doi: http://dx.doi.org/10.19052/ed.4327

Returns to Education and Skill Premiums: Estimation and Biases Associated with the Case of Argentina

María Celeste Gómez

Keywords

Education, Argentina, income, labor market

JEL Codes

C52, I25, J24, J31

Abstract

This article aims to empirically analyze returns to education and skill premiums in employed wage earners in Argentina between 2003 and 2014, under three alternative specifications of the Mincer equations. The study examines the comparative evolution of these returns during the period and identifies biases in the estimates of the proposed income equations. The final objective of this exercise is to decide which alternative is the empirically most appropriate estimate for the case of Argentina in the analysis period. Results show that the Poisson maximum likelihood model, applied to the traditional Mincer approach, generates consistent estimates of returns to the attributes of workers.

How to cite this article: Celeste Gómez, M. (2018). Returns to education and skill premiums: estimation and biases associated with the case of Argentina. *Equidad y Desarrollo* (30), 11-37. doi: http://dx.doi.org/10.19052/ed.4327

Received: June 23, 2017 • Accepted: October 2, 2017

* Product of a research carried out by the author within the framework of a doctoral thesis to obtain a PhD degree in Economic Sciences with a mention in Economics from the Universidad Nacional de Córdoba, Argentina.

** Economist at the Universidad Nacional de Córdoba, Department of Economics and Finance, Faculty of Economic Sciences, Argentina. Recipient of a dissertation write-up grant from the National Council of Scientific and Technical Research (Conicet). Email: mcelestegomez.arg@gmail.com

Retornos a la educación y premios por calificación: estimación y sesgos asociados al caso argentino

Resumen

El objetivo del artículo es analizar empíricamente los retornos a la educación y los premios por calificación en asalariados ocupados de Argentina entre 2003 y 2014, bajo tres especificaciones alternativas de ecuaciones de Mincer. Se examinó la evolución comparada de estos retornos durante el periodo y se identificaron sesgos en las estimaciones de las ecuaciones de ingresos propuestas. El propósito final del ejercicio fue decidir qué alternativa resulta la estimación empíricamente más apropiada para el caso argentino en el periodo de análisis. Los resultados muestran que el modelo de Poisson de máxima verosimilitud aplicado al enfoque de Mincer tradicional genera estimadores consistentes de los retornos a los atributos de los trabajadores.

Palabras clave

Educación, Argentina, ingresos, mercado laboral

Retornos à educação e prêmios por qualificação: estimação e vieses associados ao caso argentino

Resumo

O objetivo do artigo é analisar empiricamente os retornos à educação e os prêmios por qualificação em assalariados ocupados da Argentina entre 2003 e 2014, sob três especificações alternativas de equações de Mincer. Examinou-se a evolução comparada destes retornos durantes o período e identificaram-se vieses nas estimações das equações de renda propostas. O propósito final do exercício foi decidir qual alternativa seria a estimação empiricamente mais apropriada para o caso argentino no período de análise. Os resultados mostram que o modelo de Poisson de máxima verossimilitude aplicado ao enfoque de Mincer tradicional gera estimadores consistentes dos retornos aos atributos dos trabalhadores.

Palavras chave

Educação, Argentina, renda, mercado de trabalho

Introduction

The labor market in Argentina has undergone important changes in the last 15 years, keeping pace with the evolution of macroeconomics in general and the socioeconomic conditions of its population. Between 2003 and 2014, employment rate has increased three percentage points (pp), from 38.6% to 41.6%, while unemployment rate fell by more than eight percentage points. Salary grew by 44% in real terms during the same period. The educational level of the population was also improving at the same time. The population with completed higher education increased from 17.1% to 20.4% of the total population, while the population with completed secondary education was 20% in 2003 and reached 26% in 2014. Conversely, the number of people without education was reduced from 8% to 4.8%. If the employed population is classified based on job requirements (skills and specific knowledge of individuals who perform these jobs), four large occupational groups arise: unqualified or with operator, technical, or professional qualification (National Institute of Statistics and Censuses of the Argentine Republic [Indec], 2001). Under this criterion, in Argentina, jobs requiring operator qualification accounted for around 50% of the total number of jobs throughout the analysis period. At the same time, unqualified jobs fell from 23% to 19%, while those requiring professional and technical qualifications had little variability (from 10% to 9% and from 18% to 19% between 2003 and 2014, respectively).

In a changeable labor market, numerous factors determine the value of salary, regarding not only its average level, but also its distribution. In addition to educational level and job qualification, gender gap and the formality of employment are significant determinants of labor income. In this framework, it is important to get a precise estimation of returns to education and other attributes, in view of potential biases involved in this estimation, given the intrinsic characteristics of the labor market in Argentina (Aliaga & Montoya, 1998). To quantify wage premiums according to the attributes of workers and to detect their biases are the main objectives of this research article. This document has the following structure: The next section summarizes the theoretical framework applicable to the topic studied here. The third section describes data and methodology used in three alternative specifications of the Mincer equation. The fourth section presents and discusses the results of applying these approaches to the case of Argentina between 2003 and 2014. Conclusions are presented in the final section.

Theoretical framework

The literature on mechanisms for determining income in the labor market is abundant and varied, especially regarding labor supply, in relation to education and the level and distribution of salaried income. The model of income determination proposed by Mincer (1974) takes as a starting point the theory of human capital (Becker, 1964) and focuses mainly on years of formal education and experience as determinants of salary.

Mincer finds a positive relationship between years of education and income or salaries, with the greatest influence when the age (or experience) of the individual is included in the relationship as a control variable. Underlying this approach is the assumption that in order to achieve higher levels of income in the labor market, people will continue to invest their time in education (human capital), beyond the expected years of schooling in their formal education system.[1] The connection between labor market and the education of individuals is shown in the Mincer equations via "return rates" or "premiums" to education. A return rate to education is defined as the additional income that one more year of formal education means for the worker, which is a payment made by the employer for greater productivity derived from this investment in human capital. Galassi and Andrada (2011) show how returns guide education demand decisions in individuals.

Nevertheless, a high return for a given educational level raises the demand for education and produces a countervailing effect, which lowers its rate of return. There would be investment in education up to the point where the worker's temporary income profile (receiving progressively higher salaries) matches his cost profile (enrollment, hours of study, material, etc.) (Aliaga & Montoya, 1998).

In his work, Mincer proposes to expand the field of analysis to other attributes associated with the worker. In this area, many authors have mentioned the needs of capital or job requirements as critical factors in determining salary. In this sense, the focus would not be exclusively on the individual, but on the job and the requirements to perform it in a productive and efficient way.

In a study applied to the labor market in Argentina, although without making such a clear distinction between years of education and job qualifications, Aliaga and Montoya (1998) maintain that in a country like Argentina, with a record of

1 For a theoretical analysis of the Mincer approach and other approaches to human capital, see Galassi and Andrada (2011).

high unemployment rates, part of the return to investment in human capital comes from better job placement conditions. Introducing the concept of *employability*, they affirm that it is not just the worker who decides to participate in the labor market, but he or she is also "selected" among others to occupy a job. The connection between labor supply and demand is expressed, among other ways, in the fact that the employer pays higher salaries to workers who are more qualified.

In an analysis of the relationship between education and income inequality, Alejo (2012) attempts to reconcile two interpretations of the hypothesis that considers education as an "inequalizing" factor for income distribution, or *paradox of progress*. On the one hand, there is the hypothesis of convexity of the Mincer equation, which suggests that returns increase with educational level at increasing rates, implying that improvements in general educational levels increase salary inequality. On the other, there are models with differentiated labor income, where physical capital hires workers based on the requirements of its specific capital.

Alejo (2012) argues that the hypothesis of convexity of the Mincer curve has prevailed over income distribution in the 2000s, and identifies a relative shortage between capital requirements and the supply of qualified workers. More explicitly, Paz (2013) takes as a starting point the segmentation of labor markets in Argentina and incorporates the characteristics of the job, as well as the branch of activity of the company where the worker performs his tasks. Casal, Terán and Paz (2016) evaluate the relationship between education and income inequality in Argentina in the last 20 years, similarly incorporating job qualifications and the branch of activity as determinants, and affirm that the fall in returns to education of the most qualified workers was the decisive factor in the improvement of income distribution.

Due to its characteristics, Mincer's income equation is sufficiently flexible, which has allowed that related literature incorporate different versions of it. Argentina has the experience of a historically benefited society with an accessible educational platform and high levels of schooling, although with a history of high unemployment and low labor participation. All this has motivated the present investigation to focus on the biases associated with returns to education and premiums for skilled workers.

In relation to the above, the contribution of Aliaga and Montoya (1998) to the identification of biases in returns is a valuable precedent for Argentina. Using the Heckman two-stage selection model (Tunali, 1986), positive biases are found in the returns of the employed male population from 1990 to 1998. Galassi and Andrada (2011) present an empirical work with more recent sample selection, where they study the return rates to education in the sub-regions of Argentina in 2006. Regarding wage premiums for skills, Moncarz (2012) studies the effects of trade liberalization on these in the 1990s, applying the bias correction model proposed by Tunali (1986).

Data and methodology

Data

This study uses microdata of the permanent household survey (EPHC), in its quarterly continuous form, for the years from 2003 to 2014. The EPHC is a survey about living conditions, and it focuses mainly on the labor market, although it has a good amount of information regarding the composition and characteristics of households, and other features related to the living conditions of the population. The survey is carried out in the main urban centers of the country, and represents almost 62% of the total population of Argentina.

The survey uses a rotating panel scheme, which allows the monitoring of the same household or individual in the short run, during two consecutive years. Each household is surveyed in a total period of six quarters.[2] In order to achieve greater comparability during the 12 years of this study, a prior harmonization of databases was necessary, which involved certain methodological decisions. First, given that the EPHC was expanding its geographic coverage, the last agglomerates incorporated in 2006 were not included in the sample, in order to work with 28 urban centers during the entire analysis period.

Second, the study included a population of both sexes, between 25 and 64 years of age, full-time salaried workers (with 35 or more working hours per week). The dependent variable is monthly hourly income in the main job. As explanatory variables, years of education (completed or not), job qualification (professional, technical, operator, unqualified), age, gender, and formality of employment were included. The sample selection models also incorporated labor market (condition of employment or participation in the economically active population) and

2 For more information on the methodology of the EPHC-Indec, see Piselli (2008).

household composition variables (kinship relations, number of children and their age, quality of housing, etc.).

Third, the study included income at constant prices, using a purchasing power parity scale (Indec, 2002) as regional correction factor of the purchasing power of income between the regions of Argentina.[3] In addition, the regressions performed using ordinary least squares (OLS), *Poisson, and Probit included spatial (geographical regions of Argentina) and temporal control variables (quarters). In order to obtain population projection estimates and to incorporate corrections for non-response for income, the weighting factors included in the same databases were used (unless otherwise indicated).*

Finally, given the sample rotation scheme, the permanence of data of the same household could influence the surveyed conditions (participation in the labor market or employment, income, etc.), which could lead to inaccurate estimates and alter the results. To verify this, a restricted sample was used (without repeated observations year after year) and identical regressions to those of this study were employed. Differences between the estimates of the complete and partial samples were insignificant, which rejects this potential source of data distortion and enables a year-by-year analysis with the complete sample.

All the empirical work used microdata bases collected by the interventive management of the National Institute of Statistics and Censuses of the Argentine Republic (Indec) between 2007 and 2014, which was finished when the mandate of the previous government ended in December 2015. The new administration of the organization has developed a new consumer price index (CPI), with changes in methodology and coverage since June 2016. These databases are under review, evaluation, and recovery operations by the institute.[4]

3 Since 2007, due to the intervention of the national government on the National Institute of Statistics and Censuses of the Argentine Republic (Indec) and the loss of reliability and availability of socioeconomic indicators (mainly the consumer price index [CPI]), this study used the Gran Buenos Aires IPC, published by the Indec until 2006 to deflate revenues, linked since 2007 (with correction for purchasing power) to the IPC of the Province of Salta. The latter showed high correlations with other indexes of not intervened statistics from provincial institutions, in the periods pre- and post-intervention on the Indec.

4 For more information, see the annexed press release "Labor market: main indicators," www.indec.gob.ar.

Methodology and empirical strategy

The Mincer earnings equation (1974) shows the value that the market places (awards) on the observable characteristics of an individual. In addition to education and experience, some studies usually add other determinants, such as job training, gender, geographical region, formality of employment, branch of activity, job qualification, etc. The usual empirical strategy for the average salary level is to use OLS.

In the case of studies on income distribution, methodology can be divided into two major groups: conditional quantile regressions (for the observable characteristics of individuals) or unconditional quantile regressions (via microsimulations that compare the behavior of the conditional structure with the new configuration of characteristics) (Alejo, 2012). The starting point for the three alternative specifications used in this study is summarized in the following income equation:

$$ln\, y_i = \vartheta(k_i, \mathrm{e}_i, w_i) + ln\, u_i \tag{1}$$

Where y_i are the labor earnings of the individual i; k_i is the stock of human capital measured in years of schooling, e_i is potential work experience (approximated by the age variable); w_i are other observed attributes of the individual, and u_i is random error that captures unobserved characteristics of the individual and is distributed as $N(0,1)$. The econometric application of this equation will be performed using three methodologies: OLS, Poisson pseudo-maximum likelihood estimator, and least squares with Heckman's (1979) two-stage correction for selection bias (Tunali, 1986).

Poisson estimation of the Mincer equation

The traditional log-linear OLS model is presented in the following conditional mean equation:

$$E[y_i\, |x] = exp(\mathrm{X}_i\, \beta) \tag{2}$$

Where $y_i \geq 0$ is the hourly earnings of the individual, x_i is the matrix of observed variables, β is the matrix of coefficients, and $E[y_i\, |x] = 0$.

In relation to this approach, Santos Silva and Tenreyro (2006) suggest that in contexts where heteroscedasticity is present, OLS estimation is not the most appropriate method. With a non-constant variance of the error term, its expected value may not be statistically independent from the covariates (and, therefore, not annulled), which leads to an inconsistent and biased estimation of the parameters. Additionally, the log-linear earnings equation requires truncating the sample, and removing null observations on the dependent variable.

Given these difficulties, the authors propose the Poisson regression model (following a maximum likelihood estimation) as an alternative method. However, the use of the Poisson distribution (which is associated with problems with discrete dependent variables and non-negative integer values) is an empirical option, since the maximum likelihood estimator is equivalent to that used in the Poisson equations. Data not necessarily have this type of distribution; all that is required is to specify correctly the conditional mean (as in the equation 2).[5] Unlike the OLS estimation, the Poisson estimation allows obtaining consistent parameters. Additionally, if data have a conditional variance that is exactly proportional to the conditional mean (from which the estimators of interest arise), the proposed estimators, in turn, will be efficient.

Based on these conditions, in this section we use the maximum likelihood estimation in the context of income equations as an alternative to traditional OLS. Prior to this, following Santos Silva and Tenreyro (2006), both estimation alternatives (OLS and Poisson) are tested to verify the consistency and efficiency of their estimators. To test consistency, we used the Ramsey test (1969) or general specification test on the conditional mean of both alternatives, which allows identifying whether the values adjusted using each equation explain the dependent variable. Additionally, to evaluate efficiency in the presence of heteroscedasticity, the Park test (1966) was applied to least squares, and the Gauss-Newton regression (GNR) to the Poisson estimation.

5 See Cameron and Trivedi (2010) for an analysis of discrete variable models.

Two-stage least squares and correction for selection bias

Another source of bias is associated with data collection from individuals. The model introduced by Heckman (1979) argues for the need to identify potential bias in statistical estimations in behavioral equations, since the sample on which these estimations are performed may not be a random selection of the population under study, which would lead to biased and inconsistent estimations. In this study, the sample of employed workers (with positive salaries) has not been randomly selected from the population, since they are employed thanks to having passed a certain job selection process (Aliaga & Montoya, 1998).

Following Tunali (1986), Heckman's traditional approach can be extended to become a double-selection model. An advantage of the model developed by Tunali is that it is possible to differentiate explicitly between the probability of participating in the labor market and the probability of being employed, which is not so clear in the Heckman model, and which is relevant in the labor context of Argentina. In the first case, when incorporating a correction term for participation in the labor market, we consider those people who might not participate in the market for reasons that are far from being random.[6]

In the second case, the correction term for the probability of being employed takes into account the possibility of being selected from those who offer their labor power. The equations that express these probabilities are estimated by Probit:

$p^*_{1i} = \beta_1' x_1 i + \mathrm{u}_{1i}$ labor market participation equation (3)

$p^*_{2i} = \beta_2' x_2 i + \mathrm{u}_{2i}$ employment equation (4)

Where p^*_{1i} and p^*_{2i} are (not observed) probabilities for the individual i to participate in the labor market and to get employed, respectively; β_j are coefficient vectors; x_{ji} are covariable vectors, and u_{ji} are error terms with $u_{ji} \sim \mathrm{N}(0, \in)$. As dependent variables, the conditions participates (*does not participate*) in the labor market and

6 Among other determinants of labor participation, the discouraged worker hypothesis (Beccaria & Orsatti, 1979) indicates that "marginal" workers withdraw from job search if they consider that the labor market situation reduces their employment possibilities. In addition, from the perspective of gender, female participation is conditioned by educational level, family planning, children, and culture (Busso & Romero Fonseca, 2015).

employed (*unemployed*), respectively, are identified. The results of these estimations are summarized in the income equation by means of two correction terms, known as inverse Mills ratios (IMR) (π_{1i} and π_{2i}), which control potential bias for the probability of participating in the labor market and getting a job:

$$y_{3i} = \beta_3' x_3 i + \gamma_1 \pi_{1i} + \gamma_2 \pi_{2i} + \sigma_3 \vartheta_{3i} \qquad (5)$$

x_{3i} is identified as the covariable vector, y_{3i} as the income vector, σ_3 as an unknown scale parameter, and ϑ_3 as the error term. In this equation, the significance of the coefficients γ_1 and γ_2 will be evaluated to identify significant potentials in each of the analyzed years.[7]

The selection bias model is based on the specifications proposed by Aliaga and Montoya (1998).[8] The determinants in the participation equation are years of education, age, number of children, condition of head of household, gender, inadequate housing, and a measure of *wealth effect* (difference between individual income and family incomes). In the employment equation, the determinants are years of education, age, children, gender, and being head of household. Based on this proposal, we will compare estimates of the OLS model and estimates of Heckman's two-stage correction model.

Results and discussion

To identify the sign and degree of association between the proposed determinant variables and hourly income, Spearman's correlation coefficients were calculated independently. The results are shown in Table 1. Correlations have the expected signs and magnitudes, which are significant at 1% in all cases. Years of education, job qualification, and age are positively associated with hourly income. Similarly, year after year, correlations maintain their signs and statistical significance.

7 It should be mentioned that the non-correction of the variance-covariance matrix of coefficients and standard errors could lead to inconsistent estimations of correlation coefficients in the equations of the model (Moncarz, 2012). Although standard errors were not corrected in this work, it is left for further studies to apply an alternative correction proposed by Newey and McFadden (1994).

8 Alternately, we used specifications based on Moncarz (2012), although these did not have significant results in each year of the sample. They are available upon request.

Regarding the two central explanatory variables, years of education and job qualification, these maintain the same pattern of distribution over the years.

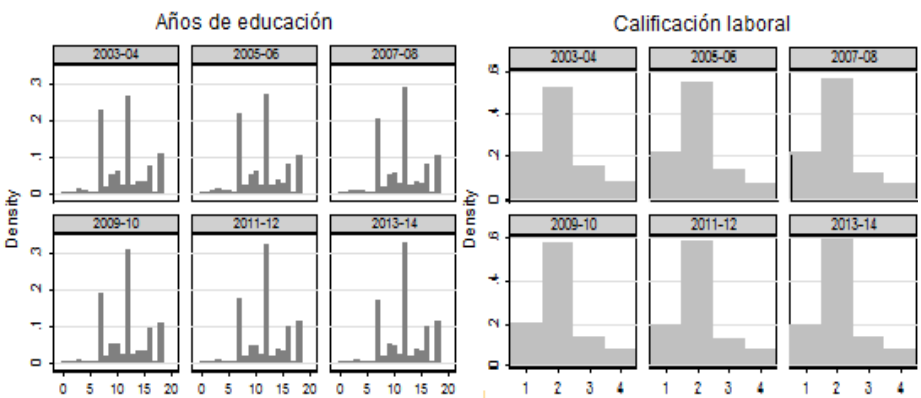
Table 1. Spearman's rank correlation between the main covariables and hourly income for the main occupation. Argentina 2003-2014

	Job qualification	Years of education	Age
Spearman's correlation (hourly income)	0.4399***	0.0581***	0.1267***

Note: Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Figure 1 indicates the distribution of workers according to their years of education and job qualification. In the case of education, distribution shows multiple modes; the three main ones are located, in order of frequency, at 12, 7, and 18 years. These periods coincide with the completion of three main levels defined in the current educational model (primary, secondary, and university level).

Figure 1. Wage earners according to years of education and job qualifications. Argentina 2003-2014



Note: Up to 18 years of education.
Note: 1 = unqualified; 2 = operator; 3 = technical; 4 = professional.
Source: Author's elaboration based on EPHC-Indec.

It should be mentioned that, for the construction of this variable, we took into account the impact of the reforms of the Argentine education system, which was modified twice in the last 20 years, with changes associated mainly with the organization of the levels and contents of school curricula. Although the study of the impact of these reforms on labor market variables widely exceeds the objectives of this work, Alzúa, Gasparini and Haimovich (2010) report that the Federal Education Law had a positive but moderate impact on the variables of employment and labor income, and zero impact when the analysis focuses on young people in condition of poverty.

Regarding job qualifications, the operator category had a floor of 54.2% of the jobs in the biennium 2003-2004, and reached its highest records at the end of the analyzed period (2013-2014) with 58.6%. On the other side, unqualified positions were relatively reduced over the years, starting with 20.8% of the total jobs in 2003-2004 and reaching 18.7% in 2014. Jobs requiring technical qualification were relatively stable in number (less than 1% of negative difference between the first and last biennium), while professional jobs were reduced by 1.4 percentage point. In the case of job qualifications, operator positions had the highest participation at the expense of unqualified jobs and professional positions.

Ordinary least squares and Poisson

When analyzing the results of the OLS estimation, Table 2 shows that the coefficients for years of education have the expected signs and are statistically different from zero, with slightly decreasing tendencies from 2003 to 2014.[9] The returns to education reported 4.9% in 2003, and they were increasing until 2006. From 2006 to 2014, the returns were reduced, and had a floor of 4% in 2013 and 4.7% in 2014.[10] This phenomenon is verified in spite of an increase in the educational level of workers, which can be interpreted as an excess of supply of qualified workers that reduced their premium in the last eight years.

9 Complete regression tables with the variables of interest and control variables are available from the author upon request.

10 The impact of discrete variables is calculated by the rule $(e^{bi}-1)*100$, where *bi* is the coefficient in tables.

Table 2. Hourly income equations by OLS. Premiums for the attributes of interest. Argentina 2003-2014

Var. Dep.												
lnrealyh_p21	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Years of education	0.0474***	0.0498***	0.0502***	0.0515***	0.0483***	0.0453***	0.0442***	0.0456***	0.0458***	0.0420***	0.0393***	0.0457***
	(0.0024)	(0.0016)	(0.0014)	(0.0013)	(0.0016)	(0.0014)	(0.0013)	(0.0013)	(0.0014)	(0.0014)	(0.0014)	(0.0012)
Operator	0.1350***	0.1500***	0.1670***	0.1810***	0.1990***	0.1830***	0.2020***	0.1690***	0.1790***	0.2010***	0.2090***	0.1940***
	(0.0202)	(0.0126)	(0.0113)	(0.0112)	(0.0129)	(0.0111)	(0.0109)	(0.0110)	(0.0110)	(0.0116)	(0.0111)	(0.0099)
Technical	0.4430***	0.3510***	0.3520***	0.3460***	0.3940***	0.3630***	0.3440***	0.3500***	0.3180***	0.3330***	0.3280***	0.3170***
	(0.0276)	(0.0183)	(0.0164)	(0.0158)	(0.0184)	(0.0159)	(0.0152)	(0.0156)	(0.0160)	(0.0157)	(0.0159)	(0.0141)
Professional	0.7020***	0.6570***	0.6210***	0.5720***	0.6350***	0.5930***	0.5970***	0.5510***	0.5190***	0.5270***	0.5340***	0.4840***
	(0.0371)	(0.0234)	(0.0219)	(0.0210)	(0.0235)	(0.0203)	(0.0191)	(0.0193)	(0.0198)	(0.0204)	(0.0190)	(0.0173)
Observations	10.923	23.918	25.913	31.380	26.291	35.487	34.442	34.564	34.846	33.551	32.991	35.238
R-squared	0.4340	0.4470	0.4560	0.4680	0.4530	0.4380	0.4400	0.4280	0.4050	0.3980	0.3850	0.3990

Note: Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Source: Author's elaboration based on EPH-Indec.

On the other hand, the coefficients for labor qualifications have the expected signs, are statistically significant and increasing with the level of qualification (operator, technical, and professional). During the period, the premium was raised for those with operator qualification (in relation to the salary that an unqualified job would provide); while in 2003, on average, a 14.7% higher salary was reported for positions in this rating, this differential reached 21.5% in 2014.

Positions requiring technical and professional qualifications saw their premiums reduced during the period (to a greater degree in the latter case). Premiums for technical qualification went from 55.4% to 37.6% and premiums for professional skills were reduced from 100% to 62.9% between 2003 and 2014. Given that in the period only the positions with operator qualification were on a path of growth, falls in premiums for technical and professional positions allow observing a smaller gap between the returns to the latter and the former. Among these qualification categories, we performed an F-test for means for each year using the bootstrap technique with 400 replications to identify whether skill premiums are significantly different from each other. Returns to professional as well as to technical and operator qualifications, and differences between their levels were all statistically significant at 1%.

Data in Table 3 show the coefficients of the income equations estimated by Poisson regression for years of education and job qualification. It can be observed that in the Poisson estimation the returns to education are lower than in the OLS estimates. Although their evolution over time is similar (with lower values in 2014 than in 2003), the Poisson version shows slightly higher oscillations. Its maximum return to education is registered in 2004 (with a premium of 6.4% for one more year of education), a year after which this measure begins to fall. With respect to skill premiums, both in the Poisson and the OLS model, coefficients are increasing with the level of qualification, with statistically significant differences at 1% between different qualification levels by the bootstrap method.

Table 3. Poisson equations for hourly income. Premiums for the attributes of interest. Argentina 2003-2014

realyh_p21	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Years of education	0.0573***	0.0617***	0.0606***	0.0599***	0.0605***	0.0519***	0.0483***	0.0509***	0.0546***	0.0481***	0.0452***	0.0479***
	(0.0037)	(0.0026)	(0.0021)	(0.0016)	(0.0056)	(0.0016)	(0.0015)	(0.0017)	(0.0017)	(0.0016)	(0.0021)	(0.0014)
Operator	0.1320***	0.1730***	0.1860***	0.1840***	0.2110***	0.1800***	0.2120***	0.1770***	0.1920***	0.2050***	0.2040***	0.2140***
	(0.0259)	(0.0165)	(0.0120)	(0.0129)	(0.0137)	(0.0129)	(0.0111)	(0.0120)	(0.0119)	(0.0116)	(0.0117)	(0.0101)
Technical	0.4560***	0.3730***	0.3500***	0.3380***	0.3620***	0.3620***	0.3430***	0.3470***	0.3140***	0.3090***	0.3370***	0.3450***
	(0.0328)	(0.0249)	(0.0185)	(0.0181)	(0.0373)	(0.0185)	(0.0163)	(0.0174)	(0.0172)	(0.0160)	(0.0222)	(0.0146)
Professional	0.6990***	0.6840***	0.6360***	0.5600***	0.5840***	0.5750***	0.6030***	0.5370***	0.509***	0.5120***	0.5030***	0.5050***
	(0.0427)	(0.0353)	(0.0264)	(0.0235)	(0.0514)	(0.0228)	(0.0200)	(0.0226)	(0.0223)	(0.0239)	(0.0207)	(0.0193)
Observations	11.491	24.753	26.566	31.982	26.826	36.143	34.952	35.097	35.353	33.980	33.450	35.677

Note: Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Source: Author's elaboration based on EPH-Indec.

In the case of education, in all the years, the pseudo-maximum likelihood estimator shows higher premiums for schooling than the least squares estimates. Premiums for each qualification level do not show as clear a difference between the two alternatives (OLS or Poisson) as those derived from returns to education. Until 2006, skill premiums estimated by Poisson regression are greater than those estimated by OLS, and later this relationship is reversed (at least for premiums for technical and professional qualifications). The year 2007 is the point of greatest differentiation, although in the rest of the period, the paths of premiums overlap.

In order to identify whether these differences between return estimates are statistically significant, we conducted a test for differences between the coefficients of these two models.[11] This showed that premium for schooling has statistically significant differences at 1% between the least squares and Poisson estimates in all years of the studied period. Regarding skill premiums, in seven of the twelve estimated years, significant differences were found for the operator qualification in both models, although with a different level of significance (of 10% as a maximum). For the technical qualification, five of the twelve annual periods registered significant differences between the two alternatives, and for the professional level, it was only possible to determine the significance of differences in six of the twelve years. Based on these results, we can affirm that there are significant differences in the estimation of different skill premiums using one or the other model.

Additionally, coefficients were evaluated in time (for each model separately) in order to identify whether the returns to education showed a significant change or maintained their values throughout the studied period. The results of the coefficient test at the beginning and end of the period (2003-2014) show a statistically significant fall at the end of the period, a result that is present in both estimation alternatives. An analysis of the subperiods shows that returns fell significantly between 2003 and 2009 (also at 1%), but between 2009 and 2014 they did not register differences. This result is verified in both specifications.

The test results of skill premiums applied at the beginning and end of the period show significant differences at 1% between 2003 and 2014, for all the qualifications in both estimation alternatives. In case of the operator qualification, there is a significant increase in its returns, opposite to what happens with technical and

11 The Seemingly Unrelated Estimation Test acts like a generalized Hausman specification test, and even overcomes some of its limitations.

professional qualifications, which "rewarded" the worker in a smaller proportion in 2014 than in 2003.

Correct specification test and heteroscedasticity pattern

To verify whether the OLS estimation leads to consistent estimates, and whether the pseudo-maximum likelihood model adequately calculates the conditional mean, this section will report the results of applying the tests proposed by Santos Silva and Tenreyro (2006). The results are shown in Table 4.

First, and aiming to test the correct specification of both models, an additional regressor—square of ($x'b$) of the original equation—was added to the respective equations. The first panel of Table 4 shows the impact of applying the Reset test. In ten of the estimated twelve years, the OLS model passed the test for correct specification, although the alternative multiplicative model (Poisson) showed a slight advantage, since throughout the studied period specification passed the additional regression test. Second, to verify heteroscedasticity pattern, the Park test was performed on the OLS model and a GNR test on the Poisson model.[12]

For the OLS model, two variables related to human capital (years of education and age) were tested, seeking to identify potential sources of heteroscedasticity. For years of education, heteroscedasticity was present in four of the estimated twelve years. The results are confirmed with the age test in all years of this study. It is concluded, then, that the OLS model is not valid due to the presence of heteroscedasticity in the errors. On the other hand, the GNR test on the Poisson estimates identifies a non-proportional variance that resulted significant in all years of this study (indicated by the lambda1 parameter). In this model, the Poisson estimator is not efficient, although still consistent (given the results of the previous test).

It is important to highlight that change in the estimation methodology has a significant impact on the value of returns to education, which does not apply to the case of skill premiums. Based on the proposal of Santos Silva and Tenreyro (2006), biases in returns are verified in the OLS model in some years of this estimation. These biases underestimate the premium for schooling. With better results, the multiplicative version proposed by Poisson correctly estimates the conditional

12 The Gauss-Newton regression (GNR) test seeks to identify whether the heteroscedasticity pattern shows equidispersion (variance proportional to the mean).

Table 4. Correct specification test and heteroscedasticity pattern

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
RESET Test (Ramsey) OLS and Poisson												
Add. OLS regressor	0.0212	0.0228	-0.0256	-0.0202	-0.0320	-0.0566**	-0.0353	-0.0516**	0.0221	-0.0255	-0.0117	0.0100
(x'b)^2	(0.0376)	(0.0251)	(0.0226)	(0.0209)	(0.0240)	(0.0250)	-(0.0234)	-(0.0255)	-(0.0278)	-(0.0316)	-(0.0318)	-(0.0268)
Add. Poisson regressor	0.0868	0.0902	-0.0085	-0.00345	0.0250	-0.0349	-0.0033	-0.0467	0.0494	-0.0322	-0.0327	0.0490
(x'b)^2	(0.0699)	(0.0718)	(0.0320)	(0.0278)	(0.0360)	(0.0299)	(0.0276)	(0.0314)	(0.0382)	(0.0489)	(0.0620)	(0.0342)
OLS - Park												
Years of education	0.1850*	0.1230***	0.1860***	0.1350***	0.0800	0.0345	0.0043	0.0484	0.0090	0.0793	-0.0657	-0.0896*
	(0.0948)	(0.0473)	(0.0462)	(0.0437)	(0.0562)	(0.0444)	(0.0496)	(0.0500)	(0.0483)	(0.0521)	(0.0543)	(0.0461)
Age	0.4460***	0.2630***	0.4690***	0.5020***	0.3110***	0.4610***	0.4670***	0.4180***	0.3610***	0.3590***	0.4020***	0.3260***
	(0.1160)	(0.0782)	(0.0739)	(0.0711)	(0.0802)	(0.0706)	(0.0723)	(0.0754)	(0.0709)	(0.0749)	(0.0763)	(0.0680)
Poisson - GNR												
Parám Lambda1	37.830**	49.770*	31.460***	34.400***	58.210**	28.980***	30.000***	29.150***	32.460***	29.480***	28.390***	26.530***
	(15.4200)	(27.0200)	(3.1570)	(2.4030)	(25.3300)	(1.7800)	(1.3110)	(1.3290)	(2.6460)	(1.6900)	(2.5100)	(2.0970)

Note: Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Source: Author's elaboration based on EPH-Indec.

mean and, although it is not the option with the lowest variance (due to which there are other more efficient estimations), it allows the consistent estimation of the returns to the observable attributes of individuals during the study period.

Least squares and two-stage correction models for selection bias

In this case, a two-stage sample selection model was adapted based on the specification given by Aliaga and Montoya (1998). The results of sample correction are summarized in the follwoing terms: IMR1 or term that corrects for participation in the labor market, and IMR2 or term that corrects for employment condition (Table 5). Additionally, the bias can be computed as percentage difference between the returns of the uncorrected least squares model and the model that incorporates Heckman's two-step correction (Aliaga & Montoya, 1998).

Table 5 shows that both model-correction terms, IMR1 and IMR2, are significantly different from zero in all the estimated years, which confirms the need to control for the probability of labor market participation and employment in the estimations. Regarding the rates of return to education, a positive bias was found in the OLS model, which indicates that it tends to overestimate returns to the aforementioned attributes. Although this bias does not have a monotonous evolution during the period (it grows until 2009, then falls again until 2011, and it recovers in 2014), it is always above 10%. Skill premiums show a different result, with no significant bias between the traditional OLS and the Heckman correction model. As with the alternative Poission estimation, the change of methodology in the correction model for selection bias effectively affects the value of returns to education, but not skill premiums.

These differences lead us to analyze how these sample bias corrections differentially affect returns to the same attributes for men and women. Therefore, and according to a good part of the literature about sample selection biases, the Heckman model was estimated for each gender separately. For reasons of space, estimations are not included here, but biases derived from gender are identified. Regarding correction factors, the term IMR1, participation in the labor market, and the term IMR2, condition of employment, resulted significant at 1% in both gender specifications and throughout the entire period of 2003-2014, which indicates the need to control for these labor market conditions in the estimations.

Table 5. Income equations with Heckman correction. Premiums to the attributes of interest and indicators of bias. Argentina 2003-2014

Vble. Dep.:												
lnrealyh_p21	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Years of education	0.0386***	0.0433***	0.0436***	0.0454***	0.0409***	0.0381***	0.0353***	0.0387***	0.0404***	0.0355***	0.0338***	0.0382***
	(0.0027)	(0.0017)	(0.0016)	(0.0015)	(0.0019)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0016)
Operator	0.1380***	0.1480***	0.1700***	0.1800***	0.1990***	0.1830***	0.2020***	0.1710***	0.1820***	0.2030***	0.2060***	0.1950***
	(0.0200)	(0.0125)	(0.0113)	(0.0112)	(0.0128)	(0.0111)	(0.0108)	(0.0111)	(0.0110)	(0.0118)	(0.0112)	(0.0100)
Technical	0.4410***	0.3440***	0.3490***	0.3440***	0.3930***	0.3650***	0.3440***	0.3500***	0.3220***	0.3350***	0.3270***	0.3190***
	(0.0271)	(0.0180)	(0.0163)	(0.0157)	(0.0184)	(0.0158)	(0.0150)	(0.0156)	(0.0160)	(0.0157)	(0.0158)	(0.0140)
Professional	0.6980***	0.6480***	0.6190***	0.5720***	0.6340***	0.5960***	0.5970***	0.5520***	0.5270***	0.5310***	0.5370***	0.4880***
	(0.0366)	(0.0233)	(0.0218)	(0.0208)	(0.0234)	(0.0202)	(0.0189)	(0.0192)	(0.0197)	(0.0203)	(0.0189)	(0.0172)
IMR1	1.0470***	1.2740***	1.2060***	0.9990***	0.8280***	0.8890***	1.3610***	0.9100***	0.6490***	0.7880***	0.4240***	0.8820***
	(0.2010)	(0.1220)	(0.1480)	(0.1360)	(0.1570)	(0.1760)	(0.1610)	(0.1260)	(0.1510)	(0.1440)	(0.1620)	(0.1430)
IMR2	-1.1470***	-1.2690***	-1.2500***	-1.0670***	-0.9990***	-1.023***	-1.4480***	-1.0270***	-0.7080***	-0.8760***	-0.5330***	-0.9780***
	(0.1780)	(0.1130)	(0.1410)	(0.1340)	(0.1620)	(0.1690)	(0.1590)	(0.1300)	(0.1500)	(0.1420)	(0.1630)	(0.1430)
Constant	0.0915	-0.2630**	0.0025	-0.0920	0.0621	0.2140*	0.3120***	0.1420	-0.0532	0.0749	0.1160	0,.520
	(0.1730)	(0.1120)	(0.1100)	(0.1000)	(0.1280)	(0.1110)	(0.1130)	(0.1110)	(0.1150)	(0.1150)	(0.1180)	(0.1070)
Observations	10.923	23.918	25.913	31.380	26.291	35.487	34.442	34.564	34.846	33.551	32.980	35.238
R-squared	0.4400	0.4530	0.4600	0.4700	0.4550	0.4400	0.4440	0.4300	0.4050	0.3990	0.3850	0.3990

Note: Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Source: Author's elaboration based on EPH-Indec.

Additionally, the bias in returns to education is positive in all the analyzed years, both for men and women, with a slightly higher value in the latter case. Again, with respect to job qualifications, it is not possible to conclude about the existence of biases of magnitude in any of the coefficients (neither for gender nor for skill type).

Comparative analysis. The three Mincer estimations

The main results of estimating the returns to education and to job qualification using Mincer equations and their application to data from Argentina (2003-2014) can be summarized graphically. Figure 2 shows the evolution of returns to education and skill premiums for the proposed models: OLS, Heckman's model, and Poisson. For the years of formal education, both least squares estimations show lower returns than those proposed by the Poisson estimation. In terms of the evolution of these returns, the long-term trend of decreasing education premiums is repeated. Due to year-to-year fluctuations in the Heckman correction model, the Poisson version is more adapted than the OLS estimation, which reflects a difference between the latter and the Heckman model, clearly manifested in the levels.

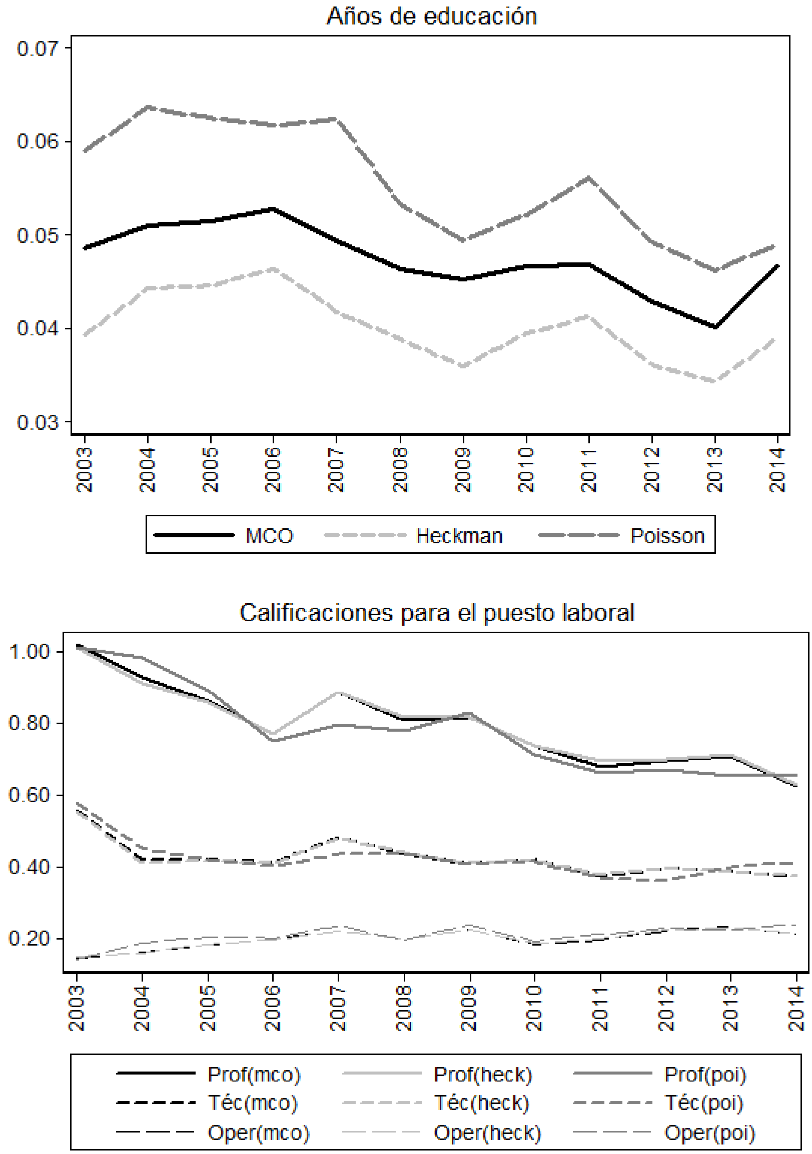
Regarding job qualifications, the return to these attributes does not differ substantially among the three alternatives. In all three models, professional and technical qualifications continue to decrease throughout the period, while operator qualification increases. The choice of a particular estimation for the Mincer equation does not seem to depend on the values of these premiums.

Conclusions

Returns to education and other premiums for the attributes of wage earners have a strong impact not only on the level of average incomes, but also on their distributions, which reinforces the need to eliminate any distortion that may affect the consistency of the estimations.

This work contributes to quantifying alternatives to the traditional OLS model, which would focus on eliminating potential biases in the value of returns to these attributes. The estimates showed higher returns to education in the Poisson model, followed in value by the coefficients of the least squares model, and, finally, by the

Figure 2. Returns to education and skill premiums. Three estimation alternatives
Wage earners Argentina 2003-2014



Source: Author's elaboration based on EPHC-Indec.

Heckman correction model. Regarding skill premiums, no differences of magnitude were found between the proposed specifications.

The two alternatives to the OLS model take as a starting point the inconsistencies of the log-linearized estimation from two different approaches. The Poisson alternative focuses on the heteroscedasticity of data as a source that would invalidate the application of the OLS model. From a more empirical perspective, the two-stage Heckman model works with distortions caused by estimations based on non-random sampling.

Beyond their starting point, both alternative specifications to the Mincer equations are a step toward the correct estimation of coefficients with respect to the basic least squares model. Nevertheless, due to its purely empirical nature, the significance of the sample bias correction model is highly dependent on the particular specifications proposed. On the contrary, the paper shows the correct specification of the Poisson model for estimating the conditional mean. Heckman's estimations and the extension of his analysis to each gender separately suggest that, in sample bias analysis, differential labor market conditions for female and male workers must be taken into account separately.

Although these reflections are based on a particular application of these three models to the Argentine case in the period of 2003-2014, they consider the utility of using the Poisson distribution for Mincer equations, especially if research in this field aims to analyze core values. For the treatment of income data, which are heteroscedastic by nature, distributive methods (either using quantile regressions or by means of indicators that capture the totality of income distributions) are good alternatives for analysis.

References

Alejo, J. (2012). *Educación y desigualdad: una metodología de descomposición basada en dos interpretaciones de la ecuación de Mincer. Evidencia para Argentina*. Anales XLVII Reunión AAEP. Retrieved from http://www.aaep.org.ar/anales/works/works2012/Alejo.pdf

Aliaga, R. & Montoya, S. (1998). Tasas de retorno a la inversión en capital humano: Argentina 1990-1998. *Revista Estudios*, *21*(86), 95-117.

Alzúa, M. L., Gasparini, L. & Haimovich, F. (2010). Educational reform and labor market outcomes: The case of Argentina's Ley Federal de Educación. *Doc. de Trabajo*, (111). Recuperado el 13 de junio de 2017 de http://cedlas.econo.unlp.edu.ar/esp/documentos-de-trabajo.php.

Beccaria, L. & Orsatti, A. (1979). Sobre el tamaño del desempleo oculto en el mercado de trabajo urbano de la Argentina. *Desarrollo Económico*, *19*(74), 251-267.

Becker, G. S. (1964). *Human capital theory*. Columbia, New York: National Bureau of Economic Research.

Busso, M. & Romero Fonseca, D. (2015). Female labor force participation in Latin America: Patterns and explanations. *Documentos de Trabajo del CEDLAS*, (187). Retrieved from http://cedlas.econo.unlp.edu.ar/esp/documentos-de-trabajo.php

Cameron, A. C. & Trivedi, P. K. (2010). *Microeconometrics using stata* (vol. 2). College Station, TX: Stata Press.

Casal, M., Terán, C. & Paz, J. (2016). Educación y desigualdad: evolución en Argentina en los últimos 20 años (1995-2015). *Anales LI Reunión AAEP*. Retrieved from http://www.aaep.org.ar/anales/works/works2016/casal.pdf

Galassi, G. & Andrada, M. (2011). Relación entre educación e ingresos en las regiones geográficas de Argentina. *Papeles de Población*, (69), 257-290. Retrieved from http://www.scielo.org.mx/pdf/pp/v17n69/v17n69a9.pdf

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, *47*(1), 153-161.

Instituto Nacional de Estadística y Censos de la República Argentina (Indec). (2001). Encuesta permanente de hogares. En *Clasificador Nacional de Ocupaciones-CON versión 2001*. Buenos Aires: autor.

Instituto Nacional de Estadística y Censos de la República Argentina (Indec). (2002). Dirección de índices de precios de consumo. En *Paridades de poder de compra del consumidor (PPCC)*. Buenos Aires: autor.

Mincer, J. (1974). Schooling, experience and earnings. *Human Behavior & Social Institutions*, (1-2), 71-93.

Moncarz, P. E. (2012). Trade liberalization and wage premium in Argentina: The role of trade factor intensity. *The Developing Economies*, *50*(1), 40-67. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1746-1049.2011.00154.x/full.

Newey, W. K. & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, *4*, 2111-2245.

Park, R. E. (1966). Estimation with heteroscedastic error terms. *Econometrica (pre-1986)*, *34*(4), 888.

Paz, J. (2013). Segmentación del mercado de trabajo en la Argentina. *Desarrollo y Sociedad* (72), 105-156. Retrieved from http://www.redalyc.org/articulo.oa?id=169129783004

Piselli, C. (2008). La encuesta permanente de hogares: fuente de datos socioeconómicos de Argentina. Instituto de Investigaciones Económicas, Facultad de Ciencias Económicas Jurídicas y Sociales, UNSa. *Reunión de Discusión*, (184). Retrieved from http://www.economicas.unsa.edu.ar/iie/Archivos/RD184.pdf

Santos Silva, J. y Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4), 641-658. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/rest.88.4.641

Tunali, I. (1986). A general structure for models of double-selection and an application to a joint migration/earnings process with remigration. *Research in Labor Economics*, *8* (Part B), 235-282.