Determinants of youth higher education enrollment in poverty and vulnerability contexts: evidence from Lima Sur.

ABSTRACT

This paper seeks to define the determinants of youth higher education enrollment in contexts of high poverty and vulnerability. We estimate a nonlinear logistic regression model with a binomial independent variable while using data of youngsters from Lima Sur compiled in INEI's 2017 Census data base. The results show that a head of household with higher education, having a private insurance, internet availability, and being a woman all increase youth enrollment. On the other hand, cohabiting at a young age, poor residence infrastructure and habitability conditions, and living in a single-parent household, decrease it.

Key words: youth, higher education, vulnerability, Lima Sur.

INTRODUCTION

Lima Sur (LS) is one of the five subregions of Metropolitan Lima (LM) which is composed of eleven districts in the southern periphery of the metropolis: Chorrillos, Lurin, Pachacamac, Pucusana, Punta Hermosa, Punta Negra, San Bartolo, San Juan de Miraflores, Santa Maria del Mar, Villa el Salvador and Villa Maria del Triunfo. At the spatial and demographic level, LS comprises 32,5% of the area (850.50 km2) and 20% of the population (1.7 million inhabitants) of LM (INEI, 2019). This subregion is characterized by a higher population density in districts with less land area, but closer to the city center, the opposite pattern to that of districts with greater land area, but farther from the center (Figure 1).

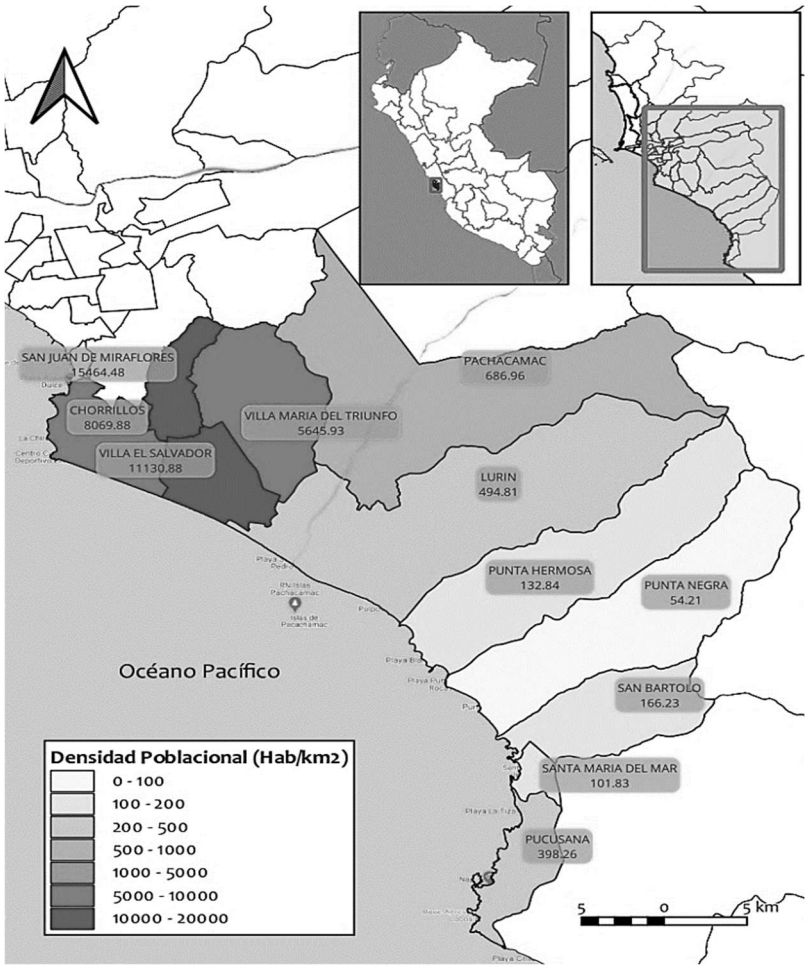

Figure 1. Lima Sur: population density map, by districts, 2018.
Source: Instituto Nacional de Estadística e Informática - INEI (2020).
Authors' elaboration.

At the socioeconomic level, LS is characterized by spaces in process of consolidation, usually occupied by poor or vulnerable population (Carrillo, Salazar, & Leandro, 2019). This characterization is congruent with the theory of spatial segregation, which explains that the most vulnerable population settles in peripheral areas, while those with more economic resources, in areas with greater urban development (Garret, Miranda, Marcos, & Christiansen, 2021). With this information in mind, it is possible to understand why most districts in LS show intra-district levels of poverty ranging between 15% and 25% (Figure 2).

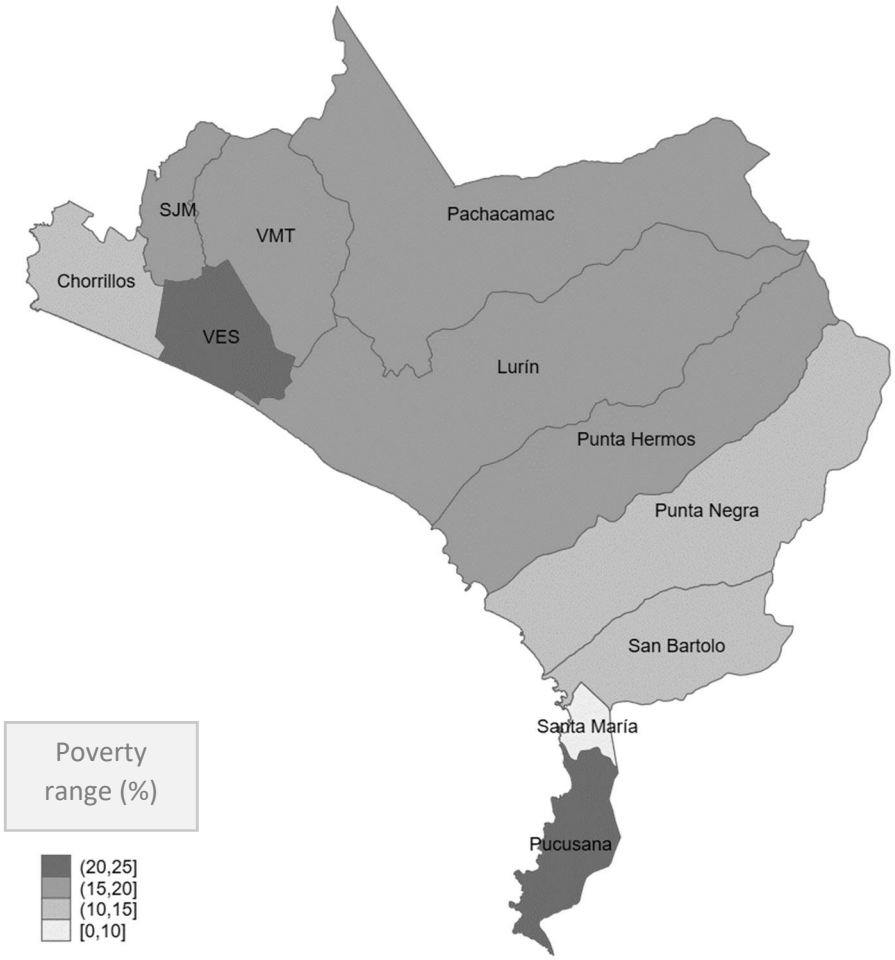

Figure 2. Lima Sur: poverty map, by districts, 2018.
Source: Instituto Nacional de Estadística e Informática - INEI (2020).
Authors' elaboration.

Within this framework, youngsters between 18 and 25 years of age represent an important part of LS' population. During 2017 alone, 14,6% of the total population in LS was composed of youngsters within this range of age (INEI, 2017). Given the areas of poverty and precariousness in which they live, it is worrying that more than half of these young people have not enrolled into the higher education system (Table 1). This figure becomes more relevant if we consider education as a tool with great potential to reduce poverty and vulnerability by increasing employability, productivity and income (World Bank, 2018).

Table 1. Lima Sur: population, aged 18 to 25 years, by educational level, 2017.

Educational level achieved	Population	Porcentage
No education	784	0,33
Primary school	7646	3,20
Secondary school	116’826	48,91
Higher education (incomplete)	78’532	32,88
Higher education (complete)	35’066	14,68
Total	238’854	100

Source: Instituto Nacional de Estadística e Informática - INEI (2020).
Authors’ elaboration.

Despite the above, enrollment into the higher education doesn’t just depend on the youngsters’ characteristics, but on many other sociocultural factors (Balarin, Alcázar, Rodríguez, & Glave, 2017). That said, this paper seeks to define the determinants of youth enrollment into higher education in contexts of high poverty and vulnerability. For this purpose, the case of young people aged 18 to 25 years in the peripheral subregion of Lima Sur is analyzed. Thus, we answer the question: What factors determine whether a young person between 18 and 25 years of age in LS enrolls into higher education? It is hypothesized that this is mainly due to family factors and the economic conditions of the household.

The distribution of this document goes as follows. First, the Literature Review section is presented, in which the national and international literature on the subject is discussed. Secondly, the Methodology section, in which the empirical strategy used is specified: econometric model, data and variables. Third, the Results section presents the statistical-inferential results obtained from the estimation of the model. Finally, the Conclusions and Recommendations formulated from the interpretation of the results are presented.

LITERATURE REVIEW

In the higher education field, there is a vast literature that reviews the different determinants that influence young people and their demand, enrollment and expenditure in education. It is on the basis of this literature that, subsequently, a group of relevant variables are selected.

At the international level, among authors who have studied these issues in developed countries, there is a certain consensus regarding the influence of individual factors - such as gender and race - and family factors - such as the income and education of the household's members - on the enrollment and permanence of young people in higher education (González & Dávila, 1998; De Pablos & Gil, 2007; Conley, 2021). Even more, De Pablos, Gil and Martínez (2010) explain that the mother's education has a positive influence on the demand for higher education and that the greater presence of children in the household and unemployed members, have a negative influence. Likewise, Booth and Joo (2009) show that birth order and family size are determinants of the allocation of resources for the education of young people in the household.

On the developing countries' side, the authors pay greater attention to family composition - two-parent families, single-parent families, number of household members - and the head of household characteristics - gender, education, employment status - as determinants of higher education demand (Ogawa & Iimura, 2010; Albert, González, & Mora, 2013; Acarenza & Gandelman, 2017; Chauca & Rodriguez, 2019). In addition, it is shown that the financing of studies through scholarships or credit programs favors access to higher education (Olavarría & Allende, 2013; Prodan, Maxim, Manolescu et al., 2015; Ogawa & Iimura, 2010), which evidences the importance of economic factors. In the same direction, demographic factors such as the type of area of residence - rural or urban - (Ogawa & Iimura, 2010) and the educational supply (Londoño, Canavire-Bacarreza, Bohórquez, & Cuartas, 2015), play an important role.

At the national level, there is a large body of literature on the subject. In line with international findings, national authors highlight the importance of individual factors - previous education and cognitive skills - and family factors – income and parental education - on the demand and enrollment in higher education (Castro, Yamada, & Arias, 2011; Guerrero, 2013; Flor, Magnaricotte, & Alba, 2020). Similarly, sociodemographic factors such as the processes of urbanization, the increase in returns to education and expectations of social mobility (Díaz, 2008), and rural conditions (Benavides, Olivera, & Mena, 2006) determine higher education demand and enrollment. In line with the above, although there are still racial gaps, gender gaps have been closed to such an extent that there are no differences between male and female youngsters who complete the educational process (Castro & Yamada, 2011).

METHODOLOGY

Econometric Model

To model the youngster's choice of enrolling in higher education, we propose a nonlinear discrete choice model with a binomial dependent variable. Following the work of Gonzales and Dávila (1998), this model combines human capital theory and discrete choice models which are derived from the random utility hypothesis (Luce, 1959; McFadden, 1984). This way, the i-th youngster has two alternatives: to enroll into higher education or not to enroll. Thus, the discrete variable Y_i is defined, which equals 1 if enrolled and 0 otherwise:

$$Y_i = \begin{cases} 1, & \text{enrolled} \\ 0, & \text{not enrolled} \end{cases} \qquad \forall i = 1, \dots, N \qquad (1)$$

Specifically, the probability that the i-th youngster enrolls in higher education (i.e., $Y_i = 1$) is explained by his or her personal and surroundings characteristics, a relationship that will be given in a nonlinear way by a probability function $F(.)$. Based on the goodness-of-fit analysis presented in Appendix 1, it is defined that $F(.)$ will follow a logistic distribution. Hence, the discrete choice nonlinear regression model with a binomial independent variable is defined as:

$$\text{Prob}(Y_i = 1|x_i) = F(x_i'\beta) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}} \qquad \forall i = 1, \dots, N \qquad (2)$$

$\text{Prob}(Y_i = 1|x_i)$ will be the probability that youngster i decides enroll into the higher education system given his or her information in x_i. This vector x_i has a $K \times 1$ order and includes K variables that capture the characteristics of the youngster and his or her surroundings. On the other hand, β is a vector of order $K \times 1$ that includes the coefficients associated with each one of the K variables contained in x_i.

The coefficients in β are estimated using the maximum likelihood method. However, given the nonlinear nature of the model, these estimated coefficients cannot be interpreted as the direct

effect of each variable on $\text{Prob}(Y_i = 1|x_i)$. As a consequence, the concept of marginal effect (ME) is introduced as the partial derivative of $\text{Prob}(Y_i = 1|x_i)$ with respect to a variable x_k, ceteris paribus. Thus, for the i-th youngster with a vector of characteristic variables $x_i = \tilde{x}_i$, the ME_{ik} will be:

$$ME_{ik} = \frac{\partial \text{Prob}(Y_i = 1|x_i = \tilde{x}_i)}{\partial x_{ik}} \qquad \begin{matrix} \forall i = 1, \dots, N \\ \forall k = 1, \dots, K \end{matrix} \qquad (3)$$

Now, because the ME_{ik} depends on the values of $\tilde{x}_i$ in each observation, the average marginal effect (AME) of each variable is calculated. This will be the simple arithmetic average of the ME_{ik} of each i youngster. Then, the AME_k of a variable x_k on $\text{Prob}(Y_i = 1|x_i)$ will be:

$$AME_k = \frac{1}{N}\sum_{i=1}^{N} EM_{ik} \qquad \forall k = 1, \dots, K \qquad (4)$$

Data

This paper utilizes the publicly available database of the XII Population Census and VII Housing Census of 2017 by Peru's Instituto Nacional de Statistical e Informática (INEI). This is a cross-sectional database which describes each citizen's individual, household and residence characteristics. We limit the data to analyze the population of young people between 18 and 25 years of age, who are children of their head of household and who reside in one of the districts in Lima Sur. Thus, we end up with information on 126'260 youngsters.

Variables

Based on the literature reviewed, the vector of explanatory variables x_i is divided into 3 groups: individual, household and residence. The detail is presented in Table 2.

Table 2. Variables

Group	Variable	Description
Dependent	Higher education	1 = enrolled; 0 = otherwise
Individuals	Woman	1 = female; 0 = male
	Age	Years of age
	Age^2	Years of age squared
	Native language	1 = native or indigenous; 0 = Spanish
	Cohabitance	1 = lives with a partner; 0 = otherwise
	Work	1 = works; 0 = doesn't work
	Health insurance	1 = private; 0 = not private or doesn't have
Household	Head - Education	1 = higher education; 0 = otherwise
	Head - Native Language	1 = native or indigenous; 0 = spanish
	Head - Age	Head's years of age
	Head – Age^2	Head's years of age squared
	Head - Works	1 = works; 0 = doesn't work
	Infants	Number of children under 5 years of age
	Older adults	Number of people over 65 years of age
	Type of husehold	2 = mother only; 1 = father only; 0 = both
Residence	Water	1 = from public network; 0 = otherwise
	Drainage	1 = connected to public network; 0 = otherwise
	Lighting	1 = from public network; 0 = otherwise
	Internet	1 = has internet; 0 = doesn't have internet
	Cooking fuel	1 = electricity or gas; 0 = otherwise
	Inadequate [1/]	1 = inadequate residence; 0 = otherwise
	Overcrowded [2/]	1 = overcrowded residence; 0 = otherwise
	Density	District's population density in logarithms

[1/] (i) an improvised residence, or (ii) a residence with walls made out of matting, or (iii) a residence with walls made out of quincha, stone with mud, wood, or similar materials, and, in all cases, without a built floor (INEI, 2018).

[2/] A residence with three or more people per bedroom (INEI, 2015).

Authors' elaboration.

RESULTS

The AME of each variable on the probability of enrollment are presented below. Each column of Table 3 shows the AME under different model specifications. Column (1) shows the specification only including the youngster's individual characteristics; column (2), only includes the household's characteristics; column (3), only the residence's characteristics; and, finally, column (4) presents the complete model. The low sensitivity of the model's estimates in column (4) compared to the other specifications demonstrates its consistency and robustness. Variables without statistical significance and those that have small AME, even with statistical significance, are omitted from the analysis.

Table 3. Average marginal effects.

Variable	(1)	(2)	(3)	(4)
Woman	0,113***			0,116***
	(0,003)			(0,003)
Age	0,385***			0,373***
	(0,012)			(0,012)
Age^2	-0,008***			-0,008***
	(0,000)			(0,000)
Native language	-0,102***			-0,033***
	(0,010)			(0,010)
Cohabitance	-0,219***			-0,134***
	(0,005)			(0,005)
Works	-0,062***			-0,032***
	(0,003)			(0,003)
Health insurance	0,249***			0,143***
	(0,007)			(0,007)

Head - Education	0,225***		0,172***
	(0,003)		(0,003)
Head - Native Language	0,003		0,020***
	(0,003)		(0,003)
Head - Age	0,027***		0,004**
	(0,002)		(0,002)
Head – Age^2	-0,000***		-0,000***
	(0,000)		(0,000)
Head - Works	0,020***		0,016***
	(0,003)		(0,003)
Infants	-0,075***		-0,059***
	(0,002)		(0,002)
Older adults	0,021***		0,012***
	(0,004)		(0,004)
Type of household (1)	-0,111***		-0,047***
	(0,004)		(0,004)
Type of household (2)	0,014***		-0,038***
	(0,004)		(0,004)
Water		0,021***	0,001
		(0,008)	(0,007)
Drainage		0,024***	0,016**
		(0,008)	(0,008)
Lighting		0,043***	0,031***
		(0,010)	(0,009)
Internet		0,192***	0,129***
		(0,003)	(0,003)
Cooking fuel		0,061***	0,055***
		(0,018)	(0,016)
Inadequate		-0,093***	-0,064***
		(0,005)	(0,005)
Overcrowded		-0,109***	-0,060***
		(0,006)	(0,006)
Density		-0,003**	-0,005***

			(0,001)	(0,001)
Observations	126’734	126’508	127’015	126’260

Note: dy/dx for factor levels is the discrete change from the base level. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Authors’ elaboration.

As for the individual characteristics that have a positive impact on the probability of enrollment, we have the following. First, being a woman, compared to being a man, increases the probability by 11,6 percentage points. Secondly, being one year older translates into an increase in the probability of 37,3%. Thirdly, having a private health insurance increases the probability by 14,3 percentage points, compared to having a nonprivate insurance or not having insurance at all. On the side of the individual characteristics with negative impacts non the probability, we have that living with a partner at an early age decreases the probability of enrollment by 13.6 percentage point.

Regarding the household characteristics that have a positive impact on the probability of enrollment, the following is found. If the head of household has enrolled at any level of higher education, the probability that the youngster enrolls into higher education increases by 17,2 percentage points. Among the household’s characteristics with negative effects on the enrollment probability, the presence of one more infant in the household decreases that probability by 5,9%. In addition, compared to biparental households, living in a single-parent household guided by the father or mother decreases youngster’s enrollment probability by 4,7 and 3,8 percentage points, respectively.

In terms of the residence characteristics that have a positive impact on the probability of enrolling, it is known that having access to internet services increase it by 12.9 percentage points. On top of that, if the residence has a cooking fuel that isn´t harmful to health (electricity or gas), the probability increases by 5,5 percentage points. On the other hand, on residence’s characteristics that have a negative impact on the probability of enrollment, it is known living in an infrastructural inadequate residence (according to INEI standards) and in overcrowding conditions of more than three people per bedroom decreases the probability in 6,4 and 6,0 percentage points, respectively.

CONCLUSIONS

This document has sought to define the determinants of youth enrollment into higher education in contexts of high poverty and vulnerability. For this purpose, data from youngsters in the peripheral subregion of Lima Sur were used. Based on the literature review, 23 possible determinants were defined and divided into three groups: individual, household and residence characteristics. As our empirical strategy, a nonlinear logistic regression model with a binomial independent variable was used together with data from INEI's XII Population Census and VII Residence Census of 2017. From the model's estimation, the results lead to the conclusions and policy recommendations presented below.

In first place, it is acknowledged that economic factors prevail as essential factors for higher education attainment in vulnerable and poor regions such as LS. This is stated due to the importance of private insurance and unharmful yet expensive cooking fuels availability on enrollment probability. As other authors say (Acevedo et al., 2008; Olavarría & Allende, 2013; Prodan, Maxim, Manolescu et al., 2015; Ogawa & Iimura, 2010), there´s a long-lasting need for scholarships and educational credit programs with low interest rates to ensure youngsters educational attainment. In the same sense, given the importance of internet availability, it is much clearer that the universalization of this service is a necessity for the youth.

In second place, on the same line with literature (De Pablos and Gil, 2007; De Pablos, Gil and Martinez, 2010; Castro, Yamada and Arias, 2011; Guerrero, 2013; Acarenza and Gandelman, 2017), a head of household with higher education is very important for youth educational enrollment. This shows the importance of role model programs to encourage youth to pursue higher education in the absence of intra-household role models. On the same note, even when residing in a single-parent household is negative for youngsters, if that household is guided by the mother, the effect is slightly less detrimental compared to the case where it is guided by the father. This strengthens the importance of female role models inside the household.

In third place, cohabiting with a partner at a young age decreases youngsters' probability of enrolling into higher education, probably due to the cost maintaining the newly formed family nucleus. This cost may well divide into two types: a monetary cost and the time cost of coping up with household chores and duties. Either way, the costs of cohabiting deter youngsters from enrolling into higher education. In this manner and also considering the negative impact of

inadequate housing and overcrowded residences, youth housing credit programs may partially help young couples confront the expenses of living together.

Finally, as in Castro and Yamada (2011), the results reveal that the educational enrollment gender gaps may be closing. This is stated considering that our results show that women are more likely to enroll into higher education than men. This said, it is important that the authorities address other gender gaps that still prevail, such as those in STEM (science, technology, engineering and mathematics) careers, which prevent women from reaching their full potential, even when they have access to higher education.

REFERENCES

Acarenza, S., & Gandelman, N. (2017). *Household education spending in Latin America and the Caribbean: Evidence from income and expenditure surveys.* Washington, DC: Inter-American Development Bank - IDB. DOI: https://doi.org/10.18235/0000651

Albert, C., González, C., & Mora, J. (2013). DETERMINANTES DE LA DEMANDA DE EDUCACIÓN UNIVERSITARIA EN COLOMBIA, 1980-2010. *Revista de Economía Institucional, 15*(29), 169-193.

Balarin, M., Alcázar, L., Rodríguez, M. F., & Glave, C. (2017). *Transiciones inciertas: Una mirada a los jóvenes de contextos urbanos vulnerables de Lima.* Lima: Grupo de Análisis para el Desarrollo - GRADE.

Benavides, M., Olivera, I., & Mena, M. (2006). *De papás y mamás a hijos e hijas: las aspiraciones sobre el futuro y rol de las familias en las actividades escolares en el Perú rural.* Lima: Grupo de Análisis para el Desarrollo - GRADE.

Booth, A., & Joo, H. (2009). Birth order matters: the effect of family size and birth order on educational attainment. *Journal of Population Economics*, pages367–397. DOI: https://doi.org/10.1007/s00148-007-0181-4

Carrillo, S., Salazar, V., & Leandro, S. (2019). *Jóvenes y educación en Lima Metropolitana y Callao.* Lima: Instituto de Estudios Peruanos - IEP.

Castro, J. F., & Yamada, G. (2011). *Brechas étnicas y de sexo en el acceso a la educación básica y superior en el Perú.* Lima: Centro de Investigación de la Universidad del Pacífico - CIUP.

Castro, J. F., Yamada, G., & Arias, O. (2011). *Higher Education Decisions in Peru: On the Role of Financial Constraints, Skills, and Family Background.* Lima: Centro de Investigación de la Universidad del Pacífico - CIUP.

Chauca, J., & Rodriguez, R. (2019). *Análisis de los determinantes sociodemográficos del acceso general a la educación superior en el Ecuador e identificación de la política de cuotas para los grupos vulnerables durante el año 2014.* Quito: Escuela Politécnica Nacional - EPN.

Conley, D. (2021). Capital for College: Parental Assets and Postsecondary Schooling. *Sociology of Education, 74*(1), 59-72. DOI: https://doi.org/10.2307/2673145

De Pablos, L., & Gil, M. (2007). Análisis de los condicionantes socioeconómicos del acceso a la educación superior. *Presupuesto y Gasto Público*, 37-57.

De Pablos, L., Gil, M., & Martínez, M. (2010). Los determinantes socioeconómicos de la demanda de Educación Superior en España y la movilidad educativa intergeneracional. *Revista de Economía Pública*, 75-108.

Díaz, J. (2008). *Educación superior en el Perú: tendencias de la demanda y la oferta.* Lima: Grupo de Análisis para el Desarrollo - GRADE.

Flor, J. L., Magnaricotte, M., & Alba, F. (2020). *Los factores que limitan la transición a la educación superior: Situación actual y recomendaciones de política pública.* Lima: Ministerio de Educación.

Garret, P., Miranda, L., Marcos, M., & Christiansen, A. (2021). Segregación socioeconómica en las escuelas de Lima Metropolitana. *Revista Peruana de Investigación Educativa* (15), 89-119.

González, B., & Dávila, D. (1998). Economic and Cultural Impediments to University Education in Spain. *Economics of Education Review, 17*(1), 93-103.

Guerrero, G. (2013). *¿Cómo afectan los factores individuales y escolares la decisión de los jóvenes de postular a educación superior? Un estudio longitudinal en Lima, Perú.* Lima: Grupo de Análisis para el Desarrollo - GRADE.

INEI (2015). *Las adolescentes y su comportamiento reproductivo 2013.* Lima: Instituto Nacional de Estadística e Informática – INEI.

INEI. (2017). *Censos Nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indígenas* [database]. Lima: Instituto Nacional de Estadística e Informática - INEI.

INEI (2018). *Perú: Perfil Sociodemográfico, 2017*. Lima: Instituto Nacional de Estadística e Informática – INEI.

INEI. (2019). *Compendio Estadístico Provincia de Lima 2019.* Lima: Instituto Nacional de Estadística e Informática - INEI.

INEI. (2020). *Mapa de pobreza monetaria provincial y distrital 2018.* Lima: Instituto Nacional de Estadística e Informática - INEI.

Londoño, M., Canavire-Bacarreza, G., Bohórquez, S., & Cuartas, D. (2015). Expansión de la educación superior y sus efectos en matriculación y migración: evidencia de Colombia. *Revista Desarrollo y Sociedad* (75), 317-348. DOI: https://doi.org/10.13043/dys.75.8

Luce, R. (1959). *Individual Choice Behavior: A Theoretical Analysis.* Nueva York: Wiley.

McFadden, D. (1984). Econometric analysis of qualitative response models. *Handbook of Econometrics* (1395–1457). New York.

Ogawa, K., & Iimura, K. (2010). Determinants of Access and Equity in Tertiary Education: The Case of Indonesia. *Excellence in Higher Education*, 3-22.

Olavarría, M., & Allende, C. (2013). Endeudamiento estudiantil y acceso a la educación superior en Chile. *Revista Española de Investigaciones Sociológicas*, 91-112. DOI: doi:10.5477/cis/reis.141.91

Prodan, A., Maxim, E., Manolescu, I., Arustei, C., & Guta, A. (2015). Access to higher education: influences and possible implications. *Procedia Economics and Finance*, 535 – 543.

World Bank. (2018). *World Development Report 2018: Learning to Realize Education's Promise.* Washington, DC: World Bank. DOI: doi:10.1596/978-1-4648-1096-1

APPENDIXES

Appendix 1. Criteria for choosing the nonlinear discrete choice model.

Based on the goodness-of-fit analysis presented in Table 4, the following can be concluded regarding the form of function F(.) in equation (2). First, the model estimated using the logistic specification presents lower values for the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) than the one estimated under the probabilistic specification. Secondly, although very similar, the Pseudo R2 value is higher for the logit model than the probit model. Finally, the value for the Area Under the Curve (AUC) in the logit model is also marginally larger than that of the logit model. Thus, it will be correct to state that the ideal specification for the nonlinear discrete choice model with a binomial independent variable will be the logit model.

Table 4. Goodness-of-fit analysis

Goodness-of-fit	Model	
	Logit	Probit
AIC	168'721,76	168'742,24
BIC	168'975,15	168'995,64
Pseudo R2	0,1152	0,1151
AUC	0,7237	0,7236

Author's elaboration.