A generalized labor market theory: inequality as labor discipline device

Adolfo Figueroa*

The standard microeconomic theory of labor market assumes that the unemployment rate operates as the labor discipline device (Shapiro and Stiglitz 1984; Bowles 1985). This theory refers basically to labor markets in advanced countries; thus the standard theory assumes implicitly that the society is underpopulated. By contrast, the development economics literature usually assumes that labor markets in developing countries are overpopulated (Lewis 1954). But then what do firms utilize as the labor discipline device in developing countries? Is it still the unemployment rate?

Arthur Lewis in his classical paper proposed an answer: the labor discipline device is given by a wage premium that firms pay above the subsistence sector income. But do labor markets in developing countries operate as the Lewis model says? Is a new model needed? An answer to these theoretical and empirical questions does not seem to exist in the development literature. This paper will seek to give an answer by constructing a new theoretical model and by confronting its predictions against a set of empirical regularities that characterize the functioning of labor markets in developing countries.

Another objective of this paper is to find common grounds in the labor market theory of the advanced countries and that of the developing countries. Inequality among workers will be examined as the possible common labor discipline device of a generalized theory of labor markets.

The paper is organized as follows. Section 1 presents the empirical regularities that characterize the functioning of labor markets in developing countries. Section 2 develops a new model that seeks to explain the behavior of labor markets in developing countries. The empirical predictions of the model include the observable equilibrium conditions and the causality relations between the endogenous and the exogenous variables of the model. The empirical predictions of the model in terms of causality relations are presented in section 3. The confrontation of the predictions of the model in terms of the observable equilibrium conditions against the set of empirical regularities is shown in section 4. A comparison of the model with conventional theoretical models is presented in section 5. Section 6 shows the generalized theory of labor markets in which inequality among workers is the common labor discipline device utilized in both advanced and developing countries, although inequality takes particular forms in each case. Section 7 concludes.

Empirical regularities of labor markets in developing countries

The simplest way to refute an economic theory is to confront its predictions against the most common empirical regularities that we observe in the real world. For the case of the behavior of labor market in developing countries, the following regularities can be stated:

Fact 1. Self employment constitutes a significant proportion of the labor force

At any given period of time, three categories of workers can be distinguished in developing countries: wage-earners, the unemployed, and the self employed. The first category refers to those that exchange their labor services for the

payment of wages in labor markets; the second refers to those that are seeking wage employment; and the third to those that are self employed in small businesses and small farms where they generate their own incomes through exchange in product markets.

Unemployment figures in developing countries show that rates are not too different from those observed in advanced countries. For instance, the average rate of unemployment in the Organization for Economic Co-operation and Development (OECD) countries was around 7% in the period 1992-2002 (United Nations Development Programme, UNDP, 2004, Table 20: 2006), whereas *urban* unemployment in Latin American countries was around 10% in the period 1998-2002 (Comisión Económica para América Latina y el Caribe, CEPAL, 2008: 11). If the latter rate is adjusted to measure total unemployment rate (as proportion of urban and rural labor force), the two rates mentioned will tend to become even closer. Differences in self employment rates are more significant: 40% in developing countries and 12% in advanced countries (International Labour Organization, ILO, 2002, Annex 2: 62-64).

Fact 2. Labor markets operate with excess labor supply, which takes two forms: unemploymentand underemployment

Another regularity in developing countries is that the average income of the self employed is smaller than the average market wage rate for a given skill level. Empirical studies measuring directly this gap are not frequent in the international literature. Two country studies can be cited at this moment. A study on Peru, based on the National Household Survey of 2003, found that the average wage rate was higher than the average income of the self employed: it was 80% higher among workers with primary level education (low skill), and 30% higher among workers with secondary level education (high skill) (Figueroa 2011, Table 5).

The other study refers to Brazil. The study is based on a sample from the national census of 1980, in which the sample size was 3% and the universe the large cities of Brazil and workers with 11 years of education or less

(thus excluding workers with post secondary education levels). This study found that the income gap between wage earners and the self-employed was around 30% (Telles 1993, Table 1: 239).

The fact that the self employed workers generate incomes that are smaller than the average wage rate, for a given level of skills, implies that self employment is a form of excess labor supply. Workers would prefer to have wage employment at the current wage rate rather than staying self employed. This situation can be defined as *underemployment*. Recognizing that some self employed workers may be generating higher incomes than the market wage rate, the statistical estimates of self employment rates that were shown above may then be taken as rough measures of underemployment.

Unemployment and underemployment therefore constitute forms of excess labor supply in developing countries. The empirical estimates show that rates of self employment are much higher than rates of unemployment. As the data that were shown above indicated, underemployment is around 40% of the labor force whereas unemployment is around 10%; hence the remaining 50% constitutes wage employment.

The common practice of using the unemployment rate as the criterion for making international comparisons about the excess labor supply is thus unwarranted. In developing countries, the rate of excess labor supply takes two forms, unemployment and underemployment, the latter being the more significant; in advanced countries, by contrast, the excess labor supply is mostly given by unemployment.

Fact 3. For a given human capital, market wage rates are higher in larger size firms than in smaller ones

There is a vast literature for advanced countries that has shown the positive effect of firm size on wage rates (*cf.* Oi 1990; Troske 1999; Abowd and Kramarz 1999; Reilly 1995; Weiss and Landau 1984; Meagher and Wilson 2004). For developing countries, empirical works are less abundant. For the case of Lima, Peru, in the period 1996-2006, and using firms with 10-99 and 100 or more workers as the threshold to separate small from large

size firms, large firms paid on average 100% above the wages paid in small firms for the case of executives, 20% for white-collar workers, and 30% for blue-collar workers (Figueroa 2009, Table 4).

Any economic theory that intends to explain the functioning of labor markets in developing countries should not be refuted by these facts. As will be shown below, conventional theoretical models are refuted by these facts. Therefore, a new model is needed, which will be presented now.

A MODEL OF LABOR MARKETS FOR OVERPOPULATED SOCIETIES

The labor market model that is presented now will assume that society is overpopulated. The model will have microeconomic foundations. The aggregation will take the form of partial equilibrium analysis. The inter-relations between the labor market and the rest of the economy will take into account only the effect of the latter upon the former, but not the vice versa effect. The analysis will firstly refer to a particular labor market, not to the aggregate labor market. This aggregation will be taken up later on.

The standard microeconomic theory of the labor market assumes that the unemployment rate operates as the labor discipline device (Shapiro and Stiglitz 1984; Bowles 1985). This is the fundamental incentive system in the functioning of labor markets. Workers who are found shirking at the work place and are then consequently dismissed will pay the cost of becoming unemployed. This theory predicts that equilibrium in the labor market is with a positive rate of unemployment. Unemployment plays a fundamental role in the functioning of labor markets and thus in the functioning of the capitalist system.

In an overpopulated society, the unemployment rate can hardly play that role. Overpopulation refers to a particular factor endowment of capital and labor in the society, such that the Walrasian real wage (the one that clears the market) would be zero or even negative in most labor markets. In this case, by definition, any labor market equilibrium with positive real wage rate would imply a very large excess labor supply; then capitalism would be viable if a subsistence sector existed. The unemployment rate alone could not be used by firms as the labor discipline device. Firms will use another device.

The new model will assume the existence of two economic sectors: the capitalist sector and the subsistence sector. Firms and workers exchange labor services for nominal wage rates in the labor markets. The social norm is such that nominal wages cannot fall. The self employed workers generate incomes in the subsistence sector, in small businesses, through exchange in product markets. The local economy or regional economy in which the particular labor market under analysis operates will produce one single good, good B, the price of which is exogenously determined. The market structure is perfect competition.

The labor demand function

Capitalist firms will produce good B using labor inputs, the services of which are bought in the local labor market at the nominal price P_h (say, for unskilled or low human capital) and $P_{h'}$ (for skilled or high human capital). These two types of labor are not substitutable for each other in production. Other inputs going to production will include material inputs, called C, at nominal price P_c, and entering in fixed quantities per unit of output. Firms are endowed with quantities of physical capital, which total K units, and sell good B at the market nominal price P_b.

The partial equilibrium analysis will refer to one of the two labor markets: take the unskilled labor market. The equilibrium condition for the individual capitalist firm (seeking to maximize profits) is that the market nominal wage rate P_h should be equal to the net value of physical marginal productivity of unskilled labor. The individual labor demand of the firm is then determined by the exogenous variables facing each firm. The market labor demand function will be obtained just by the simple aggregation of the behavior of individual firms. This function can be written as

$$D_h = H(P_h, P_b, P_c, P_{h'}, K) \qquad [1]$$
$$H_1 < 0, H_2 > 0, H_3 < 0, H_4 < 0, H_5 > 0$$

The partial derivatives indicate the effect of changes in the exogenous variables on the quantity of labor demanded by capitalist firms.

Function [1] is homogeneous of degree zero in all nominal prices; that is, if all nominal prices increased in a given proportion, which would imply maintaining relative prices constant, the quantity demanded for labor would remain unchanged. Then the function H can be reduced to three relative prices, such as P_h/P_b, P_c/P_b, and $P_{h'}/P_b$ and K, all real variables. But this transformation is appropriate if and only if the nominal prices are all flexible. If for some reason P_h could not be flexible, as the model assumes, the function can be written in nominal prices only, as it is shown in equation [1].

Real wage rate may have two definitions. As the purchasing power of workers, it is equal to P_h/P, where P stands for the price level of the economy; as the real labor cost to firms in the given region, it is equal to P_h/P_b. The model assumes that both P and P_b are exogenous to the regional economy under analysis.

The market equilibrium conditions

The structural equations and the equilibrium conditions of the model are as follows:

Labor market

$$D_h = H(P_h, P_b, P_c, P_{h'}, K) \tag{1}$$

$$H_1 < 0, H_2 > 0, H_3 < 0, H_4 < 0, H_5 > 0$$

Subsistence sector

$$Q_s = G(L_s), G' > 0, G'' < 0 \tag{2}$$

$$\pi P_h = P_b G'(L_s), 0 < \pi < 1 \tag{3}$$

Subject to

$$S_h = D_h + L_s + U \tag{4}$$

$$P_h \geq (1 + m) P_b G'(L_s) \tag{5}$$

The first equation is the labor demand function, which was already presented above. The subsistence sector is represented in the following two equations. Equation [2] says that total output (Q_s) depends upon the level of self employment (L_s) in the subsistence sector, in which the productivity of labor is subject to diminishing returns. The model assumes that there is no physical capital in the subsistence sector; hence, labor productivity is much lower in this sector compared to that of the capitalist sector. Labor productivity in the subsistence sector is too low to hire wage-earners.

Equation [3] shows the equilibrium condition in the subsistence sector: the expected nominal wage P_h^e must be equal to the *marginal income* in the subsistence sector. On the expected wage, the assumption is that workers seeking jobs in the labor market expect to receive as wage income a fraction of the current wage rate in a given period, given that finding the job takes time; hence $P_h^e = \pi \cdot P_h$. The value of π is exogenously given in the short run; then the expected nominal wage depends positively on the nominal wage rate. On the formation of expected wages, therefore, the assumption is that workers perceive more clearly changes in prices than changes in quantities in an overpopulated labor market. If the expected nominal wage were higher than the marginal income in the self employment sector, some workers would prefer to seek jobs in firms, and thus choose unemployment; if it were smaller, some workers would prefer to take the sure self employment alternative; so in equilibrium, the equality must hold.

The last two equations refer to the constraints of the system. Equation [4] is the labor supply constraint: the allocation of the total labor force (S_h) must be equal to the quantities of wage employment (D_h), self employment (L_s), and unemployment (U).

Finally, equation [5] incorporates the assumption about the particular labor discipline device that firms will use in an overpopulated economy: the market nominal wage rate must be higher than the opportunity cost of wage employment, which in this case is equal to the marginal income in the subsistence sector. The premium is given by the coefficient m $(m > 0)$, which is exogenously determined. This gap maintains both labor productivity and profits at high levels. Workers found shirking and fired from the firms

will then suffer a cost, which is given by this gap. Workers would have no incentives to respect the labor discipline if this gap were zero, if there were no costs involved. If for some reason the gap were smaller than the required, firms would seek to raise the nominal wage rate endogenously and thus seek to restore the required gap. In this model, the nominal wage rate can increase endogenously; it is sticky downwards, but not upwards.

Figure 1 depicts the equilibrium situation of the model. Total labor supply of unskilled workers is equal to the segment OO' and the two vertical lines measure monetary value of output per worker in the capitalist sector and subsistence sector. Labor supply OO' is perfectly inelastic. The nominal wage rate is given at the level P_{h0}. The labor demand curve D cuts the given nominal wage rate at point E; hence, the wage employment level is equal to the segment OA; then AO' measures the excess labor supply. The curve r'n' (measured from origin O') represents the marginal income in the subsistence sector and r'n the corresponding average income curve. The equilibrium value of self employment is given by point E', at which the expected wage rate cuts the curve r'n' (such that $\pi P_{h0} = I'$); hence, the segment O'B measures the size of self employment, with marginal income equal to BE' and average income BF. The segment AB in the horizontal axis represents the size of unemployment, as a residual quantity.

According to the equilibrium conditions stated above in equation [5], the market nominal wage must be at least equal to the opportunity cost of the wage earners plus a premium (say 30%) to generate a gap that operates as an incentive for labor discipline at the firm level. In figure 1, the curve N'R' is drawn in such a way that it lies above the curve n'r' by $(1 + m)$ times, and thus satisfies the constraint given by equation [5]. The curve N'R' then represents the effort extraction curve: the set of nominal wages that will maintain the physical labor productivity at the level given by the curve D; this set is called the set of efficiency wages. The curve N'R' consequently shows the lower bound of the set of efficiency wages. The equilibrium point E indeed lies above the curve N'R' and corresponds to an efficiency wage.

Figure **1**
Labor market equilibrium with excess labor supply in an overpopulated society
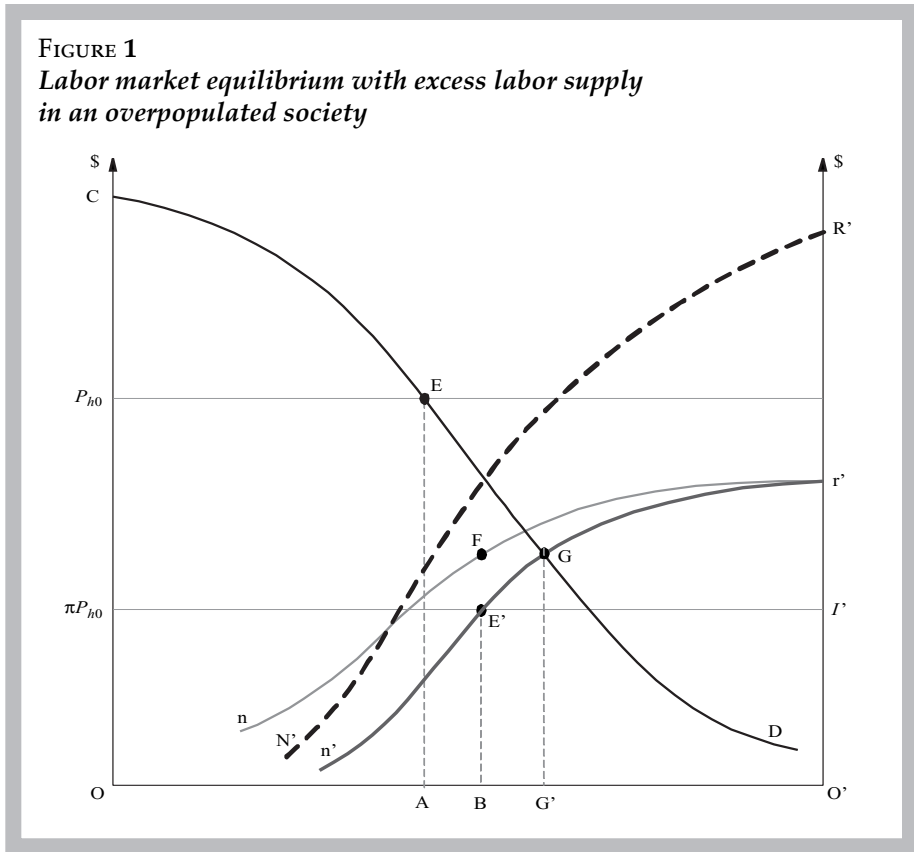
The equilibrium quantities of wage employment, self employment, and unemployment are thus shown in figure 1. No social actor has the power or the will to change this situation. Self employed workers could not compete with wage-earners accepting to work for lower wages because firms would not accept it, as the resulting labor discipline, productivity, and profits would be lower. This is a static equilibrium situation. Those equilibrium values will be repeated period after period as long as the exogenous variables of the model remain constant. Of course, behind the equilibrium quantities in the labor market, there will be labor rotations and changes in the employment situations of *individual* workers.

Introducing the effect of firm size into the model, another empirical prediction can be derived. The model now assumes that large firms use the gap between the nominal wage they pay and the nominal wage paid by small firms as labor discipline device. The implication is that nominal wages will be higher in large firms than in small firms, which in turn will be higher than the average income in self employment. The labor market operates with a hierarchy of incomes among workers of the same skill levels.

Figure 2 depicts the equilibrium situation with two groups of firms of different sizes. The labor demand curve of large firms is located at a higher level than the corresponding demand curve of small firms due to the effect of differences in the endowments of capital stock per worker. In order to maintain those levels of labor productivity, large firms pay an additional premium upon the opportunity cost of their employees. The effort extraction curve for small firms is given by the curve $N_1'R_1'$ and that of the large firms by the curve $N_2'R_2'$. Given the two nominal wages, the quantities employed in each firm group is determined (A_1 and A_2), total wage employment is thus equal to $A_1 + A_2$, and excess labor supply is also known. The expected wage is the average of both nominal wages, which determines the quantity of self employment, equal to O'B. Unemployment is just the residual. Because no social actor has the power or the will to change this situation, the equilibrium has been reached.

CAUSALITY RELATIONS

The labor market equilibrium is clearly stable, as can be seen in figure 1. Comparative statics can then be applied to derive the empirical predictions of the model. The variables of the model are

Endogenous variables: D_h, L_s, U, P_h
Exogenous variables: P_b, P_c, $P_{h'}$, K, S_h, τ

The exogenous variable τ refers to the set of state regulations in the labor market (which will not be analyzed here). In order to simplify further, the effect of the structure of firms (mix of large and small) will be ignored.

FIGURE 2
*Labor market equilibrium with large and small firms
and excess labor supply*
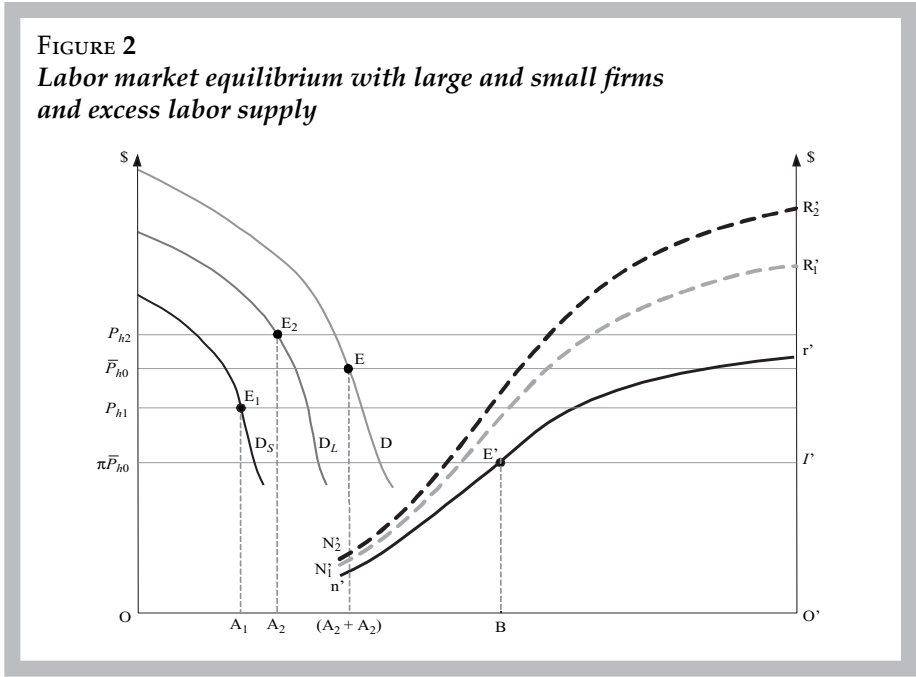
　　The effect of changes in the exogenous variables upon the equilibrium values of the endogenous variables can easily be derived from figure 1. An increase in P_b will have the effect of shifting outward the labor demand curve, the value of the labor marginal productivity curve in the subsistence sector, and thus the effort extraction curve. The labor demand expansion will increase the quantity of wage employment, whereas the nominal wage rate may remain constant or may increase. If the nominal wage is above the efficiency wage, there will be no need for the nominal wage to increase. Self employment will increase. Unemployment, the residual, will fall. Labor demand expansion will also occur as a result of a fall in P_c or $P_{h'}$.

　　In the long run, an increase of K will shift the labor demand curve outwards, but will maintain the effort extraction curve unchanged. Assuming that the nominal wage needs no endogenous changing, the quantity of equilibrium of wage employment will increase, self employment will remain unchanged and unemployment will fall. As the process of capital

accumulation continues, unemployment will disappear and self employment will fall, which implies a rise in the marginal income in the subsistence sector and the need to increase endogenously the nominal wage in order to maintain the required gap.

The effect of capital accumulation upon price and quantity in the labor market is shown in figure 3. Consider the initial equilibrium in quantities: wage employment is given by the segment OA_0, excess labor supply by A_0O'. As the demand curve expands to D_1 and D_2, due to physical capital accumulation, wage employment will rise to A_1 and A_2. Demand curve D_2 will imply a solution at point M, where the demand curve and the effort extraction cross each other; that is, the income gap between the nominal wage rate and the marginal income in the subsistence sector reaches its minimum value. Hall (2005) has recently shown a similar model of labor markets operating with sticky wages under employment fluctuations.

FIGURE 3
Effects of capital accumulation on the labor market in an overpopulated society
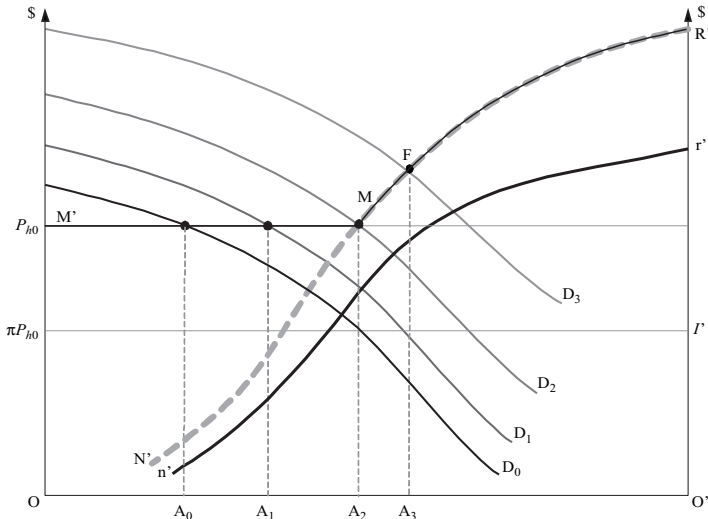
Demand curve D_3 would imply a larger quantity of wage employment and a smaller quantity of self employment and thus a higher marginal income if the nominal wage remained constant; however, the nominal wage rate could not remain constant because it would become smaller than the efficiency wage and consequently the labor productivity curve would fall and the expected profits could not be realized. Then firms will have incentives to raise the nominal wage to restore the gap required to maintain the productivity level. As a result the market nominal wage rate will increase and thus labor demand curve D_3 will imply a solution at point F, which implies wage employment level equal to OA_3. As labor demand increases further, the price and quantities will become endogenous and will move along the effort extraction curve.

In sum, as the labor demand curves expand as a result of physical capital accumulation, the equilibrium values of price and quantity in the labor market will move along the thick line M'MFR'. When equilibrium lies within the segment M'M, the effect of an upward shift in the demand for labor will cause an increase in wage employment, but the nominal wage will remain fixed; when equilibrium lies in the segment MR (along the effort extraction curve), both quantity and price will increase because the nominal wage rate becomes endogenous. Therefore, the expansion of labor demand shows that the line M'MFR' plays the role that the supply curve of labor does in Walrasian models (the curve n'r'), and then it can be called *effective supply curve*. This curve consists of two segments: the flat segment M'M, which is equivalent to the "unlimited labor supply curve" of Lewis model (Lewis 1954) and MR' that is the upward sloping segment. Hence, price and quantity at equilibrium in the labor market are determined by the interaction between the labor demand curve and the effective supply curve.

Figure 3 also shows the effect of capital accumulation upon the other endogenous variables of the model: self employment and unemployment. As long as there is unemployment, any increase in the labor demand curve will increase wage employment and reduce unemployment. Expected wage is fixed as long as the nominal wage rate is fixed; hence, the quantity of self employment will also be fixed. As demand expands, unemployment

will ultimately disappear. Beyond this point, any additional demand expansion will cause a rise in wage employment, which will come not from unemployment but from self employment. Reducing self employment implies a rising marginal income in this sector, along the line n'r'. But then the gap with the nominal wage rate will tend to close; therefore, in order to restore the efficiency wage level, the nominal wage rate must go up. So the rising segment of the effective supply curve (MR') indicates rising nominal wage rates, which is consistent with the rising of marginal income in the self employment sector.

An exogenous increase in the total supply of labor (S_h) can be represented by a segment that is longer than OO' in figure 3. Therefore, given the demand curve and given the nominal wage rate, the new equilibrium will imply an unchanged value in wage employment and self employment; that is, all the additional labor force will become unemployed.

The reduced form equations for price and quantity in the labor market are the following:

$$D_h^0 = \Phi\,(P_b, P_c, P_{h'}, K, S_h) \tag{6}$$

$$\Phi_1 > 0,\ \Phi_2 < 0,\ \Phi_3 < 0,\ \Phi_4 > 0,\ \Phi_5 = 0$$

$$P_h^0 = \Theta(P_b, P_c, P_{h'}, K, S_h) \tag{7}$$

$$\Theta_1 \geq 0,\ \Theta_2 = 0,\ \Theta_3 = 0,\ \Theta_4 \geq 0,\ \Theta_5 = 0$$

$$L_s^0 = \Psi(P_b, P_c, P_{h'}, K, S_h) \tag{8}$$

$$\Psi_1 > 0,\ \Psi_2 < 0,\ \Psi_3 < 0,\ \Psi_4 \leq 0,\ \Psi_5 = 0$$

$$U^0 = \Omega(P_b, P_c, P_{h'}, K, S_h) \tag{9}$$

$$\Omega_1 < 0,\ \Omega_2 > 0,\ \Omega_3 > 0,\ \Omega_4 \leq 0,\ \Omega_5 > 0$$

The sign of the partial derivative indicates the causality relation, that is, the relation between endogenous and exogenous variables of the model. These signs also indicate the qualitative empirical prediction of the model.

Consider the cycles of expansion and contraction of the economy, which are defined as expansion and contraction in the labor demand curve in the short run. For equilibrium situations in the flat segment of the effective labor supply curve, wage employment is pro cyclical, unemployment is counter cyclical, and self employment is neutral; by contrast, for equilibrium situations in the rising segment of the effective supply curve, wage employment is pro cyclical, self employment is counter cyclical, and the change in unemployment is negligible. This is another empirical prediction of the model.

In the long run, the model predicts that capital accumulation at a rate higher than that of population will eliminate overpopulation. In figure 3, there will exist a labor demand curve D' such that it will cross the total effective labor supply at point R'. Once equilibrium reaches this situation, there will be a regime switch in the functioning of the labor market. Firms will now use the unemployment rate as the labor discipline device.

INITIAL EMPIRICAL REFUTATION OF THE MODEL

There are two ways to refute a theoretical model. First, the observable equilibrium conditions constitute falsifiable propositions; second, the causality relations also constitute empirical predictions about the relations between the exogenous and endogenous variables of the model. The statistical test of the latter type of predictions will not be presented in this study and will have to wait for econometric work.

The following empirical predictions are derived from the model and belong to the first type. For any particular values that prices and quantities may take in the labor market, the equilibrium situation derived from the model will always imply:

1. In a given labor market (for a given level of human capital), average wage rate will be higher than mean income of the self employed.
2. Labor markets operate with excess supply, which includes unemployment and underemployment.
3. In a given labor market (for a given level of human capital), large size firms pay higher wage rates than small firms.

These predictions are consistent with the entire set of empirical regularities listed above, as Facts 1 to 3. Therefore, this model of labor markets is able to explain the functioning of the labor markets in developing countries; that is, none of the regularities or Facts refutes the empirical predictions of the model. Hence, there is no reason to reject the model and we may accept it at this first stage of our research.

The implications of the model for short run macroeconomic theory seem in order. The labor market model shown in figure 1 refers to one particular market; it presents a partial equilibrium analysis. However, it can also be interpreted as one single aggregate labor market and then integrated into the framework of a macroeconomic analysis. Under this framework, equilibrium in the labor market will be determined by interactions with other markets. It should be noted that a macroeconomic model with this type of labor market has the property that the real wage rate and the quantity of wage employment in the labor market will not be determined by real variables alone (such as capital stock or labor supply), but by nominal variables as well (such as the price level).

The behavior of the labor market that is represented in figure 1 is thus logically consistent with a macroeconomic model in which real variables and nominal variables interact. The empirical fact that changes in exogenous nominal variables (fiscal and monetary policies) affect real wages and employment levels in the labor market is empirically consistent with this model. This is another regularity that may very well be added to those listed above when dealing with macroeconomic theory in overpopulated societies. The labor market model that is represented in figure 1 is then the appropriate labor market model to be integrated into a short run macroeconomic theory that intends to explain the economic process in overpopulated societies.

Comparisons with conventional theoretical models

How do other theoretical models perform in explaining the empirical regularities?

The Lewis classical model can be put into perspective now. The model assumes that the marginal productivity of labor in the subsistence sector

is constant. Let the horizontal line starting at point I' in figure 1 represent the constant marginal labor productivity. Adding the premium rate will shift this line to the upper horizontal line that starts at point P_{h0}. The (unlimited) labor supply curve is given by this horizontal line, which crosses the demand curve D at point E and determines the wage employment level OA. The Lewis model then predicts that the labor market operates with excess supply (the segment AO'), which is constituted by self employment alone. Equilibrium exhibits zero unemployment. Fact 2 (the existence of unemployment) refutes the Lewis model.

It should also be noted that the assumption of constant marginal productivity of labor in the subsistence sector (no diminishing returns) implies that this sector has an unlimited capacity to generate incomes. The absence of unemployment is then a clear implication of this assumption. Another implication is that the gap between the average wage rate and the average income of the self employed is constant, exogenously determined. This constancy is another empirical prediction of the Lewis model, which must surely be inconsistent with facts.

A comparison can also be made with the model that was developed by Harris and Todaro (1970), another classical model in development economics. Although this model analyzed the internal migrations problems, it can also be analyzed in the light of the model that was presented in figure 1. The equilibrium condition of the Harris-Todaro model would also establish that the expected wage rate must be equal to the marginal productivity of labor in the subsistence sector. However, the expected wage rate would now depend on the market wage rate and the probability to find a job in the labor market (e); that is, in terms of our model, π would depend upon that probability. Hence, the expected wage depends on the excess labor supply rate, which in turn depends on the expected wage. Therefore, wage rate, wage-employment, unemployment, and self employment are all determined simultaneously. Let the equilibrium hold at point E' in figure 1, where the equilibrium condition would now be $(1-e)P_{h0}=I'$.

The major difference with our model is, however, that the Harris-Todaro model makes abstraction of the labor discipline device. The excess labor

supply plays no role in the functioning of the labor market and market equilibrium may even be Walrasian.

The differences in the equilibrium conditions between the model that has been developed in this study and neoclassical models can also be established with the help of figure 1. In the competitive neoclassical model, the equilibrium in the labor market will take place at point G, where the demand curve D cuts the supply curve n'r'. The equilibrium wage employment is measured by the segment OG' and the segment G'O' corresponds to self employment. Equilibrium exhibits zero unemployment, which also implies no excess labor supply. The labor market is Walrasian. Fact 2 refutes the neoclassical model.

Note that under a Walrasian labor market there would be neither unemployment nor underemployment. Moreover, the average income of the self employed would be higher than the market wage rate, as can be seen in figure 1: the value of the average productivity for the O'G' self employed workers (on the curve nr') lies above point G, which is equal to the market wage rate. This prediction of the neoclassical model is refuted by Fact 1. Neoclassical models assume Walrasian markets; thus they cannot explain the functioning of labor markets in overpopulated societies.

Neoclassical microeconomic models seek to explain unemployment in the functioning of labor markets in every type of society by introducing nonmarket factors (government interventions) (*cf.* Ehrenberg and Smith 2009). Neoclassical macroeconomic models, by contrast, seek to explain unemployment by assuming that labor markets operate with friction due to the differences among workers and jobs (*cf.* Barro 2000). The frictional and structural components of total unemployment are distinguished and analyzed in other macroeconomic models (*cf.* Krugman and Wells 2006).

Recent attempts have been made to integrate the labor market theory that assumes unemployment as labor discipline device into macroeconomic models for underpopulated societies (*cf.* Blanchard 2009). The aggregate labor market model utilized in those macro-models can also be represented in figure 1 just by reinterpreting some of its relations. Let total labor supply be now equal to OG' (to generate an underpopulation situation); let the

labor demand curve be represented by the *lower* horizontal line (constant marginal productivity of labor in the capitalist sector), now read as P_{h0}; finally, let the effort extraction curve be represented by the segment n'G, which shows, for each wage rate, the unemployment rate that firms find necessary to maintain labor discipline and thus the productivity level. Hence, equilibrium of the labor market is given at point E', where the demand curve and the effort extraction curve cross each other, with wage employment equal to OB and BG' showing the equilibrium level of unemployment. This labor market equilibrium would be able to show interactions between real and nominal variables in a general equilibrium model, a prediction that is consistent with facts, as indicated above. This is precisely what is shown in one of the most popular macroeconomics textbooks (Blanchard 2009, Appendix to chapter 6).

Using this labor market model, the macroeconomic model will be able to derive the determinants of the Non-accelerating Inflation Rate of Unemployment (NAIRU). Measured as the average unemployment rate for a large period, NAIRU is higher in Japan (2.3%) compared to the United States (6.1%) since 1970 (Blanchard 2009: 190). In the light of the labor market theory of unemployment as labor discipline device, one could interpret this difference as given by differences in the minimum unemployment rate required to assure labor discipline between these countries.

In sum, the labor market model of underemployment as labor discipline device, which has been developed in this study, can explain the basic facts in overpopulated societies, but the classical models of development economics and the neoclassical models cannot. In the case of underpopulated societies, recent macroeconomic models have integrated the labor market model of unemployment as labor discipline device and have shown consistency with facts.

A UNIFIED THEORY OF LABOR MARKETS

Do the two models, one for overpopulated and the other for underpopulated societies, have some common and consistent features? Do we have a unified or generalized labor market theory? This is a similar question that is found

in the case of physics, in which the theory of the large bodies (relativity theory) and the theory of the subatomic world (quantum mechanics theory) are inconsistent to each other, that is, they both cannot be true.

The common features of these two models are the following. Firstly, in both models, labor markets require labor discipline devices. Secondly, the labor discipline devices in both cases imply inequality among workers. In the underpopulated society, inequality is given by the gap between the wage rate and the income of the unemployed (zero income or a subsidy through the unemployment insurance that is smaller than the wage rate); in the overpopulated society, it is given by the gap between the wage rate and the marginal income of the self employed in the subsistence sector. Of course, this inequality is just part of the overall inequality that will result from the economic process at the general equilibrium solution in society. Thirdly, the use of labor discipline devices also implies labor market equilibrium with excess supply of labor. In the underpopulated society, the excess labor supply takes the form of unemployment, whereas in the overpopulated society it basically takes the form of underemployment and also unemployment.

These common features of both models can lead us to derive a generalized labor market theory, applicable to both the developing countries and the advanced countries. The generalized labor market theory states that inequality among workers plays the role of labor discipline device. This inequality refers to a given level of skills. The generalized theory predicts that when developing countries become advanced countries the labor discipline device will show a definite shift of regime: from underemployment towards unemployment.

Conclusions

The standard microeconomic theory of labor markets assumes that firms use the unemployment rate as the labor discipline device. This theory implicitly assumes that society is underpopulated and is applicable to the group of developed countries. A new theoretical model to explain the functioning of labor markets in overpopulated societies has been proposed in this paper. This model assumes that firms use the gap between the market wage rate and

the income in the subsistence sector (which generates underemployment) as the labor discipline device. This model is intended to apply to the group of developing countries.

A set of empirical predictions has been logically derived from the model, which makes the model falsifiable. At the first stage of the falsification process, the model generates empirical predictions on observable equilibrium conditions under which labor markets operate. These predictions of the model are totally consistent with the set of empirical regularities that characterize the functioning of labor markets in developing countries. A set of causality relations has also been logically derived from the model. But the falsification of these predictions must wait for econometric work. Therefore, there is no reason to reject the model and we can accept it at this stage of our research.

The model is an improvement over the other classical labor market models that were developed by Arthur Lewis and by Harris-Todaro in the field of development economics. The predictions of these classical models are refuted by some of the empirical regularities listed in this study. The predictions of neoclassical models are also refuted by the empirical regularities.

Finally, a comparison of the two labor market models -the standard and the one that has been proposed in this study- has demonstrated that, although they seek to explain different realities, they have common features: labor discipline devices are needed and inequality among workers constitutes such a general device. Thus we can conclude that we have a generalized labor market theory. The two theoretical models of labor markets are just particular models of this generalized theory.

References

Abowd, J.M.; F. Kramarz, and D. Margolis, "High wage workers and high wage firms," *Econometrica*, vol. 67(2), 1999, pp. 251-333.

Barro, R., *Macroeconomics*, Fifth Edition, Cambridge, Massachusetts, The MIT Press, 2000.

Blanchard, O., *Macroeconomics*, Fifth Edition, Upper Saddle River, NJ, Prentice Hall, 2009.

Bowles, S, "The production process in a competitive economy: Walrasian, Neo-Hobbesian, and Marxian models", *American Economic Review*, vol. 75(1), 1985, pp. 16-36.

Comisión Económica para América Latina y el Caribe (cepal), *Notas de la cepal*, Santiago de Chile, cepal, 2008.

Ehrenberg, R. and R. Smith, *Modern Labor Economics: Theory and Public Policy*, Tenth Edition, Boston, Massachusetts, Addison Wiley, 2009.

Figueroa, A. "Is education income-equalizing? Evidence from Peru", cepal *Review*, no. 102, 2011.

———, "The labor market in lima: hypotheses testing", Centrum, Catholic University of Peru, Working Paper no. 9, 2009.

Hall, R., "Employment fluctuations with equilibrium wage stickiness", *American Economic Review*, vol. 95(1), 2005, pp. 50-65.

Harris, J. and M. Todaro, "Migration, unemployment & development: a two-sector analysis", *American Economic Review*, vol. 60(1), 1970, 126-42.

International Labor Office (ilo), *Women and Men in the Informal Economy: A Statistical Picture*, Geneva, ilo, 2002.

Krugman, P. and R. Wells, *Macroeconomics*, New York, Worth Publishers, 2006.

Lewis, A., "Economic development with unlimited supplies of labor", *The Manchester School of Economics and Social Studies*, vol. 22(2), 1954, pp. 139-191.

Meagher, K. and H. Wilson, "Different firm size effects on wages for supervisors and workers", *Economics Letters*, vol. 84(2), 2004, pp. 225-230.

Oi, W., "Employment relations in dual labor markets", *Journal of Labor Economics*, vol. 8(1), 1990, pp. 124-149.

Shapiro, C. and J. Stiglitz, "Equilibrium unemployment as a worker discipline device", *American Economic Review*, vol. 74(3), 1984, pp. 433-444.

Reilly, K., "Human capital and information: the employer size-wage effect", *The Journal of Human Resources*, vol. 30(1), 1995, pp. 1-18.

Telles, E., "Urban labor market segmentation and income in Brazil", *Economic Development and Cultural Change*, vol. 41(2), 1993, pp. 231-249.

Troske, K., "Evidence on the employer size-wage premium from worker-establishment matched data", *Review of Economics and Statistics*, vol. 81(1), 1999, pp. 15-26.

United Nations Development Program (undp), *Informe sobre desarrollo humano 2004*, New York, Ediciones Mundi-Prensa, 2004.

Weiss, A. and H. Landau, "Wages, hiring standards, and firm size", *Journal of Labor Economics*, vol. 2(4), 1984, pp. 477-499.