EDUCATION, LABOR MARKETS, AND INEQUALITY IN PERU[*]

Adolfo Figueroa
Department of Economics
Catholic University of Peru
June 2007

One unnoticed paradox exists in Latin American countries: the expansion of education has been very significant in the last decades; however, the level of inequality has not been reduced. According to the data from the World Bank, the net rate of school registration in Latin America increased, between 1980 and 1997, from 85% and 94% for primary school and from 55% to 66% for secondary school (World Bank 2001, Table 6, p.285). On the other side, Latin America was the most unequal region in the world during the period 1950-1995, with a Gini coefficient of 0.50, as an average (compared to 0.33 of the South-East Asia region); moreover, the mentioned average experienced almost no variation in the whole period (Deininger and Squire 1996, Table 5; Li, Squire and Zou 1998).

The obvious question then arises: Is education an equalizing system? One of the main mechanisms for the transformation of widespread education into higher salaries and into reduced inequality is the labor market. In this sense, the search for an explanation of the paradox should, in the first place, focus on the functioning of the labor market.

The objective of this study is to solve the paradox. The study of labor markets needs the specification of the institutional framework under which they operate. For this task, the ontological universalism of standard economics will be abandoned. Instead, a theoretical approach in which the assumption that there are different types of capitalist economies, which function differently, will be adopted. Initial conditions (history) determine these differences. Three types of abstract capitalist societies that were born to capitalism under different initial conditions will be distinguished here: *epsilon* (socially homogeneous and underpopulated), *omega* (socially homogeneous and overpopulated), and *sigma* (socially heterogeneous and overpopulated). These theories are expected to explain the First World countries, the Third World countries that have weak or no colonial legacies, and the Third World countries that have strong colonial legacies. The theoretical construction of these theories and their consistency with the basic empirical facts are presented in Figueroa (2003). But labor markets analysis is still pending and will be developed now.

Labor markets will be analyzed in the context of a sigma society. This is the main focus of this study. However, in order to understand better the results, basic comparison will be made with epsilon and omega theories. In addition, some comparisons will be made with the results coming from standard economics.

Sigma theory assumes that ethnicity matters in the economic process of production and distribution. Sigma is thus a society in which labor supply is composed of a socially heterogeneous mass of workers. A model of the sigma theory is constructed here to

[*] Revised version of the paper presented at the Workshop "Inequality, Ethnicity and Human Security in Latin America", organized by the Centre for Research on Inequality, Ethnicity and Human Security (CRISE), Oxford University, and UNDP-Bolivia, held in Santa Cruz, Bolivia, September 25-26, 2006. I would like to thank the participants of the workshop for their very helpful comments.

establish the interrelations between education, human capital, wages, and incomes. In this relation, the labor market plays the main role.

The international literature has produced many empirical studies about the relations between education and incomes; however, no satisfactory answer has come up to solve the paradox. The analyses about the relations between education and incomes corresponding to Latin America have usually been done adopting the neoclassical theory, which assumes that labor markets are Walrasian, to which empirical categories have been incorporated, such as "formal" and "informal" sectors, in order to analyze issues related to the segmentation of labor market, but ignoring the ethnic factor (cf. Maloney 2004).

Sigma theory differs from the neoclassic theory because it is based on another set of basic assumptions, which also generates unconventional empirical predictions. Both theories compete against each other; therefore, both cannot be empirically valid. Therefore, if the empirical tests corroborate one, it implies the refusal of the other. Here, an empirical test of the predictions from a sigma model will be carried out for the case of Peru.

The present article is organized as follows. Section 1 presents a model of the theory of labor markets in the sigma society. The process of human capital accumulation is developed in Section 2. The relations between education and income are analyzed in Section 3. The empirical predictions of the theoretical model are presented in Section 4. These are the hypotheses that will be tested empirically for the Peruvian case. Section 5 provides a critical analysis of the database to be used in the testing process. The falsification of the hypotheses is shown in section 6. The final section presents the conclusions of the falsification process and also some of its consequences for the design of policies. The article has a Statistical Appendix, which shows the results of the statistical tests.

1. A Sigma Model of Labor Markets

The labor market will function in a particular capitalist society: the sigma society. This is considered its institutional context. The basic assumptions about sigma are presented.

A Sigma Model

Sigma is a society where individuals participating in the economic process are endowed with unequal quantities not only of economic assets but also of social assets. Social factors are thus introduced into the economic process. Certainly, social assets are special goods for they belong to the realm of people's rights and entitlements in society. They are no physical goods, nor are they marketable.

In this study, social assets will refer to political and cultural assets. Political assets are defined as the capability of exercising individual and collective rights, including the right to have rights. Inequality in the endowment of political assets generates a hierarchy of citizens in society, first-class and second-class citizens. As a result, not all individuals are equal before the law; moreover, not all individuals have the same access to public goods supplied by the state.

Cultural assets are defined as the right of groups of individuals to cultural diversity in a multicultural society. Inequality in the endowment of cultural rights generates a hierarchy of ethnic markers in society: there are first-class and second-class races, languages, religions, and customs. These markers are called cultural because their

hierarchy are socially constructed and are also transmitted from generation to generation. Inequality in cultural assets leads to social practices of segregation, exclusion, and discrimination against some ethnic groups.

Why would inequality in social assets exist? Inequality in the endowments of social assets implies the dominance of some ethnic groups by others. This domination system may be the result of a historical episode of society (a conquest, colonialism, slavery). This foundational shock generates a multiethnic and multicultural society, a hierarchical one. Capitalism was born on the legacy of that foundational shock.

The sigma society is a class society. It is also a multiethnic society, in which subaltern and dominant ethnic groups exist. Analytically, this is equivalent to assuming a society where individuals are endowed with unequal quantities of social assets, that is, with unequal political and cultural assets. The unequal distribution of social assets is historically determined. This initial inequality provides ethnic groups with either social prestige or social stigma, which leads to the existence of social exclusion (from political and cultural rights) and segregation and discrimination against subaltern ethnic groups.

Social classes, citizenships, and ethnic groups make up the social structure of sigma society. In order to derive empirically refutable predictions from sigma theory, a model of this theory must now be established. Thus, a set of auxiliary assumptions, all consistent with the primary assumptions of the theory, are introduced.

There are two social classes: capitalists and workers. There are three ethnic groups: the Blues, the Reds, and the Purples. The purples are the result of miscegenation of the other two races.[1] There are two types of citizenships: first class and second class.

Table A presents the social structure of sigma in a social matrix form. The Blues constitute the capitalist class. They concentrate the endowments of physical capital of the economy. Purples and Reds belong to the working class. Purples are endowed with skilled labor and Reds with unskilled labor. In terms of citizenship endowments, the Blues and the Purples constitute the first-class citizens, while the Reds are the second-class citizens. The Reds are the subaltern and dominated social group. The origin of this domination is historically determined. This is the initial inequality in asset endowments of individuals.

Table A. Social Structure of Sigma Society

Ethnic Group	Physical Capital	Human Capital	Citizenship	Name of the Social Group
Blues	K_b	K_{h1}	C_1	A
Purples	0	K_{h1}	C_1	Y
Reds	0	K_{h0}	C_0	Z

The social matrix shows a highly correlated society in the endowment of assets. The Blues are highly endowed with economic and political assets; the Reds are very poor in those endowments; and the Purple lie in between. For easy reference and for reasons that will become apparent later on, call the ethnic groups by the name A, Y, and Z. The

[1] In a paper that analyzes the theoretical relationships between consumer preferences and culture, Akerlof and Kranton (2000) construct an abstract world of two ethnic groups, the Greens and Reds, in which the Greens are the dominant group. To use primary colors, call them Blues and Reds here; then introduce a third ethnic group: the Purples, the result of miscegenation of the two races. As in that paper, it is assumed here that people cannot choose their ethnic identity; ethnicity is exogenous.

social group Z is ethnically Red, economically worker, and politically second rate citizen. The social group A is ethnically Blue, economically capitalist, and politically first rate citizen. The social group Y is ethnically Purple, economically worker, and politically first rate citizen.

Epsilon and omega, the other types of capitalist societies, can now be distinguished analytically from sigma. Epsilon and omega show equality in citizenship, but inequality in economic assets. There are two social classes (capitalists and workers) but there is only one citizenship class. There are three ethnic groups, but only one degree of citizenship for all. Therefore, the social matrix collapses to two social groups only: A and Y, which represent the social classes, capitalists and workers. The Reds and Purples constitute the social group Y. The social group Z does not exist. Ethnicity exists, but it does not count in the social structure. The critical asset that distinguishes sigma from epsilon and omega is citizenship.

This matrix assumes that there is a very high degree of inequality in the initial endowment of assets among individuals, a degree of inequality denominated here by the symbol delta (δ). Given these initial conditions, how production and distribution would be determined in this economy? Would the initial inequality in the endowment of individual and collective assets tend to remain or to diminish endogenously?

Given this social structure, it is possible to show that the three capitalist societies operate with mechanisms of inclusion and exclusion at the same time. The market and democracy, the two basic institutions of capitalism, constitute not only mechanisms of social integration but also have mechanisms of exclusion. Economic exclusion refers to the fact that workers are excluded (total or partial) from three *basic markets*, which play a significant role in the reproduction of inequality. These markets are: labor, credit, and insurance. Here, the labor market will be the topic of study. In Sigma society, social exclusion also exists. Z workers are excluded (total or partial, formal or informal) from the access to *basic public goods,* which also play the fundamental role in the reproduction of inequality. These are: education, health, justice, and social protection. Social exclusion can take the form of discrimination or segregation against the subordinate social group Z. These assumptions constitute the institutional foundations of labor markets.

A Labor Market Model

In the sigma economy, the basic social interactions take place in the labor market. Y-workers supply their labor to the labor market, in quantities that are exogenously determined. The labor market operates as a non-Walrasian market, in which equilibrium takes place with excess labor supply. This quantity of excess labor supply becomes, in part, self-employed in small units in the y-subsistence sector and, in part, unemployed (seeking employment and expecting to get wage income).

Z-workers are endowed with low human capital for the technology being used in the capitalist sector. Thus, their human capital endowments are not suitable for wage employment. They are not employable; therefore, they are not part of the labor supply in the labor market. Capitalist firms cannot make profits employing them, as there would be much need to invest in their training, when at the same time y-workers are in plentiful supply. It is the lack of profitability that lies behind the total exclusion of z-workers from the labor market. Z-workers conform to the definition of the underclass: workers "who are largely expendable from the point of view of the logic of capitalism" (Wright 1997, p.28). They are self-employed in small units in the z-subsistence sector.

When firms hire workers, what do firms buy in the labor services? The standard assumption in economics is that firms seek to maximize profits implies that firms buy human capital (not education) in the labor market; that is, firms are concerned with the stock of productive knowledge incorporated in the worker. This stock is denominated *human capital*. Besides technology and the capital stock of the firm, labor productivity depends on the worker's human capital. Therefore, the firm will hire workers until the marginal productivity of labor is equal to the market wage rate. Given diminishing returns, the lower the wage rate is, the higher the quantity of workers demanded. The labor demand curve will take the form of a decreasing curve. There will be one such curve, and one particular market, for each level of human capital.

According to neoclassical theory, the real wage and the quantity of employment in each labor market are determined by the interaction of pure real variables, summarized in the curves of demand and supply, and where both curves cross each other. The market wage rate is the one that clears the market. Labor markets function therefore as Walrasian markets. It is then possible to study labor markets with the method of partial equilibrium. The existence and persistence of unemployment observed in all capitalist countries is an empirical data that refutes this theory.

According to sigma theory, real wage rates and employment are not determined in each labor market. Relative prices and quantities are not determined by the interaction of pure real variables, such as productivity and labor supply, but through the interaction of real variables and monetary variables. This is due to the assumption that, according to the social norms, the nominal salary cannot decrease from its historical level; in other words, it is inflexible downwards. Therefore, the real wage rate will depend on the level of prices, which is determined endogenously by the interaction of real and monetary variables.

In sum, the level of prices and the level of employment are determined simultaneously by the interaction of the labor and monetary markets, given the nominal wage rate. The functioning of the labor market cannot be explained by the method of partial equilibrium, only by the method of general equilibrium (Figueroa 2003).

Consider a labor market model with two labor markets which correspond to two levels of human capital. Figure 1 presents the situation of equilibrium in each of the labor markets. The labor market for the high level of human capital is presented in the panel (a). The quantity supplied is given, and it is equal to 00', which refers to y-workers. The curve M"N" is the labor demand curve. The real wage rate (w_2) and the level of employment (0A) are determined in the general equilibrium by the interactions between the labor market and other markets, including monetary markets.

The resulting excess labor supply of y-workers is equal to the segment A_20'. Given the probability to find a job (π_2), which is exogenously determined, the expected income ($\pi_2 w_2$) of seeking wage employment is determined. The curve m"n" represents the labor marginal productivity in the y-subsistence sector, measured from the origin 0'. The logic of workers, who search for the maximization of their incomes, implies that the expected income of wage employment has to be equal to the marginal income of self employment in the y-subsistence sector (v_2'). This equality determines the quantity of self employment, which is equal to the segment $0'B_2$. Given the quantity of excess of labor supply, the quantity of unemployment (A_2B_2) is the residual.

The labor market for the low level of human capital is represented in panel (b). The labor supply is equal to the segment 0'0" and comes from z-workers, denominated as the group Z_1. The labor demand is given by the curve M'N' and the labor marginal productivity in the z_1-subsistence sector by the curve m'n' measured from the origin 0". The functioning of this labor market is very similar to the previous one.

There is another segment of the social group Z, named the group Z_0, endowed with the lowest level of human, considerably low to cope with the technology used by the firms. This group appears in the panel (c), with a quantity of workers equal to the segment 0"Z_0. The curve of marginal productivity of this sector of subsistence is represented by the curve mn. This group is out of the labor market and the have no choice but to stay self-employed in the z_0-subsistence sector. Firms will not have any incentive to invest in the development of human capital of this group (*on the job training*) when labor markets show an oversupply of workers.

In this model, each labor market operates as a non Walrasian market. At the market wage rate, there is an excess of labor supply. The labor market would not operate with a wage rate which clears the market, because in this case the workers would not have incentives to provide the firm the maximum effort. Shirking behavior must suffer a cost. The labor market cannot be operate in a similar manner to the potato market (as it is suggested in the neoclassical theory): while potatoes cannot change their behavior according to the buyer, workers can decide over their own behavior, over their own effort and productivity, influenced by the incentives provided by the firm.

What is the device used by firms to discipline workers in a Sigma society? Because of overpopulation, unemployment cannot be the device to ensure labor discipline, as the standard theory assumes (cf. Shapiro and Stiglitz 1984). In fact, an economy could not operate with 40% and 50% of unemployment: it would be socially unviable. The excess supply of workers would have to generate their own income in the subsistence sector.

The device to ensure discipline will be the condition that the wage rate must be higher than the cost of opportunity of wage earners, which is given by the income of self employment. Firms must therefore pay to wage-earners a wage that includes a premium above their opportunity cost. The market wage rate must play the role of *efficiency wage*; it must assure high labor productivity levels at the firm level.

In sum, the empirical prediction of the model is that the wage rate must be higher than the mean income of the self employed in the subsistence sector. In figure 1, this relation is expressed by the difference between the wage rate (w) and the value of the marginal productivity (v'), which implies that the wage rate is also higher than the average labor productivity (v). The self-employed may be called the *underemployed* because they make incomes that are below the market wage rate for the same level of human capital

Another assumption of the Sigma model of labor markets refers to the role of ethnicity. The labor demand depends on the human capital of the worker and it is not directly related to his/her ethnic origin. Firms buy human capital when they buy labor services, as shown in the labor market models depicted in Figure 1. However, the model assumes inequality in the initial endowment of human capital between y-workers and z-workers; thus firms indirectly buy also ethnicity in the labor market.

In sum, from Figure 1, where the sigma model of labor markets is presented, the following empirical hypothesis can be derived: (1) in each labor market, the market wage rate is higher than the mean income of the self employed; (2) in markets of higher levels of human capital, wage rates will be higher; that is, at individual level, the higher the human capital endowment is, the higher the market wage rate; (3) at the aggregate level, the higher the human capital endowment of workers is, the higher the mean income (addition of wages, salaries, and income of self employment).

2. Accumulation of Human Capital in Sigma Society

Up to now, the human capital endowment of each individual of each ethnic group has been considered as exogenously determined. In this section, human capital will be endogenous. What are the determinants of human capital accumulation?

The stock of knowledge about production skills constitutes an individual's human capital. This stock can also be seen as the production skills embodied in workers. People do not acquire this stock of knowledge at birth; thus, people need to invest in acquiring it through the process of education.

The Role of Initial Inequalities

In the learning process, the initial conditions brought by the individual into the educational system at each level are essential for learning. In relation to the initial endowments, in the literature of related sciences (psychology, biology, and neuroscience) the standard theory assumes that the endowment of talents matter and that those talents are multiple, the so called multiple intelligence theory (cf. Gardner 1999).

The brain plasticity theory is another important component to study the learning process. While the individual's initial talent endowments—the genetic inheritance—are exogenous (*nature)*, over time talents are endogenous because brain development depends on the social environment (*nurture*) and on the interaction between the two. As the modern literature of neuroscience the brain plasticity theory is usually stated as follows: "The brain is not a computer that simply executes predetermined programs. Nor is it a passive gray cabbage, victim to the environment influences that bear upon it. Genes and environment interact to continually change the brain, from the time we are conceived until the moment we die" (Ratey 2002, p.17)

In the aggregate, however, the endowments of initial talents based on genetics may be assumed to be normally distributed among the population (the result of a random selection), which will not b the rule for the talents generated by the social environment. The important distinction made by Rousseau (1755) points out exactly these two factors. Rousseau distinguished two types of inequalities among individuals: the *natural,* originated by the gifts of nature, natural endowments, and random; and the *artificial*, originated by the functioning of society.

A theory of human capital accumulation will be presented now. The first step consists of having a theory of the social inequality in cognitive skills among individuals. The primary assumption of this theory is that, given the random genetic endowments, there exists inequality in talents among individuals in the infancy, which is associated to differences in the socio-economic levels of families. Inequality in the infancy talents—the relevant initial conditions in the education process—is not independent of the socio-economic differences among families. Nutrition, health, early intellectual stimulation, and language are the main channels through which the wealthy can develop high levels of talents of their children compared to the poor. This theory is general, as it intends to apply to epsilon, omega, and sigma societies.

Nutrition has a direct effect on the brain development and the cognitive skills of individuals; it also has an indirect effect through illness episodes. And the degree of nutrition is positively determined by the incomes of the families.

Health condition of individuals is not neutral in relation to wealth inequality either. Therefore, on average, illness episodes will be less frequent in kids from rich families; what is more, rich families can turn into private health centers for treatment, while poor families only use public health centers, with high differences in the quality of the services.

Differences in health condition are also associated to environmental health, which is not neutral in relation to wealth inequality. Rich families can avoid problems that involves environmental cleanup (quality of water and drainage, and air pollution) through "exit" behavior (*exit*, in the known language *exit-voice* of Albert Hirschman), because they are able to construct exclusive residential neighborhoods. Poor families can only try to solve the problem using their "voice", that is, demands, protest, etc.

The access of people to the supply of basic public goods will be differentiated, depending on the type of society. In sigma societies, where spatial segregation exists among social groups, public health will not operate as universal public goods, but as local public goods, differing due to the type of citizens. Either as private provision or as a public good, health depends on the socio-economic conditions of the families.

The intellectual stimulation of children also depends on the socio-economic level of the family. In this sense the greater quantity, quality, and diversity of goods and services consumed by rich families induce them to discover the talent of their kids and to increase their intellectual stimulation.

Language is another factor of inequality in cognitive skills associated to the socio-economic level of families. There are linguistic differences among individuals. This inequality is shown in various aspects of language, such as vocabulary, syntaxes, and ways of speaking, writing and reading skills. According to the socio-linguistic theory, language inequality is due mostly to personal experiences (social environment) than to genetic factors (cf. Hudson 1996, p.204).

Inequalities in the language skill among families imply unequal cognitive skills of their children. Abstract and complex thoughts are not only language-dependent, but also complex language-dependent. Philosopher John Searle has stated clearly: "Some thoughts are of such complexity that it would be empirically impossible to think them without being in possession of symbols. Mathematical thoughts, for example, require a system of symbols. … Complex abstract thoughts require words and symbols" (Searle 1995, Chapter 3, p. 64). The implication seems to be that written language allows the individual to work with more abstract and complex thoughts than does oral language alone.

Consider an epsilon society in which people live in a written society and most of them are literate. By contrast, in a sigma society, a multilingual and hierarchical society, the y-workers live in a written culture and most of them are literate; however, z-workers live in an oral social environment and most of them are illiterate in the dominant language, and assume that their aboriginal langue is not a written language. In such a context, language skills in the dominant language will be very unequal.

Due to their illiteracy, z-populations will be limited in the use of abstract and complex thoughts. But, in addition, if z-populations come from an oral culture, where the aboriginal language is not written, those limitations will be reinforced. To be an illiterate in a written culture is different from being an illiterate in an oral culture. The handicap will be higher in the latter case. Z-populations will then show lower levels of language skills in the dominant language and thus their children will show lower levels of cognitive skills compared to the children of y-populations.

The characteristic of multicultural, multilingual, and hierarchical society makes sigma also a heteroglosic society. This term comes from socio-linguistic theory and refers to the existence of various forms or variations on the use of the dominant language, but with a hierarchy among those forms, from the ones considered socially superior to the

lowest ones. There is the problem of different accents, which will persist even at adult age due to segregation and exclusion. There will be path dependence on accents.[2]

The important consequence of heteroglosia is the unequal language skills in the dominant language. The dominant language is a second language for the z-populations and as such creates inequality in the skills in the use of this language. If the effect of the oral culture is included, the second language problem is reinforced. Z-populations could not be able to express abstract and complex thoughts in the dominant language due to two effects: the oral culture effect and the second language effect. Because complex and abstract thoughts require command on a written language, z-populations will be handicapped in this capacity. Even those that are literate in the dominant language will face this problem.

In sigma society, therefore, there will be inequality in the language skills of children, with a relatively high level for the children coming from y-families compared to those coming from z-families. In omega and epsilon societies, socially homogeneous societies, initial inequalities in language skills are still associated to socio-economic differences of families, but the effect will not be as strong as in sigma society. The assumptions made in this theory of human capital accumulation lead to the empirical prediction that linguistic inequality plays a crucial role in generating inequality in cognitive skills. A prediction of socio-linguistic theory goes even further: "Linguistic inequality can be seen as a *cause* of social inequality, but also as a *consequence* of it, because language is one of the most important means by which social inequality is perpetuated from generation to generation" (Hudson 1996, p.205).

At the end of the process of human capital accumulation, language will also count as part of the human capital of the individual. In a sigma society, a heteroglosic society, language will operate as a social marker: "Let me listen how you speak and I will tell who you are". Language can therefore give signals about his/her labor skills, positive or negative, depending on the reputation these signal have. Hence, linguistic inequality will imply inequality in human capital.

To conclude, given the random distribution of genetic talents in the population, people start the process of human capital accumulation endowed with unequal initial cognitive or learning skills, which are associated to differences in the socio-economic levels of families. It should be noticed that this proposition is not a beta proposition because learning capacity is unobservable. However, it will be used as a primary assumption (an alpha proposition) of the theory of human capital accumulation: the initial endowment of cognitive skills does matter in the learning process of individuals; moreover, this initial endowment is endogenous.

Transformation of Education into Human Capital

[2] Take the case of the Andean countries in South America, where the dominant language is Spanish. In the spoken language, "Limeño Spanish", "Standard Spanish", "Coastal Spanish", "Andean Spanish", "Amazonian Spanish", and "Indigenous Spanish" show the diversity and the hierarchy. The indigenous peoples that are literates in Spanish, besides the accent, are not able to use the written Spanish language with proficiency. They can construct sentences in first person and use verbs in present, past, and future tense, but they are unable to create sentences which express abstract reasoning, which requires the use of verbs in other modes, like subjunctive and impersonals. Quechua, an aboriginal language in South America, was not written, so most of indigenous populations come from an oral culture and still live in this oral culture. Some socio-linguistics have argued that syllogisms cannot be constructed in Quechua, which is what the Searle theory would say: abstract and complex thoughts require symbols, written language.

How can inequality in human capital among individuals and social classes be explained? People invest in schools to acquire human capital. Assume for the time being that schools are all homogeneous. Because of the initial inequality in learning capacities, the school system will generate differences in the accumulation of human capital between rich and poor families. Considering the same number of schooling years, children from rich families will acquire in average a higher human capital compared to those from poor families.

The accumulation of human capital requires financing. Rich people have greater financing capacity than poor people, which allows them to accumulate higher levels of human capital. Then, the income effect on investing in human capital is positive: The quantity of human capital demanded will positively depend upon the level of family income. Consequently, children from rich families would have a higher number of schooling years than the ones who come from poor families.

If the assumption that schools are homogenous is now abandoned, another factor of differentiation will come up. Private schools are highly equipped and have better trained teachers than public schools. The effect of such a difference is that children from rich families will assist to private schools, while the children from the poor will attend public schools. Private schools constitute in this sense a "normal good" and public schools an "inferior good".

This simple model predicts that children from rich families will acquire, on average, not only more schooling years but also a higher level of human capital for every year of schooling than the children from poor families. The transformation of education into human capital will operate differently for different social groups. The school system thus tends to reproduce the initial wealth inequality.

This theory of investment in human capital also predicts that education is not the same as human capital. In the process of transformation of education into human capital operates two types of factors: the quality of the school and the quality of the student, and both depend on the socio-economic level of families. In the aggregate, the socio-economic level of families can be substituted by the initial inequality in society (the variable δ). This transformation will operate differently in each type of society.

Consider the model of the sigma society, in which the social structure is composed by three socio-economic groups: A, Y, Z. In this society, there will be a positive relation between years of education and levels of human capital. But this relation will not be univocal; on the contrary, particular forms will emerge according to each social group. There will be a hierarchy in those relations. They are shown in the panel (a) of Figure 2, where curves A, Y, Z correspond to each of the three groups of people. For a given number of schooling years, different levels of human capital will be produced, depending on the social group.

This particular form of transformation from education into human capital, and the order of the curves, indicates not only differences in the quality of students (associated to differences in the socio-economic variable), but also differences in the quality of the school. Private and public schools will show differences in quality, but also differences will exist among public schools. The z-population will have access to public schools of the lowest quality because they are second rate citizens. Therefore, the hierarchy of the curves also shows a particular form of functioning of democracy in the sigma society. Citizens of different category have access to local public goods of different categories. The transformation of education into human capital does not take the same form for all social groups.

In epsilon and omega societies, which are socially homogeneous societies, the transformation from education into human capital will operate along two curves only. In these societies, the z-population does not exist. Therefore, the curves A and Y will show the hierarchy of the relations between schooling and human capital, reflecting the social class structure of the society.

Similar to the technology to produce machines, it is also possible to assume the existence of a technology to produce human capital. The relations shown in the panel (a) of Figure 2 refer to the transformation of schooling years into human capital levels as output. These relations represent production functions with different technologies. The quantity of human capital produced by a school will depend upon the inputs utilized in the process of producing it. According to the assumptions introduced above, these inputs include quantity and quality of school infrastructure, the quantity and quality of the staff of teachers, and the quality of the students. As assumed above, the quality of students refers to their initial cognitive skills, which in turn depends on their socio-economic background. It is usually included the effect of the interactions (externalities) among the students, called the "peer effect." Certainly, in segregated societies, such as sigma society, this effect would be small.[3]

3. Transformation of Education into Incomes

The theoretical relationship between education and incomes in each type of society is presented now. The primary assumption is that there are two processes underlying this relationship, one is the transformation of education into human capital and the other is the transformation of human capital into incomes. The first transformation has already been studied. A connection between human capital and income will be studied now. From here, the reduced-form of the theoretical system, which will connect education to incomes, will be solved.

Transformation of human capital into incomes

The transformation from human capital into incomes operates through the market system. The general relation is positive: The higher the human capital, the higher the income level will be, which reflects the positive economic return of investing in human capital. This economic return is the result of the positive effect that human capital has on labor productivity. This effect operates through two channels: one is the complementarity that exists between human capital and physical capital because human capital makes the machine more productive; the other refers to the complementarity between human capital and the adoption of new technologies because new technologies are incorporated in new machines, the operation of which requires workers with higher level of human capital.

[3] The production function of human capital is a standard concept in the neoclassic theory and it is presented as the "production function of education." Usually, it is assumed that learning at school depends on two types of variables, those related to the resources used by the school and those associated to the socio-economic origin of the students. A recent revision of both the theoretical and empirical literature on the production function of education debate, referred basically to the First World, is presented in Hanushek and Luque (2003) and Todd and Wolpin (2003). It should be remembered that in the present study the production function category belongs to the group of assumptions, not the beta propositions.

This transformation will also adopt different forms in different societies. In sigma society, for a given level of human capital, a greater average income will correspond to the children of the capitalist class (group A) compared to the children of workers (groups Y and Z). This relation is shown in the panel (b) of Figure 2. Curves A', Y', Z' represent these transformations. These curves are all sloping upward, showing positive relations between human capital and mean incomes, but they also show that the relations are hierarchical by social groups.

The hierarchy of relations is the outcome of unequal initial asset endowments (variable δ). Greater endowments of physical capital and social capital (social networks) will imply greater access to basic markets (labor, credit, and insurance markets), and greater economic opportunities for doing business.[4]

In a sigma society, Group A's incomes and human capital will travel along the line A', which is placed at the highest level of all the groups, because it has the highest endowments of these assets. Z-workers have the lowest endowments and travel along the lowest curve. They are excluded from the labor market and do not get wage incomes but self-employment incomes, which are lower than the wages. The case of y-workers lies in between. In epsilon and omega societies, only curves A' and Y' will constitute the relationships between human capital and incomes.

The market system is at work in this transformation. The model predicts that, given the initial asset inequalities, the market system is going to reproduce the outcome from the previous process that transformed education into human capital.

Transformation of Education into Incomes: The Static Reduced Form

The structural equations of the model of the theory of income and education inequalities were already represented in Figure 2. The lines of panel (a) constitute the structural equations of the relations between education and human capital and those in panel (b) constitute the structural equations of the relations between human capital and incomes. Now it is necessary to derive the equation showing the *reduced form* of the model. This derivation will be done for a static model now, in which it is assumed that the level of education of individuals is exogenously determined.

A static model of the sigma theory is considered firstly. As it is indicated in Figure 2, points m, n and r are the mean values of schooling years of the three social groups A, Y, Z, and points m', n', and r' are the correspondent mean incomes. The initial situation is such that the capitalist class has the greater level of schooling years, human capital, and incomes followed by the other two social groups. Certainly, more years of schooling of group Z will result in higher level of human capital, but along the curve Z; also, this higher human capital will be transformed into higher incomes, but along the curve Z'. The same process applies to the group Y and group A.

The reduced form can then be stated as follows: If the individual's income depends on her human capital, which in turn depends on schooling years, that income will ultimately depend upon their schooling years. Therefore, the relations showing the reduced form of the model appear. These reduced form relations will follow particular curves according to each social group. Given that the structural equations shown by the tree curves in Figure 2 are linear (only for simplicity), the curves showing the reduced-form equations will also be linear. One additional year of education will increase the

[4] A theory which explains the differences in endowments and in the accumulation of social capital, defined as social networks, between social groups A-Y-Z is presented in Figueroa (2007).

income in the same magnitude for each group, but this magnitude will be greater for population A, and smaller for population Y, and the lowest for population Z. The reduced form equations of the model are represented by the curves A, Y, and Z in Figure 3.

Analytically, the income gap between groups, such as the distance m'r' between groups Z and A as shown in Figure 2, and in Figure 3 as well, can be decomposed in three effects. They are:

(a) quantitative exclusion of education (less schooling years);
(b) qualitative exclusion of education (less accumulated human capital given the same schooling years);
(c) market exclusion (less income given the same human capital).

These effects can be separated in the same way when analyzing the income gap between groups Y and Z.

Under the initial equilibrium of sigma society shown in Figure 8.1, three cases can be taken into account to analyze the effect of exogenous changes in education on income inequality. First, one additional year of education just in group Z will increase the mean income of the group and will reduce the initial inequality, but this effect will be small. Second, one additional schooling year in all social groups will have a small effect on inequality; certainly, the mean income will increase in all groups, but it is not clear that relative incomes would change drastically. Third, an increase of schooling years in each of the three groups, with greater years in group Z, will have an ambiguous effect or a small one.

In sum, the static model predicts that inequality reduction in schooling years will not imply an important reduction of income inequality. The reason is that the difference in schooling years constitutes only one of the three effects which operate in the generation of income differences among people.

The conclusion would be different if the relations between schooling years and incomes were inseparable, that is, if just one curve existed for all groups, such as curve L in Figure 3. If the three curves were reduced to one, to a liner one (suppose the curve L were linear), one schooling year would increase the income in the same magnitude for the three social groups; thus, if differences of schooling years were reduced, income inequality would decrease as well. If curve L showed decreasing returns, income inequality would diminish even more (the relative income of the very poor would rise); if it showed increasing returns (as shown in Figure 3), inequality change would be ambiguous. The reduction of education inequality would have to be very high in order to cause a higher effect than just compensate the differences in the rate of return. The model developed here predicts, in fact, a curve L of increasing returns.

The relations presented in Figure 3 refer to mean values of the analyzed variables for each social group (not each individual). Successful cases in group Z may occur, but the prediction of the theory is that these cases will not be the rule, but the exception.

The static models of omega and epsilon theories also follow directly from the relations shown in Figure 2. The reduce form equations are represented in Figure 3 by the lines A an Y only. Each social class travels along a particular curve connecting years of schooling to incomes.

Integrating to a general equilibrium analysis, the results are simple to summarize. In the aggregate, schooling changes exogenously, which implies that human capital also changes as a result, and the labor demand curve will be shifted outward. Another general equilibrium situation will be determined. This change is similar to the effect of physical

capital accumulation, the effect of which was established in the static models of epsilon, omega, and sigma economies. Thus the new equilibrium will entail a higher output level.

Changes in schooling may also affect income inequality in the general equilibrium analysis. This is the relationship that has been studied in this section with some detail. In the three societies, the effect of increasing schooling exogenously proportionately or even favoring the poor on income inequality will be small.

Exclusion versus Discrimination in Labor Markets

Do firms buy education or human capital in the labor market? Motivated by the logic of profit maximization, firms will buy human capital in the labor market, which is the relevant factor of productivity and profits. Furthermore, in a competitive labor market, wage rates will be uniform for the same level of human capital, not for the same level of education. If differences in wage rates per schooling years were observed empirically, this fact would not contradict the model. This observation could not constitute a case of wage discrimination either (as is usually called in the literature of standard economics). A beta proposition of the sigma model is that firms will pay equal wages per equal levels of human capital, not for equal number of years of schooling.

Wage discrimination exists if firms pay different wages to workers with similar human capital. Why would this happen? Suppose a sigma economy in which y-workers and z- workers with similar levels of human capital are not paid the same wage rates; let z-workers receive less. The origin of this discrimination would come from factors such as the preferences of consumers, which would be biased against the good produced using z-workers as labor by the firm. The derived demand for z-workers would be different compared to the one for y-workers. Also, it could be originated in a problem of incomplete information that rules in the labor market. Capitalists would probably have little confidence on z-workers because of cultural differences and ethnic prejudices. In this case, transaction costs of employing z-workers would be higher in relation to y-workers. As a consequence, segmentation would appear in the labor market: given equal level of human capital, z-workers would obtain in the labor market an inferior wage rate compared to y-workers.

As in the case of mean income gap, the differences of wages between the groups of workers Y and Z can be decomposed analytically into three effects: (1) the quantitative exclusion of education (less schooling years); (2) the qualitative exclusion of education (less accumulation of human capital for equal schooling years); and (3) the wage discrimination in the labor market (less wage rate for equal human capital). The segment m'n' in Figure 2(b) can be decomposed in these effects if the horizontal axis is taken as wage rates instead of incomes.

The empirical prediction of the sigma model is that the low relative incomes (relative wages also) of z-workers relative to y-workers are generated mainly by exclusion rather than discrimination. Exclusion is the essential factor. Exclusion is the general factor while discrimination refers to special cases. The income gap between group Z and group Y cannot come from the fact that engineers of group Z receive a smaller wage rate compared to engineers of group Y; the major factors include the effect that the proportion of engineers within group Z is lower than the ones of group Y and the effect that the proportion of engineers endowed with a high level of human capital is lower in group Z compared to group Y.

Regarding omega and epsilon models, the results will be different. There will be less inequality in human capital among workers and thus less degree of inequality in wage incomes than in a sigma economy.

A Dynamic Model

The assumption now is that the quantity of education is determined endogenously in each type of society. Dynamic models of human capital accumulation will then be presented now. Consider the initial conditions of the economy, where the factor endowments (including human capital) are given and the initial inequality in the individual endowments of economic and social assets are also given. Remember that the initial inequality of learning skills among individuals is also determined by the initial inequality in asset endowments.

Consider, firstly, the sigma society. Given the initial conditions, including the initial stock of human capital, the general equilibrium of the sigma economy in the initial period will also be determined, which include the national income and its distribution of equilibrium. With these known values, the number of years of schooling of each social group will also be determined. The quantity of education of individuals will then be endogenous. New values of the stock of human capital will be determined for the next generation, which will determined national income and its distribution, which in turn will determine the quantity of education of individuals, and so on.

In the process of human capital accumulation, the income of the parents determines the income of their children, which is valid for the three social groups A, Y, Z. Even if education is supplied as public good, there is a private component in the investment in human capital. Then, richer parents with higher social position (group A) will transmit superior incomes to their children compared to what will be transmitted by poor parents with very low social position (group Z) for each additional year of education. In addition, richer parents (group A) will invest in more years of schooling for their children than poor parents do (group Z). The children of y-workers will lie in between. Therefore, the dynamic equilibrium of intergenerational income by social groups will show rising curves over time for each social group, but there will not be convergence between those curves; that is, the Z curve will not catch up with the A curve, as shown in Figure 4.

Given the initial inequalities, the degree of inequality δ, the reasons for the lack of income convergence are two: (a) On the transformation process of education into human capital, the major inequality in schooling years does not imply major equality in human capital because social groups accumulate human capital along different paths. (b) On the transformation of human capital into incomes, the access to basic markets is differentiated by social classes and ethnic groups, which does not change with the accumulation of human capital alone, which imply that human capital is transformed into income along different paths, depending on the social groups.

This dynamic model of the sigma theory predicts that children will inherit the relative position of their parents referring to income and social position. Relative incomes over time of the three social groups are subject to path dependency; in other words, the initial conditions of social groups do matter, their history counts.

The dynamic models of the omega and epsilon economies will predict the same relations. The dynamic equilibrium of intergenerational incomes for each social group (A and Y) will show rising curves over time but they will not tend to converge.

The initial inequality in the endowments of economic and social assets (δ) is the ultimate factor that determines the non-equalizing nature of the education system.

Education could not be an equalizing system in none of the three societies. Differences in years of schooling could be reduced, but this does not lead to reduction in income differences. Educational mobility is possible, but socio-economic mobility is more difficult.

In intergenerational terms, the "children" of one social group will have the tendency to inherit the economic status of the "parents" of that particular social group. The British biologist Galton established long time ago the "law of regression towards the mean" applied to the case of heights between children and their biological parents. The theoretical models of sigma, omega, and epsilon theories presented here predict that no "regression toward the mean" will occur in the case of incomes between children and their parents in the social sense.

4. Empirical Hypotheses

The empirical predictions that can be derived from the sigma model are the following:

H1. Hypothesis of the quantitative exclusion from education. A positive relation exists between the mean of years of schooling and the level of citizenship of ethnic groups.

H2. Hypothesis of the separability and hierarchy in the relations between education and incomes. A positive relation exists between the mean years of education and the mean incomes, but this relation is separable and hierarchical by ethnic groups, in the order A-Y-Z (as it is indicated in Figure 3).[5]

H3. Hypothesis of the separability and hierarchy in the relations between education and wages. A positive relation exists between the mean years of education and the mean wage rates, but this relation is separable and hierarchical by ethnic groups, in the order A-Y-Z (as it is indicated in Figure 3, if wages are measured on the vertical axis).

H4. Hypothesis of the efficiency wage. The mean wage rate is greater than the mean income of the self-employed, for a given level of education. This relation is shown in Figure 5.

H5. Hypothesis of the quantitative exclusion from the labor market. The proportion of the excess labor supply depends negatively on the level of education, but this relation is separable and hierarchical This relation is presented in Figure 6.

---

[5] This hypothesis corresponds to the reduced form equation derived from the following structural equations:

(1) $K_{hj} = u\ (E_j, X),\ j \in X,\ X=Z,Y,A,\ u_i > 0$

(2) $y_j = f\ (K_{hj}, X) = F\ (E_j, X),\ j \in X,\ X=Z,Y,A,\ F_i > 0$

The first equation shows that the human capital of the *individual j* depends positively on the schooling years and on the ethnic group X to which he/she belongs. The second equation shows that the individual income depends positively on the human capital and the ethnic group X to which he/she belongs. The function F shows the reduced form of the system and constitutes H2.

H6. Hypothesis of the asymmetric mobility in education and in incomes. There is a tendency towards convergence in years of education between ethnic groups. The differences of schooling years will be smaller between younger generations than between elder generations. However, there is no tendency towards convergence in mean incomes between ethnic groups (as it was indicated in Figure 4). Using the method of the transition matrix, this hypothesis implies a matrix of low correlation for the distribution of years of education between generations ("parents" and "children" socially speaking, not in the biological sense) and a matrix of high correlation for the distribution of incomes between generations.

Certainly, this set of empirical hypotheses would not have been derived from the neoclassic theory. The basic assumptions of the neoclassic theory differ from those of the sigma theory. The neoclassic theory assumes that the labor market is Walrasian and that there is no overpopulation. Also, it assumes a socially homogeneous society (social groups are A and Y only) and thus ignores the effect of initial inequalities in the endowments of social assets upon the process of production and distribution.

The majority of neoclassic *models* implicitly assume a unique curve which transforms education into human capital, by which the curve Y coincides with curve A and would have, in addition, a slope of 45º in panel (a) of Figure 2. This is because typically these models do not make a distinction between education and human capital. Lucas (1990), for instance, introduced human capital into the theory of economic growth, but it was based on the assumption of a socially homogenous economy and equated education to human capital.

In general, neoclassic models predict a unique relation between income and schooling years, that is, the line L in Figure 3. According to sigma theory, the distinction between education and human capital is considered an essential part of the theory. This is the alternative hypothesis to the one derived from the sigma model.

Empirical studies about the determinants of economic growth have been developed accepting these assumptions (cf. Barro and Sala-i-Martin 2005). It is true that neoclassic models have been constructed to explain ethnic differences related to incomes, especially between blacks and whites in the United States (cf. Becker ad Murphy 2001); however, for this type of explanation is commonly used the method of micro economic equilibrium which overlooks the implications of social asset inequality in the general equilibrium. In any case, these empirical studies aim at the interpretation of reality based on the neoclassic theory approach, but they do not aim at submitting the theory to the falsification test.

In the present study, Peruvian data will be used in order to confront empirically the predictions of the sigma model. Of course, this test will seek to corroborate whether Peru resembles the theoretical model, but the theory itself is not under falsification. The latter would require a large number of countries.

5. Database: Peru, 2003

The statistical testing will be will be carried out using a database constructed from the Peruvian household survey (*Encuesta Nacional de Hogares (ENAHO)*) 2002 and 2003, by the *Instituto Nacional de Estadistica (INEI*). These databases are available in internet in the following electronic address: http://www.inei.gob.pe

The sample for each survey is representative to the whole country; what is more, its particular design turns the sample into representative of various important segments of

the Peruvian economy: urban-rural and natural regions such as Coast-Andes-Amazonian. The unity of the sample is the household. In addition to the data referred to households, the survey recollected individual data about education, employment and incomes, and also data related to place of birth and place of residence, all very important variables of the model.

The size of the sample is nearly 18 thousand dwellings in 2002 and 20 thousand in 2003. The subsample used in this study is composed by individuals from 25 to older, nearly 38 thousand observations in 2002 and 39 thousand in 2003.

A critical analysis of internal consistency applied to both surveys arrived to the following results. The test for internal logic consistency of data is acceptable. The test of representativeness excluding the cases of "without information" also was acceptable. Although a high proportion of data "without information" was found in many key variables, an analysis of bivariate distributions of data permitted to distinguish data "without information" from data probably consignated as "zero" or "meaningless" at the moment of codification. Made the correction, the problems of data "without information" were reduced to an acceptable range.

Households usually count with family business, such as small stores in the urban place or a farm in the rural area. Family business generates collective incomes as the result of the activity of the family members. How to transform collective incomes into individual incomes? There are three options. First, assign this income to a particular member of the family, for example to the head of the family. This method creates biases because it overestimates the income of this person and increases the return to his/her education. Second, assign this income to all members. This option will create the problem of using appropriate criteria to do it. Third, do not attribute it to any member and maintain the income as a collective income. This alternative raises the problem of the reduction of returns to education of all individuals. Among these options, INEI seems to have used the first one. The addition of the individual incomes is then equal to the addition of family income. (Family and household seem to overlap largely in the sample.)

The INEI data is corrected by the rate inflation during the period of the sample. The period of the survey was from October to December in 2002 and from March to December in 2003. The nominal incomes have different time unities, because the interviewed declared their incomes for the habitual periods of payment or income generation. These incomes were taken then to a standardized unity of time, the trimester. Real incomes refer to prices of November 2002 and to July 2003. It is important to point out that in this period the *annual* inflation rate in Peru was low: 1.5% in 2002 and 2.5% in 2003. The correction of data on incomes for inflation has modified only a little the data recollected on nominal incomes.

One important correction had to be introduced in the INEI data however. It refers to differences in the level of prices that exist between Lima and the cities from the provinces. Therefore, data on real incomes for the Peruvian case used in this study are measured in prices of Lima of November 2002 for the first survey and prices of Lima of July 2003 for the second one. The month is the standardized time unit.

For the purpose of the present study, some categories of the survey had to be aggregated in order to construct the theoretical variables of the model. For other categories, this correction was not necessary.

In the set of basic tabulations carried out in both surveys, the finding was that statistics on central tendencies associated to key variables of the study were similar. For two surveys carried out in contiguous years, this result was expected. This result is applied also to the case of income variables (salaries, wages, and income of self employees, aggregate incomes). Tabulations and the statistical analysis were carried out for the two

surveys and the results were consistent. In this study, the results based on the survey of 2003 are presented and the only reason is that this is the most recent one.

The main variables of the theoretical models of the study include: social groups (A,Y,Z), labor category, schooling years, level of human capital (level of knowledge and age), total income, wages, salaries, net incomes of the self employed, employment (white-collar worker and blue-collar worker), self employment, unemployment and underemployment. The critical analysis of the sample and the especial tabulations derived for the purpose of this study appear in Figueroa, Cruzado and Sánchez (2007).

Empirical Estimation of Ethnic Groups

The empirical estimation of the size of ethnic groups is a complex task. Up to four criteria can be considered, each with shortcomings. *The criterion of mother tongue* has the problem of reducing ethnicity to language, so that the people who do not speak an aboriginal language change their ethnicity. *The criterion of self identification* reduces the indigenous population size in hierarchical countries where this population belongs to subordinated social groups and thus people tend to hide their ethnicity. *The criterion of rural residence* implies that people experiment an ethnic change just by migrating to the city. *The criterion of the place of birth* reduces the size of indigenous population because the children of indigenous people who are born in a different place from that of their parents change their ethnicity. In the case of Peru, the last criterion was selected because it seems the least imperfect and also because preserves the historical conditions of the formation of the Peruvian society, as it is explained in Figueroa and Barron (2005).

The social group Z is defined as the population born (independent of where they reside now) in the rural districts of the country along the three natural regions: Coast, Andes, and the Amazon. The social group A is defined as the population born in the 11 most residential districts of Lima (Barranco, Jesús María, La Molina, Lince, Magdalena, Miraflores, Pueblo Libre, San Borja, San Isidro, San Miguel, and Surco). The Y group was calculated by difference.

For the adult population (with 25 or more years of age), the estimates are the following: group A: 3.5%, group Y: 27.5% and group Z: 69%. If it is considered a more restricted definition of group Z, including those born in the rural districts of the southern Andes or the Amazon only (a common view among Peruvian middle class), the proportion will be reduced to 28%; if it is restricted even further, to include only rural districts of the southern Andes (another common view), the proportion decreases to 21% (Table 54 of the Research Report, which appears in Figueroa, Cruzado, and Sánchez (2007)).[6]

In this study, the wider concept of group Z will be used. Certainly, it is expected that the more restricted the definition of this group is, the more solid the tests applied to the hypothesis will become. This analysis of sensitivity of the results was carried out and will be reported only if a different pattern is found.

[6] The 2001 Household National Survey (ENAHO) asked a question of ethnic self identification. The result was that 43% of family heads declared to be indigenous descendants, while 34% declared to speak an aboriginal language (Hall and Patrinos 2006, tables 2.1, 2.7, 7.1). The estimates about the size of indigenous people are smaller than those shown above, but in both cases the order of magnitude is significant. For many Peruvian writers, the ethnic group that is predominant is the mestizo ("we are a mestizo country") and that the indigenous populations have almost disappeared.

6. Falsification of the Hypotheses

The empirical predictions derived from the sigma model constitute empirical hypotheses which are refutable. In this section, these hypotheses will be submitted to the statistical test. Parametric and non-parametric tests are used to test the mean differences. Both types of tests are shown in the Statistical Appendix at the end of the paper. Data refers to Peru and are based on the national household survey of 2003.

H1: Hypothesis of the Quantitative Exclusion

The empirical relation between years of schooling and the ethnic group of individuals is shown in Table 1. In terms of level of education, two thirds of the population is placed in primary and secondary level. With respect to the social groups, 3% of the population belongs to group A, whereas 26% to group Y, and 71% to Z.

The mean of schooling years varies depending on the social groups: 14 years for group A, 11 years for group Y, and seven years for group Z. The median shows the same relation: 14, 11 and five years. These data show a pronounced inequality. Group A has a mean of schooling years that doubles the one of group Z; in terms of the median, the difference is 2.8 times. Statistical tests show that these differences are statistically significant.

The variable level of education (primary, secondary, technical and university) instead of schooling years also reveals a pronounced inequality. In group A near 70% have post secondary education, while group Y shows 36%, and group Z presents 15%. The exclusion of group Z from post secondary education is remarkable.

It should be clear that such empirical findings are not the result of comparing "indigenous" populations living *at present* in rural areas with populations living in residential areas in Lima. The comparison is carried out between indigenous populations against mestizo and white populations *independently* from the place of residence; so it includes indigenous people living in Lima or living in the rural districts of Peru. The effect of rural-urban migration is thus included into those differences.

Another calculation was made involving only indigenous populations. Those living in their place of birth have a mean of five years of schooling, while those that emigrated have nine years of schooling (which is consistent with the mean of seven years for the group). Migration implies four years more of education, on average. The theoretical separation of group Z in two groups of workers presented in Section 1 above (with different levels of human capital and different areas of residence, urban or rural) seems to have a good empirical consistency.

There are few studies about the determinants of the level of knowledge acquired at different types of school in Peru. They have shown that the level of student's knowledge at the same schooling year varies according to the type of school (private-public) and depending on the location (rural-urban). The level of knowledge is greater if the student attends private schools rather than at public schools and even greater, at any urban school compared to the rural school (Rivera 1979; Cueto, Jacoby and Pollit 1997; Peru, Minister of Education 2005). Both factors of the production function of human capital (school resources and socio-economic conditions of families) seem to make a significant difference in quality of education between these two types of schools.

In terms of socio-economic factors, there is evidence of the negative role of malnutrition over the learning process at school. The study made within the ECIEL

Program, in particular, shows a negative correlation, which is statistically significant and quantitatively important, between school performance and the degree of malnutrition based on a sample of the population of students in Lima, Puno-city and Puno-countryside (Rivera 1979). Because the social group Z is predominant in rural areas, it follows that the transformation of education into human capital is as shown in the structural equations of the model.

In sum, Table 1 shows a significant inequality in the endowment of schooling years between ethnic groups in 2003. These results are consistent with the hypothesis. After 180 years of Republican life, and despite the spreading of the educational system, Peru shows yet a significant educational exclusion.

H2: Hypothesis of separability and hierarchy in the relation education-incomes

Table 2 presents data on incomes by level of education and by social group. In this case population refers to the economically active population and those employed (the unemployed are not considered). The mean of schooling years increases in one year and the median in two years with respect to the values shown in Table 1. These averages increase also for the social groups, in particular for group Z.

The table suggests the existence of a positive relation between the mean income and the level of education for the whole population. In addition, this positive relation can also be observed in each of the three social groups.

The question is whether the observed relation between incomes and education is statistically separable for each social group and whether it shows a hierarchal order between A-Y-Z, as the hypothesis proposes. In general, table 2 shows that for each level of education the mean income is greater in group A, which is greater than in group Y, which in turn is greater than in group Z. The hierarchy of incomes goes as the hypothesis says.

The statistical test to assess these differences should take into account the results of hypothesis H1. Given the existence of quantitative educational exclusion, which corroborated hypothesis H1, the test must be carried out only between *comparable* levels of education. Figure 7a shows the nature of the empirical relation found between incomes and education. In group Z, people are concentrated in the first levels of education, very small fraction, less than 10%, reach post secondary levels. In group Y, there is a very small population at the level illiterate. Illiterates or people with only primary level were not observed within group A.

The comparable levels of education and social groups are the following: at primary level, groups Y-Z; at secondary level, the three groups A-Y-Z; and at technical and university levels, the groups A-Y. The statistical tests showed that all these differences in the means are statistically significant.

The slopes of the lines reflect an increasing trend in the three social groups. The slope of curve A probably grows more rapidly than the one of Y, and this one even more compared to Z. The standard method of regressions could be applied to this data so as to obtain an econometric test on the relations between the slopes, using the social groups A-Y-Z as fictitious variables (*dummy variables)*. However, data reveals secondary education as the unique comparable level. An implicit assumption of the standard model of regressions is that the fictitious variables have to cover all the range of observations of independent variables; if exclusion exists, as in this case, the model is not applicable. In fact, the econometric models assume implicitly the absence of exclusions in the relations between variables. The exclusion phenomena cannot then be analyzed under this method.

An analysis of independent regressions for each social group showed that, in effect, the slopes differ and that such differences are statistically significant. The result appears in Figure 7b. In sum, the data does not refute hypothesis H2.

H3. Hypothesis of separability and hierarchy of relations education-wages

In the empirical study on the relations between levels of education and labor remunerations, it is necessary to distinguish two labor categories: white-collar workers and blue-collar workers. The two categories have differences due to the level of education; but also they differ qualitatively because the category of white-collar workers has a higher social status than the blue-collar workers. In sum, two separated markets are considered for each group.

The tables 3a and 3b present data on education and remunerations related to the two markets, the one of white-collar workers and the other of blue-collar workers. The mean of schooling years is 13.7 years for white-collars and 8.5 for blue-collars; the median values are 14 and 10. The levels of education appear as one factor of difference between the two groups. In consequence, the proportion of Z-workers in the blue-collar worker market is 20% and in the white-collar worker market is only 14%; whereas, the proportions within group Y are 21% and 36%, and 8% and 62% in group A. Figure 8 shows the results.

Table 3a presents data for the blue-collar workers group. The majority of this group has primary and secondary education. Illiterates do not participate; neither does the group with high education. Therefore, the social group A is not part of the labor supply in this market. Within the group of blue-collar workers, it can also be observed that those with secondary education get the higher wages. This relation also works for every social group. Given the primary level, the social group Y receives a higher salary than the group Z; similar results are observed in the secondary level. These differences are, in general, statistically significant.

Table 3b presents data on levels of education and mean wages for white-collar workers. The mean and median of the schooling years of this group show little variation between ethnic groups. Then, white-collar workers seem to constitute a relatively homogenous group with respect to the endowment of schooling years; that is, most of them completed post-primary education and near 70% acquired post-secondary education.

Within the group of white-collars, the relation being corroborated is the one showing that the higher the level of education is, the higher the wage rates. The same positive relation between wages and education is observed for each social group. The relation that establishes this hypothesis is that, for a given level of education, the wages should show a hierarchy concerning the social groups A-Y-Z. The data demonstrate that, in effect, there is such hierarchy at the university level; however, the differences observed at the levels of secondary and technical are small. The statistic tests point out that differences are in general significant.

In sum, the empirical evidence presented here shows that the wages or salaries depend positively on the level of education of workers, both inside the group of white-collars and blue-collars. As the level of education of white-collars is higher than the one of blue-collars, it is also observed that the mean wage is higher than the mean salary. Finally, for a given and comparable level of education, wages depend on ethnicity, that is, there is a hierarchy, given by the order A-Y-Z; the same relation is observed in the case of salaries. In sum, the empirical data presented do not refute the hypothesis H3.

H4. Hypothesis of efficiency wage

Table 4 presents data on mean incomes of the wage-earners and that of the self-employed, by ethnic groups. The hypothesis of the efficiency wage refers to each social group; therefore, the relation will be examined considering the three social groups.

In the social group Z, it is observed that the majority of workers are either self employed or blue-collar. The hypothesis establishes that mean incomes of the first group must be inferior compared to the second one. This relation is corroborated for the aggregate. It is observed that this relation is also valid for each level of education, except for the university level. However, it should be noticed that a relative small quantity of group Z exist at university level, which generates a weak comparison in statistical terms. The statistical test shows that the observed differences are all significant in the relevant levels, which are: primary, secondary, and technical, and also for the aggregate.

In the social group Y, the majority of workers are either self employed or white-collar. It is observed that the mean income of the self employed is less than the mean wage for each level of education and also for the aggregate. The statistical test shows that the observed differences are all significant for the comparable or relevant levels, which are: primary, secondary, technical and university; as well as for the aggregate.

In the group A, the majority are self employed or white-collar. In this group the levels of no-education and primary are irrelevant. In the rest of the levels, the relation assumed by the hypothesis is corroborated for all the cases, except at secondary level. Given the small size of the sample, these differences were not subject to statistical test.

It is also important to point out that, in Table 4, a generalized positive relation is observed between incomes and levels of education. In other words, in every column of the table, independently from the social group and the labor category, this relation is being observed.

Figure 9 summarizes the results. The statistical test of the hypothesis of efficiency wage is first applied to Y-workers; in effect, it is corroborated that the mean wage is greater than the mean income for the aggregate and for the relevant levels of education. The statistical test of the hypothesis of the salary of efficiency is then applied to Z-workers; in effect, it is corroborated that the mean salary is greater than the mean income of the self employed, both in the aggregate and in the relevant levels of education. Therefore, the labor markets seem to operate with efficiency wages and salaries, as predicted by the sigma model, where the self employed constitute part of the excess labor supply.

H5. Hypothesis of the quantitative exclusion from the labor market.

Table 5 shows data on unemployment and underemployment by social groups. The "unemployed" include the workers who are looking for a job actively. The group of "underemployed" is defined as those who are self employed and have, on average, smaller incomes than the dependent workers, for a given level of education.

In the aggregate, the employment rate is 7% and the rate of self employment is 51%, implying an estimation of 58% of the rate of excess labor supply in Peru in 2003. As for the differences in the level of education, data show that the rate of excess labor supply is correlated negatively, as it predicts the theoretical model, with the level of education: the higher the level of education is, the smaller the rate. The rates of unemployment do not

vary excessively between the levels of education. The theoretical model in fact does not have definite predictions on this rate.

As for the differences by social groups, the results indicate that the rate of excess labor supply are highly pronounced: 66% in the group Z, 45% in Y, and 32% in A. Given that the differences in the unemployment rates are not important, the difference is mainly due to the underemployment rates.

In the social group Z, an empirical counterpart regarding the two groups of workers defined in the theoretical model (Section 1) can be found. If those who did not reached secondary education are defined empirically as group Z_0, this group involves nearly 2.9 millions of people representing 35% of the national labor force (which is 8.4 millions). Then, the excess labor supply could be decomposed into 7% of unemployment, 16% of underemployment (partial exclusion from the labor market) and 35% for total exclusion of the labor market (Z_0 workers).

The crucial test of the hypothesis is whether the rate of labor exclusion is lower in social groups, in the hierarchical order Z-Y-A, for a given level of education. As for the level of no-education, a relation between groups does not exist because the size is very small for Y and A. At primary level, the comparable groups are Z and Y: the rates of exclusion are 80% and 62%. At secondary level, the three groups can be compared and the resulting rates are 58%, 50% and 39%. As for the technical level, the comparable groups are Y and A, with observed rates of 31% and 26%. At university level, the comparable groups are also Y and A, with rates of 33% and 29%.

In all cases, the observed rates are consistent with the hypothesis. The statistical tests indicate that, in effect, the observed differences in these rates are statistically significant.

H6. Hypothesis of high mobility in education and low mobility in incomes

Tables 6a and 6b present the ratios of education and incomes by inter generational groups of workers and by social groups. Age brackets are used as measure of different generations. In table 6a, an important increase can be appreciated in the schooling years between generations of the social group Z: the "children" (in the social, not biological, sense) have the double of years of schooling than their "parents". The increase generated in other social groups is smaller. In the social group A, it seems that there is an educational threshold which is 14 years. The consequence is that the educational gaps between the group Z and the other group have shortened considerably.

The estimated changes in income gaps are presented in Table 6b. In social group Z, it is observed that the mean income of this group in relation to social group A has not changed much: the relative income of the "parents" is 30% and that of the "children" is 40%. This change is small compared to the changes occurred in education attainment shown in the previous table. The situation of social group Y seems to have a similar pattern. Compared to the income of the group A, the relative income of the "parents" and "children" is almost constant, around 40% and 50%, except for the eldest group in which the figure is 100%. The latter result is certainly a paradox. The small sample size of adult people in the group A is something that may explain this paradox.

The conclusion to highlight about the convergence tendencies between social groups of different generations is that, while there is a tendency to reduce the differences in schooling years, there is no similar tendency to reduce income inequality. This result is consistent with the hypothesis.

From the data base utilized in this study, an estimate of the degree of inequality has been obtained. The income distribution by ventiles of the population is shown in Table 7. The corresponding Lorenz Curve is shown in Figure 10. The degree of concentration is significant as can be seen in several points of the distribution: the poorest third of the population receives 5% of total income, whereas the top decile receives 45%. As to the position of social groups in the income pyramid, Table 7 shows that the poverty incidence is higher for the group Z (43% are found in the poorest third), whereas the wealth incidence is very high for the group A (40% are found in the top decile).

The associated Gini Coefficient is equal to 0.59 for Peru in 2003. This figure must underestimate the true Gini coefficient because incomes of the Peruvian economic elite are not included in the sample. It is well known that economic elites are absent or underrepresented in household sample surveys. Strictly speaking, the empirical social group A refers mostly to the middle class (executives, administrators, and top professionals) and does not correspond to the capitalist class of the model. Theoretically, in the dual social structure of capitalists and workers, the middle class is usually included in the capitalist category (Wright 1997). This is also the criterion followed in this study.

Previous estimates of the Gini Coefficient for Peru were higher than the estimate found in this study, around 0.62 (Figueroa 2003, Table 6.2). In any case, Peru still shows a relatively high degree of inequality and remains in the group of the most unequal countries of the world. Education does not seem to be, as the sigma model predicts, an equalizing system.

7. Conclusions and Policy Implications

The statistical relations presented in this study do not refute the predictions of the sigma model. In the education-income relations in Peru, indeed, people's ethnicity matters. In the process of generation and reproduction of inequality in Peru, its history counts. In sum, Peru resembles a sigma society; the empirical predictions of the sigma model are consistent with the observed data.

The role of ethnicity in the education-income relations has been scarcely studied in the international literature, as it is shown in the revision of the literature carried out by Loaiza (2006). However, there are few empirical studies about the First World that can be compared with the results of this study. As for hypothesis H3, the works by Shapiro (1984) and Trejo (1997) for the United States and Chiswick (1980) for England show that, in the relation education and incomes, ethnicity matters. For Peru, the study by Ñopo et al (2004) found that, in urban wage differentiation, ethnicity matters.

As to the hypothesis of the persistence of intergenerational inequalities, the studies by Coach and Dunn (1997) for the United States and Germany, and Solon (1992) and Zimmerman (1992) for the United States find that intergenerational mobility is too weak: children tend to inherit the relative position of their parents in schooling years, occupation, or incomes. The study by Borjas (1984) about the United States introduces the ethnic variable and finds that it is an explanatory variable in the regression. With less quantitative methods, Bowles, Gintis and Groves (2005) arrive to the same conclusion. For Latin America, Dahan and Gaviria (1999) and Behrman, Gaviria, and Székeley (2001) show less conclusive results, but ethnic variables are ignored. In sum, the results obtained for Peru are not contradictory to those obtained for other parts of the world.

Therefore, there is no reason to reject the sigma theory at this stage of the research. This result implies that neoclassic theory, the competitor, would not be able to explain these facts.

Education is not an equalizing system in sigma type societies. But it is the *current* education system that fails; not any education system. How to change the current education system? Certainly, by changing the exogenous variables of the theory. Just to make sure, the exogenous variable in sigma theory is the initial inequality in the endowment in economic and political assets among ethnic groups. Once sigma theory is accepted, it can guide the discussion on policies. Turning to objectives, one precise objective of long term policies should be to reduce inequality, not only in education between ethnic groups (which is under way), but most importantly in human capital.

The standard development economics usually refers to the need for the implementation of *equal opportunity* policies, the content of which is unclear. With the help of Figure 2, it is possible to provide analytical content to this proposal for the case of education policies. In a sigma society, equal opportunity in education means to shift the Z and Y curves into curve A in Figure 2(a). If equal opportunity education policy is defined as the opportunity for workers to have access to higher schooling years only, inequality would not change because that would imply that the poor should continue to move along the curve Z in Figure 2(a). The equalizing education system implies, on the contrary, the displacement of the curves of human capital accumulation upwards.

The other common education policy is called *affirmative action*. It consists in applying measures of discrimination (positive) in favor of the poor so that their children can have access to school. To give subsidy to poor families subject to the condition that their children should remain at school is an extended policy in Latin America now. Again, this policy implies that the poor will increase years of education but, if nothing else is changed, they will move along the curve Z. Affirmative Action would imply a different package of measures if it is a policy by which there is intended a displacement of the curve Z towards curve Y and both towards curve A.

Health services constitute another essential factor in the accumulation of human capital. In this study, health services play, in effect, a very important role in this process. One of the factors underlying the relations shown in Figure 2(a) is the supply of health services. The analytical distinctions mentioned above about policies associated to both equal opportunity and affirmative actions also apply to the case of health service supply. One way to displace the curves of human capital accumulation is by the elevation of quantity and quality of health services offered to the poor.

In the short term, one of the main objectives of the economic policy would have to focus on reducing the problem of excess labor supply. For the case of Peru, the database used reveals a high level of excess labor supply, near 58% of the labor force. This rate is composed by 7% of unemployment, 16% of underemployment which is the partial exclusion from the labor market (mainly mestizo population), and 35% of total exclusion from the labor market, except for some temporal wage employments (mainly indigenous population).

In relation to this structure of the excess labor supply, the employment policy cannot focus only on the unemployed group, which is the conventional view. It is important to include among the objective group the underemployed. In any case, demands for employment come generally from the mestizos, who have political voice. The new perspective should be to include in the employment policies also indigenous workers, both from the rural or urban area, which in the case of Peru is composed by small scale farmers (*campesinos*), craftsman, and merchants. Due to their political exclusion, they are voiceless in the priorities of public policies; however, they are also part of the excess labor supply, as this study has shown.

The problem of employment refers to the existence of great excess of labor supply, which varies depending on the ethnic group. In sum, the objective of public policies

should be the reduction of this excess of labor supply. Once "the problem of employment" is defined in this manner, the objectives and the instruments of public policies would have to be different from the current ones.

References

Akerlof, George y Rachel Kranton. 2000. "Economics and Identity." Quarterly Journal of Economics, 115(3), 715-733.

Barro, Robert y Xavier Sala-i-Martin. 1995. Economic Growth. McGraw Hill.

Becker, Gary y Kevin Murphy. 2001. Social Economics: Market Behavior in a Social Environment. Harvard University Press.

Behrman, Jere, Alejandro Gaviria y Miguel Székely (2001). "Intergeneracional Mobility in Latin America". Working Paper 452, Junio. Washington DC: BID.

Borjas, George. 1992. "Ethnic Capital and Intergenerational Mobility." Quarterly Journal of Economics, 107 (1), 123-150, February.

Bowles, Samuel, Herbert Gintis y Melissa Groves (eds.). 2005. Unequal Chances: Familiy Background and Economic Success. Princeton University Press.

Chiswick, Barry. 1980. "The Earnings of White and Coloured Males Immigrants in Britain." Economica, 81-87, February.

Couch, Kenneth y Thomas Dunn. 1997. "Intergenerational Correlation in Labor Markets Status: A Comparison of the United States." Journal of Human Resources, 32 (1), 210-232, Winter.

Cueto, Santiago, Enrique Jacoby y Ernesto Pollit. 1997. "Rendimiento escolar de niños y niñas en zonas rurales y urbanas del Perú." Revista de Psicología de la Universidad Católica del Perú, Vol. XV, No. 1.

Dahan, Momi y Alejandro Gaviria. 1999. "Sibling Correlations and Social Mobility in Latin America." Working Paper 395, Office of the Chief Economist. Washington DC: BID.

Deininger, Klaus y Lyn Squire. 1996. "A New Data Set Measuring Inequality." The World Bank Economic Review, Vol. 10, No. 3, September.

Figueroa, Adolfo. 2003. La sociedad sigma. Una teoría del desarrollo económico. Lima and México: Coedition Fondo Editorial de la Pontificia Universidad Católica del Perú and Fondo de Cultura Económica.

Figueroa. Adolfo y Manuel Barrón. 2005. Inequality, Ethnicity and Social Disorder in Peru. Working Paper 8, CRISE, University of Oxford, March.

Figueroa, Adolfo, Viviana Cruzado, and Claudia Sánchez. 2007. "El problema del empleo en una sociedad sigma: Anexo metodológico y base de datos." Documento de Trabajo 256. Departamento de Economía, Pontificia Universidad Católica del Perú, Lima, May.

Figueroa, Adolfo. 2005. "Competition and the Circulation of Elites." Paper presented to the Conference on Regulation, Competition, and Income Distribution: Latin American Experiences, organized by the University of Illinois, University of Manchester, and University of Sao Paulo, Paraty (Brazil), November 18-21, 2005. Departamento de Economía, PUCP (manuscrito). Forthcoming in The Quarterly Review of Economics and Finance, 2007, June.

Gardner, Howard. 1999. Intelligence Reframed: Multiple Intelligences for the 21st Century. Basic Books.

Hall, Gillete and Anthony Patrinos, 2005. Indigenous Peoples, Poverty, and Human Development. Palgrave.

Hanushek, Eric y Javier Luque. 2003. "Efficiency and Equity in Schools Around the World." National Bureau of Economic Research, Working Paper 8949, May, Cambridge, Massachusetts.

Hudson, R.A. 1996. Sociolinguistics. Cambridge University Press.

Li, Hongyi, Lyn Squire y Heng-fu Zou. 1998. "Explaining International and Inter-temporal Variations in Income Inequality". Economic Journal 108 (1), pp. 26-43, January.

Loaiza, Carlos. Revisión de la literatura internacional sobre educación y mercados laborales. Lima: Departamento de Economía, Pontificia Universidad Católica del Perú, febrero (manuscrito)

Lucas, Robert. 1990. "Why Doesn't Capital Flow from Rich to Poor Countries." American Economic Review, Vol. 80, No. 2.

Maloney, William. 2004. "Informality Revisited." World Development (22) 7, pp. 1159-1178, July.

Ñopo, Hugo, Jaime Saavedra y Máximo Torero (2004). "Ethnicity and Earnings in Urban Peru". Discussion Paper 980, Institute for the Study of Labor, IZA, Bonn.

Perú. Ministerio de Educación. 2005. Evaluación Nacional 2004. Lima (Press Release Note of 10/28/2005)

Ratey, John. 2002. A User's Guide to Brain. Vintage Books.

Rivera, Iván. 1979. "Los determinantes de la calidad de la educación en el Perú. Un Estudio ECIEL". Publicaciones CISEPA. Documento de Trabajo No. 44. Lima: Departamento de Economía, Pontificia Universidad Católica del Perú.

Rouseau, Jean-Jacques. 1755. A Discourse on Inequality. Traducción del francés por Maurice Cranston. Penguin Books, 1984.

Searle, John. 1995. The Construction of Social Reality. New York: The Free Press.

Shapiro, Carl y Joseph Stiglitz.1984. "Equilibrium Unemployment as a Worker Discipline Device." The American Economic Review, (74) 3, June.

Shapiro, David. 1984. "Wage Differentials Among Black, Spanic, and White Young Men." Industrial and Labor Relations Review , 37 (4), 570-581, July.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." The American Economic Review, 82 (3), 393-408, June.

Todd, Petra y Kenneth Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievements." Economic Journal, 113 (45), February, pp. F3-F33.

Trejo, Stephen. 1997. "Why Mexican American Earn Low Wages?." Journal of Political Economy, 105 (6), 1235-68, December.

World Bank. 2001. World Development Report. Attacking Poverty. Washington DC.

Wright, Erik. 1997. Class Counts. Comparative Studies in Class Analysis. Cambridge University Press.

Zimmerman, David. 1992. "Regression Toward Mediocrity in Economic Stature." The American Economic Review, 82 (3), 409-19, June.

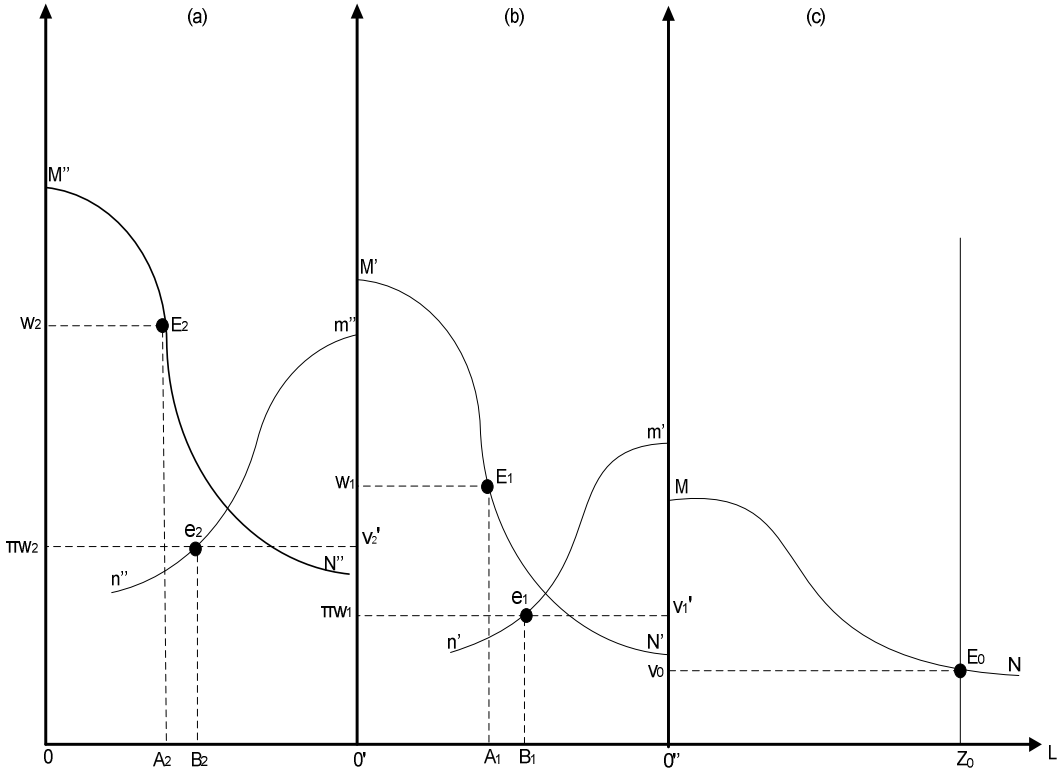

Figure 1. Labor markets in a sigma economy.

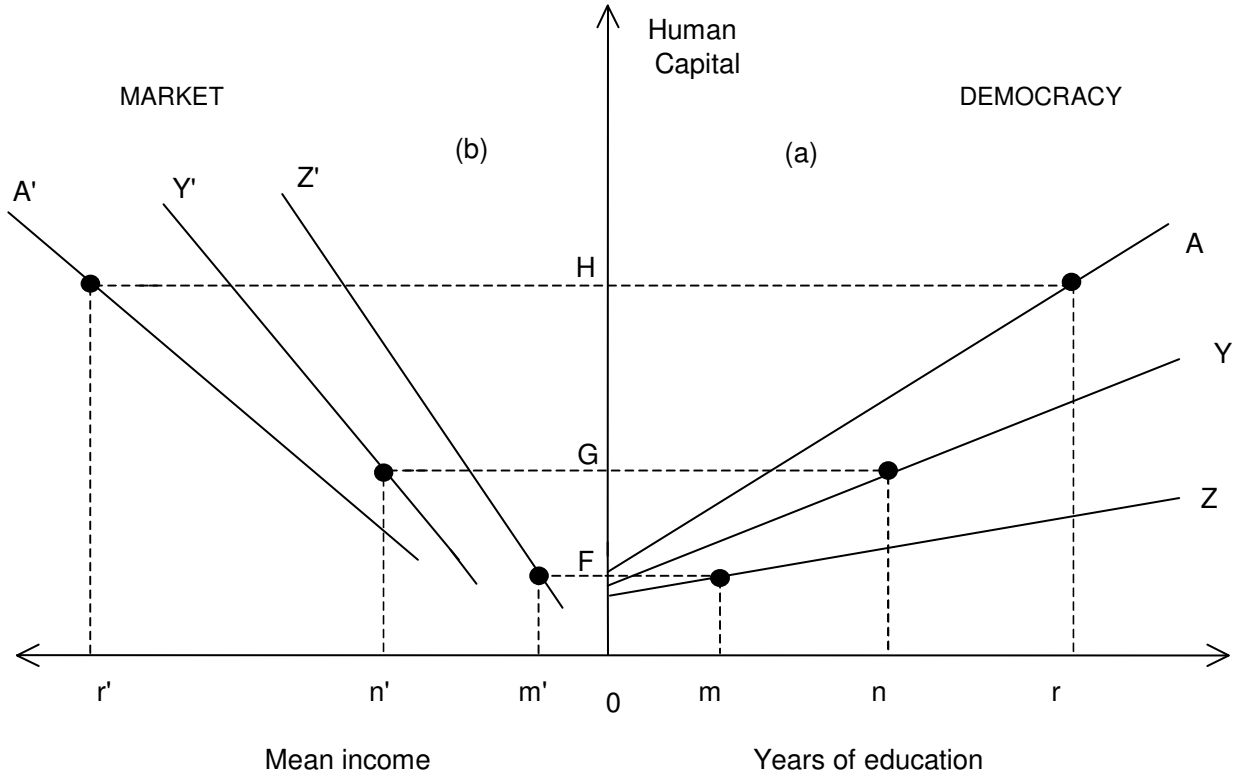

Figura 2. Theoretical relations between education, human capital and mean income.

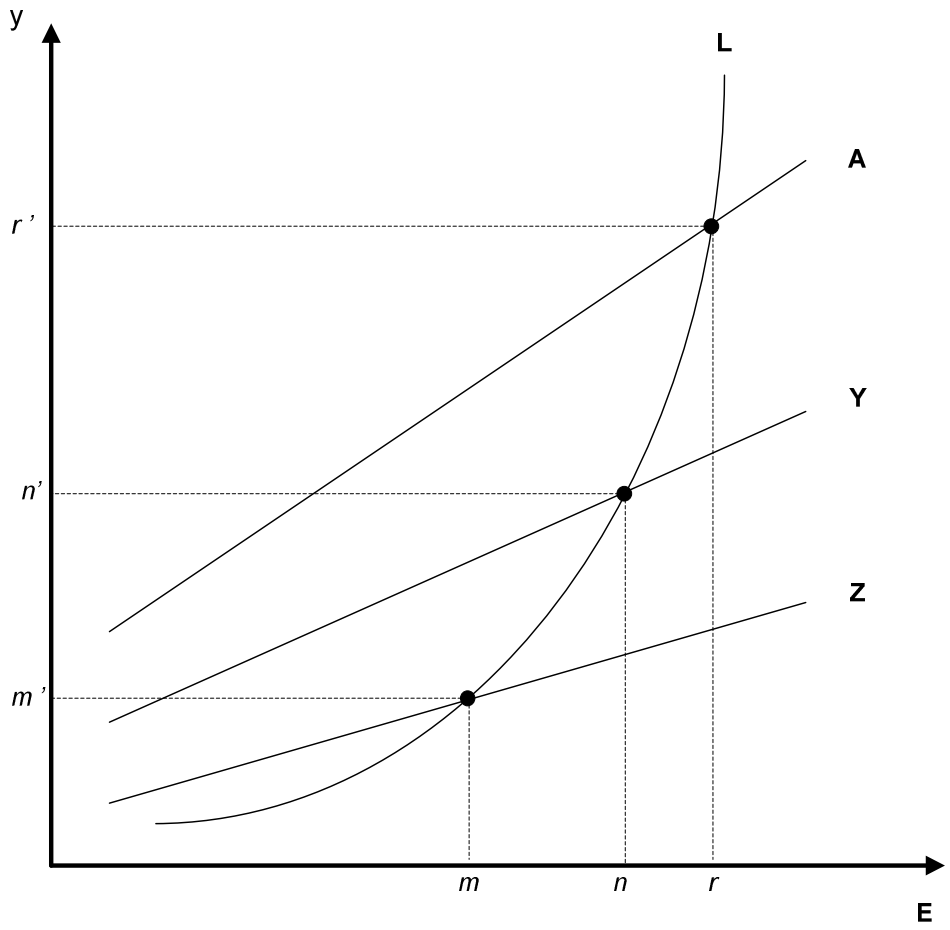

Figure 3. Hypothesis on the relationship between education, mean income and ethnic groups.

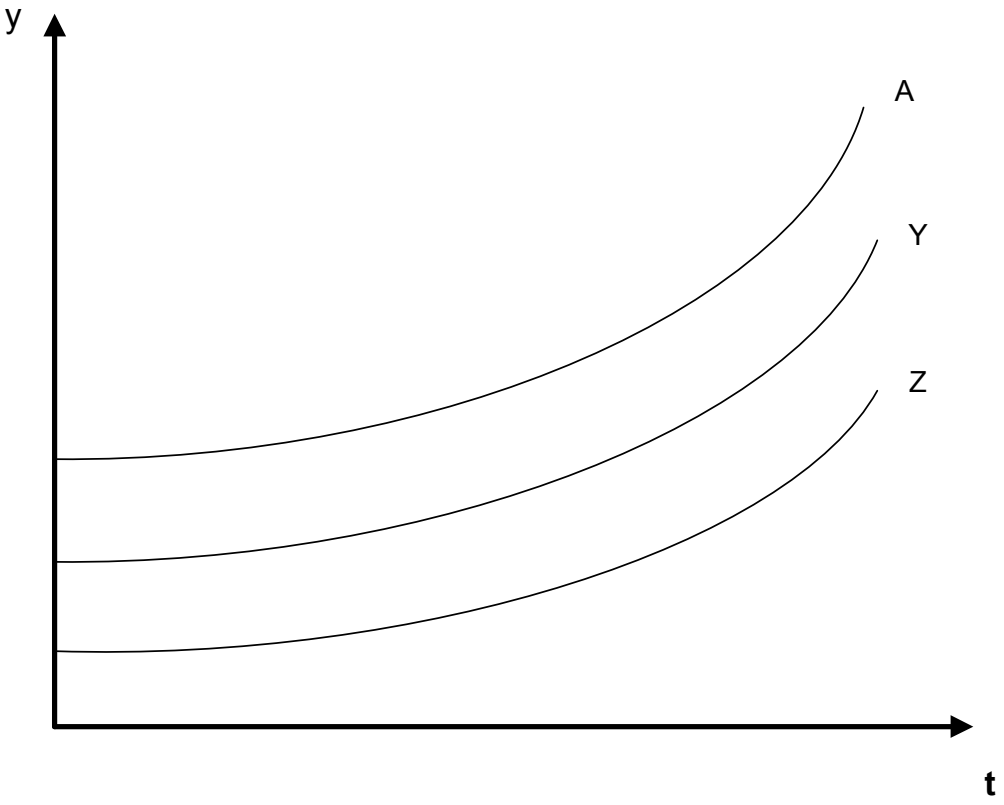

Figure 4. Hypothesis on the trajectories of mean income by ethnic groups.

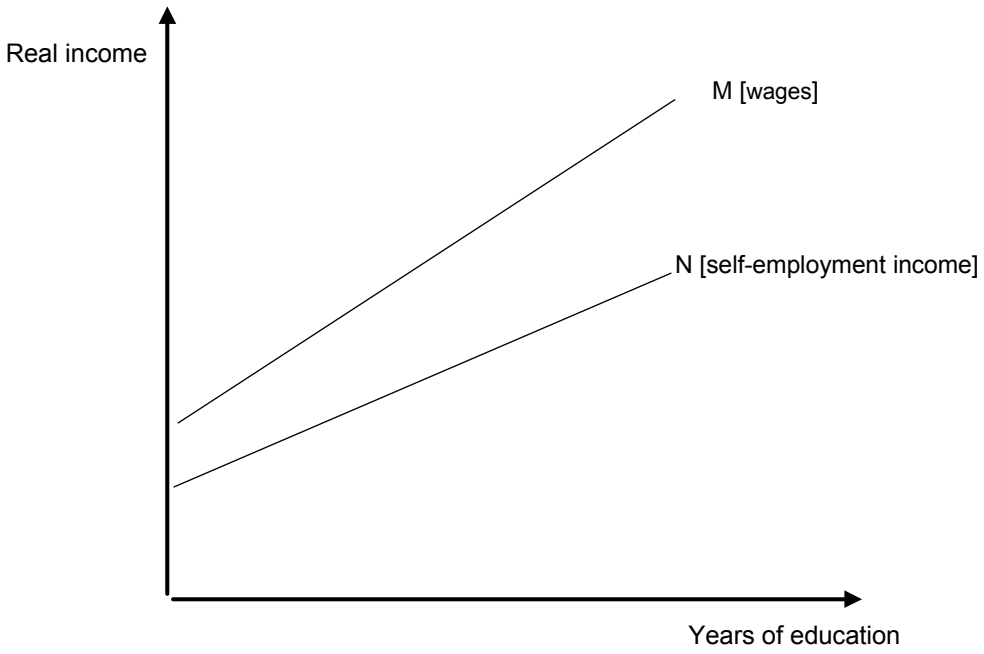

Figure 5. Hypothesis on the relationship between wages, self-employment income and education for a given ethnic group.

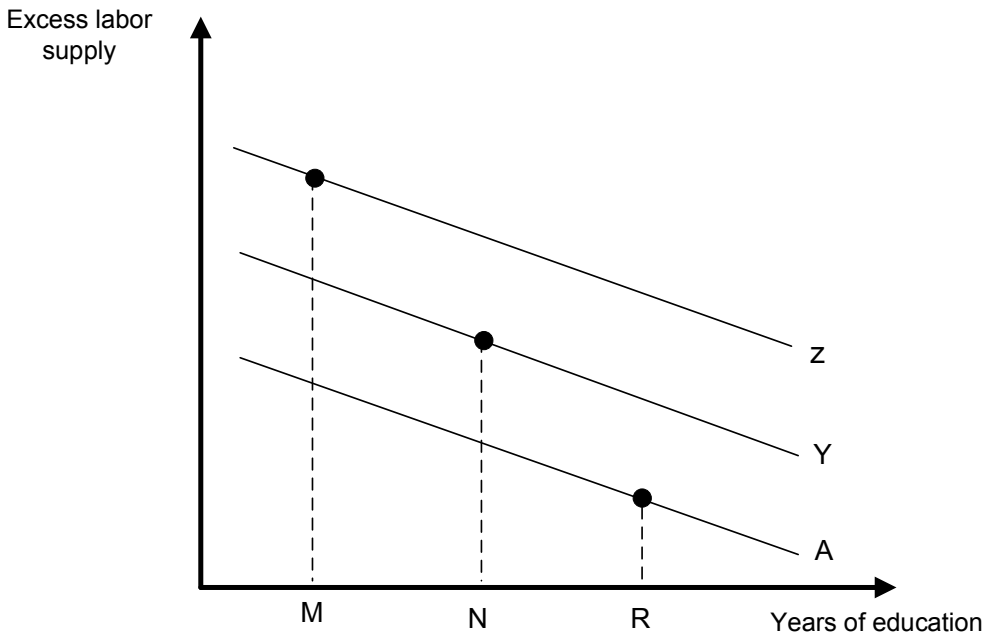

Figure 6. Hypothesis on the relationship between excces labor supply and education by ethnic groups.

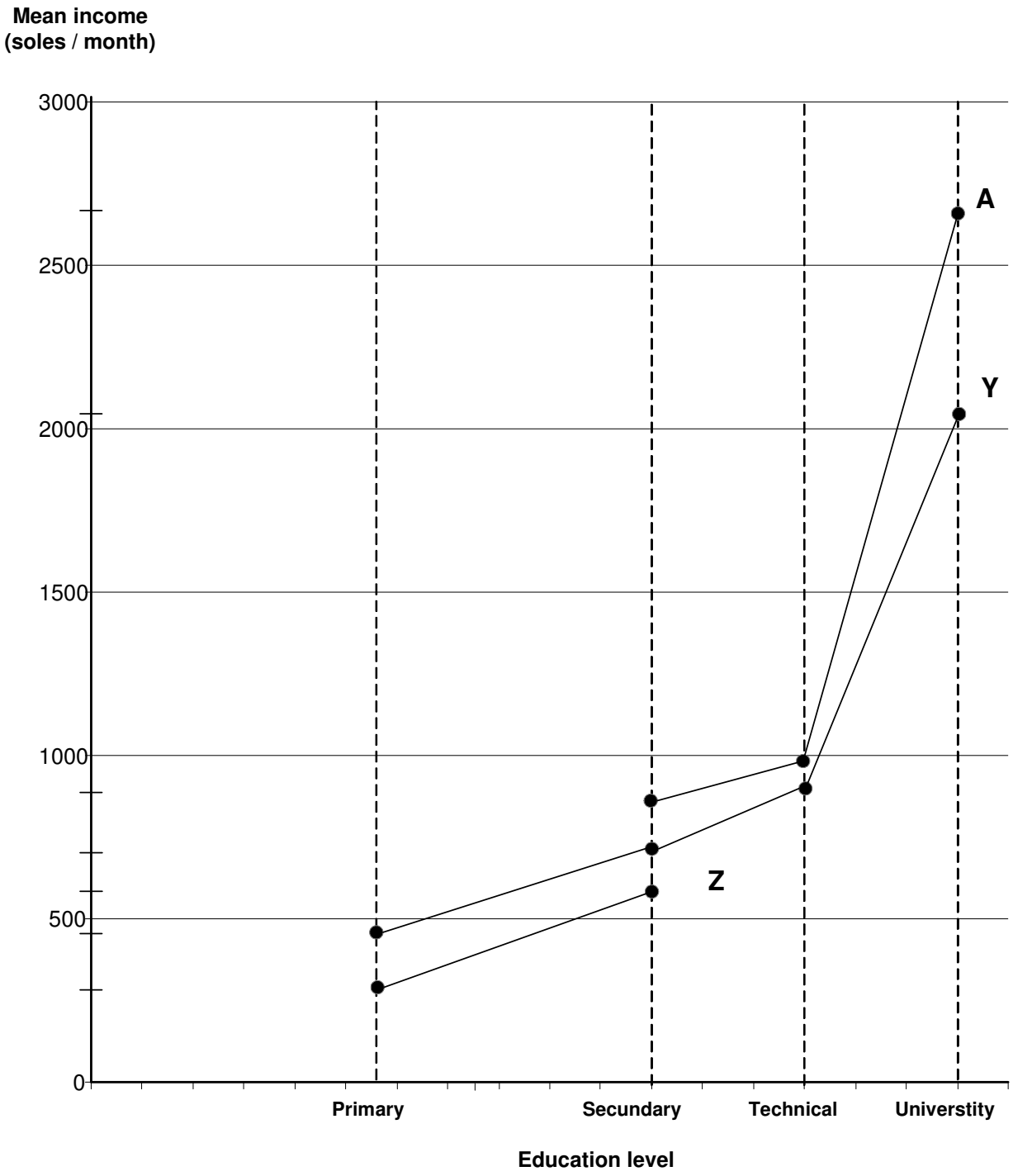

Figure 7a. Peru: Empirical relation between mean income and education level y ethnic groups (source: table 2).

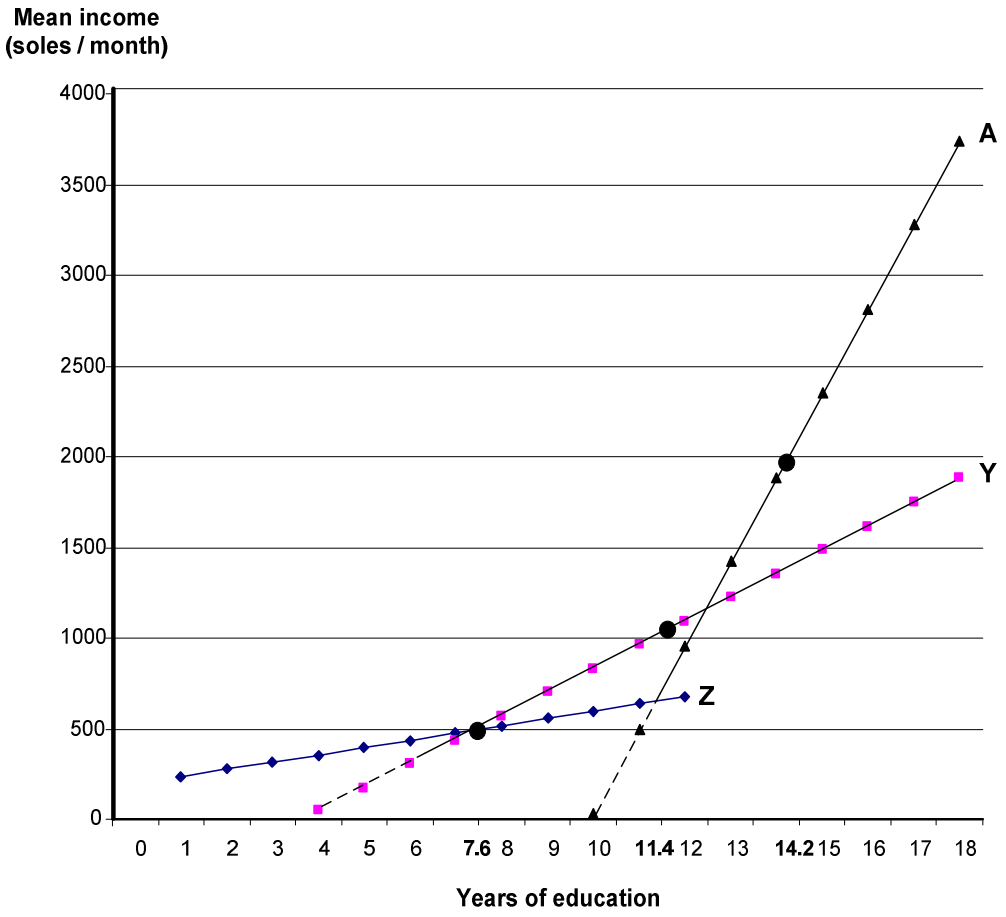

Figure 7b. Peru: Regresion lines on the relationship between mean income and education by ethnic groups.

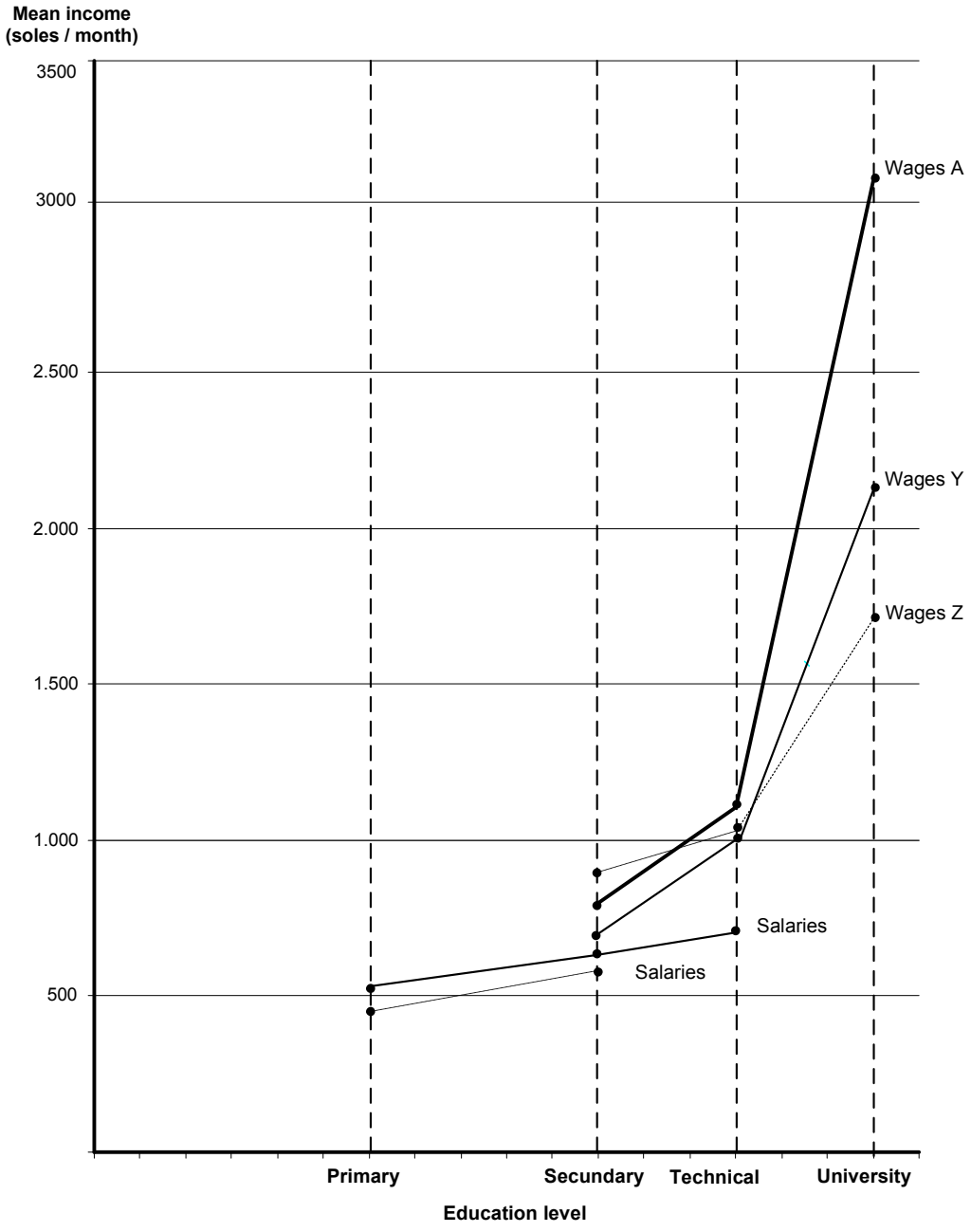

Figure 8. Peru: Empirical relations between wages, salaries and education by ethnic groups (source: tables 3a, 3b).

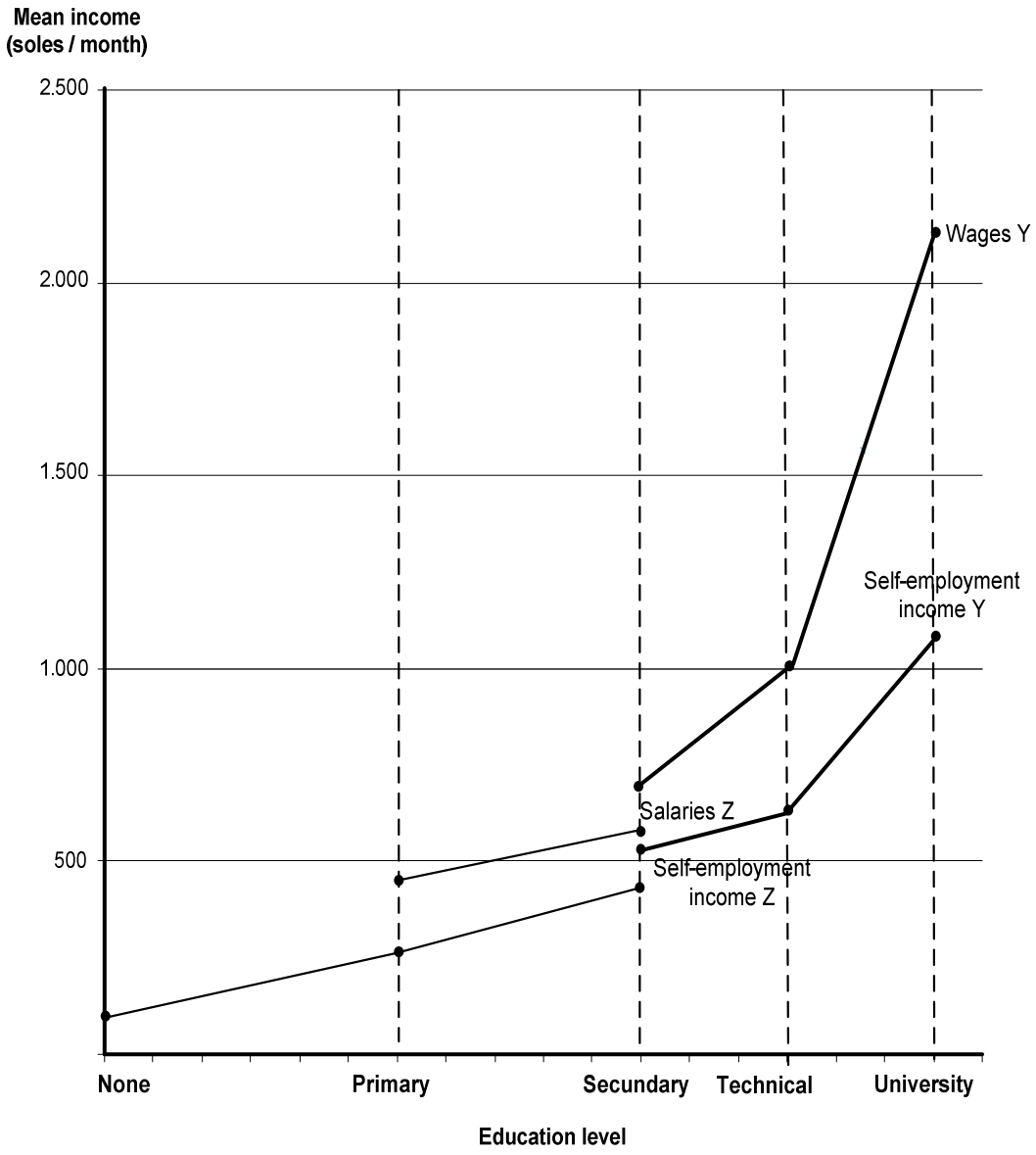

Figure 9. Peru: Empirical relations between mean incomes from wages, salaries and self- employment by ethnic groups.

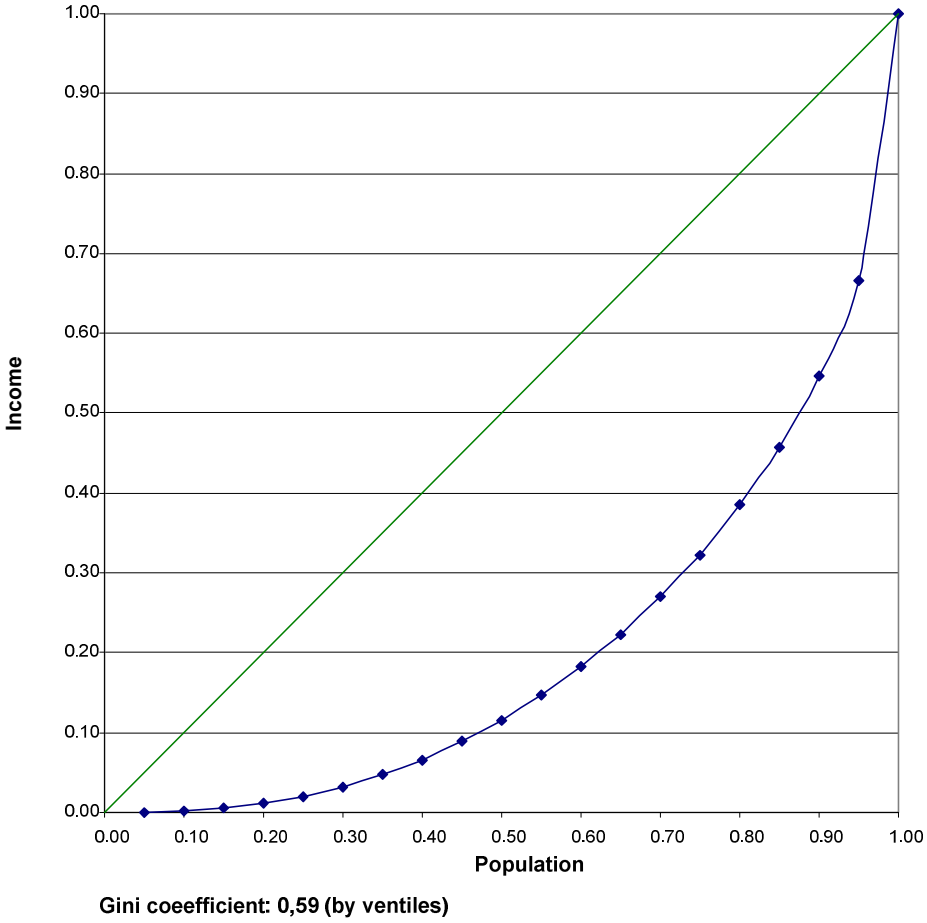

Figure 10. Peru: Lorenz Curve, 2003 (source: Table 7).

Table 1. Peru: Education level by social groups, 2003 (1000 people and percentages).

Education level	Social group						Total	
	Z		Y		A			
	N	%	N	%	N	%	N	%
None	1,283	14.1	110 *	3.3	0.7 *	0.2	1,394	10.9
Primary	3,903	43.0	556	16.5	10.1 *	2.8	4,470	34.9
Secondary	2,611	28.7	1,475	43.8	94.9	26.9	4,180	32.6
Technical	725 *	8.0	530	15.7	73.4	20.8	1,328	10.4
University	562 *	6.2	697	20.7	174.4	49.3	1,433	11.2
Total N	9,083	100.0	3,368	100.0	353.4	100.0	12,805	100.0
Percentage (%) horizontal	70.9		26.3		2.8		100.0	
Years of education								
mean	6.9		10.8		13.7		8.2	
median	5.0		11.0		14.0		9.0	

Notes:

N Population, 25 or more years of age (sample expansion).

* Población relativamente pequeña, 10% o menos de su grupo.

Source: Based on ENAHO 2003.

Table 2. Peru: Mean income by education level and social groups 2003 (1000 people, soles/month, and percentages).

Nivel educativo	Social group						Total		
	Z		Y		A				
	N (%)	IM	N (%)	IM	N (%)	IM	N	%	IM
None	9.6	171 *	1.9	297 *	--	--	604	7.1	180
Primary	41.0	330	13.1	479	1.0	137 *	2,704	32.0	346
Secondary	32.7	592	44.0	710	22.7	798	2,998	35.4	637
Technical	9.6	844 *	17.2	912	15.8	965	1,001	11.9	876
University	7.1	1,529 *	23.8	2,005	60.5	2,721	1,148	13.6	1,943
Total	100.0	535	100.0	1,015	100.0	1,981	8,464	100.0	717
Total N	5,841		2,330		293		8,464		
%	69.0		27.5		3.5		100.0		
Years of education									
mean	7.6		11.4		14.2		9.0		
median	8.0		11.0		15.0		11.0		

Notes:

N Population, 25 or more years of age (sample expansion).

IM Monthly mean income, soles of Lima, October 2003.

-- No population exist in this category.

* Small number, 10% or less.

Source: Based on ENAHO 2003.

Table 3a. Peru: Mean wages (blue-collar workers) by education level and social group, 2003 (1000 people, soles/month, and percentages).

Education level	Social group						Total	
	Z		Y		A			
	N (%)	IM	N (%)	IM	N (%)	IM	N %	IM
None	6.0	306 *	1.8	349 *	--	--	4.7	310
Primary	37.5	489	21.6	518	--	--	32.4	490
Secondary	46.7	616	59.4	651	78.2	569 *	50.8	627
Technical	8.1	695 *	13.8	712	21.8	743 *	10.0	703
University	1.7	763 *	3.4	796 *	--	--	2.1	778
Total	100.0		100.0		100.0	606 *	100.0	579
Total N	1,163		482		23		1,668	
%	69.7		28.9		1.4		100.0	
% of EAP (from table 2)	19.9		20.7		7.8 *		19.7	
Years of education								
mean	8.0		9.7		11.3		8.6	
median	9.0		11.0		11.0		10.0	

Notes:

N Population of blue-collar workers, 25 or more years of age (sample expansion)

IM Monthly mean income, soles of Lima, October 2003.

-- No population exist in this category.

* Small number, 10% or less.

Source: Based on ENAHO 2003.

Table 3b. Peru: Mean salaries (white-collar workers) by education level and social group (1000 people, soles/month, and percentages).

Education level	Social Group						Total	
	Z		Y		A			
	N (%)	IM	N (%)	IM	N (%)	IM	N %	IM
None	0.1	357 *	0.5	431 *	--	--	0.3	415
Primary	4.6	703 *	1.8	641 *	--	--	2.8	685
Secondary	29.2	871	28.1	707	14.2	763	27.2	788
Technical	32.9	1,045	26.2	1,009	17.7	1,102	28.4	1,033
University	33.2	1,780	43.4	2,203	68.1	3,182	41.3	2,212
Total	100.0	1,222	100.0	1,433	100.0	2,469	100.0	1,442
Total N	807		833		181		1,821	
%	44.3		45.7		10.0		100.0	
% of EAP (from table 2)	13.8		35.7		61.8		21.5	
Years of education								
mean	13.5		13.7		14.9		13.7	
median	14.0		14.0		16.0		14.0	

Notes:

N Population of white-collar workers, 25 or more years of age (sample expansion)

IM Monthly mean income, soles of Lima, October 2003.

-- No population exist in this category.

* Small number, 10% or less

Source: Based on ENAHO 2003.

Table 4. Peru: Mean wage, mean salary, and mean income for self-employment by social group, 2003 (soles/month).

Social group	Education level					
	None	Primary	Secondary	Technical	University	Total
Group Z						
Salary	357 *	703 *	871 *	1,045 *	1,780 *	1,222
Wage	306 *	483	618	695 *	763 *	557
Self-employment income	136	260	451	568 *	818 *	327
Group Y						
Salary	431 *	641 *	707	1,009	2,203	1,433
Wage	349 *	518 *	651	712 *	796 *	630
Self-employment income	221 *	384 *	550	696	1,180	621
Group A						
Salary	--	--	763	1,102	3,182	2,469
Wage	--	--	568 *	743 *	--	608
Self-employment income	--	137 *	1,032	508 *	1,135	1,013

Notes:

-- There is no people in this category.

* Small number, 10% or less of the social group.

Source: Based on ENAHO 2003.

Table 5. Peru: Rate of excess labor supply (unemployment and underemployment) by education level and social group, 2003 (percentages).

Education level	Social group			Total
	Z	Y	A	
None	88.1	73.7*	--	87.1
	(11.7)	*(12.3)*		*(11.7)*
Primary	79.7	61.8	100.0*	77.7
	(6.5)	*(7.2)*	*0.0*	*(6.6)*
Secondary	57.6	49.8	39.3	54.4
	(5.4)	*(8.2)*	*(11.4)*	*(6.5)*
Technical	35.0*	31.1	25.7	32.9
	(5.6)	*(7.3)*	*(12.4)*	*(6.6)*
University	29.3*	32.7	28.8	31.0
	(7.0)	*(8.7)*	*(2.7)*	*(7.2)*
Total	65.6	44.5	31.5	58.4
	(6.6)	*(8.1)*	*(6.4)*	*(7.0)*

Notes:

-- There is no people in this category.

* Small number, 10% or less of the social group.
Unemployment rate in parenthesis; underemployment es given by the difference with total.

Source: Based on ENAHO 2003.

Table 6a. Peru: Mean years of schooling by age bracket and social group, 2003.

	Age bracket								Total	
	25-34		35-44		45-54		55-65			
Social group	N	e	N	e	N	e	N	e	N	e
Z	1,551	11	1,636	8	1,260	7	850	5	5,297	8
Y	912	12	670	12	449	11	218	9	2,249	12
A	146	14	111	14	29	14	6	12	291	14
Total	2,609	12	2,417	10	1,738	8	1,074	6	7,837	9
Ratio Z/A	0.8		0.6		0.5		0.4		0.6	
Ratio Y/A	0.9		0.8		0.8		0.7		1.2	

Notes

N. Occupied economically active population (million people)

e: Mean years of schooling

Source: Based on ENAHO 2003.

Table 6b. Peru: Mean income by age bracket and social group, 2003.

	Age bracket								Total	
	25-34		35-44		45-54		55-65			
Social group	N	IM	N	IM	N	IM	N	IM	N	IM
Z	1,551	530	1,636	596	1,260	623	850	473	5,297	563
Y	912	771	670	987	449	1,195	218	1,626	2,249	1003
A	146	1,456	111	2,455	29	2,982	6	1,648	291	1991
Total	2,609	666	2,417	790	1,738	810	1,074	713	7,837	743
Ratio Z/A	0.4		0.2		0.2		0.3		0.3	
Ratio Y/A	0.5		0.4		0.4		1.0		0.5	

Notes

N. Occupied economically active population (million people).

IM: Mean income (soles/month)

Source: Based on ENAHO 2003.

Table 7. Cumulative income distribution by ventile, 2003.

	Total		Z	Y	A
Ventiles	**Fi**	**Qi**	**Fi**	**Fi**	**Fi**
1	0.05	0.01	0.07	0.01	0.01
2	0.10	0.01	0.13	0.03	0.01
3	0.15	0.01	0.20	0.05	0.02
4	0.20	0.01	0.26	0.07	0.03
5	0.25	0.02	0.32	0.10	0.04
6	0.30	0.03	0.38	0.14	0.06
7	0.35	0.05	0.43	0.18	0.08
8	0.40	0.07	0.49	0.22	0.09
9	0.45	0.09	0.54	0.27	0.12
10	0.50	0.12	0.59	0.32	0.13
11	0.55	0.15	0.64	0.38	0.15
12	0.60	0.18	0.68	0.45	0.18
13	0.65	0.22	0.73	0.52	0.27
14	0.70	0.27	0.77	0.58	0.32
15	0.75	0.32	0.81	0.64	0.40
16	0.80	0.39	0.85	0.71	0.48
17	0.85	0.46	0.90	0.78	0.54
18	0.90	0.55	0.94	0.84	0.61
19	0.95	0.67	0.98	0.92	0.74
20	1.00	1.00	1.00	1.00	1.00
Gini (ventiles)	0.59				

Notes:

Fi: Accumulated relative frequency of the occupied economically active population (25 or more years of age).

Qi: Accumulated income share

Source: Based on ENAHO 2003

APPENDIX: SUMMARY OF STATISTICAL TESTING

Hypothesis 1: Differences in mean years of schooling

Social groups	Variance equality test		Difference of means test			
	T Statistics	p-value	Difference	Confidence interval	F Statistics	p-value
Z - Y	354.56	0.00	-3.96	[-4.10 -3.81]	2024.40	0.00
Z - A	150.21	0.00	-6.87	[-7.20 -6.55]	1620.48	0.00
Y - A	53.21	0.00	-2.92	[-3.26 -2.58]	258.30	0.00

Social groups	U Test of Mann Withney			
	Sum of ranks		Z statistics	p-value
	Z	Y		
Z - Y	490,300,000	218,000,000	-61.49	0.00
Z - A	420,400,000	10,058,291	-24.62	0.00
Y - A	38,984,586	2,479,585	-12.57	0.00

The statistical testing refers to mean years of schooling differences between social groups. The first test uses F statistics and the second the (non parametric) U Test of Mann Withney. The results show that the observed mean differences are statistically significant.

Hypothesis 2: Mean income differences

Level of education	Social groups	Variance equality test		Difference of means test			
		T Statistics	p-value	Difference	Confidence interval	F Statistics	p-value
Primary	Z - Y	4.24	0.04	-126.41	[-164.86 -87.97]	33.04	0.00
Secondary	Z - Y	0.63	0.43	-102.02	[-162.35 -41.68]	6.35	0.01
	Z - A	0.15	0.70	-217.56	[-513.60 78.47]	2.11	0.15
	Y - A	0.07	0.80	-115.54	[-416.41 185.32]	0.35	0.55
Technical	Y - A	0.96	0.33	-41.85	[-307.86 224.17]	0.15	0.69
University	Y - A	21.48	0.00	-938.12	[-1677.40 -198.83]	3.30	0.07

Level of education	Social groups	U Test of Mann Withney			
		Sum of ranks		Z Statistics	p-value
		Z	Y		
Primary	Z - Y	34,224,589	4,808,441	-9.83	0.00
Secundary	Z - Y	23,098,507	10,714,470	-7.61	0.00
	Z - A	17,017,958	289,828	-2.76	0.01
	Y - A	3,033,103	113,184	-1.20	0.23
Technical	Y - A	553,451	31,370	-0.72	0.47
University	Y - A	785,177	103,934	-0.66	0.00

The test now refers to mean income differences between social groups. The parametric and non parametric tests show that the observed differences are statistically significant at both the primary and university levels of education. For the other levels of education, the parametric test does not pass the test of variance equality; hence, parametric tests are the relevant one. This test shows that differences are not statistically significant in the case of the technical level; whereas in the secondary level, the differences are significant in all groups, except in the Y-A groups. In four out of six cases, the hypothesis is not refuted by facts.

Hypothesis 3: Differences in mean salaries and mean wages

White collar workers: Salaries

Level of education	Social groups	Variance equality test		Difference of means test				
		T Statistics	p-value	Difference	Confidence interval		F Statistics	p-value
Secondary	Z - Y	8.73	0.00	164.01	[93.99	234.03]	21.08	0.00
	Z - A	1.71	0.19	107.76	[-143.00	358.52]	0.71	0.40
	Y - A	0.39	0.53	-56.25	[-307.71	195.21]	0.19	0.66
Technical	Z - Y	4.75	0.03	35.85	[-67.13	138.83]	0.47	0.50
	Z - A	0.70	0.40	-56.94	[-378.31	264.43]	0.12	0.73
	Y - A	0.00	0.97	-92.79	[-421. 73	236.15]	0.31	0.58
Universtity	Z - Y	21.64	0.00	-422.79	[-1011.39	165.82]	1.98	0.16
	Z - A	60.13	0.00	-1402.16	[-2212.04	-592.28]	11.52	0.00
	Y - A	16.44	0.00	-979.37	1-1822.26	-136.48]	5.19	0.02

Level of education	Social groups	U Test of Mann Withney			
		Sum of ranks		Z Statistics	p-value
		Z	Y		
Secondary	Z-Y	476,879	271,598	4.27	0.00
	Z-A	284,144	8,852	1.63	0.10
	Y-A	125,362	6,994	0.28	0.78
Technical	Z-Y	667,938	391,302	3.39	0.00
	Z-A	405,470	13,601	1.30	0.19
	Y-A	175,132	10,004	0.35	0.73
University	z-y	646,406	660,130	-1.85	0.06
	Z-A	363,848	56,139	-5.15	0.00
	Y-A	344,454	52,041	-4.10	0.00

Given the relative size of social groups (shown in the first hypothesis on exclusion), the relevant comparison is between groups Z-Y at secondary level of education, Z-Y at technical level and Z-Y-Z at the university level. The last comparison can be made by parametric and non parametric testing. Due to the failure in the variance equality test, the other two can be done by using non parametric test alone. The results show that the observed differences are statistically significant.

Blue collar workers: Wages

Level of education	Social groups	Variance equality test		Difference of means test				
		T Statistics	p-value	Difference	Confidence interval		F Statistics	p-value
Primary	Z - Y	0.38	0.54	-35.08	[-115.23	45.07]	0.74	0.39
Secondary	Z - Y	4.01	0.05	-34.34	[-91.07	22.40]	1.41	0.24

Level of education	Social groups	U Test of Mann Withney			
		Sum of ranks		Z Statistics	p-value
		Z	Y		
Primary	Z-Y	1,139,292	212,899	-0.93	0.35
Secondary	Z-Y	1,653,321	774,385	-2.71	0.01

The relevant comparison is between groups Z-Y at primary and secondary level of education. The parametric test cannot be used because the variance equality test fails. The non parametric test shows that the observed differences are statistically significant for secondary but not for primary.

Hypothesis 4: Mean differences between wage incomes (salary incomes) and incomes from self employment

Social group Z

Level of education	Variance equality test		Difference of means test				
	T Statistics	p-value	Difference	Confidence interval		F Statistics	p-value
Primary	72.56	0.00	443.13	[287.90	598.36]	31.31	0.00
Secondary	24.64	0.00	420.17	[365.23	475.12]	224.64	0.00
Total	*495.77*	*0.00*	*894.33*	*[757.69*	*1030. 97]*	*164.58*	*0.00*

Level of education	Social groups	U Test of Mann Withney			
		Sum of ranks		Z Statistics	p-value
		Self employed	Employed		
Primary	Z-Y	16,417,007	603,688	-11.93	0.00
Secondary	Z-Y	4,943,900	2,010,686	-24.30	0.00

Social group Y

Level of education	Variance equality test		Difference of means test				
	T Statistics	p-value	Difference	Confidence interval		F Statistics	p-value
Secondary	0.42	0.52	157.72	[80.83	234.61]	16.17	0.00
Technical	0.21	0.65	313.21	[70.70	555.73]	6.41	0.01
University	20.77	0.00	1022.85	[502.82	1542.88]	14.86	0.00
Total	*140.96*	*0.00*	*812.48*	*[586.29*	*1038.68]*	*49.57*	*0.00*

Level of education	U Test of Mann Withney			
	Sum of ranks		Z Statistics	p-value
	Self employed	Employed		
Secondary	772,113	492,733	-12.58	0.00
Technical	64,839	285,864	-13.97	0.00
University	84,954	479,825	-12.06	0.00

In the case of Z-workers, the relevant comparisons include primary and secondary levels of education. Both the parametric and non parametric tests show that the observed differences are statistically significant. In the case of Y-workers, the relevant comparisons include secondary, technical, and university levels of education. The non parametric test shows that the observed differences are statistically significant in all cases. The parametric test is applicable to the university case only, and there it shows that the difference is significant as well.

Hypothesis 5: Mean differences in the exclusion rates from labor market

Level of education	Social groups	Chi square	p - value
Primary	Z – Y	6,855	0.00
Secondary	Z - Y - A	4,855	0.00
Technical	Z – Y	1,034	0.00
University	Z – Y	819	0.00

In order to test the differences in the proportions of exclusion from the labor market between social groups, the chi square test is utilized. The test shows that the observed differences are statistically significant for all levels of education and the relevant social groups in each level.