

# Ethnicity and Earnings in Urban Peru<sup>1</sup>

Hugo Nopo, Jaime Saavedra and Máximo Torero

## Abstract

In this paper we study the relationship between ethnic exclusion and earnings in Urban Peru. Our approach to the concept of ethnicity involves the usage of instruments in many of its several dimensions: mother tongue, parental background, religion, migration events and race. In order to approximate what can be called racial differences in a context like the Peruvian in which “racial mixture” is the main characteristic of the population, we use a score-based procedure to capture both the differences and the mixtures. By means of this procedure each individual is assigned intensities by pollsters in each of the four categories that correspond to the most easily recognized distinct racial groups in the Peruvian society: Asiatic, White, Indigenous, and Black.

We find that the multidimensional race indicator is correlated with several human capital and physical capital assets, as well as with access to public services. Using Blinder-Oaxaca (B-O) decompositions we find that a substantial part of the earnings differences between racial groups cannot be explained by differences in individual characteristics. To take into account the fact that B-O doesn't consider the probability distribution of the individual characteristics, and specifically race in our case, we also use a semi-parametric technique for the estimation of differences in hourly earnings. This estimation treats the typical wage equations in a linear fashion but let estimators for the racial intensity effects to interact freely, without restricting them to a functional form. The results suggests that among wage earners after controlling for a large set of characteristics, there are racially related earnings differences in favor of predominantly White individuals. In the case of the self-employed, none of the empirical distributions of earning differences attributable to race is substantially above zero.

Keywords: Race discrimination, minorities, wage differentials, semi-parametric.

JEL Classification: J15, J31, J71.

---

<sup>1</sup> Paper prepared as part of the global project on Social Exclusion in Peru undertaken with the support of the Inter American Development Bank. The authors would like to thank Gissele Gajate and Martin Moreno for their excellent assistance and Jorge de la Roca for his collaboration during the survey. Comments from Santiago Cueto, Fernando Andrade, Javier Escobal, and other researchers at GRADE; Dante Contreras, Juan José Díaz, Luojia Hu and Christopher Taber; participants at the 2001 Annual Meeting of the Network of Inequality and Poverty; and a seminar held at the IDB helped us significantly to improve this work. Nopo (h-nopo@northwestern.edu) is at Northwestern University. Saavedra (Jaime@grade.org.pe) and Torero (mtorero@grade.org.pe) are Senior Researchers at GRADE.

## 1. Introduction

The outcomes of discrimination and exclusion related to ethnicity, culture, physical appearance and religion are very notorious. At the same time, the mechanisms by which those operate are rather subtle. Indigenous or ethnic minorities are more likely to be poor than any other group. While overall the poverty rate is 54%, according to the 2000 Living Standards Measurement Survey (LSMS), the poverty rate of the population whose mother's tongue is Quechua, Aymara or other native language is 70%. Moreover, more than 75% of this group can be found in the three bottom deciles of the income distribution.

Ethnic and race discrimination in Peru has been studied, usually through the analysis of case studies, rather than in a systematic general approach with some pretension of statistical significance. Ethnic discrimination has also been studied, usually through case studies. Callirgos(1993) gives a global overview of the origins and particular characteristics that Peruvian racism may have. Oliart (1989), Pozzi-Escot (1989), Callirgos (1993), and Mendoza (1993) tried to tackle ethnic and cultural discrimination. Finally, using case studies Sulmont (1995) has documented some of the elements of social exclusion that may be present in Peruvian labor markets.

Despite the obvious importance of the topic for a country like Peru, there are very few data sources that can capture ethnic discrimination, and empirical work that tries to tackle exclusion and discrimination issues from a quantitative perspective is scarce. Most of the empirical literature has approximated racial and ethnic discrimination with supposedly easily observable variables, in most of the cases, mother tongue as in several of the World Bank studies. MacIsaac (1993), for instance, finds that more than 80% of non-indigenous people - defined as those whose mother tongue is Spanish- have access to public water supply or access to electricity while less than 45% of indigenous people -defined as those whose mother tongue is Quechua or Aymara- has access to the same type of public goods; and that years of schooling is 8.1 for non-indigenous people while it is only 5.5 for indigenous people, just to mention two of the most striking differences. Psacharopoulos and Patrinos (1994) find that *ceteris paribus*, individuals whose mother tongue is Quechua have earnings that are 8% lower than the average of the population.

However, the approximation of ethnicity using mother tongue is clearly incomplete, as there are other ethnic differences within the Spanish and Quechua speaking populations<sup>2</sup>. In this paper, in measuring the effect of ethnic exclusion over earnings, we improve over the measuring of discrimination by approximating ethnicity using variables related to several dimensions of the concept, such as mother tongue, parental background, race, and religion. All these variables are linked to differences among individuals that may have measurable consequences over economic opportunities.

One of the major concerns when measuring race is that the Peruvian population is neither predominantly Indigenous nor White nor Black nor Asian, but is a continuum of different degrees of mixtures that is difficult to treat empirically. In order to approximate what can be

---

<sup>2</sup> Other attempts to approximate ethnic characteristics are found in some household surveys in Peru. As an example, the 2000 Living Standards Measurement Survey (LSMS 2000) inquires about racial characteristics, and more than 98% of the rural populations is self-reported as "*Mestiza*" (mixed race), and only a tiny percentage of the sample self-report themselves as Indigenous, Asian, Black or White.

called racial intensities, a score-based procedure is used. In such procedure every individual receives a score (in an ordinal scale ranging from 0 to 10) for each of the four categories representing the groups that are most easily recognized by people as distinct racial groups: Asian, White, Indigenous, and Black. We construct indicators based on self-reported data and pollster's data.<sup>3</sup> Several papers have looked at the effects of racial differences in a similar fashion. For instance; Arce, Murguía and Frisbie (1987) use in their study two phenotypical dimensions: skin color, ranging from very light to very dark and physical features, from Very European to Very Indian, both on a 1 to 5 scale. Johnson and Farrell (1995) look at differences in income between light-skinned and dark-skinned black males in Los Angeles. Darity and Mason (1998) review studies that use "skin shade", i.e. dark skinned black males with light skinned black males as categories of analysis. Using Latin American data, Silva (1985) compares earnings of blacks, mulatos and whites in Brazil, and Telles and Lim (1998) use a classification of black, *morenos* and whites to analyze earning differentials in Brazil combining self-reported data with data reported by pollsters.

The instruments that proxy ethnicity are then used in two econometric set-ups in order to explore their interaction with earnings. First, in a completely parametric set-up, we compare earnings among three groups of individuals (predominantly White, predominantly indigenous and Mestizos)<sup>4</sup> computing racial wage gaps, and decomposing those gaps into explained and unexplained terms according to the Blinder-Oaxaca methodology as in Blinder (1973) and Oaxaca (1973). The second set-up does not classify the individuals into racial groups but estimates the treatment effects associated with different intensities in the racial ordinal scales by means of a semi-parametric specification as in Robinson (1988).

The rest of the paper is organized as follows: In Section 2 we introduce our approach to measure ethnicity in Peru, showing our findings for the newly created data. Section 3 shows an analysis of the interplay among race and ethnic variables with a set of characteristics related to the performance of individuals in the labor market. Sections 4 and 5 are devoted to the two econometric setups we propose to measure the role of ethnicity on earnings: the parametric analysis using the Blinder-Oaxaca decompositions and the semi-parametric analysis with intensity treatments respectively. Finally conclusions are presented in Section 6.

## 2. Measuring Ethnicity

In anthropology literature, the concept of ethnicity is defined in a general way as the community of individuals that share cultural elements and that organize their daily life around it. Generally, it is associated with the idea of native communities that are isolated or that have reduced contact with other communities. In urban settings, ethnic characteristics are associated, in a complex and passionately debated interaction, with culture, religion, language, traditions and race, among other dimensions.

As mentioned above, we use information on mother tongue, religion and parental background to approximate ethnic differences. A more complex issue is the use of race indicators as an additional dimension of ethnicity. Several disciplines debate the complex interplay that exists between race and ethnicity. Here we just recognize that race, together with other ethnic

---

<sup>3</sup> See Angel and Gronfein (1998) and Anderson, Silver and Abramson (1998) for previous use of this methodology.

<sup>4</sup> We did not concentrate on Asian and Black characteristics because the sample sizes for these populations are relatively small in the LSMS.

characteristics, may generate differences among people that may have measurable consequences over economic opportunities. In order to approximate racial characteristics a score-based procedure is used, in which each individual received a score for each of the four categories: Asiatic, White, Indigenous, and Black; which are groups that are more easily recognized by people as distinct racial groups. Scores were given by each individual and independently by the pollster. This score ranges from 0 to 10 in each category; with zero meaning that the individual did not have physical characteristics that resembled a typical individual of the respective racial group and 10 that she had mostly characteristic features of that group. With this multi-dimensional racial intensity indicator, we are able to characterize a person as a *Mestizo*, but within those *Mestizos*, there is still racial variability<sup>5</sup>.

The self-report of race has been used in other countries with some success due to the fact that the classification of races in other places tends to be easier or more direct (see Hirshman and Alba, 1998 and Telles and Lim, 1998). However, in Peru, the majority of the population tends to define themselves as “*Mestiza*”, category that includes people who actually have very different characteristics and are perceived by the others also as very different. The use of a second source of information, as it is the pollsters’ perception is a technique supported by arguments as those exposed by Angel and Gronfein (1988) and Anderson, Silver and Abramson (1998). Even though this method might be also subjected to various criticisms, we tried to reduce the problems associated to the inter-observer variability with an intensive pre-fieldwork training, as suggested by Boergerhoff-Mulder and Caro (1985)<sup>6</sup>.

There are three additional aspects through which we capture ethnic characteristics. First, we use language, a variable that typically has been used as the sole indicator of ethnicity in many labor studies in Latin America. Here we use language information for the individual and for his/ her parents. The second aspect is migration. In the data set, we consider both short-term migration as well as migration from place of origin, which is important due to the migratory process that has taken place in Peru in the last 50 years. Finally, since the religion of the individual might be relevant, we consider this variable as well.

### *The Data.*

The data used comes from the urban households of the LSMS survey for 2000 and from an additional module carried out by GRADE in 2001. The latter was applied to a significant fraction of urban household members eighteen years or older and was designed to explore in depth racial and ethnic characteristics<sup>7</sup>. A novel and interesting feature of this additional

---

<sup>5</sup> For certain econometric procedures, it was also useful to dichotomize our measure in order to identify three different groups: “Indigenous”, “Whites” and “*Mestizos*” (Also Asian and Black were identified but the sample sizes were too small).

<sup>6</sup> In addition, following Allport and Kramer (1946), Scodel and Austrin, (1957), and Toch, Rabin and Wilkins (1962); we used pictures as instruments to standardize the pollster’s reports. We carried out an intensive training in order to minimize inter-observer variability, homogenizing scoring criteria among all pollsters.

<sup>7</sup> The module covered 70% of the original people surveyed (5,700 individuals). The lost by attrition of 30% of the cases did not represent significant differences in the main individual and household characteristics between this sub-sample and the total sampled population. In addition, the special module elaborated for this project included questions regarding physical characteristics, linguistic uses, geographic origin, religious habits, and information related to parents (mother tongue, geographic origin, religion and education). The survey also included a section dedicated to questions about credit, access to social capital of the people surveyed and cultural consumption. Finally, questions related to discrimination episodes were included.

module is the way the “race” variable was surveyed. As previously mentioned, the race of each individual had been considered as a four-dimensional vector (White, Indigenous, Black and Asian) with an ordinal measure of intensity, ranging from 0 (lowest) to 10 (highest), in each dimension independently.<sup>8</sup> For example, a predominantly white individual could be one with intensities 7, 1, 0 and 1 for the categories White, Indigenous, Black and Asian respectively. An example of a predominantly indigenous individual could be one with intensities 2, 8, 0 and 1 in the same dimensions. With this feature, we are able to capture better the racial diversity and the different degrees of “mestizaje” present in the Peruvian society.

Graph 1 shows the distribution of the population according to the intensities in each race category, using the perceptions of the pollsters. As observed, the White intensity distribution is skewed towards the right, suggesting that the majority of individuals are characterized by pollsters as having some white characteristics, but are not predominantly white; on the other hand the indigenous intensity distribution is more skewed towards the left. Because populations with a strong Asian or Black ancestry are relatively small, the LSMS is not necessarily representative of these groups. Still, a small number of individuals have, according to pollsters’ perceptions, racial features that resemble observable characteristics of these groups.

An immediate implication of the distribution of the population according to racial intensities will be our inability to establish statistical regularities regarding these latter groups. Therefore, we will concentrate the discussion on the consequences on earnings of racial differences among people with white and indigenous traits. In some statistical application where it is informative to divide the sample into groups, we use predominantly White, predominantly Indigenous and *Mestizo* as analytical categories.

As it is documented in the literature and shown in Graph 2, there are significant differences between the race variable self-reported by the individuals surveyed and the same variable reported by the pollster (for example see Tellez (1998) for the Brazilian case).

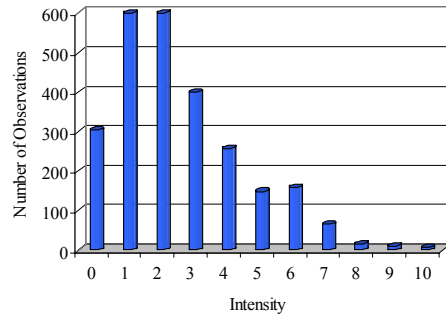
---

<sup>8</sup> We did not impose ex-ante any condition on the values that the 4-dimensional vectors of racial characteristics may attain. That is, we left room for any of the  $11 \times 11 \times 11 \times 11 = 11^4$  possible combinations of race intensities. Interestingly, the results obtained show low variability in terms of the sum of the four racial intensities for each individual.

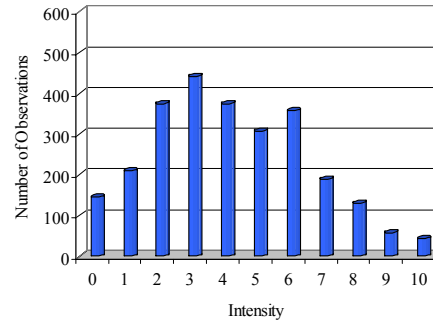
# Graph 1

## Racial Intensity Distributions in Urban Peru - Pollster's Perception

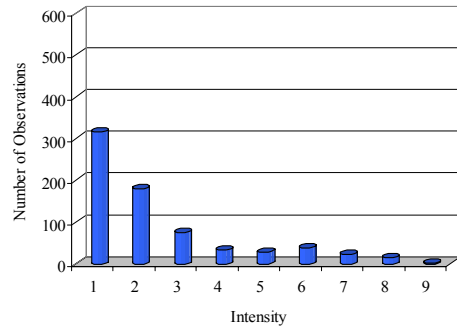
Indigenous Intensity Distribution



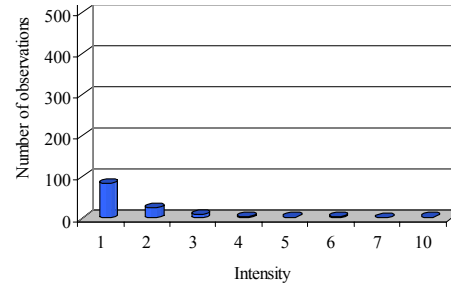
White Intensity Distribution



Asian Intensity Distribution



Black Intensity Distribution

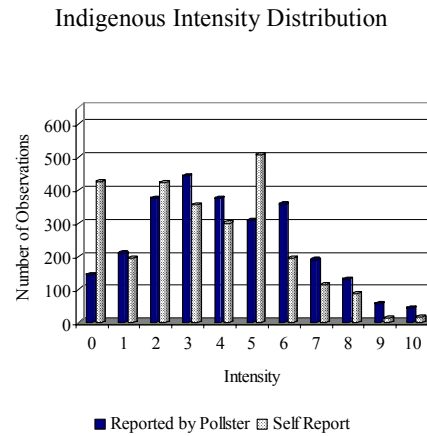
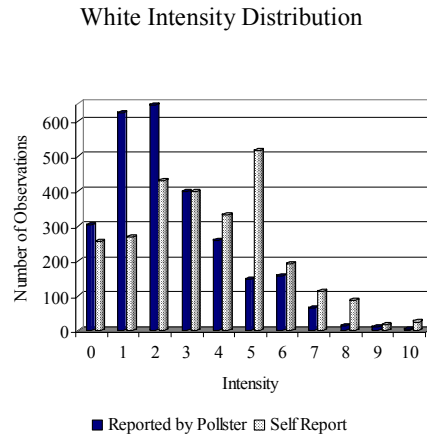


Elaboration: Own

Source: Living Standard Measurement Survey (ENNIV 2000) and Additional Ethnic Module

Note: We are not reporting "0" for the intensity distributions in the Asian and Black dimension given that this value represent the vast majority of observations.

**Graph 2**  
**Comparison between Race Intensity as Reported by Pollster and Self Reported**



The self-reported White intensity distribution is skewed to the right of the same distribution reported by the trained pollster, while the self-reported indigenous intensity distribution is skewed to the left of the same distribution reported by the pollster. Overall, respondents tend to score themselves with higher values of white intensity and lower values of indigenous intensities than pollsters. That is, individuals consider themselves “less indigenous” than they are actually perceived by pollsters.

Given that the paper's main objective is to identify labor market exclusion given observable ethnic characteristics and not self-exclusion, we will concentrate on the pollster scores rather than the self reported scores. However, differences between both types of scores could have strong implications on the quantification of the racial earnings gaps.<sup>9</sup>

<sup>9</sup> As it is shown by Telles and Lim (1998) for the Brazilian case, while the White-Brown gap is around 26% using the pollster perception, it is reduced to 17% if the self-report is used (both gaps are calculated controlling for human capital and labor market characteristics).

We are relying on the fact that people associate the words White, Indigenous, Black and Asian with different sets of phenotypic characteristics. Given other individual traits, these characteristics as perceived by a third person may or may not be associated with other socioeconomic variables or outcomes. Race, together with other ethnic-related characteristics, may have real effects in the labor markets, which we will try to approximate here, though we do not dwell into the specific sociological or economic mechanisms that may cause these effects.

### 3. Ethnicity and Individual Characteristics

Table 1 shows descriptive statistics for a set of demographic and ethnic-related variables emphasizing their interaction with race intensity indicators along the White and Indigenous dimensions. Quintiles are defined dividing the sample according to the ranking implicit in the intensity scores given by the pollster. Individuals perceived as predominantly White report higher levels of education, a smaller family size and fewer children. Individuals that are predominantly Indigenous are less educated and have more children. Individuals perceived as having more Indigenous features report more frequently that their mother tongue is a native language, and are slightly more likely to be Christian non catholic, and are much more likely to be migrants. Regarding parental background, as the White intensity increases, mother's education is higher and the likelihood of the mother having a native language as a mother tongue is lower. Moreover, when we look at labor-related characteristics as the log of hourly income, the quintiles with higher White intensity had higher incomes than the quintiles with higher Indigenous intensities. Also, among the whiter quintiles there are more professionals and technicians, as well as executive staff.

Graph 3 presents the relationship between racial intensities and some key variables.<sup>10</sup> It should be noted that the pollster records her own perception about racial intensities independently of the answers that the respondent gives about her characteristics<sup>11</sup>. Years of schooling are positively correlated with the White intensity indicator and negatively correlated with the Indigenous intensity. Similarly, the same pattern is observed regarding attending a private educational institution, access to phone lines and access to health insurance. On the other hand, migrant status and family size are positively correlated with the Indigenous intensity indicator and negatively correlated with the White intensity one.

---

<sup>10</sup> We do not report White intensities of 9 and 10 as the number of observations for these cells is too small.

<sup>11</sup> Moreover, the main LSMS survey was applied a few months before people were re-interviewed for the additional ethnic module.

**Table 1 Descriptive Statistics by Quintiles of Racial Intensity**

	White					Indigenous					Total
	Quintile of Racial Intensity					Quintile of Racial Intensity					
	First	Second	Third	Fourth	Fifth	First	Second	Third	Fourth	Fifth	
<b>PERSONAL AND FAMILY CHARACTERISTICS</b>											
Gender (Male)	61.0%	63.0%	57.0%	54.0%	55.0%	55.0%	56.0%	58.0%	61.0%	61.0%	58.0%
Age (Years)	41.4	39.2	38.4	37.7	39.5	38.1	38.3	38.2	39.3	42.8	39.1
Marital Status (Married)	72.0%	66.0%	64.0%	59.0%	58.0%	57.0%	62.0%	60.0%	69.0%	73.0%	63.0%
Education (Years of schooling)	8.8	9.8	10.8	10.9	11.3	11.5	10.8	10.3	9.8	8.8	10.4
Family size	5.9	5.7	5.6	5.6	5.3	5.5	5.7	5.7	5.6	5.7	5.6
Number of children	2.9	2.7	2.7	2.8	2.5	2.5	2.7	2.8	2.8	2.8	2.7
Racial Diversity Measure	2.7	2.5	2.5	2.6	3.1	3.2	2.8	2.5	2.5	2.5	2.6
<b>ETHNIC NON - LABOUR RELATED CHARACTERISTICS</b>											
<b>MOTHER TONGUE</b>											
Spanish	82.0%	89.0%	91.0%	92.0%	96.0%	97.0%	97.0%	94.0%	82.0%	78.0%	90.0%
Native language	18.0%	11.0%	9.0%	8.0%	4.0%	3.0%	3.0%	6.0%	18.0%	22.0%	10.0%
<b>RELIGION</b>											
Catholic	87.0%	90.0%	90.0%	89.0%	92.0%	93.0%	92.0%	89.0%	87.0%	87.0%	89.0%
Christian non Catholic	10.0%	9.0%	8.0%	9.0%	6.0%	5.0%	7.0%	10.0%	11.0%	11.0%	9.0%
Other religions	2.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	2.0%	1.0%
No religion	1.0%	0.0%	0.0%	1.0%	2.0%	2.0%	1.0%	1.0%	1.0%	1.0%	1.0%
<b>BIRTHPLACE</b>											
Migrant	43.0%	40.0%	36.0%	37.0%	35.0%	31.0%	35.0%	38.0%	43.0%	46.0%	39.0%
Born in a rural or semi rural area	16.0%	14.0%	12.0%	11.0%	8.0%	8.0%	10.0%	11.0%	15.0%	19.0%	13.0%
<b>MOTHER'S ETHNIC CHARACTERISTICS</b>											
<b>EDUCATION</b>											
Primary School	55.0%	66.0%	74.0%	75.0%	84.0%	86.0%	79.0%	72.0%	62.0%	49.0%	71.0%
Secondary School	11.0%	16.0%	22.0%	26.0%	39.0%	38.0%	26.0%	20.0%	17.0%	8.0%	22.0%
College	2.0%	3.0%	4.0%	6.0%	11.0%	10.0%	6.0%	4.0%	4.0%	1.0%	5.0%
<b>MOTHER'S MOTHER TONGUE</b>											
Native Language	46.0%	30.0%	21.0%	21.0%	11.0%	8.0%	13.0%	21.0%	38.0%	54.0%	26.0%
<b>BIRTH PLACE</b>											
Born in a rural or semi rural area	24.0%	21.0%	19.0%	16.0%	13.0%	16.0%	15.0%	17.0%	22.0%	24.0%	19.0%

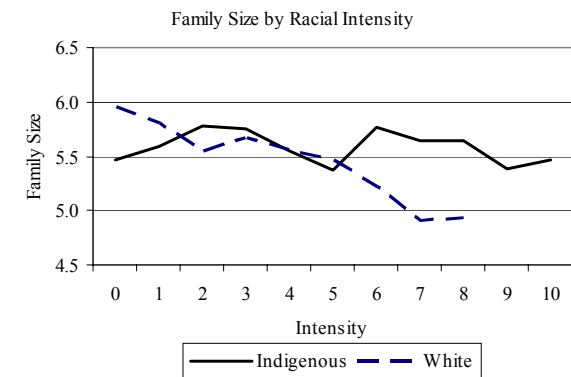
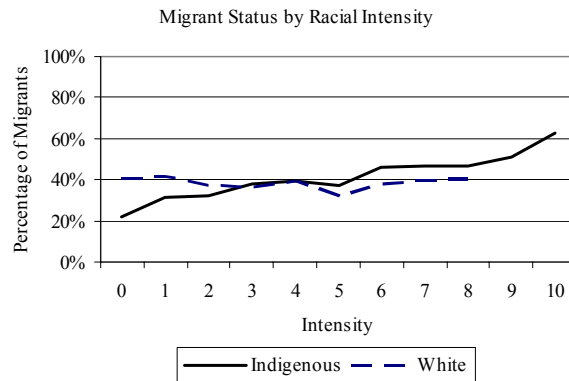
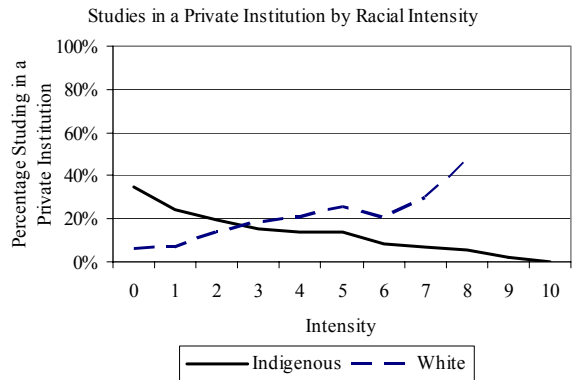
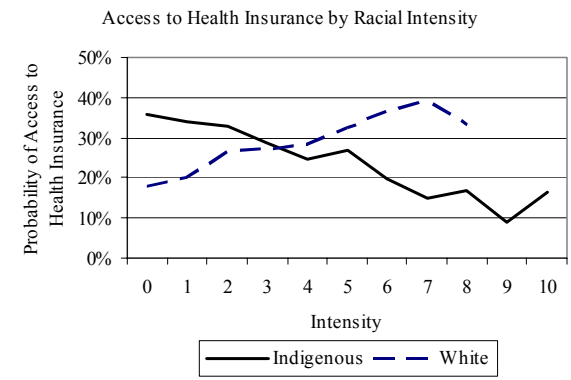
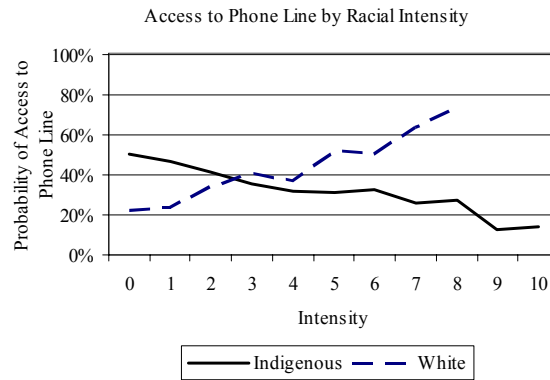
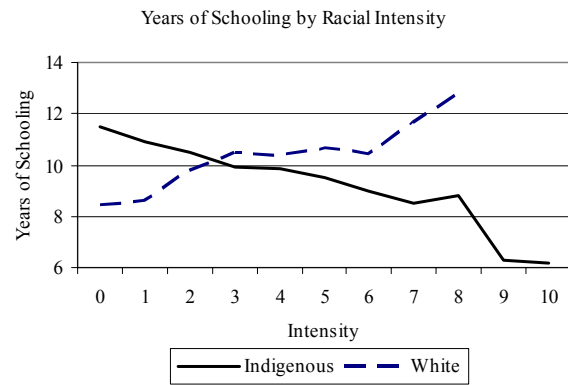
(...continued)

	White					Indigenous					Total
	Quintile of Racial Intensity					Quintile of Racial Intensity					
	First	Second	Third	Fourth	Fifth	First	Second	Third	Fourth	Fifth	
<b>LABOR RELATED CHARACTERISTICS</b>											
Ln of hourly income	0.6	0.9	0.9	1.0	1.1	1.1	0.9	0.9	0.8	0.7	0.9
Tenure (Years)	9.3	8.1	7.8	7.3	8.2	7.6	7.8	7.1	8.3	10.4	8.1
<b>TYPE OF EMPLOYMENT</b>											
Private wage earner	31.0%	34.0%	35.0%	39.0%	38.0%	39.0%	40.0%	38.0%	31.0%	27.0%	35.0%
Public wage earner	13.0%	17.0%	18.0%	15.0%	16.0%	16.0%	16.0%	16.0%	17.0%	14.0%	16.0%
Self - employed	54.0%	46.0%	45.0%	45.0%	44.0%	43.0%	41.0%	45.0%	49.0%	57.0%	46.0%
<b>OCCUPATIONS</b>											
Professionals and technicians	11.0%	18.0%	22.0%	21.0%	28.0%	26.0%	23.0%	19.0%	18.0%	14.0%	20.0%
Officials and Managers	2.0%	1.0%	1.0%	1.0%	3.0%	3.0%	2.0%	1.0%	1.0%	2.0%	2.0%
Executive staff	3.0%	4.0%	5.0%	7.0%	9.0%	10.0%	6.0%	5.0%	3.0%	3.0%	6.0%
Merchants and Salespersons	27.0%	24.0%	26.0%	30.0%	23.0%	24.0%	25.0%	26.0%	27.0%	29.0%	26.0%
Workers of the Services	13.0%	14.0%	12.0%	14.0%	13.0%	11.0%	12.0%	16.0%	14.0%	12.0%	13.0%
Agricultural workers	13.0%	7.0%	7.0%	4.0%	4.0%	3.0%	5.0%	7.0%	9.0%	12.0%	6.0%
Non agricultural workers	31.0%	31.0%	26.0%	24.0%	21.0%	25.0%	27.0%	26.0%	29.0%	28.0%	26.0%
<b>ECONOMIC ACTIVITIES</b>											
Extractive activities	12.0%	6.0%	6.0%	4.0%	4.0%	3.0%	5.0%	7.0%	9.0%	11.0%	6.0%
Mining	0.0%	1.0%	1.0%	0.0%	0.0%	0.0%	1.0%	1.0%	0.0%	1.0%	1.0%
Manufacturing consumption goods	9.0%	10.0%	9.0%	9.0%	9.0%	9.0%	10.0%	8.0%	9.0%	8.0%	9.0%
Manufacturing capital goods	5.0%	5.0%	3.0%	4.0%	5.0%	5.0%	4.0%	4.0%	4.0%	3.0%	4.0%
Construction	6.0%	5.0%	4.0%	5.0%	5.0%	4.0%	5.0%	5.0%	5.0%	6.0%	5.0%
Trade	33.0%	31.0%	34.0%	36.0%	31.0%	33.0%	32.0%	31.0%	33.0%	36.0%	33.0%
Electricity	9.0%	9.0%	8.0%	9.0%	8.0%	9.0%	8.0%	9.0%	8.0%	8.0%	8.0%
Finances	3.0%	4.0%	4.0%	5.0%	9.0%	8.0%	5.0%	4.0%	3.0%	3.0%	5.0%
Other services	24.0%	30.0%	31.0%	28.0%	28.0%	28.0%	29.0%	31.0%	29.0%	23.0%	29.0%
<b>FIRM SIZE</b>											
1 worker	2.0%	3.0%	1.0%	2.0%	1.0%	1.0%	1.0%	2.0%	3.0%	1.0%	2.0%
2 to 10 workers	16.0%	17.0%	21.0%	19.0%	19.0%	18.0%	23.0%	18.0%	17.0%	15.0%	18.0%
11 to 50 workers	9.0%	9.0%	7.0%	11.0%	11.0%	12.0%	8.0%	11.0%	8.0%	7.0%	9.0%
51 or more workers	4.0%	6.0%	6.0%	7.0%	7.0%	8.0%	7.0%	6.0%	4.0%	4.0%	6.0%

Elaboration: Own

Source: Living Standard Measurement Survey (ENNIV 2000) and Additional Ethnic Module

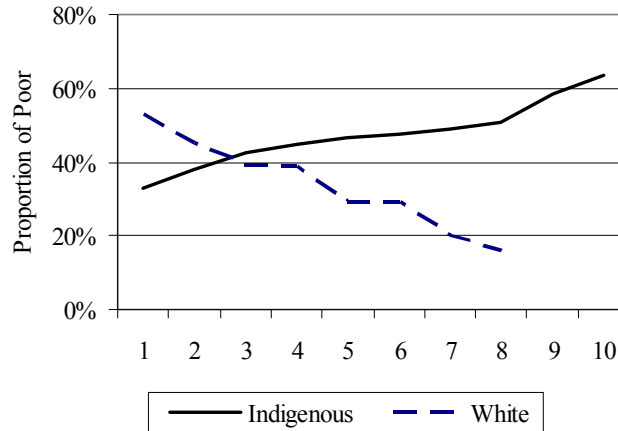
### Graph 3 Characteristics by Racial Intensity



Elaboration: Own  
Source: Living Standard Measurement Survey (ENNIV 2000) and Additional Ethnic Module

Graph 4 shows the relationship between racial intensities and poverty. The horizontal axis indicates the race intensities as perceived by the pollster and the vertical axis shows the proportion of poor individuals. It is clear that the higher the intensity of White the less poor are households and the higher the intensity of Indigenous the poorer they are.

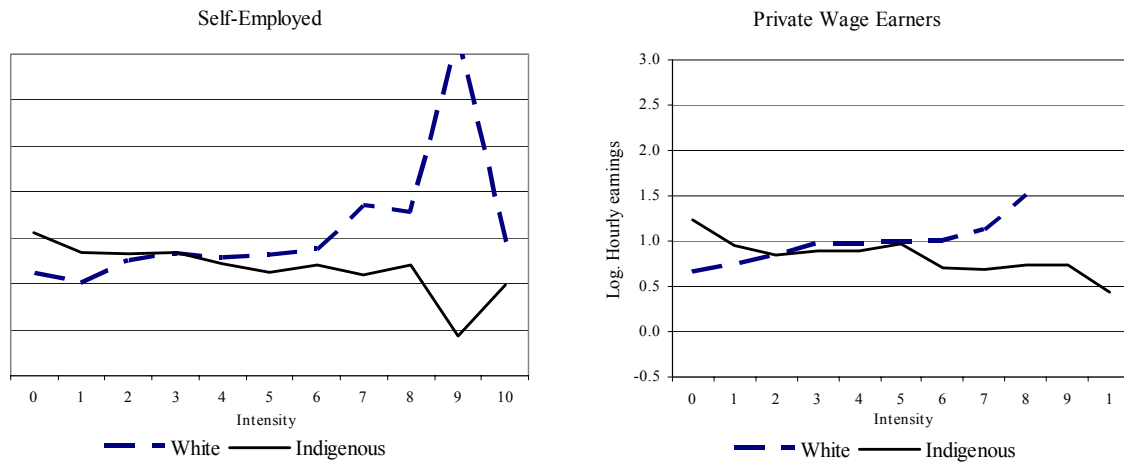
**Graph 4**  
**Poverty Index by Racial Intensity**



Analysing raw averages for the self-employed and private wage earners, the log hourly wage is positively correlated with the White intensity indicator and negatively correlated with the Indigenous intensity. In both cases and at the same time the average levels of earnings are lower for the self-employed than for the private wage earners (see Graph 5).

Mother tongue seems to have a strong correlation with raw earnings but only among the self-employed. Also, among the self-employed there is a small difference in earnings depending on migrant status. There are differences in log hourly wages between workers of different religions. Finally, being born in a rural area is correlated with lower earnings, independent of the type of job, as shown in Table 2. All these variables, however, may be correlated with human capital and demographic variables, so further analysis is required, as presented in the next two sections.

**Graph 5**  
**Hourly Earnings by Racial Intensity and Type of Employment**



Elaboration: Own

Source: Living Standard Measurement Survey (ENNIV 2000) and Additional Ethnic Module

**Table 2**  
**Log hourly earnings by other ethnic characteristics and type of employment**

	Private Wage	
	Earners	Self Employed
Average of Log hourly earnings	0.89	0.70
Mother Tongue		
Spanish	0.87	0.76
Native language	0.82	0.46
Religion		
Catholic	0.88	0.74
Christian non catholic	0.80	0.53
Other religions	1.02	0.45
No religion	0.85	1.08
Migrant Status		
Migrant	0.88	0.76
Non migrant	0.87	0.69
Birtplace		
Born in rural or semirural area	0.67	0.33
Born in an urban area	0.89	0.79

Elaboration: Own

Source: Living Standard Measurement Survey (ENNIV 2000) and Additional Ethnic Module

**4. Race and earnings: Decomposing the earnings gap using a parametric approach.**

In this section we analyze the racial earnings gaps to determine to what extent these can be explained by differences in individual characteristics that the labor market rewards (provided that, as we have seen in the previous section, these characteristics differ substantially according to race). We use the traditional Blinder-Oaxaca decomposition,

based on the estimation of earnings equations.<sup>12</sup> Such decomposition works by separating the earnings gap into two additive terms, one that is attributed to differences in average observable characteristics between the groups being compared and another attributed to a mix of the existence of unobservable characteristics and discrimination in the labor market.

Given that the race variable is measured as intensity rather than as a taxonomical classification of individuals into groups, there is no natural classification of the population into clearly differentiated comparison groups, as required by the Blinder-Oaxaca approach. For that reason, arbitrarily, we partition the sample into three racial groups: (predominantly) Whites, (predominantly) Indigenous and (neither White nor Indigenous) Mestizos. In order to proceed with such partition we need a criterion to define the “frontiers” among racial groups. That choice of appropriate frontiers is not a trivial task. After all, in some way we will “clusterize” observations, and as it is widely documented, a potential lack of information is an important issue to take into account.

Given the ordinal nature of the race intensity variables, we determine the frontiers by means of a cut-off rule defined in the following way:

- If an individual has her/his Indigenous intensity variable greater than or equal to a cut-off “ $c$ ” and her/his White intensity variable smaller than the same cut-off “ $c$ ”, she/he will be considered as an Indigenous.
- Analogously, if an individual has her/his White intensity variable greater than or equal to a cut-off “ $c$ ” and her/his Indigenous intensity variable smaller than the same cut-off “ $c$ ”, she/he will be considered as a White.
- An individual that is considered neither Indigenous nor White will be considered as a Mestizo.

Having specified a cut-off rule like this, and given that the race intensity variable takes only integer values from zero to ten in the four dimension where it is defined, we are able to explore empirically the whole set of alternatives for cut-off points<sup>13</sup>. Also, the higher the cut-off the more pronounced the differences in race intensities between groups, and hence, the higher the number of individuals classified as mestizos (or equivalently, the lower the number of individuals classified as Indigenous or Whites). We also find, empirically, that the higher the cut-off values the higher the wage gap between Indigenous and Whites<sup>14</sup>. Taking these issues into account, we decided to explore and report our results for three different cut-off values: 3, 4 and 5.

In the Peruvian labor market, we find statistical evidence that the main variables playing roles in the earning equations, as hourly wage, education and occupations are distributed

---

<sup>12</sup> Blinder (1973) and Oaxaca (1973).

<sup>13</sup> It should be noted that not all integer values between 0 and 10 may work as reasonable cut-off values. The cut-off can not be too high so that few individuals will have race intensities above it. On the other hand, it cannot be too low so that few individuals will have race intensities below it (that is, too high or too low cut-off values may imply no White or no Indigenous individuals in the proposed classification).

<sup>14</sup> This is because at higher cut-off values we are comparing “extremely White” individuals with “extremely Indigenous” individuals.

differently among the self-employed than among the private wage earners. On the other hand, interactions in the labor market are substantially different for both groups. While the former is facing directly the final consumers of their products or services, the latter is facing its own and intermediary firms, which as a consequence of their profit maximization objectives might have different mechanisms through which discrimination may operate. For these reasons we decided to perform our analysis separately for self-employed and private wage earners. In Table 3 we show the analysis for these two groups and the three partitioned criteria.

**Table 3**  
**Sample distribution and Average Earnings for Different Partition Criteria**

**A. Distribution of the sample (%)**

	Self Employed			Private Wage-Earners		
	Criterion 5	Criterion 4	Criterion 3	Criterion 5	Criterion 4	Criterion 3
White	11	15	16	13	17	19
Mestizos	46	33	32	56	43	38
Indigenous	43	52	53	31	39	43
Total	100	100	100	100	100	100

**B. Average Log of Hourly Wage**

	Self Employed			Private Wage-Earners		
	Criterion 5	Criterion 4	Criterion 3	Criterion 5	Criterion 4	Criterion 3
White	0.9972	0.9461	0.9474	1.1714	1.0945	1.0732
Mestizos	0.7899	0.8099	0.8007	0.8977	0.9201	0.9529
Indigenous	0.6024	0.6223	0.6296	0.7890	0.7876	0.7734

**C. Earning Gaps**

	Self Employed			Private Wage-Earners		
	Criterion 5	Criterion 4	Criterion 3	Criterion 5	Criterion 4	Criterion 3
White vs. Indigenous	0.3948	0.3239	0.3178	0.3824	0.3069	0.2998
White vs. Mestizos	0.2072	0.1362	0.1467	0.2737	0.1744	0.1203
Mestizos vs. Indigenous	0.1876	0.1876	0.1711	0.1087	0.1325	0.1795

Among the self-employed, the results obtained from cut-offs 3 and 4 are rather similar in terms of shares of the population that result from the partition. With a stricter criterion of 5 as a cut-off value<sup>15</sup>, we identify a distribution that is 11% Whites and 43% Indigenous in the population. In the private wage sector figures are different. While a light cut-off criterion of 3 leads us to 19% of Whites and 43% of Indigenous, a stricter cut-off

<sup>15</sup> As mentioned previously, a higher cut-off value corresponds to more pronounced differences in racial intensities.

criterion of 5 provides us figures of 13% and 31% of Whites and Indigenous respectively. A clear regularity is that the share of Indigenous workers is larger among the self-employed.

Differences in the average logarithm figures for hourly wages according to different criteria are small; the average wages obtained from cut-offs 3 and 4 are very similar while the average wages of the White group obtained with the cut-off 5 are slightly higher. The racial wage gap (Whites vs. Indigenous) is on average 32% for self-employed individuals and 31% for private wage earners using criteria 3 and 4 as cut-offs, while the usage of criterion 5 for cut-off results in racial wage gaps of 39% and 38% for self-employed and private wage earners respectively. For all cut-off criteria, raw measures of average earnings for the *Mestizo* group were between the other two groups and relatively equidistant to both<sup>16</sup>.

Next, the average earnings gaps between racial groups are decomposed into two parts: a component explained by the differences in individual characteristics and another by the differences in returns to those characteristics. In order to explore the differences in characteristics between the groups being compared, we include in the regressions three sets of variables, human capital variables (years of schooling, years of experience at the same occupation and its square), personal and family variables (sex, age, age squared, days sick, religion, migrant status, mother's educational level, mother born in rural areas, an index of racial diversity and a dummy for social networks) and labor market variables (dummies for economic sector, occupations, firm size and the inverse Mills ratio correction for selectivity bias of being working as a self-employed or as a private wage earner)<sup>17</sup>.

---

<sup>16</sup> Also, we find that log hourly earnings in the wage sector are between 11% and 19% higher than in the self-employment sector for all racial groups and cut-off criteria; simultaneously, the percentage of indigenous participation is higher among the latter. This may be consistent with a sorting mechanism that plays against the indigenous population, segregating them in the lower earnings sector.

<sup>17</sup> Some dummy variables that resulted statistically non-significant (at a 10% significance level) in the estimation of the earnings equations were ignored.

**Table 4 - Earnings Gap Decomposition**

	Criterion 5			Criterion 4			Criterion 3		
	White vs. Indigenous	White vs. Mestizos	Mestizos vs. Indigenous	White vs. Indigenous	White vs. Mestizos	Mestizos vs. Indigenous	White vs. Indigenous	White vs. Mestizos	Mestizos vs. Indigenous
<b><u>Self-Employed</u></b>									
<i>Wage Gap</i>	<i>0.3948</i>	<i>0.2072</i>	<i>0.1876</i>	<i>0.3239</i>	<i>0.1362</i>	<i>0.1876</i>	<i>0.3178</i>	<i>0.1467</i>	<i>0.1711</i>
<b>Differences in Characteristics (%)</b>	<b>0.207</b>	<b>0.132</b>	<b>0.114</b>	<b>0.197</b>	<b>0.089</b>	<b>0.127</b>	<b>0.175</b>	<b>0.074</b>	<b>0.145</b>
Education and Experience	0.074	0.056	0.033	0.063	0.026	0.032	0.035	0.073	0.013
Personal and Family	0.062	0.037	0.045	0.057	0.003	0.069	0.066	0.000	0.077
Labor	0.072	0.039	0.036	0.077	0.061	0.026	0.073	0.001	0.055
<b>Differences in Returns (%)</b>	<b>0.187</b>	<b>0.075</b>	<b>0.074</b>	<b>0.127</b>	<b>0.047</b>	<b>0.060</b>	<b>0.143</b>	<b>0.073</b>	<b>0.026</b>
<b><u>Private Wage Earners</u></b>									
<i>Wage Gap</i>	<i>0.3824</i>	<i>0.2737</i>	<i>0.1087</i>	<i>0.3069</i>	<i>0.1744</i>	<i>0.1325</i>	<i>0.2998</i>	<i>0.1203</i>	<i>0.1795</i>
<b>Differences in Characteristics (%)</b>	<b>0.244</b>	<b>0.155</b>	<b>0.046</b>	<b>0.194</b>	<b>0.139</b>	<b>0.069</b>	<b>0.156</b>	<b>0.128</b>	<b>0.055</b>
Education and Experience	0.054	0.046	0.038	0.044	0.026	0.033	0.031	0.036	0.016
Personal and Family	0.013	0.012	-0.007	-0.009	0.028	-0.002	-0.026	0.048	-0.015
Labor	0.178	0.098	0.015	0.159	0.084	0.039	0.151	0.044	0.054
<b>Differences in Returns (%)</b>	<b>0.138</b>	<b>0.118</b>	<b>0.063</b>	<b>0.113</b>	<b>0.036</b>	<b>0.063</b>	<b>0.144</b>	<b>-0.007</b>	<b>0.125</b>

Note: Education and experience include years of schooling, years of experience at the same occupation and its square). Personal and family variables include sex, age, age squared, days sick, religion, migrant status, mother's educational level, mother born in rural areas, an index of racial diversity and a dummy for social networks. Labor market variables include dummies for economic sector, occupations, firm size and the inverse Mills ratio correction for selectivity bias

Table 4 shows, for different cut-off criteria, the earnings gap decomposition from the pairwise comparison of three groups (White, Indigenous and Mestizos) according to the sets of characteristics previously mentioned. Among the self-employed about two thirds of the gap is explained by differences in characteristics, while the rest is explained by differences in returns and unobservables (or discrimination). Labor-related characteristics (economic activity, occupation and firm size) constitute the more important explanatory category. Occupational differences are an important source of earnings differentials, which suggests that some sort of occupational segregation has significant correlation with the wage gap. Summarizing, among the self-employed, there is an important share of the earnings gap not explained by characteristics and that could be related to discrimination or unobservable differences. Also, with these decompositions we show that characteristics differ dramatically and explain a large part of the earnings differentials, suggesting that some exclusion mechanisms may operate at pre-labor market stages as well as in the labor market.

Among wage earners, the largest part of the earnings gap is explained by differences in characteristics of the three groups. However, when we analyze the decomposition within these differences, it appears that personal and human capital related characteristics are important explanatory variables for the earnings gap between *Mestizos* and Indigenous, but labor markets characteristics are important in explaining the gap between Whites and *Mestizos*.

Among the labor-related characteristics that we are controlling for, the variables that are driving the results are those related to the economic activity and occupation of the individuals. In particular, services related occupations are playing an important role in the determination of this share of the gap for the self-employed, while non-service related occupations are the determinants of this share of the gap for the private wage earners<sup>18</sup>. This may be suggesting some sort of occupational segregation that has significant correlation with the wage gap. It can be argued that behind the labor-related characteristics of the individuals (e.g. employment, occupation and firm size) there are also some sorting mechanisms operating in the labor market that discriminate against certain groups.

Emphasizing these potential sorting mechanisms in the labor market we decided to explore further the occupational segregation using the Duncan Index of occupational segregation.<sup>19</sup> Again, the computation was made for the three cut-off points we have been considering.

We find lower levels of occupational segregation by race among the self-employed than among the private wage earners. While the Duncan Index for Whites vs. Indigenous is in the range 0.22-0.25 for self-employed individuals, it is in the range 0.40-0.46 for private wage earners. Although it should be noted these different levels of occupational segregation have different implications on the wage gaps; the components of the wage gap that can be explained by differences in occupations are roughly the same for self-employed and private wage earners. That is, for the same levels of wage differentiation due to occupational segregation, we find significantly higher levels of segregation in the private sector. Another look at the segregation indices, considering the *Mestizos* population, will show us another interesting fact among both the self-employed and the private wage earners: the segregation

---

<sup>18</sup> The details of these results are available from the authors.

<sup>19</sup> For a discussion of occupational segregation as well as for a contemporaneous analysis of this issue in three Latin-American countries, see Deutsch et al (2001).

Roughly, the Duncan Index of Occupational Segregation can be interpreted as the percentage of individuals of one of the comparing groups that would have to change their occupations (to another in which their racial group is under-represented) such that the labor force can show similar empirical distributions of individuals according to occupations for the two groups being compared, that is, in order to obtain a level of zero segregation.

between Whites and Mestizos is more pronounced than the segregation between Mestizos and Indigenous<sup>20</sup>.

**Table 5. Duncan Index of Occupational Segregation**

		Self-Employed	Private Wage Earners
<b>Criterion 5</b>	White vs. Indigenous	0.2498	0.4586
	White vs. Mestizos	0.2606	0.3371
	Mestizos vs. Indigenous	0.1677	0.2504
<b>Criterion 4</b>	White vs. Indigenous	0.2212	0.4109
	White vs. Mestizos	0.2251	0.3074
	Mestizos vs Indigenous	0.1673	0.2315
<b>Criterion 3</b>	White vs. Indigenous	0.2172	0.4020
	White vs. Mestizos	0.1693	0.3127
	Mestizos vs. Indigenous	0.1439	0.2666

The Blinder-Oaxaca approach has been an object of criticism as it takes into account only average values of individual characteristics without considering the probability distributions of such characteristics. Specifically, the Blinder-Oaxaca decomposition compares the average wages of the whole sample of Whites and Indigenous, being the case that not all individuals in these samples are “comparable”. There are White individuals with individual characteristics that are not reached by Indigenous (for example, high levels of education and many years of experience) as well as there are some Indigenous individuals with personal characteristics such that there is no White individual comparable to them (for example, individuals with no education and many years of experience). As it is pointed out by DiNardo et al (1996), Barsky et al (2001) and Nopo (2002), the misspecification resulting from the failure of recognizing these facts could lead to the overestimation of the “differences in returns” component, similar to the lack of an adequate control group when measuring the effect of treatment on the treated. Coincidentally, their strategies to overcome these issues rely on non-parametric techniques, minimizing the potential failures in capturing the correct linear specification. Here, we propose, as detailed in the next section, a semi-parametric approach that treats some variables in the earnings equations in a linear fashion but let some others (specifically, the race intensity variables) to interact freely, without restricting them to a functional form. This allows us to better measure the difference between comparable individuals as well as to better identify the impact of intensities of race over wage differentials.

<sup>20</sup> Given the importance of the Black population in Urban Peru and having the presumption that this group is facing the same sort of exclusion mechanisms that the Indigenous population faces in the labor market, we also considered a second exercise of partitioning the population in which we added the Black population to the Indigenous. The criterion to classify an individual as Black is analogous to the criterion defined above (by means of cut-off values) and hence, the criterion to classify an individual as Mestizo changes in the natural way: an individual is considered Mestizo if she/he is neither White nor Indigenous nor Black. We did not find statistically significant differences between the gaps computed from the comparison of Whites vs. Indigenous and from the comparison of Whites vs. the combination of Indigenous and Blacks. These two groups are equally under-performing in the labor market in terms of wages and this result is consistent for all cut-off criteria. The interested reader will find the details of our results for the comparison between Whites and the combination of Indigenous and Blacks in the Annex 1.

## 5. Ethnic characteristics and earnings. A semi-parametric approach

To formally estimate the differentials in per-hour earnings according to race and ethnic differences controlling for individual characteristics, we used a semi-parametric technique that allowed us to obtain linear parametric estimators for the typical Mincerian wage equations coefficients and non-linear, non-parametrical, estimators for the racial intensity effects.

Since our race indicators are constructed in ordinal scales, it is not possible to treat them parametrically performing arithmetic operations (which means we are not able to introduce multiplicative interacted effects, nor general polynomials of race effects in a typical earnings equation). Thus, we use a semi-parametric technique for the estimation of differences in hourly earnings according to race and ethnic differences that allow us to obtain linear parametric estimators for the typical wage equations and non-linear, non-parametrical, estimators for the racial intensity effects. In the linear component of the wage equation we include gender, age, age squared, tenure, tenure squared, marital status, years of schooling, occupation and economic sector controls, firm size dummies, social networks variables, and a correction for selectivity bias in the decision to work in the wage sector or as a self-employed.

The non-parametrical estimators for the effects explained by race require some additional explanation.<sup>21</sup> Given the observed distribution of race intensities in the four dimensions and the lack of observations in the Black and Asian categories, we decided to consider only the White and Indigenous dimensions and construct two variables for each individual. These new variables will be denoted as  $Z_W$  and  $Z_I$ , and for each individual they represent the intensity quintile in which the individual is situated in the White and Indigenous intensity distribution respectively. In both cases, the first quintile is the one with the lowest intensity. In this way, if an individual is in the first intensity quintile according to the White distribution and in the fourth intensity quintile according to the Indigenous distribution, he will receive the treatments 1 and 4 for  $Z_W$  and  $Z_I$  respectively.

With the construction of these two variables we proceeded to estimate the model:

$$y = \beta x + \varphi(z_W, z_I) + \varepsilon \quad \dots(1)$$

Where  $y$  is the hourly earnings rate, the first term of the right-hand side constitutes the linear specification of a typical earnings equation in which  $x$  represents individual characteristics and  $\beta$  represents the returns to these characteristics; the second term is the non-parametrical estimator for differences in the hourly earnings rate accrued to racial differences of the individuals,  $\varphi$ , which can be interpreted as a “combination of racial quintiles” fixed effect; finally, the third element is an error term.

Next, we summarize the procedure to estimate (1). Applying conditional means to both sides of equation (1) and assuming that the expected value of the error term conditional on any quintiles combination is zero, we obtain

$$E[y | z_W, z_I] = \beta E[x | z_W, z_I] + \varphi(z_W, z_I) \quad \dots(2)$$

---

<sup>21</sup> There is one parametric alternative: using dummy variables for each possible combination in the “four-dimensional space of race”. Unfortunately it would be necessary to give up many degrees of freedom and, as we will see by the end of this section, there are also some statistical limitations to this approach.

Additionally, subtracting equation (2) from equation (1) the “racial component”,  $\varphi(Z_W, Z_I)$ , will vanish and hence we will have the transformed equation

$$y - E[y | z_W, z_I] = \beta(x - E[x | z_W, z_I]) + \varepsilon \quad \dots(3)$$

This is a linear equation in deviations from conditional means that can be estimated by OLS (with no intercept). Let  $b$  be the OLS estimator obtained for equation (3). Finally, from (2), it is possible to recover  $\varphi(Z_W, Z_I)$ :

$$\varphi(z_W, z_I) = E[y | z_W, z_I] - bE[x | z_W, z_I] \quad \dots(4)$$

It has been shown by Robinson (1988) that the linear estimators obtained by using this procedure are root- $n$  consistent<sup>22</sup>. For the non-linear component of our equation it is possible to construct an empirical joint distribution for all the possible 25 effects by means of a re-sampling technique as bootstrap.<sup>23</sup> Having constructed this empirical distribution we are able to test hypotheses about the significance of the difference between any pair of different effects. For this purpose, at every re-sampling exercise we compute the differences between any two effects and find the empirical distribution of these new random variables. The empirical probability of having positive values for such difference variables will constitute our bootstrap estimators for the confidence levels.

As an example, say we are interested in the difference between race effects for individuals at the fifth quintile in the White intensity distribution and simultaneously at the first quintile in the Indigenous intensity distribution (that is, those pronouncedly “White”) with respect to individuals at the first quintile in the White intensity distribution and simultaneously at the fifth quintile in the Indigenous intensity distribution (those pronouncedly “Indigenous”). We proceed as follows: for each bootstrap iteration we compute the difference between effects,  $\varphi(5,1) - \varphi(1,5)$ , such that we are able to obtain as many pseudo-realizations of the difference as we have bootstrap iterations. With these pseudo-realizations we construct the empirical distribution for the variable we are interested in and, in particular we can find the (empirical) probability of having such a variable with a positive value.

There is a parametric specification that is “almost” equivalent to the semi-parametric approach proposed in this section. It involves the usage of 24 dummies, one for each quintile combination (and taking one of those quintile combinations as a base group). By means of this specification it is possible to obtain point estimators and standard errors for the effect of each quintile combination (that is, estimators that will have the same interpretation as  $\varphi(Z_W, Z_I)$  in the semi-parametric specification). However, we are interested not only in those point estimators, but also in the difference of effects between any two groups (pair differences of these point estimators). Since the point estimators obtained by the usage of  $\varphi$  and by the usage of the set of dummies have the same expected values, the estimators for the differences in effects will have also the same expected value. The differences arise in the estimation of standard errors for these estimators. The typical arguments in favor of the bootstrapping techniques over the usage of asymptotic theory apply here. Namely, computing the empirical distributions of the effects allows us to take into account not only the first and second moments of the random variables (as the asymptotic theory works) but also their higher

<sup>22</sup> Provided our model trivially satisfies assumptions (i)-(x) of Theorem on page 939 in Robinson (1988).

<sup>23</sup> For a general discussion of the bootstrap technique see Efron (1991, 1993) or Horowitz (2001).

moments that in some applications happen to be significant. Partial results of the parametric part of our estimations are shown in Table 6.

**Table 6 (continuation)**  
**Urban Peru: Selected Parametric Components of the Wage Equation for Private Wage Earners and Self-Employed**

	Private Wage Earners			Self-Employed		
	Without controls by race	Includes controls by quintile of race intensity	Semi-parametric estimation	Without controls by race	Includes controls by quintile of race intensity	Semi-parametric estimation
<b>ETHNIC NON - LABOR RELATED CHARACTERISTICS</b>						
<b>MOTHER TONGUE</b>						
Speak native language	0.008 (0.08)	-0.025 (0.23)	-0.023 (0.21)	-0.060 (0.65)	-0.053 (0.57)	-0.067 (0.69)
<b>RELIGION</b>						
No religion	-0.233 (1.03)	-0.195 (0.85)	-0.207 (0.89)	0.215 (0.84)	0.208 (0.80)	0.245 (0.90)
Christian non Catholic	0.030 (0.38)	0.016 (0.19)	0.008 (0.09)	-0.039 (0.44)	-0.022 (0.25)	-0.023 (0.25)
Other religions	0.099 (0.43)	0.106 (0.46)	0.113 (0.48)	-0.206 (0.90)	-0.207 (0.90)	-0.177 (0.74)
<b>MIGRANT STATUS AND BIRTHPLACE</b>						
Migrant	-0.019 (0.39)	-0.021 (0.43)	-0.024 (0.48)	0.184 (3.27)**	0.176 (3.10)**	0.192 (3.25)**
Born in a rural or semi rural area	-0.003 (0.03)	-0.010 (0.11)	-0.019 (0.21)	-0.205 (2.24)*	-0.212 (2.29)*	-0.226 (2.35)*
<b>RACIAL DIVERSITY MEASURE</b>						
	-0.033 (2.79)**	-0.034 (2.73)**	-0.036 (2.88)**	-0.027 (1.84)	-0.028 (1.83)	-0.027 (1.68)
<b>MOTHER'S ETHNIC CHARACTERISTICS</b>						
Mother's Mother Tongue	0.009 (0.16)	0.045 (0.76)	0.053 (0.88)	-0.004 (0.06)	0.025 (0.33)	0.037 (0.47)
Mother was born in a rural or semi rural area	-0.015 (0.21)	-0.010 (0.15)	-0.011 (0.15)	-0.084 (1.06)	-0.082 (1.02)	-0.089 (1.07)
<b>Mother's Education</b>						
Primary School	0.029 (0.47)	0.022 (0.35)	0.026 (0.42)	0.161 (2.48)*	0.158 (2.42)*	0.185 (2.72)**
Secondary School	0.141 (2.57)*	0.129 (2.32)*	0.132 (2.35)*	0.165 (1.88)	0.137 (1.52)	0.126 (1.35)
College	0.013 (0.13)	0.004 (0.04)	-0.004 (0.04)	0.041 (0.25)	0.058 (0.36)	0.056 (0.33)
Observations	1080	1080	1080	1422	1422	1422
Adjusted R-squared	0.29	0.30	0.27	0.17	0.17	0.17

Note: Includes controls for gender, age, age squared, tenure, tenure squared, marital status, years of schooling, occupation and economic sector controls, firm size dummies, social networks variables, and a self selection correction for the decision to work in the wage sector or as a self-employed. t-statistics between parenthesis.

In the first column we show results without any racial controls; in the second we include racial controls using simple dummies that represent the quintile of the race intensity quintile to which the individual belongs in the White and Indigenous scale; the third column show the results including the non-parametric component explained above. We do not find any effects on ethnic-related variables such as religion, birthplace, or native tongue<sup>24</sup> over earnings. So

<sup>24</sup> This is somewhat surprising, as in the previous literature, native tongue implies a negative premium in earnings equations. This result holds when we use the whole sample of the LSMS 2000, but disappears when we

the earnings gaps that we find when analysing raw data disappear when demographic and human capital controls are introduced. There is, however, a positive effect of mother's education. In the case of wage earners, what makes a difference in earnings is the fact that the mother reached secondary education, while among the self-employed, the threshold seems to be only primary education. A racial diversity indicator shows that among wage earners, the more racially diverse the household, the lower the salary of the individual.<sup>25</sup> The migrant status has a positive and significant effect on earnings only among the self-employed, while being born in rural or semi-rural areas have a negative effect over earnings but only among the self-employed. Interestingly, this effect does not vary with the inclusion of race-related variables.

In Graph 6 we report the non-parametric effects of belonging to different racial intensity groups over earnings; after controlling for personal characteristics, other ethnic variables, mother's characteristics, occupation, sector of economic activity and firm size. Since the domain of the function that measures the racial intensity effects is two-dimensional,  $(Z_W, Z_I)$ , a complete report of the expected values of combined racial effects will require plotting a three-dimensional surface for every earnings equation (one for private wage earners and one for self-employed individuals). In order to avoid that visual complexity, we are reporting our estimators for the expected value of the function  $\varphi(Z_W, Z_I)$  at different transversal sections of the function (at  $Z_W=1$ ,  $Z_W=3$  and  $Z_W=5$  in the left panels and at  $Z_I=1$ ,  $Z_I=3$  and  $Z_I=5$  in the right panels).

Among private wage earners (upper panels) the earnings effect is larger for workers in the fifth White intensity than for those in the first White intensity. Among the latter, the effect is smaller the higher they are in the Indigenous intensity scale. Looking at the same effect but for different levels of the Indigenous scale, it appears that the earnings effect increases with the White intensity scale. Among the self-employed (lower panels), no clear patterns emerge.

The significance level of the differences observed in the previous graphs can be estimated by the bootstrap technique. We compute empirical distributions for several pairs of differences between earnings effects. The distributions are shown in Graph 7. Among wage earners, the difference  $\varphi(5,1) - \varphi(1,5)$ , where  $\varphi(5,1)$  is the earnings effect for the predominantly White and  $\varphi(1,5)$  is the earnings effect for the predominantly Indigenous, is significantly greater than zero with a confidence of 97.1%. The difference  $\varphi(4,2) - \varphi(2,4)$ , is above zero in 37% of the bootstrapped cases and the difference in racial effects between predominantly White individuals and Mestizos,  $\varphi(5,1) - \varphi(3,3)$ , is positive in 75.2% of the cases. This suggests that after controlling for a large set of characteristics, there are racially-related earnings

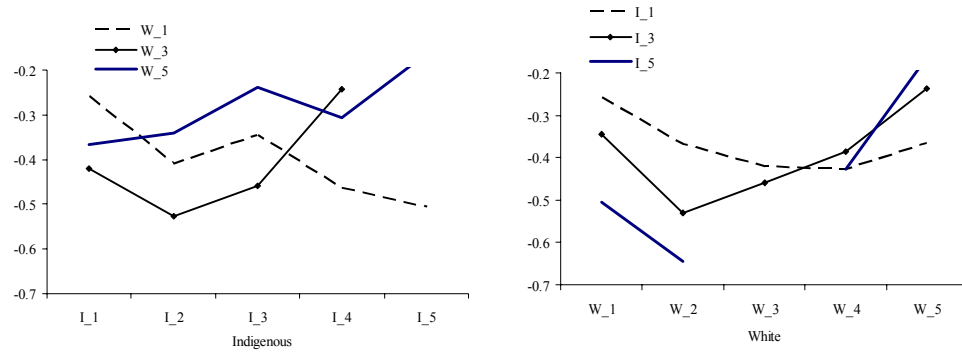
---

limit the sample to urban areas. This could be because in most urban regions the language used is Spanish rather than native which is essentially present in rural regions.

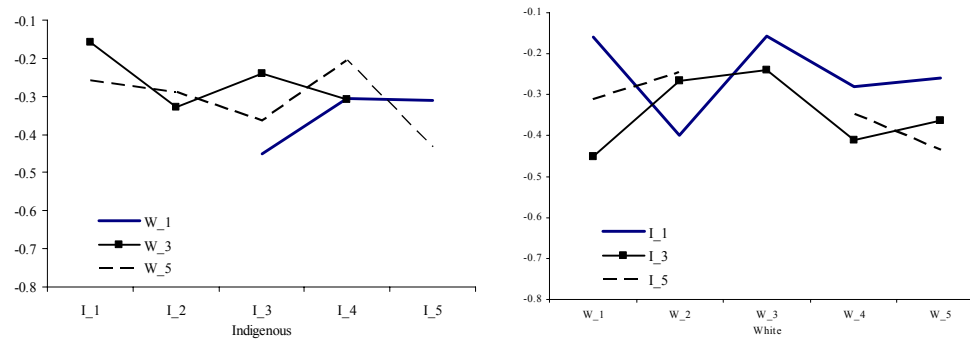
<sup>25</sup> Based on the information about race intensities reported by the pollsters, we constructed a racial diversity indicator measuring degrees of heterogeneity at every household. We proceeded in this way under the beliefs that individuals from racially diverse households suffer also some sort of exclusion. Interestingly, this measure turned to be negative and significant among the wage earners and negative but not significant among the self-employed.

differences in favor of predominantly White individuals. In the case of the self-employed, none of the empirical distribution of differences statistically differs from zero in any case.

**Graph 6**  
**Non-Linear Estimation**  
**Private Wage Earners**

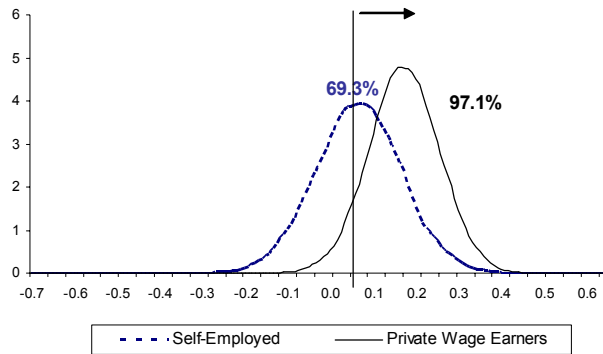


**Self-Employed**



## Graph 7

Empirical Distribution for the Difference of Racial Effects  
(White=5, Indigenous=1) compared to (White=1, Indigenous=5)



Empirical Distribution for the Difference of Racial Effects  
(White=4, Indigenous=2) compared to (White=2, Indigenous=4)



Empirical Distribution for the Difference of Racial Effects  
(White=5, Indigenous=1) compared to (White=3, Indigenous=3)



## 6. Conclusions

Given that Peru is an extremely diverse country where ethnic groups cannot be easily identified, we approximate the ethnic diversity of the country using a large set of variables like language, religion, origin and race. The use of race indicators in a country like Peru is complex. Here we recognize that race, together with other ethnic characteristics, may generate differences among people that may have measurable consequences with regard to economic opportunities. In order to approximate racial characteristics we use a score-based procedure in which each individual received a score for each of the four categories: Asiatic, White, Indigenous, and Black; which are groups that are more easily recognized by people as distinct racial groups. Scores were self-reported by each individual and were also assigned, independently, by the pollster. With this multi-dimensional racial intensity indicator, we are able to characterize a person as a *Mestizo*, but within those *Mestizos*, there is still racial variability that we can capture and explore.

The self-reported White intensity distribution is skewed to the right of the same distribution reported by the trained pollster, while the self-reported Indigenous intensity distribution is skewed to the left of the same distribution reported by the pollster. Overall, respondents tend to score themselves with higher values of White intensity and lower values of Indigenous intensities than pollsters score them. The empirical analysis shows that our race indicators are clearly related to poverty variables and specific assets. For instance, individuals who have higher intensities in the White scale have a lower poverty index, more years of schooling, greater access to phone lines, and more access to health insurance and to private education.

Using a decomposition technique, the earnings gap calculated in the pair-wise comparison of three groups (White, Indigenous and *Mestizos*), we find that for the self-employed, two thirds of the raw race-earnings gap is explained by differences in individual characteristics, while the rest is explained by differences in returns and unobservables (or discrimination). Labor-related characteristics (economic activity, occupation and firm size) constitute the most important explanatory category. Occupational differences are an important source of earnings differentials, which suggests that some sort of occupational segregation has significant correlation with the wage gap. However, these effects may be overestimated.

Among wage earners, the largest part of the earnings gap is explained by differences in characteristics. When we analyze the decomposition within these differences, it appears that personal and human capital characteristics are important explanatory variables for the earnings gap between *Mestizos* and Indigenous, but labor-market characteristics are important in explaining the gap between Whites and *Mestizos*.

We then use a semi-parametric approach that treats some variables in the earnings equations in a linear fashion but lets others (specifically, the race-intensity variables) interact freely, without restricting them to a functional form. For the non-linear component of our equation it is possible to construct an empirical joint distribution for all

possible effects related to different racial combinations by means of a re-sampling technique as bootstrap. Having constructed this empirical distribution we are able to test hypotheses about the significance of the differences between any pair of different effects.

We do not find any effects on ethnic-related variables, such as religion, birthplace, or native tongue, over earnings. Among the self-employed, controlling for personal, labor-market (occupation, sector and firm size) and other ethnic characteristics, we find no significant differences in earnings explained by race. However, for private wage earners there are significant differences between predominantly White and predominantly Indigenous workers. Non-parametric estimators of the differences of racial effects of different groups are consistent with discrimination among wage earners but not among the self-employed. On the other hand, while occupational segregation by race among the self-employed falls in the same range in which we find the levels of occupational segregation by gender in Latin America (Deutsch et.al., 2001) the same figures for wage earners are considerably higher.

## 5. Bibliography

Allport, Gordon, and Bernard Kramer; (1946). "Some roots of prejudice." *The Journal of Psychology*, 22 (1946), pp.9-39.

Anderson Barbara A., Brian D. Silver and Paul R. Abramson; (1988). "The effects of the Race of the Interviewer on Race-Related Attitudes of Black Respondents in SRC/CS National Election Studies." *Public Opinion Quarterly*, Vol 52, No. 3, pp. 289-324.

Angel, Ronald and William Gronfein; (1988). "The Use of Subjective Information in Statistical Models." *American Sociological Review*, Vol. 53, pp. 464-473.

Arce, C.H., E. Murguia and W.P. Frisbie; (1987). "Phenotype and Life Chances Among Chicanos." *Hispanic Journal of Behavioral Sciences*, 9:19-32.

Barsky Robert, John Bound, Kerwin Charles and Joseph Lupton; (2001). "Accounting for the Black-White Wealth Gap: A Nonparametric Approach." NBER Working Paper #8466.

Becker, Gary; (1971). "The Economics of Discrimination." Second Edition. Chicago: The University of Chicago (1971).

Blinder, A.; (1973). "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal of Human Resources*, VII, 4, pp. 436-455.

Boergerhoff-Mulder and T. M. Caro; (1985). "The Use of Quantitative Observational Techniques in Anthropology." *Current Anthropology*, Vol. 26, pp. 323-335.

Callirgos, Juan Carlos; (1993). "El racismo: la cuestión del otro (y de uno)." -- Lima : DESCO, 235 p.

Chiswick, Barry; (1988). "Differences in Education and Earnings across Racial and Ethnic Groups: Tastes, Discrimination, and Investment in Child Quality." *The Quarterly Journal of Economics*, 103(3): 571-597, 1988.

Darity, William and Patrick Mason; (1998). "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *The Journal of Economics Perspectives*, 12:63-90.

Deutsh, Ruthanne; Andrew Morrison, Claudia Piras and Hugo Nopo; (2001). "Working Within Confines: Occupational Segregation by Gender in Three Latin American Countries." In *Women at Work: a Challenge for Development*. Santiago, Chile. 2001.

DiNardo, John; Nicole Fortin and Thomas Lemieux; (1996). "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, Volume 64, No. 5 (Sep., 1996), 1001-1044.

Efron, B.; Tibshirani, R.; (1991). "Statistical Data Analysis in the Computer Age" In *Science*, 1991, Vol. 253, pp. 390-395.

- “*An Introduction to the Bootstrap*” (New York: Chapman & Hall, 1993).
- Hirshmann, Charles and Richard Alba; (1998). “The Meaning and Measurement of Race in the U.S. Census: Glimpses into the Future.” Paper presented at *The Annual Meeting of the American Sociological Association*. 1998.
- Horowitz, Joel; (2001). "The Bootstrap." In *Handbook of Econometrics, Vol. 5*, J.J. Heckman and E. Leamer, eds., Elsevier Science B.V., 2001, pp. 3159-3228.
- Johnson, J.H. and W.C. Farrell; (1995). “Race Still Matters.” In *Chronicle of Higher Education*, Vol XLI, No. 43 (July 7),: A48.
- Longhofer, Stanley D.; Peters, Stephen R.; (1998). “Self-Selection and Discrimination in Credit Markets.” Federal Reserve Bank of Cleveland. Originally published July 1998. Revised June 2000.
- MacIsaac, Donna; (1993). "Peru." In *Indigenous People and Poverty in Latin America: An Empirical Analysis*. George Psacharopoulos and Harry Patrinos. (Ed) The World Bank. Latin American and the Caribbean technical Departments. Regional Studies Program. Report No. 30. pages 179 - 224.
- Mendoza Arroyo, Amalia, ed.; (1993). "Cuestión de piel: testimonios de racismo en el Perú." Lima: ADEC-ATC, 1993. 132 p.
- Nopo, Hugo; (2002). “The Gender Wage Gap in Peru 1986-2000. Evidence from a Matching Comparisons Approach.” Unpublished Mimeo. Northwestern University. 2002.
- Oliart, Patricia; (1989). "El cristal con que se mira. Algo sobre género, raza y clase en el Perú." Lima : PUCP, 1989. Mimeo. 20 p.
- Oaxaca, Ronald; (1973). “Male-Female Wage Differential in Urban Labor Market.” *International Economic Review*, Vol. 14, No. 3, (Oct. 1973), pp. 693-709.
- Psacharopoulos, George and Patrinos, Harry A., eds.; (1994). *Indigenous Peoples and Poverty in Latin America: An Empirical Analysis*. Washington, DC: The World Bank.
- Pozzi-Escott, Ines. “La Discriminación Etnico Cultural en la Escuela Peruana”. Lima: P.U.C.P. 1989.
- Robinson, P.M.; (1988). “Root-n-consistent Semiparametric Regression.” *Econometrica*, 56, 931-954.
- Saavedra, Jaime; (1997). "¿Quiénes Ganan y Quiénes Pierden con una reforma estructural?: cambios en la distribución en la dispersión de ingresos según educación, experiencia y género en el Perú urbano." Notas para el debate No. 14. GRADE. Lima., August 1997, pages 9-77.
- Scalera, Domenico; Zazzaro, Alberto; (2001). “Group reputation and persistent (or permanent) discrimination in credit markets.” *Journal of Multinational Financial Management*, vol. 11 (2001) pp 483–496.

Scodel, Alvin and Harvey Austrin; (1957). "The perception of Jewish photographs by Non-Jews and Jews." *Journal of Abnormal and Social Psychology*, 54, March, pp. 278-280.

Silva, N.V.; (1985). "Updating the Cost of Not Being White in Brazil." In P.-M. Fontaine, ed., *Race, Class and Power in Brazil*. Los Angeles: University of California.

Silva, N.V.; (1992). "Racial Discrimination in Income." Paper presented at the International Seminar, *Labor Market Roots of Poverty and Inequality in Brazil*. August 12-14, Rio de Janeiro.

Sulmont Samain, Denis; (1995). "Exclusión social y empleo: notas para un debate"/ Denis Sulmont Samain.

Telles, Edward and Nelson Lim; (1998). "Does it matter who answers the race question? Racial classification and income inequality in Brazil?." *Demography*, 35 (4).

Toch, Hans, Albert Rabin, and Donald Wilkins; (1962). "Factors Entering into Ethnic Identifications: an Experimental Study." *Sociometry*, 25, September 1962, pp.297-312.

World Bank; (1999). "Peru, Education at a Crossroads: Challenges and Opportunities for the 21<sup>st</sup> Century". Human Development Department. LACR. Report No. 19066.

ANNEX 1

Decomposition Exercise including predominantly Black individuals in the sample.

**Comparison among Different Partitions**

**A. Distribution of the sample (%)**

	Self Employed			Private Wage-Earners		
	Criterion 5	Criterion 4	Criterion 3	Criterion 5	Criterion 4	Criterion 3
White	11	15	14	13	17	18
Mestizos	43	30	31	52	41	37
Indigenous and Blacks	46	55	55	35	42	45
Total	100	100	100	100	100	100

**B. Average Log of Hourly Wage**

	Self Employed			Private Wage-Earners		
	Criterion 5	Criterion 4	Criterion 3	Criterion 5	Criterion 4	Criterion 3
White	1.0059	0.9494	0.9586	1.1616	1.0837	1.0669
Mestizos	0.7920	0.8182	0.8118	0.8810	0.9144	0.9570
Indigenous and Blacks	0.6136	0.6292	0.6311	0.8296	0.8095	0.7832

**C. Earning Gaps**

	Self Employed			Private Wage-Earners		
	Criterion 5	Criterion 4	Criterion 3	Criterion 5	Criterion 4	Criterion 3
White vs. Indigenous and Blacks	0.3923	0.3202	0.3275	0.3320	0.2742	0.2837
White vs. Mestizos	0.2139	0.1313	0.1468	0.2806	0.1693	0.1099
Mestizos vs. Indigenous and Blacks	0.1784	0.1889	0.1807	0.0514	0.1049	0.1738

### Earnings Decomposition

	Criterion 5			Criterion 4			Criterion 3		
	White vs. Indigenous and Blacks	White vs. Mestizos	Mestizos vs. Indigenous and Blacks	White vs. Indigenous and Blacks	White vs. Mestizos	Mestizos vs. Indigenous and Blacks	White vs. Indigenous and Blacks	White vs. Mestizos	Mestizos vs. Indigenous and Blacks
<b><u>Self-Employed</u></b>									
<i>Wage Gap</i>	<i>0.3923</i>	<i>0.2139</i>	<i>0.1784</i>	<i>0.3202</i>	<i>0.1313</i>	<i>0.1889</i>	<i>0.3275</i>	<i>0.1468</i>	<i>0.1807</i>
<b>Differences in Characteristics (%)</b>	<b>0.177</b>	<b>0.127</b>	<b>0.085</b>	<b>0.183</b>	<b>0.096</b>	<b>0.127</b>	<b>0.146</b>	<b>0.073</b>	<b>0.121</b>
Education and Experience	0.079	0.059	0.035	0.065	0.024	0.035	0.040	0.068	0.017
Personal and Family	0.044	0.022	0.029	0.045	0.000	0.067	0.060	-0.007	0.072
Labor	0.054	0.045	0.021	0.074	0.073	0.024	0.047	0.011	0.031
<b>Differences in Returns (%)</b>	<b>0.215</b>	<b>0.087</b>	<b>0.093</b>	<b>0.137</b>	<b>0.035</b>	<b>0.062</b>	<b>0.181</b>	<b>0.074</b>	<b>0.060</b>
<b><u>Private Wage Earners</u></b>									
<i>Wage Gap</i>	<i>0.3320</i>	<i>0.2806</i>	<i>0.0514</i>	<i>0.2742</i>	<i>0.1693</i>	<i>0.1049</i>	<i>0.2837</i>	<i>0.1099</i>	<i>0.1738</i>
<b>Differences in Characteristics (%)</b>	<b>0.245</b>	<b>0.157</b>	<b>0.030</b>	<b>0.176</b>	<b>0.136</b>	<b>0.059</b>	<b>0.128</b>	<b>0.136</b>	<b>0.044</b>
Education and Experience	0.045	0.044	0.025	0.038	0.027	0.028	0.021	0.036	0.010
Personal and Family	0.030	0.007	-0.009	-0.002	0.019	0.004	-0.020	0.045	-0.010
Labor	0.170	0.106	0.013	0.140	0.090	0.027	0.127	0.055	0.044
<b>Differences in Returns (%)</b>	<b>0.087</b>	<b>0.124</b>	<b>0.022</b>	<b>0.098</b>	<b>0.034</b>	<b>0.045</b>	<b>0.156</b>	<b>-0.026</b>	<b>0.129</b>

### Duncan Index of Occupational Segregation

		<b>Self-Employed</b>	<b>Private Wage Earners</b>
<b>Criterion 5</b>	White vs. Indigenous and Blacks	0.2546	0.4543
	White vs. Mestizos	0.2540	0.3249
	Mestizos vs. Indigenous and Blacks	0.1550	0.2532
<b>Criterion 4</b>	White vs. Indigenous and Blacks	0.2211	0.4185
	White vs. Mestizos	0.2138	0.3074
	Mestizos vs Indigenous and Blacks	0.1577	0.2369
<b>Criterion 3</b>	White vs. Indigenous and Blacks	0.2132	0.4074
	White vs. Mestizos	0.1840	0.2964
	Mestizos vs. Indigenous and Blacks	0.1612	0.2912