Robust standard errors in parenthesis. Significance levels are denoted by * 0.10 ** 0.05 *** 0.01, respectively.

Discussing IIA assumption. IIA "implies that adding another alternative or changing the characteristics of a third alternative does not affect the relative odds between alternatives. This implication is implausible for applications with similar alternatives" (Wooldridge, 2010, p. 502). According to this description, IIA restricts staying in the same degree, changing degrees, or dropping out of school so as not to be affected by each other occurrence. However, is quite reasonable to presume that IIA requirement would not be met in a multinomial logit context because a new educational decision (i.e. suspend academic studies for a while) is restricted by the possibility of doing any of the remaining options. Data supports this intuition. In Table A2 of the Appendix, I show the results from Hausman & McFadden (1984) tests for a multinomial logit model. In most cases, I reject the null hypothesis of IIA so estimation by multinomial probit seems to be the most reasonable alternative.

Comparability. In the context of our study, the binary logit model estimated in question 1 captures the marginal effects between staying or dropping out of school, while the multinomial probit recently estimated captures the marginal effects of covariates on the choice between staying, changing the degree or leaving school. Consequently, results are not directly comparable. In addition, we must recall that the identification of β is different in each model. Finally, Wooldridge (2010) points out that in general, we can not make any comparison between coefficients from a logit or a probit model because there are differences in the distributional assumptions made before estimation. However, we should consider that, depending on the analysis context, if qualitative comparisons must be carried out for illustration purposes only then they might be done as long as we take some precautions at the moment of the analysis.

2.3 Question 3

Discussion on the Conditional Logit Model: Pros & Cons

So far we have focused on non-linear discrete choice models which allow the inclusion of individual-specific covariates, however, we can also make a twist to the analysis of these models by including alternative-specific covariates. McFadden (1974) proposed a special type of model that fits a conditional logistic model for matched individual–control data, also known as a fixed-effects logit model. A conditional logistic model fits maximum likelihood models with a binary dependent variable. In this sense, the analysis behind conditional logistic models is somewhat different from its "unconditional" version (as we do in empirical question $\#1$) in that the data are now grouped. Hence, the analysis is across groups and the likelihood is calculated relative to each cluster; that is, a conditional likelihood.

In terms of the pros of implementing these types of models in our setting is that we will control for alternative-specific covariates, such as the time that takes to do any paperwork or the tax payment. In this sense, the economic theory of individual preferences would suggest that benefits and costs associated with any choice (not necessarily individual-specific traits) have a significant impact on decision-making. In our setting, the optimal decision is to stay in school, because it has zero costs in terms of money and time, so any rational consumer of education (seen as a final good) would rather go for this option. On the other hand, in terms of disadvantages, these types of models are somewhat invariant to individual-specific traits, so they consider that variation in student characteristics is equally important, which is a very strong assumption. Then, estimates are probably non-extendable to many settings where the subject of analysis considers individual heterogeneity as the centerpiece, as might be the case for our example case.

Conditional Logit Model: Estimation & Marginal Effects

As argued before, now we are interested in studying alternative-specific variation. In this sense, we must create two new regressors, as follows:

Table 5: Rules for alternative-specific regressors

Regressor	Stay	Change degree	Dropout
Time	0 days	2 days	7 days
Tax	0 CLP	30k CLP	100k CLP

Creating the regressors above implies rectangularizing the dataset first. Before we had the dataset at the individual-choice level but now the dataset is at the student-alternative level, so every individual faces the three alternatives (staying, changing degree, and dropping out) and there is an additional binary variable that takes the value of 1 for the case where the individual made its choice, and 0 for the remaining alternatives. This adjustment will triplicate the number of observations from $338,336$ to $1,015,008$ ($338k \times 3$ educational options).

Table 6 displays the results from estimations of the conditional logit model including coefficients for the alternative-specific covariates (time and tax), as regressions for the alternatives. Notice that for the latter case, "staying in the same degree" is the baseline category, so results are not reported. Standard errors are not reported for some coefficients because the Hessian matrix is imprecisely estimated as including the intercept in the model makes the log-likelihood function to be not concave. However, I opted for this estimation strategy so I could get alternative-specific marginal effects to make them comparable with those I report in question # 4 down below.

Table 6: Estimates from Conditional Logit model

	$\mathbb{P}(\text{Choice} = 1)$
Covariates	
Time	30.551
Tax	-2.151
	(54,062.81)**
Alternatives	
Change Degree	
Constant	0.925
	(124.08)**
Dropout	
Constant	0.255
N	1,015,008

Note: Standard errors in parenthesis. Significance levels are denoted by * 0.10 ** 0.05 *** 0.01, respectively.

Marginal effects and interpretation. Once again, we can not give any direct interpretation to coefficients estimated from discrete choice non-linear models. We must get marginal effects instead. An *ad-hoc* definition for these marginal effects might be given by:

$$\frac{\partial \mathbb{P}(\text{Alternative}|x_g)}{\partial x_g} \equiv \frac{\partial \mathbb{P}(\text{Alternative} = j|x_g, \text{Alternative} = i)}{\partial x_g}$$

Where $i, j \in \{1, 2, 3\}$, with j representing the new alternative to be chosen, and i the alternative that was chosen before. Notice that, as an abuse of notation I define x_g to distinguish between alternative-specfic covariates from individual-specific covariates, but $g = 1, 2, 3$. In this case, my interpretation of the marginal effects is the average change in the probability that a student chooses one out of the three alternatives (stay, switch, dropout) conditional on a unit change in the alternative-specific covariate (time or tax) and assuming she has already chosen the i-th alternative before. Table 7 displays results for marginal effects of the conditional logit. The interpretation here is extensive since we have nine combinations because of the three available alternatives. Next, I discuss the meaning of some of these coefficients randomly selected, but it certainly will not be complete.

- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_g,\text{Alternative}=\textit{Stay})}{\partial x_g}$. Let us focus on the first three rows of the table. For instance, an increase of 1 day in the duration of paperwork dramatically increases the probability of staying in the same degree next year by ≈ 6.6 percentage points, assuming the student was already choosing "staying", and holding constant all other variables. In addition, an increase of one thousand Chilean pesos (CLP) in the tax increases the probability of dropping out of school by ≈ 0.38 percentage points –which is somewhat counter-intuitive–, assuming the student already chose to stay in the same degree and holding all other variables constant. An analog interpretation holds for the remaining outcomes.

- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_g,\text{Alternative}=\mathit{Change})}{\partial x_g}$. Now, let us take a look at the rows four to six of the table. For instance, an increase of 1 day in the duration of paperwork decreases the probability of dropping out of school next year by ≈ 0.44 percentage points, assuming the student was already choosing to change from the degree program, and holding constant all other variables. In addition, an increase of one thousand CLP in the tax increases the probability of staying at the same degree next year by ≈ 0.08 percentage points assuming the student was already choosing to stay at school and holding constant all other variables. An analog interpretation holds for the remaining outcomes.
- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_g,\text{Alternative}=\mathit{Dropout})}{\partial x_g}$. Finally, let us focus on the last three rows of the table. An increase of 1 day in the duration of the paperwork decreases the probability of changing from a degree program by ≈ 0.44 percentage points assuming the student was already choosing to drop out and with all other variables being constant. On the other hand, an increase of one thousand CLP in the tax increases the probability of staying in the same degree next year by ≈ 0.38 percentage points assuming the student was already choosing to drop out before and holding constant all other variables. An analog interpretation holds for the remaining outcomes.

Table 7: Marginal Effects from Conditional Logit model

Alternative	Time	Tax
$1 \times$ Stay	6.566***	-0.462***
	(719.74)	(-728.77)
$1 \times$ Change	-1.182***	0.0833***
	(-149.13)	(149.07)
$1 \times$ Dropout	-5.383***	0.379***
	(-499.57)	(503.79)
$2 \times$ Stay	-1.182***	0.0833***
	(-149.13)	(149.07)
$2 \times$ Change	1.624***	-0.114***
	(151.11)	(-151.11)
$2 \times$ Dropout	-0.441***	0.0311***
	(-138.42)	(138.56)
$3 \times$ Stay	-5.383***	0.379***
	(-499.57)	(503.79)
$3 \times$ Change	-0.441***	0.0311***
	(-138.42)	(138.56)
$3 \times$ Dropout	5.824***	-0.410***
	(521.20)	(-526.22)
N	1,015,008	

Note: Standard errors in parenthesis. Significance levels are denoted by * 0.10 ** 0.05 *** 0.01, respectively.

2.4 Question 4

Discussion on the Mixed Logit Model: Pros & Cons

Mixed logit models are an extension of previous work of McFadden (1974). These models constitute a more attractive specification when compared to its canonical version. One of the main advantages relies upon the fact that now we are not only accounting for variation in cluster-specific covariates but also for unobservable individual heterogeneity and how this affects variation in preferences. Hence, independent variables come in two forms: alternative-specific and individual-specific. Alternative-specific varies between groups (among all the alternatives), while individual-specific characteristics vary within groups (among all the students). On the other hand, one of the main disadvantages of estimating these types of non-linear models is their computational complexity, which is even more evident when it comes to assuming a normal distribution for the random term of the latent equation. But also, interpretation is more challenging not only because of the non-linear nature of the model but because is not that easy to disentangle unobserved effects from the alternative-specific component and unobserved heterogeneity of the individual-specific.

Mixed Logit Model: Estimation & Marginal Effects

In this case, I implement a mixed logit model to estimate the election between the three educational outcomes (staying, changing degree program, or doping out from university) conditional on the time and tax implemented for every choice, but also as a function of individual covariates that might affect student preferences. In this case, I will restrict the number of covariates to only a few so that computational complexity for computing marginal effect substantially decreases and also reduces the number of parameters to be estimated by the model. I opted for implementing a mixed logit model instead of a mixed probit due to computational efficiency solely. Table 8 shows estimations of coefficients only from a mixed conditional logit. Once again, in these types of models, $\hat{\beta}$ does not have any direct interpretation, because they are on a log-odds scale, so we might be interested in analyzing marginal effects to get a clue about heterogeneous effects on the probability of choosing any educational outcome. Nonetheless, coefficient estimations give us a first insight into the significance and potential direction of the effects.

Table 8: Estimates from Mixed Logit model

		Individual-Specific	
Covariate	(1) $\mathbb{P}(\text{Choice} = 1)$	(2) $\mathbb{P}(\text{Change Degree} = 1)$	(3) $\mathbb{P}(\text{Dropout} = 1)$
Time	21.21***		
	(140.20)		
Tax	-1.490***		
	(-140.71)		
Female		-0.128***	-0.221***
		(-8.48)	(-27.50)
RM		-0.0793***	-0.0459***
		(-5.15)	(-5.61)
Beneficiario		-0.214***	-0.638***
		(-13.94)	(-77.00)
N	1,015,008		

Note: Standard errors in parenthesis. Significance levels are denoted by * 0.10 ** 0.05 *** 0.01, respectively.

Marginal effects and interpretation. As argued before, we can not retrieve any conclusions from raw coefficients, so we need to compute marginal effects. In a mixed logit model, the marginal effects are expressed in percentage points so that a one-unit change in the alternative-specific covariate or the individual-specific covariate, leads to a percentage-point increase/decrease in the alternative. Alternative-specific covariates are "time" and "tax", while student-specific covariates are "gender" $\mathbb{1}(\text{mujer} = 1)$, "location" for the region where the university is located $\mathbb{1}(\text{Región Metropolitana} = 1)$, and "beneficiario" if the student was granted any scholarship in 2018 $\mathbb{1}(\text{becario} = 1)$. "Staying in the same degree" is the baseline category. Table 9 displays results for marginal effects of the mixed logit. Again, the interpretation can turn quite extensive, so I delimit the analysis by picking up one or two covariates randomly.

<u>**Alternative-Specific**</u>

- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_g,\text{Alternative}=\textit{Stay})}{\partial x_g}$. Let us focus on the first three rows and columns (1) and (2) of the table. For instance, an increase of 1 day in the duration of paperwork dramatically increases the probability of staying in the same degree next year by ≈ 6.6 percentage points, assuming the student was already choosing "staying", and holding constant all other variables. In addition, an increase of one thousand Chilean pesos (CLP) in the tax increases the probability of dropping out of school by ≈ 0.38 percentage points, assuming the student already chose to stay and holding all other variables constant. An analog interpretation holds for the remaining outcomes.
- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_g,\text{Alternative}=\textit{Change})}{\partial x_g}$. Now, let us take a look at rows four to six and columns (1) and (2) of the table. For instance, an increase of 1 day in the duration of paperwork decreases the probability of dropping out of school next year by ≈ 0.44 percentage points, assuming the student was already choosing to change from the degree program, and holding constant all other variables. In addition, an increase of one thousand CLP in the tax increases the probability of staying at the same degree next year by ≈ 0.08 percentage points assuming the student was already choosing to stay at school and holding constant all other variables. An analog interpretation holds for the remaining outcomes.
- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_g,\text{Alternative}=\textit{Dropout})}{\partial x_g}$. Finally, let us focus on the last three rows and columns (1) and (2) of the table. An increase of 1 day in the duration of the paperwork decreases the probability of changing from a degree program by ≈ 0.44 percentage points assuming the student was already choosing to drop out and with all other variables being constant. On the other hand, an increase of one thousand CLP in the tax increases the probability of staying in the same degree next year by ≈ 0.38 percentage points assuming the student was already choosing to drop out before and holding constant all other variables. An analog interpretation holds for the remaining outcomes.

<u>**Individual-Specific**</u>

- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_i,\text{Alternative}=\textit{Stay})}{\partial x_i}$. Let us focus on the first three rows and columns (3) to (5) of the table. Being a female increases the probability (with respect to male students) of dropping out of school by ≈ 0.04 percentage points, assuming the student was already choosing "dropout", and holding constant all other variables. In addition, attending an institute in the Metropolitan Region (with respect to the remaining regions in Chile) increases the probability of dropping out of school by ≈ 0.01 percentage points, and being granted a scholarship increases the probability (with respect to students without any benefit) of staying by ≈ 0.1 percentage points, assuming the student was already choosing to stay and holding all other variables constant. An analog interpretation holds for the remaining outcomes.
- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_i,\text{Alternative}=\textit{Change})}{\partial x_i}$. Now, let us take a look at rows four to six and columns (3) to (5) of the table. Being a female increases the probability (with respect to male

students) of choosing "dropout" by ≈ 0.003 percentage points assuming the student was already choosing "change degree" and holding all other variables constant. Studying in an institution located in the Metropolitan Region or being granted a scholarship decreases the probability of changing degree programs assuming the student was already choosing the same alternative by 0.004 and 0.01 percentage points, with respect to students that are studying out of the RM or without any benefit, respectively. An analog interpretation holds for the remaining outcomes.

- **Marginal effects for** $\frac{\partial \mathbb{P}(\text{Alternative}=j|x_i,\text{Alternative}=\textit{Dropout})}{\partial x_i}$. Finally, let us focus on the last three rows and columns (3) to (5) of the table. Being a female, studying in the Metropolitan Region or being an scholar increases the probability of choosing "change degree" assuming that the student was already selecting "dropping out" by $0.001, 0.001$, and 0.003 percentage points, respectively. These results are with respect to being a male, studying outside the RM, or being a scholar, respectively, and holding all other covariates constant. An analog interpretation holds for the remaining outcomes.

Table 9: Marginal Effects from Mixed Logit model

	Alternative-Specific		Individual-Specific		
	(1)	(2)	(3)	(4)	(5)
	Time	Tax	Female	Region Metropolitana	Beneficiario
1 × Stay	6.594***	-0.463***			
	(690.39)	(-697.06)			
1 × Change	-1.216***	0.0854***	0.00505***	0.00311***	0.00835***
	(-148.98)	(148.95)	(8.44)	(5.16)	(13.88)
1 × Dropout	-5.378***	0.378***	0.0386***	0.00798***	0.110***
	(-479.81)	(482.83)	(27.46)	(5.62)	(78.37)
2 × Stay	-1.216***	0.0854***			
	(-148.98)	(148.95)			
2 × Change	1.658***	-0.116***	-0.00689***	-0.00424***	-0.0114***
	(151.04)	(-151.06)	(-8.44)	(-5.16)	(-13.88)
2 × Dropout	-0.442***	0.0311***	0.00318***	0.000656***	0.00911***
	(-137.82)	(137.95)	(26.91)	(5.61)	(68.07)
3 × Stay	-5.378***	0.378***			
	(-479.81)	(482.83)			
3 × Change	-0.442***	0.0311***	0.00184***	0.00113***	0.00303***
	(-137.82)	(137.95)	(8.44)	(5.16)	(13.86)
3 × Dropout	5.820***	-0.409***	-0.0418***	-0.00863***	-0.119***
	(498.99)	(-502.60)	(-27.46)	(-5.62)	(-78.37)
N	1,015,008				

Note: Standard errors in parenthesis. Significance levels are denoted by * 0.10 ** 0.05 *** 0.01, respectively.

Comparing results. In this case, we can only compare alternative-specific coefficients, because these are the coefficients we obtained in both problems 3 and 4. Coefficients for "time" and "tax" reported here mirror those we obtained before, not only qualitatively (both of them preserve the sign) but they do not change that much in magnitude as well. Significance is mostly preserved. In general, we can conclude that in our very specific example case, there are no major differences in this case between computing these estimates jointly or separately in conditional or mixed logit format, respectively. However, we should always be cautious on what could be the most precise model to implement according to our setting. My intuition obtained from this exercise is that mixed logits offer a more comprehensive insight of the problem by including both, alternative- and individual-specific heterogeneity.

References

Ayala Reyes, M., & Atencio Abarca, I. (2018). Retención en la educación universitaria en chile. *Revista de la Educación Superior*, *47*(186), 93–118. 10, 11

Barrios, A. (2013). Deserción universitaria en chile. *Revista del Centro de Investigación Social de Un Techo para Chile*, (pp. 59–72). 10, 11

Bordón, P., Canals, C., & Rojas, S. (2015). Retención en los programas e instituciones de educación superior. *Estudios de Política Educativa*, (2), 174–201. 10, 11

Cabrera, A. (1994). *Logistic regression analysis in higher education: An applied perspective.*, vol. 10 of *Higher education: Handbook of theory and research*. 9

Colin Cameron, A., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press. 7

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, (52), 1219–1240. 15, 23

Hu, S., & John, E. P. S. (2001). Student persistence in a public higher education system: Understanding racial and ethnic differences. *The Journal of Higher Education*, *72*(3), 265–286. 9, 10

McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior*. Frontiers in Econometrics. New York: Academic Press. 16, 19

MINEDUC (2012). Deserción en la educación superior en chile. Serie Evidencias 9, Centro de Estudios Ministerio de Educación. 10, 11

Paulsen, M. B., & John, E. P. S. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *The Journal of Higher Education*, *73*(2), 189–236. 9

Ruud, P. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press. 4

Valenzuela, J. P., & Kuzmanic, D. (2023). Dropouts and transfers: socioeconomic segregation in entrance to and exit from the chilean higher education. *Higher Education Research Development*, *8*(42), 2048–2065. 13

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. 2, 5, 14, 15, 16

A Appendix

A.1 Additional Tables

Table A1: Fraction of missing RUN

Accessed through	Count	Fraction (%)
Foreign student	185	68.3
Directly (regular student)	74	27.3
Other	12	4.4
Total	271	100.0

Note: This table reports the number of RUN that are missing and were dropped before estimations. The first column shows the mechanism through which the student got access to the undergrad degree. This table is discussed in subsection 2.1.

Table A2: Hausman & McFadden (1984) test

Statistic	(1)	(2)	(3)
Model 1			
χ^2	1,765.78	26.18	407.25
p-value	0.000	0.000	0.000
Model 2			
χ^2	728.85	1.07	187.05
p-value	0.000	0.301	0.000

Note: This table reports Hausman tests. H_0 is IIA, while H_a is absence of IIA. Column (1) excludes the category "stays in the same degree". Column (2) drops the category "changes degree". Column (3) excludes the category "drops out". Model 1 excludes quintile of income as a covariate, while Model 2 includes it. This table is discussed in subsection 2.2.