Labor Surplus Economies

Gustav Ranis

Follow this and additional works at: https://elischolar.library.yale.edu/egcenter-discussion-paper-series

Recommended Citation
https://elischolar.library.yale.edu/egcenter-discussion-paper-series/908
LABOR SURPLUS ECONOMIES

Gustav Ranis
Yale University

December 2004

Notes: Center Discussion Papers are preliminary materials circulated to stimulate discussions and critical comments.

This paper can be downloaded without charge from the Social Science Research Network electronic library at http://ssrn.com/abstract=631206

An index to papers in the Economic Growth Center Discussion Paper Series is located at: http://www.econ.yale.edu/~egcenter/research.htm
Labor Surplus Economies

Abstract

The labor surplus economy model has as its basic premise the inability of unskilled agricultural labor markets to clear in countries with high man/land ratios. In such situations, the marginal product of labor is likely to fall below a bargaining wage, related to the average rather than the marginal product. The reallocation of such disguisedly unemployed workers by means of “balanced” intersectoral growth ultimately permits the entire economy to operate on neo-classical principles. Finally, the paper introduces open economy dimensions, indicates the existence of other labor surplus sub-sectors and briefly responds to neo-classical critiques on both theoretical and empirical grounds.

Key Words: Development theory, labor markets

JEL Codes: O10, O12, O17
I. Introduction

Labor surplus economies are closely associated with the concept of economic dualism, i.e., the existence of organizational heterogeneity as between major sectors of an economy. The basic premise is that there exist some sectors or sub-sectors in which, in the presence of a large endowment of unskilled labor and the absence of sufficient cooperating land or capital, with a given technology and a wage level bounded from below, labor markets cannot clear. A full employment, neoclassical “wage equals marginal product” solution would drive remuneration below socially acceptable, possibly subsistence, levels of consumption. Consequently, a labor surplus exists in the sense that a substantial portion of the labor force contributes less to output than it requires, i.e., its marginal product falls below its remuneration, set by bargaining. The “labor surplus” designation then arises from the fact that if such workers were reallocated to other, competitive, or neoclassically functioning sectors, such reallocation would eliminate the aforementioned inefficiency and thus materially enhance the total output of the system.

The prime location for such surplus labor has traditionally been developing countries’ agricultural sectors, concentrated especially in subsistence agriculture, characterized by family farms, i.e., excluding commercialized plantation agriculture which consists of profit maximizing entities able to hire and fire workers following well-known neoclassical principles. Surplus labor makes its appearance in the context of owner-operated extended family networks, communes, villages or similar tenurial arrangements, all configurations in which income or output shares are determined via bargaining in relation to (though not necessarily equal to) the average rather than the marginal product of labor. Wage determination is thus based on a sharing principle, a function of the fact that when high man/land ratios are part of the initial conditions low marginal
productivity workers cannot be dismissed or otherwise eliminated. In Section II, we present the static version of the labor surplus economy. In Section III, we describe the conditions for balanced growth. Section IV introduces open economy dimensions, and Section V cites some extensions and responds to some critiques.

II. The Static Labor Surplus Economy

Diagram 1 illustrates the situation of relatively scarce land, intensively cultivated, yielding extremely low increments of output at the margin. Labor is measured on the horizontal and land on the vertical axis, with production contour lines indexed as $M, M'$, and $M''$ in diagram 1a. Given technology, fixed land, at $ON$, and labor endowment at $OS = OS' = OS''$ in diagrams 1a, b, c, the total product curve is $ODQ_A$ in diagram 1b and the marginal product of labor, depicted by curve $ABC$ in diagram 1c, approaches very low levels, substantially below the bargaining or institutional wage or income share $OW$, which is related to (again not necessarily equal to) the average product (slope of $OQ_A$ in diagram 1b). Under these conditions, we can locate the proportion of the total agricultural labor force which is “in surplus,” in the sense that it is “disguisedly unemployed” or “underemployed,” as $ST'$ in diagram 1c. This includes all those whose marginal product lies below their consumption or income share. They represent the “labor surplus” phenomenon or what Rosenstein-Rodan (1943) and Nurkse (1953) long ago designated as “hidden rural savings” which could be mobilized via reallocation to higher productivity activities elsewhere in the economy.
Diagram 1
The Static Labor Surplus Economy
It should be emphasized that “labor surplus” therefore does not mean, as has often been asserted, that a substantial portion of the agricultural labor force can be withdrawn without loss of output. Such a zero marginal product condition constitutes a statistically highly unlikely razor’s-edge event but, partly because it has been assumed for purely diagrammatic and/or mathematical convenience by Lewis, Fei-Ranis, and others, it has drawn extensive and often intemperate critical comment in the literature. Schultz (1964), for example, cited the fact that output in India declined with a decline in the agricultural working population due to an influenza epidemic as proof that surplus labor was a “false doctrine.” As Sen (1967) pointed out in rebutting Schultz on this point, when some workers with low (or even zero) marginal productivity are withdrawn, some of those left behind are likely to adjust by working harder. Or, put more broadly, any withdrawal of labor from agriculture is very likely to be accompanied by a reorganization of production arrangements on the part of those left behind, i.e., technology change. This would be equivalent to an upward shift of the $ODQ_A$ curve in diagram 1b and of the $ABC$ curve in diagram 1c.

III. Balanced Growth in the Labor Surplus Economy

Dynamically, the labor surplus condition can thus be seen as permitting an increasing number of agricultural workers and an increasing volume of agricultural surplus, defined as the difference between total agricultural output and what is needed to satisfy the remaining agricultural population’s consumption requirements, to move out and support the expansion of commercialized activities, industry and services, rural and urban. This labor surplus condition of the economy then ultimately comes to an end when increases in agricultural productivity, freeing up workers and generating agricultural surpluses, accompanied by increases in productivity in the expanding commercialized sector, enhancing the demand for workers, have proceeded in a
more or less “balanced” fashion long enough, and at a rate exceeding population growth, to mop up the disguisedly unemployed, i.e., all those whose marginal product lies below their wage or consumption standard.

This critical concept of the need for “balance” between the non-commercialized and commercialized components of the labor surplus economy has really three ingredients: one, the most obvious, is that the release of labor from non-commercialized agriculture be roughly in balance with its absorption by commercialized non-agriculture. Another, focused on the product rather than the organizational dimension of dualism, suggests that relative advances in productivity in the two sectors proceed in such a fashion that the inter-sectoral terms of trade are not substantially affected, i.e., that the system does not encounter food shortages or, less likely, food surpluses, in the course of the development process. Thirdly, the financial intermediation network, primitive at first, more sophisticated later, represents a crucial link as it must be capable of transforming non-commercialized sector surpluses, joined by commercialized sector profits, into efficient investment, mainly in the commercialized sector.

Turning, first, to more specifics on the inter-sectoral labor market, it should be noted that the unskilled real wage in the commercialized sector will tend to be tied to, though certainly not equal to, the non-commercialized real wage. A substantial unskilled labor wage gap is indeed likely to be required, partly to induce the typical agricultural worker to overcome her attachment to soil and family, partly to meet transport costs, and partly as a consequence of such institutional factors as commercialized sector minimum wage legislation, unionization, public sector wage setting, etc., all of which usually do not extend into non-commercialized activities. Once these two wage levels are given within a general equilibrium context, the release of labor by the non-
commercialized sector and its absorption by the commercialized sector represents an essential ingredient of balanced growth in the labor surplus economy.

It should also be noted that both wages may be expected to rise over time, in part because, as agricultural sector labor productivity increases, there is also likely to be some upward adjustment of the bargaining wage which is tied to the rising average product; moreover, the intersectoral wage gap may rise as a consequence of a change in the extent of commercialized sector interventions via minimum wage increases, enhanced union bargaining power, etc. The two unskilled real wage patterns over time may thus be conceived of as a step function, horizontal at any point in time, reflecting the labor surplus condition, but at a slightly higher level, again horizontal, in the next period. All this will, of course, yield a gently rising labor supply curve over time, giving way to a sharply rising pattern once the labor surplus has been exhausted and remuneration is determined neoclassically, i.e., by the marginal product. Meanwhile, the existence of a relatively constant or gently upward sloping real wage over time in both sectors, with a possibly growing gap between them, can be expected to induce labor intensive technology choices and, more importantly, labor using technology change in both the non-commercialized and commercialized sectors of the labor surplus economy.

Secondly, an understanding of the workings of the intersectoral commodity market is required for an assessment of the contribution of the non-commercialized sector to the rest of the economy. This can be seen in terms of the net real resources transferred, i.e., the difference between the shipments of food and raw materials delivered to the commercialized sector and the shipments of goods and services sent in the opposite direction. The agricultural sector’s export surplus may thus be viewed as the contribution of that sector to both the labor reallocation and overall growth process over time.
The main participants in the dualistic commodity market are thus, on the one hand, the owners of the agricultural surplus and, on the other, the newly allocated workers who may be thought of as receiving wage income in the form of non-agricultural goods and anxious to trade some of these for the food “left behind.” Once this transaction is completed, the reallocated worker finds herself in possession of the agricultural goods needed to at least maintain her consumption standard—most likely to increase it because of the aforementioned inter-sectoral wage gap. In this fashion the dualistic commodity market is indispensable for transforming the consumption bundle of the agricultural labor force into a wages fund for the newly allocated non-agricultural workers. At the same time the owners of the agricultural surplus, e.g., the landlords and/or the government via land taxes, obtain a claim against a portion of the newly formed non-agricultural capital stock; the other portion results from the reinvestment of profits by commercialized sector entrepreneurs. The above underlines the importance of the product, along with the organizational dimension of balanced growth in the closed labor surplus economy, rooted in the fact that food and non-agricultural products cannot readily be substituted for each other. Agriculture is thus a necessary condition for non-agriculture, while the converse does not strictly hold. In the open economy, food imports, of course, become possible, thus helping the system avoid premature food shortages, as illustrated by Japan’s historical experience in the early decades of the 20th century (see Hayami and Ruttan 1970).

Thirdly, the financial counterpart of the real resources contribution of the non-commercialized to the commercialized sector over time is effected through the workings of the intersectoral financial market. As we have seen, the savings of the agricultural sector become a claim against non-agriculture, the magnitude of which is determined by the size of its export surplus. These savings must somehow be channeled into non-agricultural investment, i.e., what
is left of the agricultural surplus that is not siphoned off by consumption or intermediate input requirements must find its way into capital formation in the rest of the economy.

The dynamics reflecting all the main facets of such a balanced growth path can be illustrated by reference to diagram 2 within a simplified setting—i.e., without intermediate input flows between the two sectors. Total population \( L \) is shown on the horizontal axis in quadrant II, moving from right to left, with agricultural output and the institutional consumption standard \( c = w_a \), measured in terms of agricultural goods, on the vertical axis. The curve \( OQ' \) describes per capita food availability for the total population, or \( Q/L \), at a given level of technology, for various possible proportions, \( \theta \), of the total population already allocated to other activities, \( B \), i.e., \( (\theta = B/L \geq 0) \). One equilibrium point along a balanced growth path may then be defined as
follows:  let initial consumption $c = w_a^0$, and the terms of trade between $w_a^0$, the “wage in terms of agricultural goods” ($Q_a$), and $w_{na}^0$ the “wage in terms of non-agricultural goods” ($Q_{NA}$) be given.  For simplification only, we assume that there is no wage gap between unskilled agricultural and non-agricultural workers.  The price-consumption curve ($PC$) in quadrant I of Diagram 2 then indicates all possible points of tangency between changing terms of trade and a given typical worker’s consumer preference between agricultural and non-agricultural goods.  Point $e$ is the consumption equilibrium point for the typical worker, given the terms of trade shown, regardless of whether she is engaged in agricultural or non-agricultural activities.  $B$ is the population outside of agriculture and the remaining agricultural population $V$ ($L = B + V$) produces enough food to meet everyone’s consumption requirements at the institutional wage.

The auxiliary 45° line in quadrant III transposes workers $B^0$, already allocated to non-agricultural work, onto the vertical axis, i.e., $OB^c$.  The consistent equilibrium point for employment in the non-agricultural sector is then point $d^o$, located at the intersection between the “horizontal” supply curve of non-agricultural labor, at wage level $w_{na}^0$, and the demand curve for non-agricultural labor, or the marginal productivity curve corresponding to a particular level of the capital stock and technology in that sector.  This describes an equilibrium position $abc$ in both the intersectoral labor and commodity markets.

Turning to the definition of balanced growth over time and assuming no upward adjustment of the agricultural real wage and, thus, of the non-agricultural real wage which is “tied” to it, balanced increases in agricultural and non-agricultural productivity resulting from capital accumulation and technology change can be shown by a shift of the per capita food availability curve to $OQ_a^1$ in quadrant II, with $Lb^1$ or $OB^1$ workers now allocated, as well as of
the marginal productivity of non-agricultural labor curve $d^1$ in quadrant IV. This would result in a new equilibrium position $a'b'd^1$ where, once again, the two intersectoral markets clear. Such a growth path would clearly meet the labor market equilibrium condition, and a little more work would permit us to demonstrate that equilibrium in the commodity market sense, as previously defined, also continues to be achieved, permitting agricultural and non-agricultural workers to exchange some of the goods they produce for the goods they need, at the given terms of trade, enabling everyone to remain at the same equilibrium point $e$.

Turning to the intersectoral financial market, the landlords and/or the government, whoever owns the agricultural surplus, would end up with a claim against some part of the non-agricultural capital stock. This, plus the reinvested industrial profits represented by the shaded area in quadrant IV of Diagram 2 would be invested in the non-agricultural sector, causing, along with technology change, the indicated shift of the marginal productivity curve. The investment fund for the next period is thus composed of this period’s savings out of the agricultural surplus plus the savings out of non-agricultural profits. For the sake of convenience, we have made the assumption of no leakage into consumption by either landlords or capitalists. The allocation of the society’s investment fund plus its innovative energies, as between the sectors, would then be guided by the relative shortages of agricultural and non-agricultural goods, as reflected, in the case of a market economy, by changes in the intersectoral terms of trade. In a non-market economy the role of changes in the terms of trade as a signaling device would be taken over by evidence of unplanned shortages or surpluses in the material balances sense. We have here again made a simplifying, but not critical, assumption that technology change is responsible for agricultural productivity change, while all the investment funds are allocated to non-agriculture.
As we have already noted, the entire transition process must not only be balanced but also proceed at a pace in excess of population growth if the initial reservoir of surplus labor is to ultimately be exhausted and neoclassical wage determination is to take over. Moreover, if balanced growth, as indexed by the rate of labor reallocation, only marginally exceeds the rate of population growth on average, the length of time it takes to arrive at the commercialization point, marking the end of labor surplus, must also be politically acceptable.

The real world, of course, does not quite operate in such a smooth fashion. There are times when, under the impetus of an “industry first” strategy, non-agricultural productivity increases for some time at a rate in excess of agricultural productivity growth, leading to food shortages, the shifting of the terms of trade in favor of agriculture, and an increase in the non-agricultural real wage. The reverse can also occur, although empirically there seems to be less danger of that. Most successful labor surplus societies (e.g., historical Japan and post-war South Korea, Taiwan and Thailand) have, in fact, experienced something approaching constancy in the terms of trade.

In any case, progress along a balanced growth path at a rate in excess of population growth—and sufficiently in excess to guarantee a politically acceptable time perspective—is essential to a society’s successful transition into a modern growth regime. Success is defined as the end of labor surplus, i.e., the end of organizational dualism in the labor market. Once balanced growth has proceeded long enough and fast enough labor surplus gives way to labor shortage in both sectors, which means that the marginal productivity calculus of wage determination takes over. At this point organizational dualism disappears; and, given considerable increases in per capita incomes and the workings of Engel’s Law, product dualism also atrophies over time as agriculture gradually becomes an appendage to the economy, or just
another symmetrical sector within the system’s input-output matrix. Increasingly the economy is then ready to perform according to the rules of modern economic growth as described by Simon Kuznets (1966).

IV. Open Economy Dimensions

Thus far we have discussed the development of the labor surplus economy mainly in a closed economy context. The open economy or trade related dimensions of development in the labor surplus economy are, of course, important enough to warrant substantial amendment of the analysis presented here. During the early colonial, or open agrarian, phase of development the economy may well be tied to foreign markets by virtue of some of the labor force being weaned away from food production and into land-based export oriented activity, e.g., minerals and other primary products of interest to foreign investors. This typically leads to a triangular relationship among the cash crop export sector, the foreign sector, and the food producing domestic agricultural sector. But once the economy moves out of its colonial or “overseas territory” phase and into a national development-oriented effort our analysis must be amended to take “openness” into account.

To do so, we must, first, recognize that the export-oriented cash crop agricultural sub-sector continues to generate foreign exchange earnings but that these are now used, in addition to possible food imports, to assist in the construction of a new, domestically oriented, non-agricultural sector producing previously imported non-durable consumer goods, i.e., to fuel so-called primary or “easy” industrial import substitution. These raw material-intensive exports thus provide a second source of agricultural surplus which, converted into industrial capital goods imports, and possibly supplemented by the inflow of foreign savings, helps finance non-agricultural growth in the same balanced growth context. In this way a new triangular
relationship between two kinds of commercialized activities, one agricultural, and one non-agricultural, plus the food producing non-commercialized agricultural hinterland, replaces the colonial triangle.

What happens at the end of this primary import substitution phase is critical, i.e., once domestic markets for the non-durable consumer goods are exhausted, it is apparent that relatively natural resources rich labor surplus countries have a tendency to continue on with import substitution, now shifting from labor intensive light industries to the more capital intensive durable consumer goods, the processing of raw materials, and the production of capital goods. At the same time, in the minority of countries which have a relatively poor natural resources base we observe a shift from a domestic to an export market orientation for the same labor intensive non-durable consumer goods. In that case the export sector now constitutes a powerful new production function available to the economy through which traditional and, later, non-traditional exports can be converted into imported capital goods and raw materials. Moreover, the openness of the economy permits foreign capital to provide additional finance in support of the balanced growth process. Finally, an important potential advantage of the economy’s openness is, of course, the whole range of additional technological alternatives now made available, which, hopefully with modifications and adaptations, can help increase the efficiency and speed of the balanced growth process.

The open economy, in other words, not only permits the labor surplus economy to harvest the normal gains from trade, to benefit from the vent for surplus of previously under-utilized resources—in this case not only raw materials but also unskilled labor—but also, dynamically, to affect the direction of technology change and thus introduce competitive forces and ideas from
abroad which are able to diffuse throughout the economy and are undoubtedly of considerable
importance in determining the success of the labor surplus economy’s transition efforts.

V. Extensions and Critiques

Up to now we have focused exclusively on owner-operated agriculture as the typical
representative of the non-commercialized sector of the labor surplus economy. It should,
however, be recognized that there are very likely to exist substantial portions of non-agricultural
activities, both rural and urban, and both industrial and services-oriented, which are labor surplus
in the way we have defined the condition. This time, the cooperating factor in short supply is
capital. Most relevant is the so-called informal sector—both rural but most heavily urban—
which occupies a large, often dominant, position in many developing countries. Family and
cooperative ventures in this setting are characterized by the same sharing of total income, i.e., a
bargaining wage, coupled with low marginal productivity, that we encountered in subsistence
agriculture. We are here including not only the substantial portions of both the rural and urban
populations engaged in distributive trades and services—ranging from the vendors of tea,
flowers and cigarettes to barbers, bootblacks and car watchers but also to the blacksmith, metal
working, and repair shops which dominate the landscape in most labor surplus developing
countries. Some portions of this informal sector, especially its urban branch, are likely to be
static and of the labor absorptive “sponge” variety; others may be capable of technology change,
of subcontracting arrangements with the urban formal or commercialized sector as well as of
generating surpluses for investment in that sector. Thus, organizational dualism is quite
pervasive in both rural and urban non-agriculture, even as product dualism now loses its
distinctive characteristic.
As development over the past half century has proceeded apace, some initially labor surplus countries, including Taiwan, South Korea and Thailand, have graduated from their initial labor surplus condition, evidenced by gently rising unskilled wages in both sectors, finally giving way to rapid and sustained increases, as secular labor shortages make their appearance. Such a turning point was reached around 1968 in the case of Taiwan, around 1973 in the case of South Korea and around 1993 in the case of Thailand. It is also true that many developing countries, starting with up to 80% of their population and 50% of their output in food producing agriculture, have gradually shifted substantially into non-agricultural pursuits, with services retaining their dominant position, even as their composition has changed radically, in the commercialized direction. As a consequence, the number of contemporary developing countries with typical initial labor surplus characteristics has been declining. Nevertheless, a large preponderance of the developing world, certainly by weight of population, continues to find itself in a labor surplus condition. This holds, for example, for China and India, huge countries both currently engaged in a vigorous balanced growth effort, as well as for other parts of South Asia, much of Central America, the Caribbean and parts of Latin America. Even some countries of sub-Saharan Africa, once considered land surplus by some observers, may, as a consequence of population growth and the loss of land to the Sahara, be approaching labor surplus status—though, given the AIDS epidemic, this remains a more controversial issue.

It should, finally, be noted that the fundamental concept of the labor surplus economy has come under increasing attack by the dominant neoclassical school of economics. While still viewed as relevant in the South and wherever heavy population pressure on scarce cultivable land remains a feature of the landscape, most Northern economists in the Becker micro-econometric tradition find it difficult to accept the notion of an exogenous or bargaining wage in
the non-commercialized sectors instead of one determined endogenously by the customary interaction between demand and supply. The crux of the critique is based on the rejection of the notion that initial conditions, i.e. a highly unfavorable ratio of people to cooperating land or capital, can lead to the subsidization of some members of the society by others, in lieu of ejecting them.

The work of Rosenzweig and associates (e.g., Rosenzweig 1988), presenting evidence of rising labor supply curves in a cross-section of such heavily populated agricultural sectors as India’s, typifies current mainstream rejection of the “unlimited supply of labor” condition underlying the labor surplus economy construct. Yet we would contend that such efforts capture an expressly static snapshot picture, addressing cross-sectional labor/leisure decisions across households already working at full capacity (i.e., with little leisure to spare), while labor surplus models are concerned with the conditions governing intersectoral labor reallocation over time.

The exogenous agricultural wage assumption underpinning labor surplus economies, so troubling to neoclassical economists, gets support from anthropologists like Geertz (1963) and Scott (1976), as well as from economists like Lewis (1972), Ishikawa (1975), Fei/Ranis (1964), Osmani (1991), Ohkawa (1972) and others. Fafchamps (1992) provides an overview of the principles underlying the “solidarity network” among peasants as depicted in anthropological evidence. Ishikawa (1975), an astute long-time observer of Asian economic development, endorses the concept of a “minimum subsistence level of existence” (MSL), one version of the institutional real wage. His work indicates the prevalence of a “community principle of employment and income distribution which promises all families... an income not less than MSL.” Hayami and Kikuchi (1982), basically neoclassical in outlook, find that in Indonesia “wages do not adjust on the basis of labor’s marginal product, but according to the subsistence
requirements of the time and social conventions.” Only over time is there a tendency to adjust, but even then it does not necessarily occur by altering wages to equal the marginal product, which could reduce the wage below subsistence. Instead, in Java, harvest contracts are adjusted to include weeding duties without a complementary rise in the wage rate, thereby not threatening the MSL but moving institutionally towards equilibrium. Even Ken Arrow (1986), one of the high priests of neoclassical economics, states that it may take a considerable period of time before equilibrium is reached. Osmani (1991) presents a model of downward rigidity of the sharing rule insisted on by the workers themselves. Current work in what is called behavioral economics may also prove to be of help in developing a theoretical structure to rationalize cross-worker subsidization in the absence of assured reciprocity—especially as some members of the group are likely to be leaving agriculture over time.

Perhaps even more relevant, there is evidence, not only for Taiwan, Korea, and Thailand but also for post-enclosure England between 1780 and 1840 and for post-Restoration Japan between 1870 and 1920, indicating substantial increases in agricultural labor productivity, while both agricultural and non-agricultural unskilled real wages were rising only gently, until commercialization was reached and wages began to rise steeply in line with rising marginal productivity. Thus, both historical and twentieth century development patterns are inconsistent with the neoclassical school’s one-sector full employment equilibrium assumptions.

In the final analysis, what is relevant is whether the labor surplus model provides a better fit for the observed empirical pattern of successful labor abundant developing countries; whether the model is better suited to analyze relative agricultural neglect in failure cases; whether it is better able to explain changing patterns of technology choice and the direction of technology change; whether, in sum, it makes better sense than to assume away the initial existence of
underemployment and disequilibrium before the one-sector, fully commercialized modern
growth epoch can be reached.

Gustav Ranis
Frank Altschul Professor of International Economics
Yale University
References


