Article

Gender Dissimilarities in Human Capital Transferability of Cuban Immigrants in the US: A Clustering Quantile Regression Coefficients Approach with Consideration of Implications for Sustainability

Aleida Cobas-Valdés * and Javier Fernández-Macho [ORCID]

Department of Quantitative Methods, University of the Basque Country, 48015 Bilbao, Spain; javier.fernandezmacho@ehu.eus
* Correspondence: aleida.cobas@ehu.eus; Tel.: +34-6017049

Abstract: Female participation in the labor market has been increasing over time. Despite the fact that the level of education among women has also increased considerably, the wage gap has not narrowed to the same extent. This dichotomy presents an important challenge that the United Nations Sustainable Development Goals with respect to gender inequities must address. Hispanics constitute the largest minority group in the US, totaling 60.6 million people (18.5% of the total US population in 2020). Cubans make up the third largest group of Hispanic immigrants in the US, representing 5% of workers. This paper analyzes the conditional income distribution of Cuban immigrants in the US using the clustering of effects curves (CEC) technique in a quantile regression coefficients modeling (QRCM) framework to compare the transferability of human capital between women and men. The method uses a flexible quantile regression approach and hierarchical clustering to model the effect of covariates (such as years of education, English proficiency, US citizenship status, and age at time of migration) on hourly earnings. The main conclusion drawn from the QRCM estimations was that being a woman had the strongest negative impact on earnings and was associated with lower wages in all quantiles of the distribution. CEC analysis suggested that educational attainment was included in different clusters for the two groups, which may have indicated that education did not play the same role for men and women in income distribution.

Keywords: transferability of human capital; earnings distribution; immigrant workers; quantile regression; discrimination by gender

1. Introduction

Information in the literature about the gender gap is essentially based on transferability of human capital (schooling and labor experience), the family division of labor, selection into the labor force, and discrimination [1]. From a human capital [2] perspective, predominantly male occupations pay more than predominantly female occupations. Similarly, women are viewed as choosing occupations in which their skills will depreciate less rapidly during spells of absence from the labor market. In this article, we propose the use of the clustering of effects curves (CEC) technique in a quantile regression coefficients modeling (QRCM) framework to analyze how the skills of men and women are grouped into a cluster and to detect possible differences that could help to better understand why the gender pay gap exists. This paper, to our knowledge, marks the first attempt to use cluster analysis in a flexible quantile regression context to analyze skills and gender gaps in labor markets for immigrants.

According to the theory of occupational crowding, male jobs pay more because women excluded from them by discrimination are shunted into other occupations with no (or less) discrimination, and the resulting increased supply of labor (or crowding) lowers

their wages [3]. When women are concentrated into particular occupations due to their preferences or according to their family burdens, the negative effects of increased female presence in those fields could be a costly compensating differential [4].

On the other hand, many studies have agreed that wage inequality is socially established and that work in women's occupations is undervalued by reason of institutionalized bias against women [5], even if the skills required for lower-paying, female-dominated jobs are comparable to those in better paid male-dominated jobs [4]. Most of the segregation and all of the stratification in the US occurs because women work in different and lower-paying occupations compared with men of similar characteristics. At the same time, many empirical research studies have found that the proportion of females in an occupation can have a net negative effect on wages [6–9].

With regard to immigrants, some structural approaches suggest that disadvantaged group members were employed in low-wage occupations due to entry barriers to high-wage occupations [10–12]. The effects of any type of segregation on the labor market—especially if the segregation implies stratification of groups—are further aggravated when they do not result from differences in the accumulation of human capital [13].

Hispanics constitute a large and rapidly growing segment of the US population. From 2018 to 2028, the US Bureau of Labor Statistics (BLS) projects the number of Hispanics in the labor force to increase by about 7.4 million, more than any other ethnic group [14]. Within Hispanics, Cubans were the third largest group of Hispanic immigrants in the US, representing 5% of overall workers. If we consider migration from Cuba as a labor migration, we have to take into account that this kind of migration is highly selective on the basis of age and sex to yield inflows of working age males [15], which may force the concentration of Cuban immigrant women into certain sectors.

Cuban immigrants have positively self-selected in terms of educational level, in their migration decision to move to the US; that is, people with the highest levels of education tend to migrate [16]. On the other hand, under the Cuban Adjustment Act (CAA) of 1966, Cuban immigrants in the US have enjoyed a migrant status with more facilities for integration than any other group of immigrants has had, which has been a determining factor in their integration and assimilation process. For these reasons, Cuban immigrants in the US are an immigrant group of special interest in analyses of skill transferability and gender differences in the labor market.

Sustainability means the capacity to maintain some entity, outcome, or process over time [17]. However, extensive studies in the literature on development also used the concept to refer to improving and sustaining a healthy economic, ecological, and social system for human development [18]. Studies about gender differences in the labor market are of crucial importance, not only for countering gender inequality but also as a way to contribute to sustainable development. As mentioned in [19], equal opportunities between women and men contribute to enhancing the competitiveness of the economy and better economic performance.

Given these trends, and the relative importance of Cubans as a share of the US Hispanic population, it is undoubtedly of interest to analyze the whole distribution of earnings of Cuban immigrants in the United States by gender and quantify the effects of socioeconomic variables at different points on the earnings' distribution. Furthermore, it is important to investigate and compare the process of skill transferability between men and women, taking into account income distribution, to identify possible differences between both groups in the transferability of human capital.

Our contribution, with respect to previous studies, is three-fold: first, this paper uses a robust regression analysis technique, in a field of study (i.e., the study of inequality) within which the continued use of the classical Ordinary Least Squares (OLS) regression is no longer justified by the distinct limitations in efficiency of this method due to the fact that the assumptions of no covariance and constant variance are usually unreasonable and OLS estimators may be deficient in linear models with non-normal errors. Moreover, it is a method referring only to the analysis of the conditional mean, as pointed out in [20–22].

Second, we study the rates of returns on labor market skills at the different locations of the conditional income distribution and detect the existence of gender differences in the contribution of skills to income, while taking into account the differences between the highest and the lowest income earners. Third, we use the CEC technique proposed in [23] from a quantile regression coefficients modeling (QRCM) framework developed in [24] to find similar curves of covariate effects on earnings in a quantile regression framework and detect dissimilarity by gender in human skills' transferability. In this third contribution, we focus on education and the role it plays in predicting earnings.

The quantile regression model [25] proposes different regression lines for the different quantiles of the earnings' distribution. QR allowed us to describe the conditional distribution of earnings on the covariates at different points of the distribution and hence offer a more comprehensive overview of the link between the response variable and the covariates. Meanwhile, the QRCM framework focuses on model selection when estimating the conditional quantile function, using information on all quantiles simultaneously.

The data used in this study are from the American Community Survey (ACS) database provided by the Integrated Public Use Microdata Series [26]. The main conclusions to be drawn from the QRCM estimations were that females earned less, regardless of their position in the income distribution. The CEC algorithm suggested that education was not a competitive advantage for Cuban-born female immigrants, as compared with Cuban-born male immigrants in the US. This raised the possibility of a gender-based penalty to wages.

2. Materials and Methods

2.1. Linear Quantile Regression

Quantile regression (QR) is a well-known method introduced in [25]. It allowed us to study the dependence relationship between a response variable and its covariates along the whole conditional distribution of the response variable. QR has been used more frequently in analyses of the relationship between a response variable and covariates due to its robustness in the presence of outliers or extreme values. Another reason for the increased use of QR is that the effect of covariates cannot be assumed to be the same in each of the quantiles of the response variable's distribution, as is the case with the typical OLS method. In this sense, the coefficients estimated using QR provide more information than can be obtained with the simple OLS mean regression method. For any $p \in (0, 1)$, a linear quantile regression model can be written as:

$$y_i = x_i' \beta_p + \varepsilon_{pi} \tag{1}$$

where y_i denotes the observation of the response variable for individual i, and the given covariate vector for observation i is defined by $x_i = (1, x_{i1}, x_{i2}, \cdots x_{ip})^T$. The quantile-specific linear effects are denoted by $\beta_p = (\beta_{p0}, \beta_{p1}, \cdots \beta_{pn})^T$, where $\beta \in \mathbb{R}^n$. No specific assumptions are made for the error term (ε_{pi}), apart from ε_{pi} and ε_{pj} being independent for $i \neq j$ and considering that the distribution function at 0 is p.

The quantile function $Q_{Y_i}(p/x_i)$ of the response variable y_i, conditional on the covariate vector x_i at a given quantile parameter p, is given by:

$$Q_{Y_i}(p/x_i) = x_i^T \beta_p \tag{2}$$

In a standard QR model, the different quantile coefficients, β_p, are estimated one at a time for a specific quantile p, and they can be interpreted as rates of change of the response variable conditional to each covariate.

2.2. Quantile Regression Coefficients Modeling (QRCM)

QRCM, as developed in [24], facilitates modeling of the corresponding quantile regression coefficient functions, $\beta(p)$, as parametric functions that depend on the order of the quantile p. The QRCM framework overcomes the limitations of linear QR: (i) the entire conditional quantile function (QF) is estimated immediately and without any distributional

assumptions, and semiparametric models can be implemented by letting $b(p)$ be a flexible function of p, and (ii) estimation and inference are simplified, and a better interpretation of the results is possible, increasing efficiency. The QRCM framework also focuses on model selection when estimating the conditional QF at the same time, using information on all the quantiles simultaneously. Using QRCM, the effect of each covariate is a curve in the space of the quantiles. In the parametric quantile models, this approach assumes that the QF is known up to the parameter vector θ. The quantile function defined in Equation (2) needs to be reformulated as:

$$Q_{Y_i}(p/x_i) = x_i^T \beta\,(p) \tag{3}$$

where $\beta(p)$ is a q-dimensional vector that is a function of $p \in (0, 1)$ and depends on a finite-dimensional parameter θ:

$$\beta(p/\theta) = \theta b(p) \tag{4}$$

where $b(p) = [b_1(p), b_2(p), \cdots b_k(p)]^T$ is a set of k known functions of p, which can be defined as any set of functions, including polynomials, splines, and trigonometric functions, with the only requirement being that $b(p)$ must induce a well-defined quantile function (QF) for some θ, as in [24]. θ is a matrix with the θih entries being associated with the *ith* covariate and the *hth* function, $i = 1, \cdots, q$ and $h = 1, \cdots, k$. The QR coefficient associated with the *jth* covariate is:

$$\beta_i(p/\theta) = \theta_{j1} b_1(p) + \cdots \theta_{jk} b_k(p), \quad i = 1, \cdots q \tag{5}$$

The conditional quantile function can be formulated as:

$$Q_{Y_i}(p/x_i, \theta_i\,) = x_i^T \theta\, b(p) \tag{6}$$

2.3. Cluster of Effects Curves (CEC)

The estimation of a single set of coefficients often provides a misrepresentation of the true dependence structure between a response variable and a set of covariates. A common solution involves clustering units and then estimating different models for each group. This solution does not identify a group's impact on the response variable, and it requires suitable tools for comparing the models estimated for different samples. An alternative solution would be the simultaneous use of QR and cluster analysis, as proposed in [27,28].

Clustering is an unsupervised method used to discover natural groups in datasets based on the similarity or dissimilarity of the members without any background knowledge of the data's characteristics. In our study, the hierarchical agglomerative clustering method was applied to group the socioeconomic trends for the different slopes of quantiles in the income distribution. It was also applied, specifically, to analyze gender dissimilarities in education as the cornerstone to the transferability of human capital.

Following [23], the algorithm used in this paper allowed the clustering of distribution curves based on a new method for finding curve similarities in a QRCM framework. In this new method, the effect of covariates on a response variable is represented by curves in the space of percentiles. Using a QRCM approach, each covariate's effect is a curve in the multidimensional space of the percentiles. By collecting all the curves and describing the effects of each covariate on the response variable, we were able to assess whether one or more covariates had the same effects on the response variable [29].

2.4. Descriptive Characteristics of the Sample

Our aim was to model the relationship between the (log) of gross hourly earnings, y, for individual i using total pre-tax wage and salary income (expressed in contemporary dollars). The latter referred to money received as an employee for the previous year, as the measure of earnings, and a set of regressors or covariates x: an indicator for gender (x_1) (female = 1), an indicator for civil status (x_2) (married = 1), an indicator for US citizenship (x_3) (US citizenship = 1), an indicator for English proficiency (x_4) (proficiency in English = 1), age at time of migration (x_5), years of education (x_6), and potential work

experience (x_7). The covariates considered were some of the most important used in relevant literature about the economic assimilation of immigrants.

We used data from repeated cross-sectional samples between 2010 and 2017, taken from a random 1% sample of the American Community Survey (ACS) obtained through the Integrated Public Use Microdata Series (IPUMS) at the University of Minnesota [26]. We considered Cuban-born immigrants in the US. The sample consisted of 22,077 respondents, of whom 9642 were women and 12,435 were men.

We restricted our sample to individuals aged between 25 and 55, working 60 h or less per week during the year preceding the census, and who entered the US when they were aged between 17 and 49. This last criterion corresponded to the age group most likely to migrate for economic reasons [30].

Table 1 summarizes the descriptive statistics of the variables included in the study for Cuban immigrants. In the sample, 44% of the immigrants were female. We observed that almost 49% of men in the sample had US citizenship, which was 7% more than women. Notably, in both samples, about half were proficient in English and almost 97% worked. No gender differences occurred for age at migration or number of weeks worked in the year prior to the survey.

Table 1. Descriptive statistics of relevant variables for Cuban workers in the US.

| Variables	Mean		1st Quartile		Median		3rd Quartile	
	Women	Men	Women	Men	Women	Men	Women	Men
Hourly Earnings	15.26	18.79	7.74	9.62	11.15	13.94	16.83	21.15
Years of Education	13.10	12.64	12.00	12.00	12.00	12.00	16.00	14.00
Potential Experience	25.76	26.60	18.00	20.00	26.00	27.00	33.00	33.50
Age at Migration	31.04	31.30	24.00	25.00	30.00	31.00	38.00	37.00
Working Hours per Week	36.78	40.03	35.00	40.00	36.78	40.00	40.00	40.00
Weeks Worked per Year	47.33	47.81	52.00	52.00	52.00	52.00	52.00	52.00
Is an American Citizen	0.49	0.42						
Has Proficiency in English	0.52	0.51						
Is Married	0.54	0.58						
In Labor Force	0.97	0.97						
Gender	0.44	0.56						

For dummy variables to denote socioeconomic characteristics, e.g., citizenship, proficiency in English, marital status, gender, and participation in the labor force, the mean offers the percentage of individuals within these conditions. Potential work experience has been calculated as Age–Years of Education–6.

Women and men were similar in terms of years of education at the first and second quartiles, as Table 1 illustrates. In both groups, 50% of individuals had 12 years of education or more. In contrast, similarities through years of education distribution were not replicated in terms of their respective income distribution at the same quartiles.

Taking into account the mean, woman had more years of education than men. Of note is the fact that, at the third quartile, women had two more years of education than their male counterparts. Figure 1 shows the density of (log) hourly earnings for the two groups and Figure 2 shows the symmetric boxplot for gross hourly earnings. Differences in the patterns of income were easily appreciated at each location of the earnings' distribution.
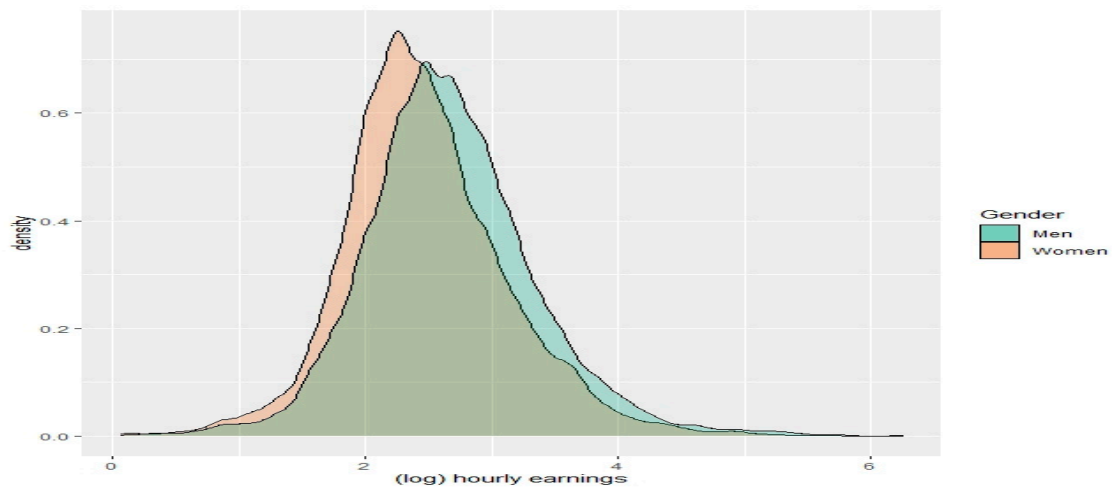
Figure 1. Kernel density estimates for (log) hourly earnings. Central values for men are higher; thus, when superimposing one density function on the other, the tail of the distribution of men is slightly asymmetric to the right, while the density function of women is slightly asymmetric to the left.

(**a**)   (**b**)
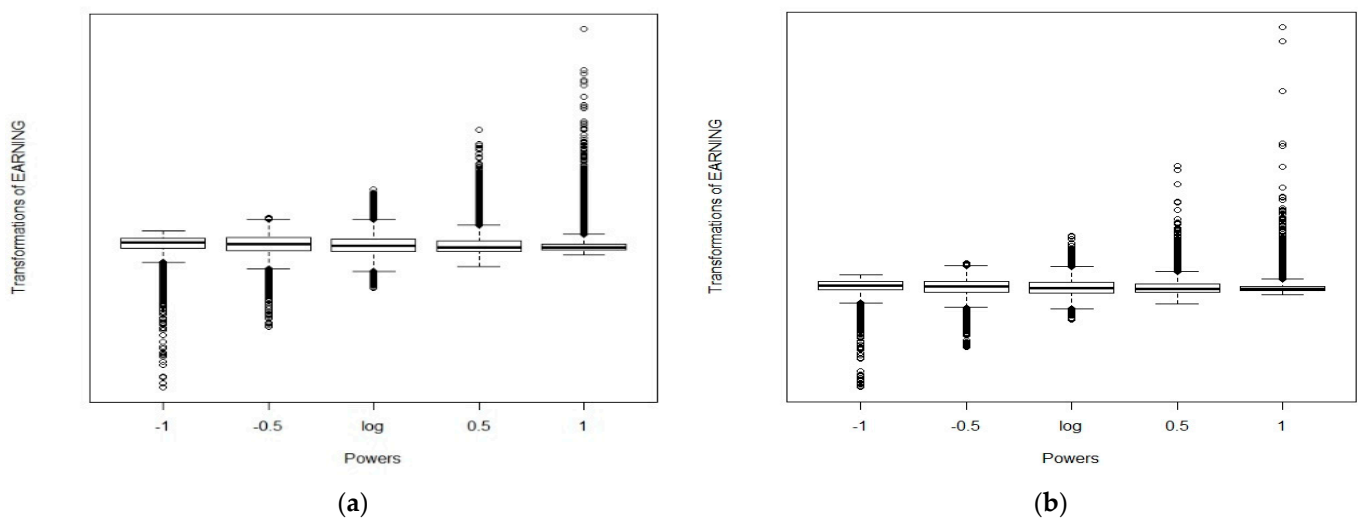
Figure 2. Symmetry boxplots for hourly earnings. (**a**) Boxplot for men in the sample. (**b**) Boxplot for women in the sample.

Figure 3 shows the considerable dissimilarity between females and males in terms of the behavior of the means of the variables' hourly earnings and years of education. The between-group differences for the two variables were analyzed and confirmed through respective Wilcoxon nonparametric tests. In accordance with the National Committee on Pay Equity, Equal Pay Day in the US in 2021 was Wednesday, 24 March. The date is used to symbolize how far into a year women must work to earn what men earned in the previous year. The US Census Bureau (2020) pointed out that, in 2019, females earned 82.3 cents for every dollar earned by males [31].
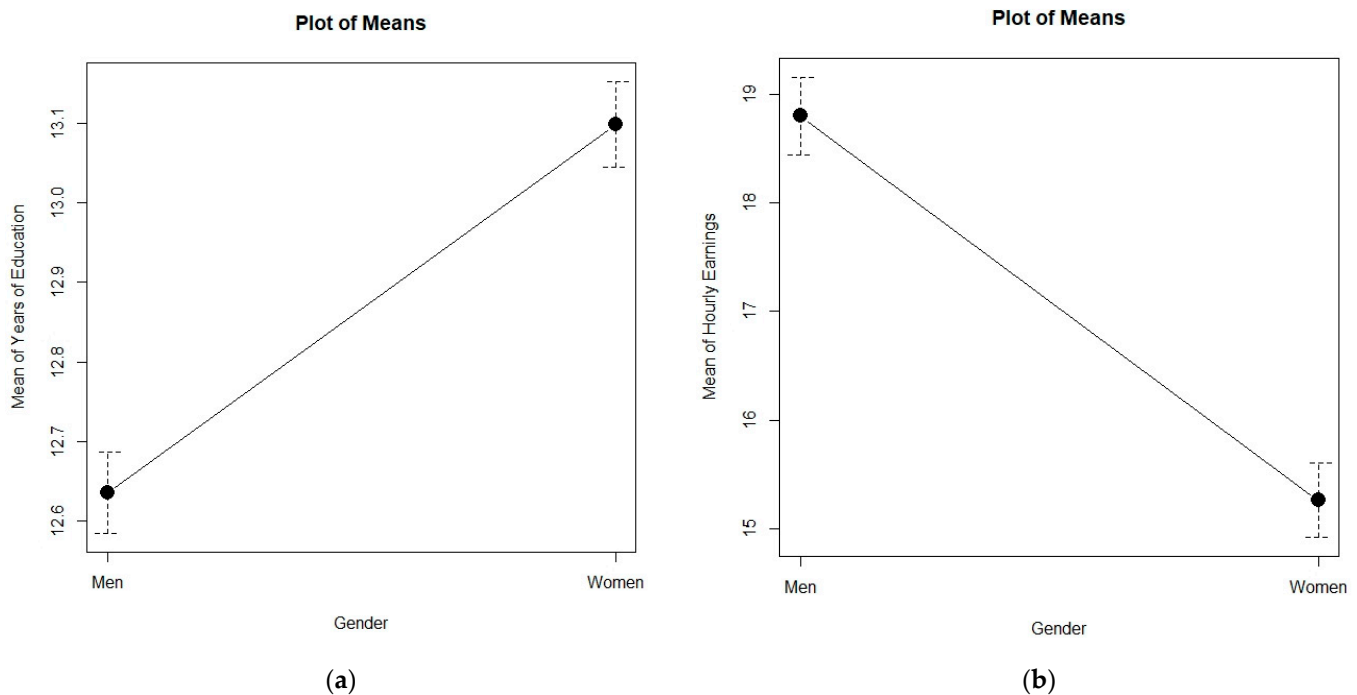
Figure 3. Plots of means considering gender as a variable of classification. (**a**) Plot of means for years of education (Wilcoxon test $p < 0.01$). (**b**) Plot of means for (log) hourly earnings (Wilcoxon test $p < 0.01$) .

3. Empirical Results

As a first step, we estimated a QRCM model. Using this approach, as in [23], we aimed to estimate the coefficient functions $\beta_0(p/\theta)$, $\beta_1(p/\theta)$, $\cdots \beta_q(p/\theta)$, namely, effects curves on (log) hourly earnings. This approach led us to assess the q curves that described the effects of each socioeconomic characteristic considered as a covariate on the response variable, in particular, education. The intercept β_0 and the coefficients associated with the covariates were modeled by a shifted Legendre polynomial of degree five in order to create an orthogonal spline. These curves were subsequently clustered based on similarities of effects, following [23], as a variable selection procedure.

Table 2 reports the QRCM estimations and Figure 4 shows the estimated quantile coefficients for the covariates together with the corresponding 95% confidence intervals. We were able to observe the relationship between the associated regression coefficients and the order of quantiles, considering a Legendre polynomial. Results from this estimation also suggested that all socioeconomic characteristics included were statistically relevant for all percentiles.

Gender (specifically, being female) and age at time of migration had a negative effect on wages. Being female had the strongest negative impact on earnings and was associated with lower wages in all quantiles of the distribution, with the greatest negative impact on the central and upper part of the earnings' distribution. This could indicate that, between men and women earning the same income, women required higher qualifications, solely because they are women. The intercept values in the Legendre polynomial for obtaining the coefficient associated with the covariates 'is woman' and 'age at migration' were negative at all the values for $\beta_i(p)$. The existence of significant negative coefficients on women at each location of the earnings' distribution can be taken as evidence of wage discrimination against women.

Table 2. Quantile regression coefficients modeling (QCRM) estimations.

	Intercept	p	p^2	p^3	p^4	p^5
	Dependent Variable: (log) Hourly Earnings					
Intercept	0.6817 (0.1164)	0.7888 ** (0.0407)	–0.0549 ** (0.0415)	0.3338 ** (0.0408)	–0.0088 ** (0.0316)	0.1354 ** (0.0294)
Is Woman	–0.1537 ** (0.0338)	–0.0662 ** (0.0126)	0.0194 ** (0.0131)	0.0044 ** (0.0130)	–0.01283 *** (0.0099)	0.0003 *** (0.0091)
Is Married	0.0649 ** (0.0340)	–0.0087 ** (0.0127)	0.01553 ** (0.0131)	0.0154 ** (0.0131)	–0.0001 *** (0.0099)	0.0015 *** (0.0092)
Is an American Citizen	0.1655 ** (0.0394)	0.0075 ** (0.0154)	–0.0213 ** (0.0160)	–0.0229 ** (0.0163)	–0.0109 ** (0.0123)	0.0015 ** (0.0117)
Proficiency in English	0.0619 ** (0.0372)	0.0884 ** (0.0141)	–0.0104 ** (0.0146)	–0.0004 ** (0.0145)	–0.0189 ** (0.0110)	–0.0083 ** (0.0104)
Age at Migration	–0.0035 *** (0.0025)	–0.0037 *** (0.0009)	0.0022 *** (0.0010)	–0.0004 *** (0.0010)	–0.0009 *** (0.0008)	–0.0007 *** (0.0007)
Years of Education	0.0314 *** (0.0074)	0.0122 *** (0.0024)	–0.0004 *** (0.0025)	0.0004 *** (0.0023)	0.0025 *** (0.0018)	0.0009 *** (0.0017)
Potential Experience	0.0056 *** (0.0019)	0.0048 *** (0.0008)	0.0022 *** (0.0008)	0.0012 *** (0.0008)	0.0019 *** (0.0006)	0.0011 *** (0.0006)

Standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$ indicate coefficients significant at the 10%, 5%, and 1% significance levels, respectively. Regression coefficients associated with any covariates, $\beta_i(p)$, including the intercept, based on a fifth-degree Legendre polynomial, $\beta_i(p) = \theta_0 + \theta_1 p + \theta_2 p^2 + \theta_3 p^3 + \theta_4 p^4 + \theta_5 p^5$; $p \in (0,1)$.
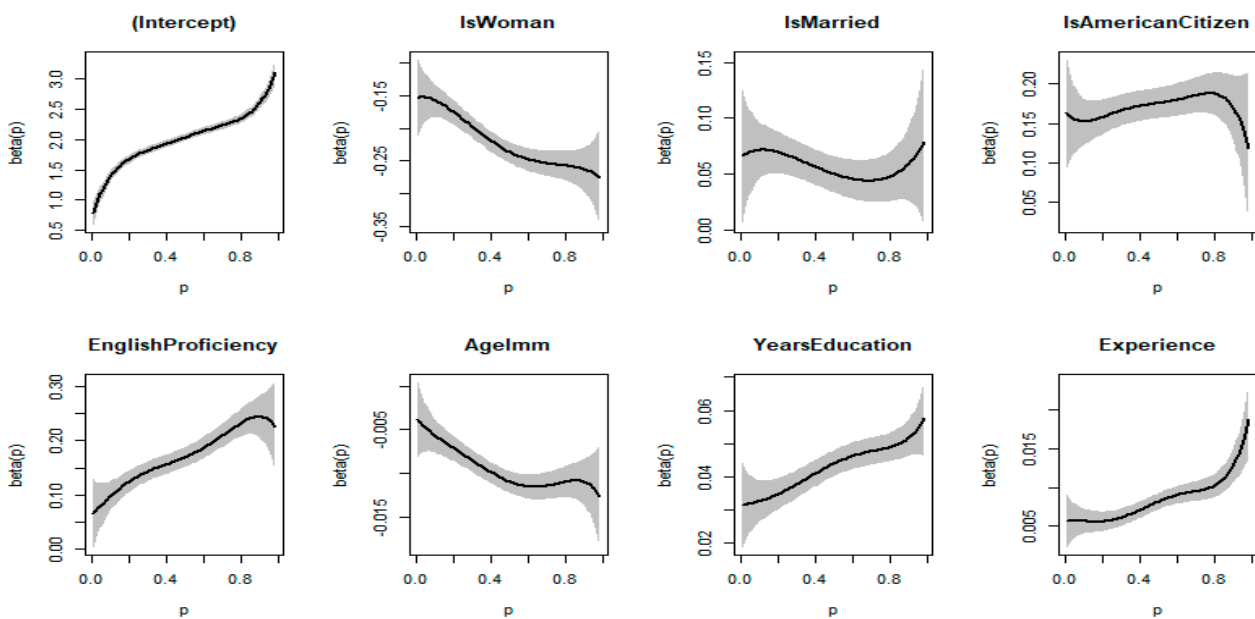
Figure 4. Quantile regression coefficients modeling (QRCM) estimation. 95% confidence intervals are shown in shaded areas.

Regarding age at time of migration, the effect was negative for all quantiles and highest (in absolute terms) at the upper part of the distribution, i.e., people who earn more. This result demonstrated that workers who were older at the time of migration were likely to receive lower earnings in the US than their younger counterparts. It also suggested that the age at which an individual migrates to the US could be a potentially important determinant of how that immigrant will eventually perform in the labor market [32].

The significantly positive coefficients for all quantiles of being an American citizen, being married, and being proficient in English were indicative of comparative advantages in favor of individuals who have these characteristics. The return to proficiency in English was higher for people who earned more; for the 90th percentile, speaking English well or very well was associated with a 25% increase in earnings, compared to around 7% for the

5th percentile. Being an American citizen had relatively homogeneous effects on hourly earnings from the 5th to the 80th percentiles, with a less significant impact among people who earned more. The curve associated with positive married status exhibited an inverted U-shape, meaning that the effect of this variable on hourly earnings was similar at the extremes of the distribution.

Years of educational attainment and a higher level of potential experience both had an increasing positive effect on earnings. This suggested that earnings for all individuals, regardless of their respective positioning in income distribution, were positively impacted by these variables. Specifically, the returns for education were higher at the top of the conditional earnings distribution but lower than expected. The authors of [20] found that, in 1995, the returns for education for male workers born in the US were 3.9% for the 10th percentile and 7.9% for the 90th percentile. This means that the returns for education for Cuban-born workers in the 2000s were similar to those for US-born workers 15–22 years ago.

In our second step, we used the CEC algorithm [23], with a view to finding similarities in the effects of the covariates on the (log) hourly earnings in the two groups. Specifically, we focused on the covariate associated with years of education, because education is considered a maximizer of individuals' skills and the primary provider of human capital [33]. In this sense, we established three clusters in each group to allow for comparability. The results are summarized in Figure 5.
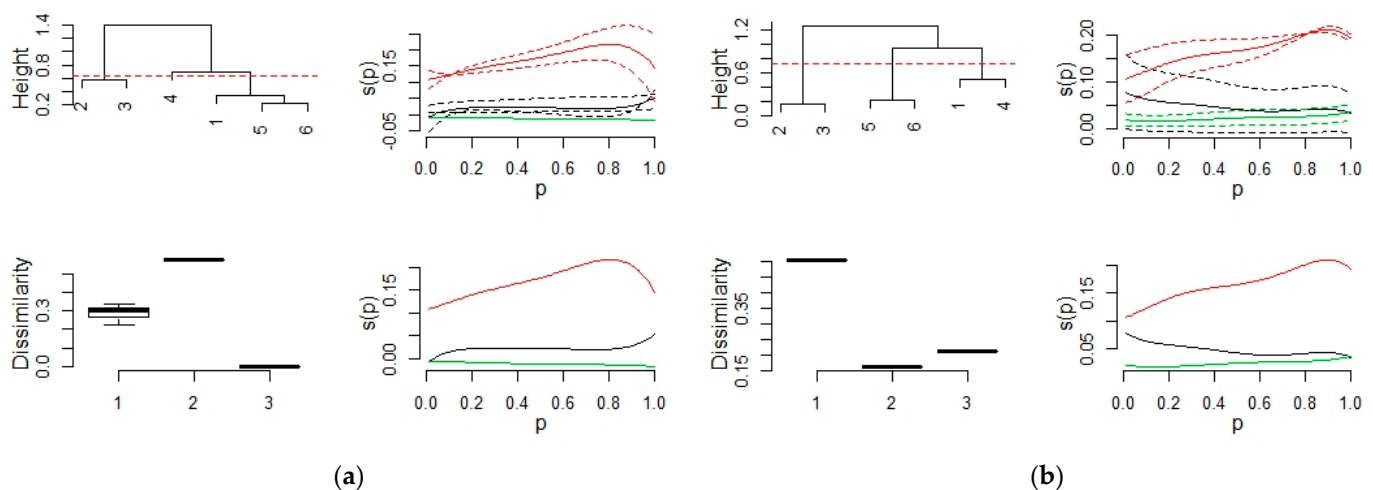
Figure 5. Output of the ClustEff algorithm. The left panel shows the dendrogram and the boxplot of the average dissimilarity within each cluster. The right panel shows the seven curves associated with each covariate, clustered in three groups. (**a**) (Women); (**b**) (Men).

The clusters can be described as follows:

1.　For Women

　　(i)　Cluster 1 contained the effect of the variables, indicating that the individual is a US citizen and proficient in English. This cluster had the greatest positive impact on earnings. This effect increased toward the higher locations of the earnings' distribution.

　　(ii)　Cluster 2 contained the effect of the curve associated with age at the time of migration. This cluster had an almost constant and negative effect throughout the conditional earnings' distribution.

　　(iii)　Cluster 3 contained the effects of variables associated with years of education and potential work experience, as well as the effect of the variable indicating that the individual is married. The cluster effect was positive and almost constant but close to zero. The positive impact was more important for individuals at the upper tail of the earnings' distribution.

2. For Men

(i) Cluster 1 contained the effects of the variables indicating that an individual is a US citizen and proficient in English, as for women. This cluster had the greatest positive impact on earnings. This effect increased toward the higher locations of the earnings' distribution. However, unlike what was observed for women, the trend change of the curve for the upper tail of the distribution was smoother.

(ii) Cluster 2 contained the effects of the curve associated with age at the time of migration and marriage status. The cluster effect was constant, close to zero, and basically conditioned by the negative impact of age at time of migration on earnings.

(iii) Cluster 3 involved the functions related to the covariates associated with years of education and potential work experience. The observed cluster effect was positive but almost constant and close to zero. The positive impact was more important for individuals at the bottom of the earnings' distribution.

The CEC estimation showed that in both groups, the functions associated with the covariates that mark an individual as a US citizen and fluent in English were included in the same cluster and shared no similarities in terms of form and distance with the remaining functions associated with other covariates. It is worth noting that this cluster had the greatest positive impact on earnings, well above the positive return of education on earnings. We also observed that, for women, the function associated with the covariate related to age at time of migration was selected as an independent cluster that was not similar in the shape of the curve or distance with any other cluster.

Educational attainment was included in different clusters for the two groups, despite the fact that, on average, women, and especially those who were in the third quartile of the distribution of years of education, had more education than men. This suggested that education acquired in Cuba was not a differentiating component in the transfer of human capital for Cuban women immigrating to the US. It is also important to note that the cluster in which education for women was included had fewer benefits for women earning the lowest salaries (that is, the poorest women). In the case of men, the cluster in which education was included was more favorable to those who earned lower wages. On the other hand, in the case of women in the cluster where education was included, the variable associated with marriage status was also included. This evidence could confirm what was proposed by the authors of [34], who stressed that, because of their family obligations, married women often face multiple constraints in their choice of labor supply. In such cases, the employer may exploit the inelasticity of the supply curve and pay workers below the competitive wage. Similarly, the authors of [35] also concluded that the female labor supply to a given firm may be less elastic than the male supply, and thus, the employer may pay them lower wages compared to men.

4. Conclusions

This study represents an important step in the analysis of gender discrimination among immigrants. Results were consistent with the notion of wage discrimination. With the aim of furthering our understanding of the transferability of skills (specifically education) of immigrants in the US, this study used quantile regression coefficients modeling (QRCM), which allowed us to model the quantile regression coefficient functions (p) as parametric functions that depended on the order of the quantile p. Moreover, we made use of the clustering of distribution curves (CEC) algorithm, based on a new method for finding similarities between the curves of effects of covariates on earnings. The use of these two QR methods marked a major contribution to the study of the process of immigrant assimilation in host countries. In particular, it enabled the use of flexible techniques to describe the process of transferability of human capital. We utilized a new perspective to study similarities in the effects of covariates on wages in a QR context. This new approach considered the shape of the curves and the distance between them.

Three findings were deemed particularly important. First, being a woman had the strongest negative impact on earnings and was associated with lower wages in all quantiles of the distribution, with the greatest negative impact at the central and upper parts of the earnings' distribution. Second, for women and men, education had a positive effect on earnings, even though it remained a non-differentiating element because its effect was similar to other variables included in the cluster. Moreover, in the case of women, education was in the same cluster as two other variables (potential work experience and marriage status), while in the case of men, education was in the same cluster as only one other variable (potential work experience). Third, for women, the function associated with the covariate related to age at time of migration was selected as an independent cluster that shared no similarities (in terms of form or distance) with the remaining functions associated with the other covariates.

Despite having higher educational attainment than men, women end up in occupations with lower compensation because of institutional bias. In order to achieve the levels of equality called for in the United Nations Sustainable Development Goals, nations must invest considerable resources to assure educational outcomes, and in addition, initiatives must be put in place to amend institutional practices and arrangements.

These findings confirmed that the transferability of skills among immigrant populations is gendered. If individuals do not bring their full potential in acquired and developed skills to the host society, full sustainability cannot be guaranteed. As was pointed out in [36], the process of women's assimilation and the recognition of their important role in society is imperative to attain sustainability. At present, what distinguishes organizations from each other are intangible resources that are capable of becoming necessary competitive advantages that are sustainable over time. Among these intangible resources, human capital plays a determining role in building a sustainable economy. Societies replacing women or not using them to their fullest capacity does not serve this goal [37].

References

1. Blau, F.D.A.K.; Lawrence, M. Gender Wage Gap: Extent, Trends, and Explanations. *J. Econ. Lit.* **2017**, *55*, 789–865. [CrossRef]
2. Becker, G.S. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 2nd ed.; University of Chicago Press: Chicago, IL, USA, 1980; pp. xviii, 268 and Ill.
3. Bergmann, B.R. Occupational Segregation by Sex: The Root of Women's Disadvantage. In *Economic Emergence of Women*, 2nd ed.; Palgrave Macmillan: New York, NY, USA, 2005; pp. 61–84.
4. Addison, J.T.; Ozturk, O.D.; Wang, S. The Occupational Feminization of Wages. *ILR Rev.* **2018**, *71*, 208–241. [CrossRef]
5. Magnusson, C. Mind the Gap: Essays on Explanations of Gender Wage Inequality. Ph.D. Thesis, Institute of Social Research, Stockholm University, Stockholm, Sweden, 2010.
6. England, P. *Comparable Worth Theories and Evidence*; Routledge: London, UK, 2017.

7.   Reid, L.L. Devaluing women and minorities—The effects of race/ethnic and sex composition of occupations on wage levels. *Work Occup.* **1998**, *25*, 511–536. [CrossRef]

8.   Cohen, P.N.; Huffman, M.R.L. Individuals, jobs, and labor markets: The devaluation of women's work. *Am. Sociol. Rev.* **2003**, *68*, 443–463. [CrossRef]

9.   Smith-Doerr, L.; Alegria, S.; Fealing, K.H.; Fitzpatrick, D.; Tomaskovic-Devey, D. Gender Pay Gaps in US Federal Science Agencies: An Organizational Approach. *Am. J. Sociol.* **2019**, *125*, 534–576. [CrossRef]

10.   Weeden, K.A. Why do some occupations pay more than others? Social closure and earnings inequality in the United States. *Am. J. Sociol.* **2002**, *108*, 55–101. [CrossRef]

11.   Kalleberg, A.L. Precarious Work, Insecure Workers: Employment Relations in Transition. *Am. Sociol. Rev.* **2009**, *74*, 1–22. [CrossRef]

12.   DiPrete, T.A.; Eller, C.C.; Bol, T.; van de Werfhorst, H.G. School-to-Work Linkages in the United States, Germany, and France. *Am. J. Sociol.* **2017**, *122*, 1869–1938. [CrossRef]

13.   Gradin, C. Segregation of Women into Low-paying Occupations in the United States. *Appl. Econ.* **2020**, *52*, 1905–1920. [CrossRef]

14.   US Census Bureau. *Geographical Mobility and Migration Main*; U.S. Department of Commerce: Washington, DC, USA, 2019.

15.   Massey, D.S. Immigration Policy Mismatches and Counterproductive Outcomes: Unauthorized Migration to the US in two eras. *Comp. Migr. Stud.* **2020**, *8*, 21. [CrossRef]

16.   Cobas-Valdés, A.; Sainz, A.F. Cuban Migration to the United States and the Educational Self Selection Problem. *Int. J. Cuba. Stud.* **2014**, *6*, 41–54. [CrossRef]

17.   Basiago, A.D. Sustainable development in Indonesia—A case-study of an indigenous regime of environmental-law and policy. *Int. J. Sustain. Dev. World Ecol.* **1995**, *2*, 199–211. [CrossRef]

18.   Mensah, J. Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature review. *Cogent Soc. Sci.* **2019**, *5*, 1653531. [CrossRef]

19.   Sawicka, J.; Lagoda, J. Gender and Sustainability in the Economic Development: Equal Chances for Women at the Labor Market. *Oeconomia* **2015**, *14*, 115–125.

20.   Martins, P.S.; Pereira, P.T. Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Econ.* **2004**, *11*, 355–371. [CrossRef]

21.   Kneib, T. Beyond Mean Regression. *Stat. Model.* **2013**, *13*, 275–303. [CrossRef]

22.   Haupt, H.; Lösel, F.; Stemmler, M. Quantile Regression Analysis and Other Alternatives to Ordinary Least Squares: A Methodological Comparison on Corporal Punishment. *Methodol. Eur. J. Res. Methods Behav. Soc. Sci.* **2014**, *10*, 81–91. [CrossRef]

23.   Sottile, G.; Adelfio, G. Clusters of effects curves in quantile regression models. *Comput. Stat.* **2019**, *34*, 551–569. [CrossRef]

24.   Frumento, P.; Bottai, M. Parametric Modeling of Quantile Regression Coefficient Functions. *Biometrics* **2016**, *72*, 74–84. [CrossRef]

25.   Koenker, R.; Bassett, G. Regression Quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]

26.   Ruggles, S.; King, M.L.; Levison, D.; McCaa, R.; Sobek, M. *Machine-Readable Database, Version 5.0*; In Integrated Public Use Microdata Series (IPUMS); IPUMS: Minneapolis, MN, USA, 2019.

27.   Davino, C.; Vistocco, D. *Quantile Regression for Clustering and Modeling Data. Advances in Statistical Models for Data Analysis*; Springer: Cham, Switzerland, 2015; pp. 85–95. [CrossRef]

28.   Geraci, M. Modelling and estimation of nonlinear quantile regression with clustered data. *Comput. Stat. Data Anal.* **2019**, *136*, 30–46. [CrossRef] [PubMed]

29.   Sottile, G.; Francipane, A.; Adelfio, G.; Noto, L.V. A PCA-based clustering algorithm for the identification of stratiform and convective precipitation at the event scale: An application to the sub-hourly precipitation of Sicily, Italy. *Stoch. Environ. Res. Risk Assess.* **2021**, 1–15. [CrossRef]

30.   Bertoli, S.; Fernandez-Huertas Moraga, J.; Ortega, F. Crossing the border: Self-selection, earnings and individual migration decisions. *J. Dev. Econ.* **2013**, *101*, 75–91. [CrossRef]

31.   *Technical Report*; US Bureau of Labor Statistics, Wage and Hour Division: Washington, DC, USA, 2020.

32.   Friedberg, R.M. *The Labor Market Assimilation of Immigrants in the United States: The Role of Age at Arrival*; Working Papers; Brown University: Providence, RI, USA, 1992.

33.   Fitzsimons, P. Human Capital Theory and Education. In *Encyclopedia of Educational Philosophy and Theory*; Springer: Cham, Switzerland, 2015; pp. 1–5.

34.   Ponthieux, S.; Meurs, D. Gender Inequality. *Handb. Income Distrib.* **2015**, *2*, 981–1146.

35.   Ransom, M.R.; Oaxaca, R.L. New Market Power Models and Sex Differences in Pay. *J. Labor Econ.* **2010**, *28*, 267–289. [CrossRef]

36.   Choi, S.H.; Jang, H. How Can Marriage Immigrants Contribute to the Sustainability of the Host Country? Implications from the Leisure and Travel Patterns of Vietnamese Women in South Korea. *Sustainability* **2021**, *13*, 1039. [CrossRef]

37.   Del Valle, I.D.; Castillo, M.A. Human capital and sustainable competitive advantage: An analysis of the relationship between training and performance. *Int. Entrep. Manag. J.* **2009**, *5*, 139–163. [CrossRef]