CUBAN IMMIGRANTS IN THE UNITED STATES: WHAT DETERMINES THEIR EARNINGS DISTRIBUTION?

Aleida Cobas-Valdés[1]

In terms of education level, Cuban immigrants positively self-select in their migration decision, that is, the people with the highest levels of education are those who tend to migrate. In general, migrants have more years of study than the mean years of study of workers in Cuba (Cobas and Fernández, 2014) and therefore they are workers with extensive skills in their source country. Since it is the most highly skilled people who migrate, it is undoubtedly of interest to analyze the whole distribution of earnings and quantify the effect of socioeconomics variables in different location of the distribution of earnings.

The United States has been the main destination for migrants from Cuba, and other Latin American countries. The U.S. Census Bureau reports that in 2013, 54 million people in the United States (17% of the population) were of Hispanic origin (U.S. Census Bureau, 2013); among Hispanic-origin groups, Cubans represented 3.7%. It is well known that an important proportion of the labor force in the U.S. is made up of immigrants. In 2014, 16% of the 146.3 million employed people in the U.S. comprised persons of Hispanic or Latino ethnicity (U.S. Bureau of Labor Statistics, 2015), with Cubans making up 4%, making them the third largest group of Hispanic workers in the U.S.

Upon arrival in the U.S., immigrants can be expected to be at an earnings disadvantage with respect to natives because they lack certain skills and information that natives have (Friedberg, 1992). Over time, they may increase their income as they improve their English language skills and adapt to the specific skill sets of the country. In terms of benefits to the host country, migration increases its production capacity and technological capabilities, and from an economic point of view entails no significant costs in terms of social services, which would be required if the majority of those who migrated had less education (Cuecuecha, 2005). Moreover, the migration of higher educated individuals has a negative impact on countries of birth since most educational investment on migrants is lost (Aupetit and Gérard, 2009).

This paper describes the earnings of Cuban immigrants in the U.S. using Ordinary Least Squares (OLS) and Quantile Regression (QR). QR is a method for estimating the relationship between a response variable and a set of explanatory variables for the whole conditional probability distribution of the response variable. The explanatory variables considered are those deemed as most important in the literature: years of education, age on arrival in the U.S., potential job experience, sex, marital status, ethnicity, citizenship status and proficiency in English. The applications of QR range from the field of education (Rangvid, 2007) to biostatistics (Terry et al., 2007) and economics (Chamarbagwala, 2010).

Some researchers, using cross-section models, have tended to confuse the true assimilation of immi-

1. Editor's Note: This essay was awarded First Prize in the 2016 ASCE Student Competition for graduate students.

grants in the U.S. labor market. Chiswick (1978) found that the wage of Cubans immigrants in the U.S. who arrived in 1965–1969 increased by about 37% within the first 10 years after immigration. This could lead to the conclusion that the earnings of immigrants grow rapidly, and this rapid growth enables immigrants to overtake the earnings of native workers within 10–15 years of immigration. Borjas (1985) compared different cohorts, at the same points of their U.S. life cycle, using cross-section models within immigrant cohorts, and found that for the same period, 1965–1969, the earnings of Cuban immigrants decreased by about 25%.

Some recent studies about earnings distribution of Latin immigrants in the U.S. reveal that they face considerable disadvantages with respect to native-born persons. Chiswick et al. (2008) found that immigrants from non-English speaking countries had mean hourly earnings of around 12% lower than those of native born workers in the U.S.; Elliott and Lindley (2008) concluded that differences in human capital endowment and socioeconomic characteristics explained some of the lack of earnings assimilation; Hunt (2012) concluded that non-English-speaking immigrants earned less than their native counterparts. Borjas (2015) found that Cuban immigrants who arrived in the United States between 1995 and 1999 experienced a wage decrease of 4.4% in their first 10 years in the U.S.

The contribution of this article is twofold. Firstly, it uses the QR method to estimate the effects of different socioeconomic characteristics on the conditional probability distribution of the earnings of Cuban immigrants in the U.S. to analyze the distribution of earnings of Cuban immigrants in the U.S. and quantify the effect of this socioeconomic variables. Secondly, it illustrates the differences in the impact of these characteristics between higher and lower income workers. Taking into account that a QR model proposes different regression lines for the different quantiles of the earnings distribution, the contribution of the socioeconomics characteristics on different levels of earnings can be compared.

METHODOLOGY

Many of the issues that social researchers are currently analyzing are related to the values of variables of interest located at the tails of the distribution of such variables of interest. The QR method proposed in this paper measures the effect of the explanatory variables at various points of the (log) hourly earnings distribution.

A robust OLS estimation on Equation (1) employing the heteroskedasticity robust HC4 estimator of Cribari-Neto, 2004 was initially carried out. This estimator improves sample performance when the error term is independent but heteroskedastic, especially in the presence of influential observations (Zeileis, 2004):

$$\mathbf{w_i} = \mathbf{x_i'\beta} + \boldsymbol{\varepsilon_i} \qquad \mathbf{i = 1, \ldots, n} \tag{1}$$

where w_i is the logarithm of gross hourly earnings for individual i, using total pre-tax wage and salary income (expressed in current currency units, e.g.,dollars), that is money received as earnings by an employee for the previous year; x_i is a vector of socioeconomics characteristics of Cuban immigrants in the U.S. including an intercept; ε_i is the error term; and $\beta \in \mathbb{R}^m$ is a vector of unknown parameters.

Next, the proposed model in Equation (1) under the conditional QR is estimated. For any t v (0,1) a linear quantile regression model can be written as:

$$\mathbf{w_i} = \mathbf{x_i'}\,\boldsymbol{\beta_{\tau i}} + \boldsymbol{\varepsilon_{\tau i}} \tag{2}$$

The quantile function $\mathbf{Q_{wi}(\tau|x_i)}$ of the response variable $\mathbf{w_i}$ conditional on covariate vector $\mathbf{x_i}$ at a given quantile parameter$\boldsymbol{\tau}$ is given by

$$\mathbf{Q_{wi}(\tau|x_i)} = \mathbf{x_i'\boldsymbol{\beta}_\tau} \tag{3}$$

No specific assumptions are made for the error term, apart from $\varepsilon_{\tau i}$ and $\varepsilon_{\tau j}$ being independent for $i \neq j$ and to consider that the distribution function at 0 is τ.

Since all continuous explanatory variables are centered at their median value, we use the centercept concept (Wainer, 2000), which has a more meaning-

Table 1. Statistics of covariates: Mean and Standard Deviation

Variable	Description	Mean	Standard Deviation
Woman	Dummy variable: 1 if woman	0.4264	0.4946
Black	Dummy variable: 1 if black	0.0309	0.1730
Married	Dummy variable: 1 if married	0.6283	0.4833
AmericanCitizen	Dummy variable: 1 if an American citizen	0.4461	0.4971
EnglishProficiency	Dummy variable: 1 if proficiency in English	0.5314	0.4990
AgeImm	Age of the individual (in years) at time of migration	29.81	8.2119
YearsEducation	Years of Education of the individual	12.46	2.9781
Experience	Potential Experience of the individual[a]	27.43	11.5426
Log Hourly Earnings	Response Variable	2.6219	0.7724

Source: Computed by the author from 2011 American Community Survey (ACS) in the U.S. provided by IPUMS (2011).

a. Potential Experience of the individuals = Age-Years of Education-6

ful economic interpretation than the usual intercept because it provides the value of the (log) hourly earnings for the median individual.

EMPIRICAL RESULTS

The data used in this paper come from the random sample of 1% of the 2011 American Community Survey (ACS) provided by Integrated Public Use Microdata Series (IPUMS, 2011). We include only individuals who entered the U.S. at the age of 17 years or over. This approach is intended to exclude people who completed their education in the U.S (Lowell et al., 2008). In addition, only individuals who entered the U.S. when under 50 years old are considered, since the assumption is that the group aged between 16 and 49 years consists of those most likely to migrate for economic reasons (Bertoli et al., 2013). Only individuals in work and those aged between 25–64 at the time of the survey are considered.

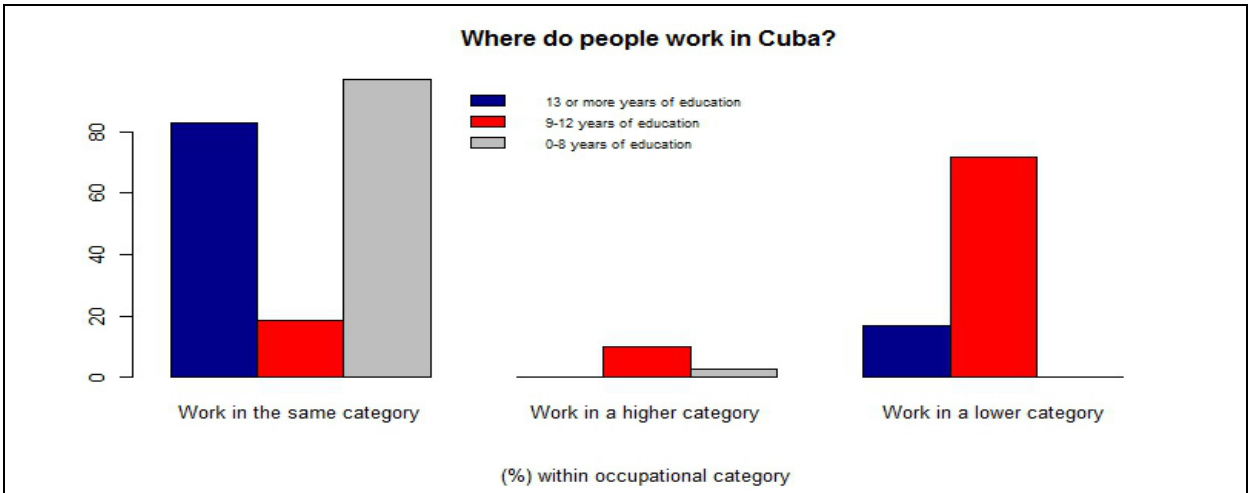
The sample consists of 19,079 observations. Table 1 gives a summary of the variables considered in the estimation models for Cuban immigrants in the U.S. As can be seen, 43% of the persons in the sample are women, 63% are married, 45% are naturalized citizens in the U.S., 53% speak English well or very well, and approximately 3% are black. Fifty percent of Cubans in the U.S. were 30 years old or more at the time of their entry into the U.S., have 12 or more years of education and earn $8.33 per hour or more. On average they have 27 years of potential experience.

Figures 1 and 2 show where individuals work depending on their level of education. A comparison of the occupations of Cubans in Cuba and Cubans in the United States reveals that a higher percentage of Cubans in U.S. with 13 or more years of education work in jobs that require lower skills than they actually possess. Figure 3 presents a histogram and box plot of (log) hourly earnings for Cuban immigrants. The histogram of (log) hourly earnings gives a good picture of the presence of positive skewness. Table 2 contains the sample correlations between the explanatory variables. As can be seen, there are significant correlations between Proficiency in English and Citizenship Status (r = 0.3387), that is, Cubans with greater Proficiency in English are more likely to be U.S. citizens. Age on arrival in the U.S. and proficiency in English are negatively associated (r = - 0.3611); this means that younger individuals seem to be the most predisposed to learn English.

Nearly 60% of Cubans in the United States who have 13 or more years of education take jobs with low skill requirements, and 78% of those with 9–12 years of education work as skilled workers (Cobas and Fernández, 2014). Many Cubans begin their working life in U.S. in jobs with skill levels much lower than, and very different from, those that they held in Cuba (Cattan, 1993).

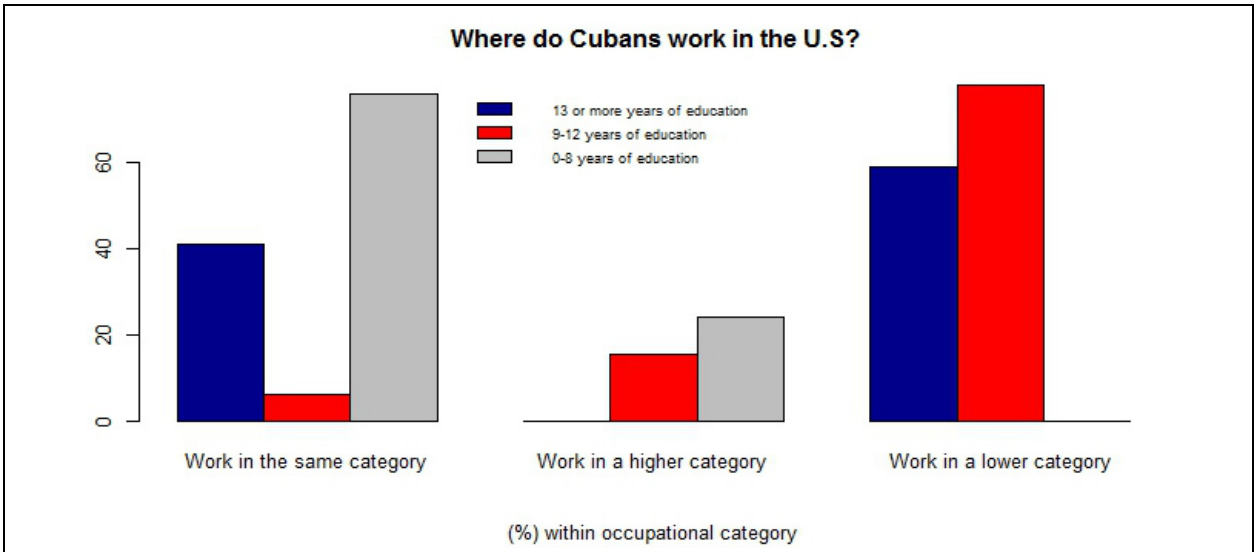
As can be seen in the first column of Table 3, in the OLS estimation there is only one non-significant explanatory variable (Black) to explain the (log) hourly earnings, at the significance level of 5%. The remaining explanatory variables are statistically significant. Note that being married, being a naturalized American, speaking English well or very well, years of education and potential job experience show a positive

Figure 1. Job Category in Cuba



Source: Computed by the author based on the Population and Housing Census of Cuba 2002 provided by IPUMS-International (2011).

Figure 2. Job Category in the U.S.



Source: Computed by the author based on the 2011 American Community Survey (ACS) in the U.S. provided by IPUMS (2011).

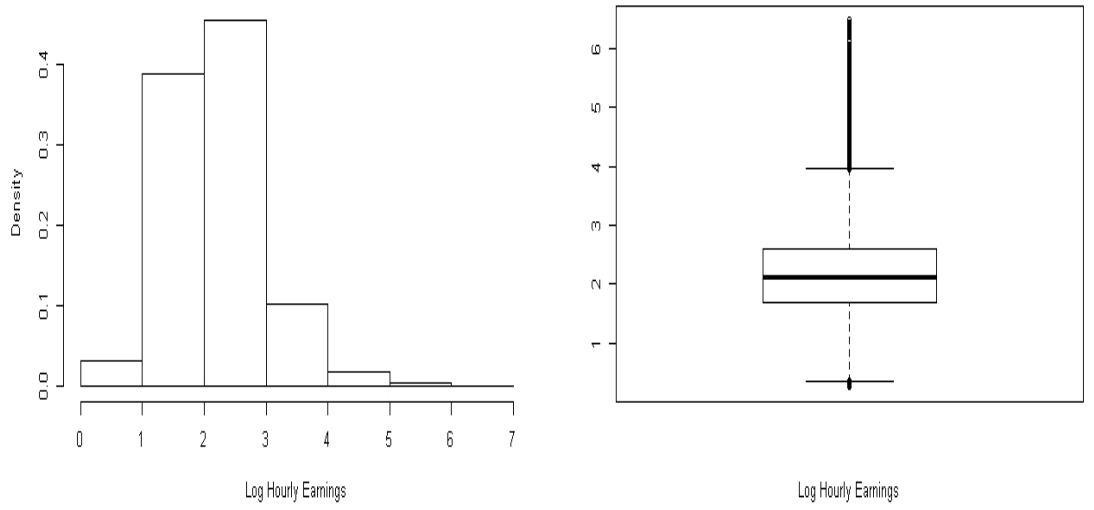
link to (log) hourly earnings, while being a woman and age on entry in the U.S. show a negative link to the response variable.

The estimate of the return to years of education on the mean earnings is lower than expected, equal to 6.23%. An additional year of potential job experience is shown to increase mean earnings by 0.46%. Being a woman decreases mean earnings by 28.39% which is the greatest decrease in mean earnings. The model explains approximately 15% of the variance in (log) hourly earnings.

Figure 4 shows the QR results and Table 3 presents the QR estimations for five values of τ-quantiles: 0.10, 0.25, 0.50, 0.75 and 0.90. Increments in earnings associated with the different socioeconomic characteristics vary across the earnings distribution.[2]

2. Tests on whether these coefficients differ significantly from one point in the distribution to another have been performed and the null hypothesis of equality of coefficients is rejected at the 5% significance level.

Figure 3. Histogram and Box Plot of (log) Hourly Earnings



Source: Computed by the author based on the 2011 American Community Survey (ACS) in the U.S. provided by IPUMS (2011).Figure 3. Histogram and Box Plot of (log) Hourly Earnings.

Table 2. Sample Correlation of Variables

	Woman	Black	Married	American Citizen	English Proficiency	AgeImm	Years of Education	Experience
Woman	1.000	-0.020	-0.0353	0.1118	0.0198	-0.0945	0.0443	0.0324
Black		1.000	-0.0483	-0.0401	-0.0073	0.0311	-0.0278	-0.0111
Married			1.000	0.0622	-0.0145	0.0439	0.0121 *	0.0638
American Citizen				1.000	0.3387	-0.2659	0.1208	0.4033
English Proficiency					1.000	-0.3611	0.2916	-0.0115
AgeImm						1.000	-0.0404	0.2292
Years of Education							1.000	-0.2925
Experience								1.000

Note: All correlations are significant at the 1% significance level. * means significant only at the 10% significance level.

Table 3. OLS and Linear Quantile Regression

	Dependent variable: Log Hourly Earnings					
	OLS	10%	25%	50%	75%	90%
Centercept	2.1266 (0.0134)***	1.3763 (0.0180)***	1.6915 (0.0143)***	2.0707 (0.0142)***	2.4683 (0.0184)***	2.9478 (0.0298)***
IsWoman	-0.2839 (0.0108)***	-0.2626 (0.0141)***	-0.3128 (0.0109)***	-0.3256 (0.0109)***	-0.2889 (0.0144)***	-0.2355 (0.0230)***
IsBlack	-0.0377 (0.0323)	-0.0646 (0.0266)**	-0.0992 (0.0344)**	-0.0344 (0.0354)	-0.0310 (0.0326)	0.0946 (0.0824)
IsMarried	0.0654 (0.0108)***	0.0709 (0.0143)***	0.0710 (0.0114)***	0.0683 (0.0109)***	0.0688 (0.0145)***	0.0547 (0.0226)**
IsAmericanCitizen	0.0884 (0.0133)***	0.1029 (0.0194)***	0.1111 (0.0141)***	0.0898 (0.0135)***	0.0871 (0.0171)***	0.0024 (0.0288)
EnglishProficiency	0.1564 (0.0118)***	0.0992 (0.0162)***	0.1514 (0.0123)***	0.1577 (0.0121)***	0.1432 (0.0157)***	0.1824 (0.0260)***
AgeImm	-0.0110 (0.0008)***	-0.0087 (0.0011)***	-0.0113 (0.0008)***	-0.0121 (0.0008)***	-0.0114 (0.0010)***	-0.0119 (0.0017)***
Years of Education	0.0623 (0.0023)***	0.0425 (0.0030)***	0.0525 (0.0023)***	0.0653 (0.0024)***	0.0727 (0.0029)***	0.0751 (0.0049)***

Note: Significance Codes.*** 0.001 ** 0.05 * 0.10. Standard Errors in parentheses.

Table 3. OLS and Linear Quantile Regression (Continued)

	Dependent variable: Log Hourly Earnings					
	OLS	**10%**	**25%**	**50%**	**75%**	**90%**
Experience	0.0046 (0.0006)***	0.0009 (0.0009)	0.0038 (0.0006)***	0.0053 (0.0006)***	0.0054 (0.0008)***	0.0075 (0.0013)***
Years of Education Sqd.	0.0051 (0.0003)***	0.0029 (0.0004)***	0.0041 (0.0003)***	0.0054 (0.0004)***	0.0063 (0.0004)***	0.0064 (0.0008)***
Experience Sqd.	-0.0003 (0.00004)***	-0.0003 (0.00005)***	-0.0003 (0.00004)***	-0.0003 (0.00004)***	-0.0003 (0.00006)***	-0.0002 (0.00009)**

Note: Significance Codes.*** 0.001 ** 0.05 * 0.10. Standard Errors in parentheses.

Being a woman causes a decrease in wages in all quantiles of the distribution; the decrease is highest from the 25th percentile to the 75th percentile. On average, women in the United States are paid just 77 cents for every dollar paid to men. For Latinas, the gap is larger. Latinas in the United States are paid, on average, just 54 cents for every dollar paid to white, non-Hispanic men (Fry and Taylor, 2013).

Being black is not a variable that explains the variability of (log) hourly earnings from the central part of the distribution to the upper part, since it is not a significant predictor in any percentile from the median at a significance level of 5%. For people who earn less it is a significant predictor and it causes decreases in earnings. Being married is a variable that produces a homogeneous effect on the different quantiles of the distribution. Being an American citizen has little effect on hourly earnings and only a minor effect in the upper quantiles.

The return to proficiency in English is higher for people who earn more. For the 90th percentile, it gives rise to an increase in hourly earnings of about 18% compared to around 10% for the 10th percentile. The payoff for immigrants from learning the English language is likely to depend on the frequency with which they will use those skills in their everyday interactions (Lazear, 1999). Borjas (2015) shows that more recent immigrants to the U.S are improving their English language skills at a far lower rate than earlier immigrants.

Regarding age on entry in the U.S, being younger is more important in the upper quantiles of the distribution of earnings. The return to potential experience is almost zero across the distribution of earnings of workers, but it is greater for people who earn more. The number of years of education has a smaller impact on earnings in the lower quantiles (4.25% for 10th percentile and 5.25% for 25th percentile) than in the upper quantiles (7.27 % for the 75th percentile and 7.51% for the 90th percentile).

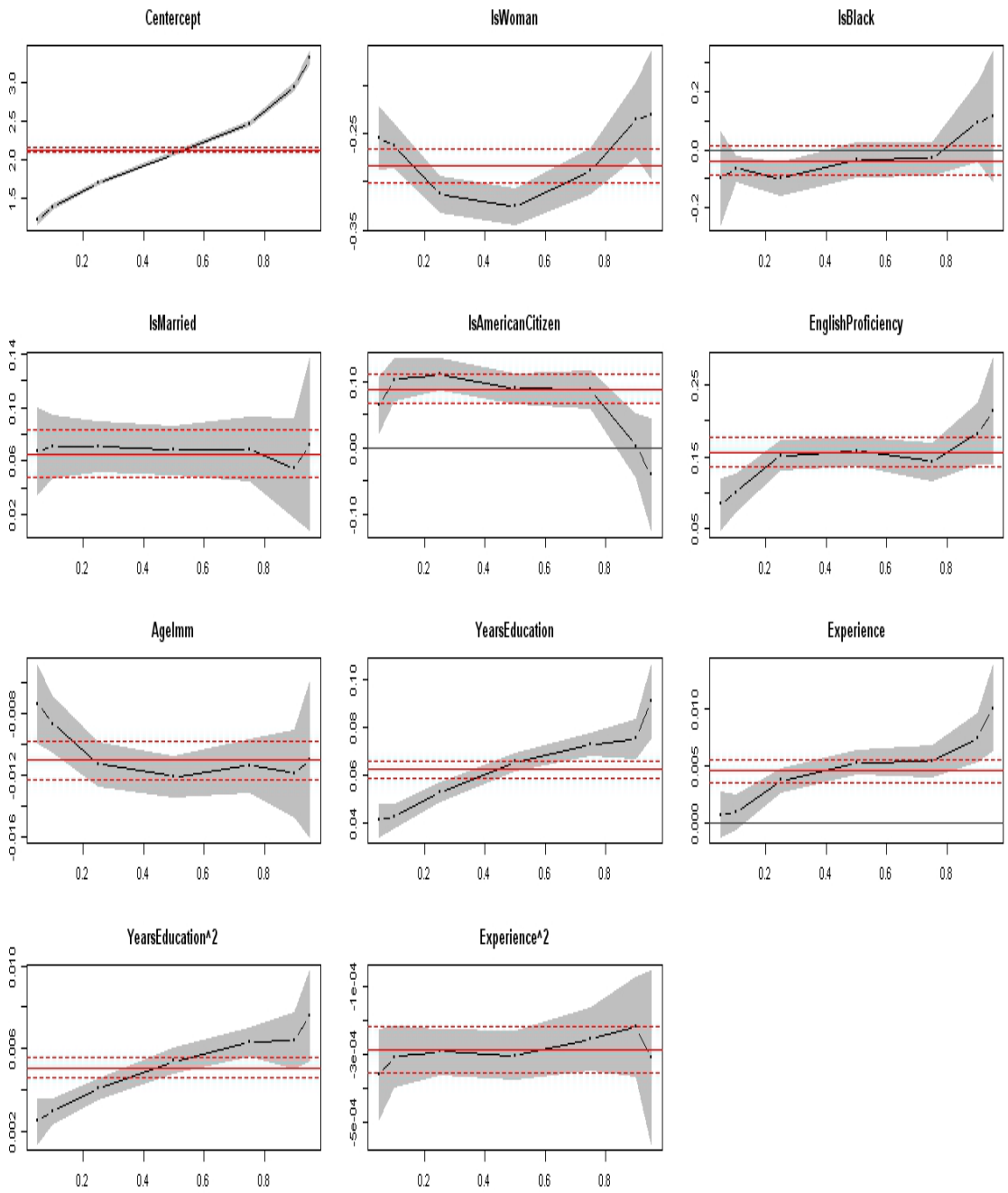
The returns to education are higher at the top of the conditional earnings distribution but lower than expected. One possible explanation for this situation is related to over-education. Lower earnings for over-educated workers increase the skill dispersion of pay by extending the lower tail of the wage distribution of the highly educated (Martins and Pereira, 2004). Immigrants may also encounter difficulties in obtaining good jobs due to language barriers, less extensive network connections and lags in cultural adjustment (Bohon, 2005). Another reason may be the fact that Cuban workers are not as highly valued in the U.S market now as they were years ago, due to increases in the emigration of highly skilled people. Skilled workers, earning relatively higher wages, have become relatively more abundant and, as a result, their relative wages have decreased (Machado and Mata, 2005).

CONCLUSIONS

This paper has analyzed the distribution of earnings of Cuban immigrants in the U.S in terms of certain observable characteristics: years of education, potential job experience, age at time of emigration, ethnicity, marital status, sex, citizenship status and proficiency in English. It uses the 2011 American Community Survey (ACS) of the U.S. Within the sample, only workers aged between 25 and 64 years who immigrated to the United States when they were between 17 and 49 years old are considered.

For the analysis, OLS estimation with standard errors calculated using the heteroskedasticity robust HC-4 estimator of Cribari and Neto (2004) was first em-

Table 4. Quantile Regression



Source: Calculated by the author based on the results of QR estimation.

ployed and then QR, a technique which allows us to characterize the whole distribution of earnings of Cuban immigrants in U.S was proposed.

When we use OLS estimation with heteroskedasticy-robust standard errors, all explanatory variables except being black prove to be significant at the 5% significance level. All other variables being constant, speaking English well or very well is the variable that produces the biggest increase in mean hourly earnings (about 15.64%); being a woman produces a decrease in mean hourly earnings of 28.39%; being married produces an increase of about 6.54%,; being

a U.S citizen produces an increase of 8.84%; one more year of study increases expected hourly earnings by 6.23% and one year of potential experience increases expected hourly earnings by 0.46%.

However, the application of QR shows how the influence of the different variables considered on hourly earnings varies across the earnings distribution. With this type of estimation method differences between highly-skilled and low-skilled Cuban immigrants in the labour market can be detected.

The main conclusions of this article are the following: being a woman decreases hourly earnings at all points of the distribution, with the decrease being greater for individuals in the central part of the earnings distribution. The return to proficiency in English is greater for those people who earn more and the returns to education has a smaller impact on earnings in the lower quantiles and a greater impact at the top of the conditional earnings distribution but even then it is lower than expected. In particular, this article demonstrates how the return to education on earnings is less for Cuban born workers than expected.

REFERENCES

Aupetit, Sylvie Didou and Gérard, Etienne (2009). "Fuga de cerebros, movilidad académica, redes científicas." In: *Perspectivas latinoamericanas* (primera ed.) México, DF: Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional.

Bertoli Simone; Fernández-Huertas, Jesús and Ortega, Francesc (2013). "Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions." *Journal of Development Economics*, Vol. 101, p. 75–91.

Bohon, Stephanie A. (2005). "Occupational Attainment of Latino Immigrants in the United States." *Geographical Review*, Vol. 95, No. 2, p. 249–266.

Borjas, George J. (1985). "Assimilation, Changes in Cohort Quality and the Earnings of Immigrants." *Journal of Labor Economics,* Vol.3, No.4, p. 463–489.

Borjas, George J. (2015). "The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again." *Journal of Human Capital,* Vol.9, No 4, p. 483–517.

Cattan, Peter (1993). "The diversity of Hispanics in the U.S. work force." *Monthly Labor Review*, Vol. 8, p. 3–15.

Chamarbagwala, Rubiana (2010). "Economic liberalization and urban-rural inequality in India: a quantile regression analysis." *Empirical Economics*, Vol. 39, No. 2, p. 371–394.

Chiswick, Barry R. (1978). "The effect of Americanization on the earnings of foreign-born men." *The Journal of Political Economy*, Vol.86, No 5, p. 897–921.

Chiswick, Barry R. and Miller, Paul W. (2008). "How immigrants fare across the earnings distribution in Australia and the United States." *Industrial and Labor Relations Review*, Vol.61, No.3, p. 353–373.

Cobas Valdés, Aleida and Fernández Sainz, Ana (2014). "Cuban Migration to the United States and the Educational Self-Selection Problem." *International Journal of Cuban Studies*, Vol.6, No.1, 2014, p. 41–54.

Cribari-Neto, Francisco (2004). "Asymptotic Inference Under Heteroskedasticity of Unknown Form." *Computational Statistics & Data Analysis*, Vol. 45, No.2, p. 215–233.

Cuecuecha, Alfredo (2005). "The Immigration of Educated Mexicans: The Role of Informal Social Insurance and Migration Costs." Available at http://economics.ca/2005/papers/0298.pdf.

Elliott, Robert and Lindley, Joanne (2008). "Immigrant wage differentials, ethnicity and occupational segregation." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 171, No. 3, p. 645–671.

Friedberg, Rachel (1992). "The Labor Market Assimilation of Immigrants in the United States: The Role of Age at Arrival." Working Paper. Brown University.

Fry, Richard and Taylor, Paul (2013). *Hispanic high school graduates pass whites in rate of college enrollment*. Washington DC: Pew Research Center.

Hunt, Priscillia (2012). "From the bottom to the top: a more complete picture of the immigrant-native wage gap in Britain." *IZA Journal of Migration*, Vol.1, No. 9, p. 1–18.

Integrated Public Use Microdata Series (IPUMS) (2011), Steven Ruggles, J.; Trent,A.; Genadek, K.; Goeken R,; Schroeder M.B. and Sobek, M., Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Integrated Public Use Microdata Series, International (IPUMS-International) (2011). Version 6.1 [Machine-readable database]. Minnesota Population Center. Minneapolis: University of Minnesota.

Lazear, Edward P. (1999). "Culture and Language." *Journal of Political Economy*, 107, No. 6S, p. 95–125.

Lowell, B. Lindsay; Pederzini, Carla; and Passel, Jeffrey (2008). "The Demography of Mexico/U.S Migration." In *Mexico-U.S Migration Management: A Binational Approach*. Maryland, Lexington Books. Agustín Escobar y Susan F. Martin, editors.

Machado, José A.F and Mata, José (2005). "Counterfactual Decomposition of Changes in Wage Distribution Using Quantile Regression." *Journal of Applied Econometrics*, Vol. 20, No.4, p. 445–465.

Martins, Pedro S. and Pereira, Pedro T. (2004). "Does education reduce wage inequality? Quantile regression evidence from 16 countries." *Labour Economics*, No.11, p. 355–371.

Motel, Seth and Patten, Eileen (2012). "Hispanic of Cuban Origin in the United States." Pew Hispanic Center,http://www.pewhispanic.org/files/2012/06/2010-Cuban-Factsheet.pdf

Rangvid, Beatrice S. (2007). "School composition effects in Denmark: Quantile regression evidence from PISA 2000." *Empirical Economics*, 33 (2), 359–388.

Terry, Mary B.; Wei, Ying; and Esserman, Denise (2007). "Maternal, Birth and Early-Life Influences on Adult Body Size in Women." *American Journal of Epidemiology*, Vol.166, No.1, pp. 5–13.

U.S. Bureau of Labor Statistics (October 2015). "Labor Force Characteristics by Race and Ethnicity, 2014". BLS Reports.

U.S Census Bureau, (2013). Geographical Mobility/ Migration. U.S Department of Commerce.

Wainer, Howard (2000). "The Centercept: An Estimable and Meaningful Regression Parameter." *Psychological Science,*Vol. 11, No. 5, p. 434–436.

Zeileis, Achim (2004). "Econometric Computing with HC and HAC Covariance Matrix Estimators." *Journal of Statistical Software*, Vol. 11, No.10, p.1–17.