Chapter 10

Rural and Urban

10.1. Overview

10.1.1. The structural viewpoint

The literature on economic growth, a good part of which we studied in Chapters 3 and 4, might tempt you to view economic development (or economic growth anyway) as a process that transforms all incomes and all sectors of the economy in some harmonious and even fashion. But our study of inequality, poverty, and population growth in later chapters does alert us to the possibility of *uneven* growth—growth that first proceeds by benefiting some groups in society. The same is often true of various *sectors* of the economy. More often than not, economic development entails the rapid growth of some parts of the economy, while other parts are left behind to stagnate or even shrink. The structural transformation that inevitably accompanies change is an integral part of the development process, and to study it we must look at economies in more disaggregated form.

By far the most important structural feature of developing countries is the distinction between the rural and the urban sector. As we saw in Chapter 2, a significant fraction of the population of a typical developing country lives in the rural areas. Of this fraction, a sizable majority is connected to agriculture as a way of life. Not surprisingly, agriculture often accounts for a large share of national output as well (although the proportions are usually lower than the corresponding share of people) and almost always accounts for more than its fair share of total poverty.

Quite apart from these features intrinsic to agriculture, there is another worth emphasizing: the interconnection between agriculture and the rest of the economy, notably the industrial sector. As economic development proceeds, individuals move from rural to urban areas: agriculture acts as a supplier of labor to industry.

At the same time, labor supply isn't all that is at stake. If international trade in food grain is not an option (and it often isn't, because governments fear lack of self-sufficiency in food production), then a nonagricultural sector can come about only if agriculture produces more food than its producers need for their *own* consumption. That is, agriculture must be capable of producing a *surplus* that can be used to feed those who are engaged in non-agricultural pursuits. Thus agriculture is also a supplier of food to industry. These twin resources—food and labor—need to move in tandem if development is to proceed. We take up this point in more detail later. Before we do so, here is an overview of the nonagricultural and agricultural sectors.

10.1.2. Formal and informal urban sectors

Begin with the nonagricultural sector; that is, economic activity in urban or semi-urban areas. People who live in these areas are involved in industrial enterprises, both at the production and managerial level, and in various service sectors, such as retailing, trade, or tourism. Once we start to disaggregate our economy, it is useful to take the process a bit further, and a further division naturally

suggests itself. In all developing countries, two forms of urban economic activity are apparent (although the line between them is not very finely drawn). There are firms that operate under the umbrella of accepted rules and regulations imposed by government. Often, the workers of these firms belong to a union, and collective bargaining between firms and workers is not uncommon. These firms are required to pay minimum wages and must conform to certain standards of safety, rules of compensation for workers, pension schemes, and the like. Such firms pay taxes, may receive infrastructural facilities, such as access to subsidized electricity, and may have access to foreign exchange quotas or the right to import certain inputs. Although these norms and regulations vary from country to country, the point is that such firms adhere by and large to such regulations and receive, in turn, the benefits of state economic support.

Think of these firms as the *formal sector* of the economy. The formal sector bears a close resemblance to economic activity in developed countries. Because this sector is set up in a way that permits the creation and maintenance of records, firms in the formal sector are relatively tangible entities: they can issue shares and pay out dividends, they can be audited, and they are protected by bankruptcy laws and implicit or explicit forms of insurance. To be sure, entry into the formal sector is typically costly. Perhaps only a certain minimum size of economic activity warrants the setup costs: a license may be required for operation, tax records need to be kept, pension schemes need to be set up for employees, and so on.

In contrast, the urban *informal sector* is a loose amalgam of (usually small-scale) organizations that escape the cover of many of these regulations and do not receive access to privileged facilities. The informal sector usually does not adhere to norms of minimum wages, retirement plans, or unemployment compensation. They do not pay taxes and they receive little government support. These firms or businesses are not illegal in the strict sense, but there is a shadowy penumbra within which they live, and it is often convenient for the government to look the other way. It is difficult to implement the rule that a peanut vendor pay his taxes, in part because it is impossible to ascertain how much he earns. The same goes for the hawker, the teenager who shines shoes, domestic servants, professional beggars, owners of tea stalls, rickshaw pullers, and the young boy who assists in selling bus tickets or carrying your shopping bags.

As we have seen already, an enormous fraction of the labor force comes under this classification. Setup costs are relatively low: the business or trade is usually small scale, and license fees and advance tax payments are unnecessary (although the occasional bribe may be needed).

Bolivia's Formal and Informal Sectors[1]

In 1986, Bolivia's official labor force numbered 1.6 million, which was about half of the economically active population or a quarter of the total population. The informal sector was large and the lack of proper accounting here suggests that the figure of 1.6 million is probably a significant underestimate.

In the late 1980s, nearly half of all workers were in agriculture. Industry accounted for another 20% and the rest went to services. As we noted in Chapter 2, the rapid growth of the services sector in developing countries is often an indicator that the agricultural sector is releasing labor faster than the industrial sector can soak it up. Bolivia is no exception. The services sector has grown steadily (and mainly at the expense of agriculture) since 1950. Urban workers were clustered in the cities of La Paz (40%), Santa Cruz (20%), and Cochabamba (20%). Urban incomes significantly exceeded rural incomes; the lowest incomes were in the southern highlands. Bolivia has a significant history of strong organized labor: labor unions were powerful and strong, and strikes or demonstrations are not infrequent.

Of course, most nonunionized labor was in the informal sector. This sector included nonprofessional, self-employed, unpaid family workers, domestic servants, and businesses with five or fewer employees. La Paz was the center of the informal sector, but

there was also an illegal component linked to the coca industry.

The informal sector was characterized by ease of entry, the use of credit from noninstitutional sources, and nonadherence to government regulations, especially regarding the sale of smuggled goods. We can imagine, then, that the variation in informal incomes was quite high. Owners of small businesses might average an income as much as twelve times the minimum wage. In contrast, salaried workers and domestic servants made much less: typically around half the minimum wage.

Informal activities included transportation (usually unregistered buses or taxis), laundry, electrical services, black market currency transactions, money lending, family grocery stores, and the sale of food, clothing, and smuggled consumer items. Industrial workers in the informal sector included seamstresses, weavers, carpenters, and butchers.

10.1.3. Agriculture

What about agriculture? In most cases agriculture is a giant informal sector in itself if we go by the preceding definition. Tax authorities have no way to observe how much output a farmer produces, and even if they do, they cannot prove it in a court of law, so agriculture often goes untaxed. Likewise, it is very difficult, if not impossible, to implement minimum wages for rural labor. Pension plans, unemployment insurance, and organized old- age security don't exist, by and large. Nonetheless, a collection of informal institutions creates substitutes for these missing sources of support, as we shall see in the chapters to come. These substitutes are necessary: people in agriculture are often very poor and they face high levels of risk. Without these informal substitutes, no city in a developing country could withstand the consequent flow of rural–urban migration that might result.

The primary occupation in agriculture is, of course, farming. The great staples, such as wheat and rice, are farmed both for self-consumption and market sales. A variety of other crops are produced, and the degree of self- consumption varies with the nature of the crop. Cash crops, such as cotton, sugar, and luxury varieties of rice, are the most market oriented and are produced largely for market sale.

Production is organized in many ways. There are family farms that farm their own land, often largely for self-consumption. There are large owner–cultivators or capitalist farms that produce crops using modern techniques and large quantities of hired labor. There are tenant farmers who lease land from other nonfarming (or partially farming) landowners and pay rents to these landowners. Finally, there are laborers who work for wages or a commission on the land of others. Laborers may be casual employees (e.g., hired just for the duration of the current harvest) or long-term permanent employees.

As we will see, the notion of risk and uncertainty is central to the concept of agricultural organization in developing countries. In more than one developing country, the state of the weather affects macroeconomic stability, the balance of payments, and even political fortunes, all because it affects the harvest. To farmers, a good harvest means a high income in a given year, but the next year could be totally different. The weather also affects the incomes of agricultural laborers, even if they do not farm their own land, because the scale of agricultural employment, is weather-dependent. Thus agricultural uncertainty is a fundamental fact of life that plays a key role in the development process.

To give you a better idea of agricultural activity in developing countries, I am going to introduce the ICRISAT sample: a well-studied set of villages in India.

10.1.4. The ICRISAT villages

Much of agriculture in the Third World is carried out in regions known as the semi-arid tropics, which are characterized by rainfall dependence (although precipitation is scanty and uncertain both in timing and volume), primitive technology and labor-intensive cultivation, poor infrastructure, and often extreme population pressure on the land. Systematic and reliable data on such regions are not widely available; most of the data collected so far (which aren't much) are erratic and often unreliable. The International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) in Hyderabad, India, is a welcome exception to this rule. Since 1975, ICRISAT has put together detailed data that track the behavior and fortunes of certain representative sample households from eight villages in the Indian semi-arid tropics.[2] The volume of data now available from ICRISAT provide rich insight into the functioning of typical rural economies.[3]

Soil fertility

There are considerable differences in soil texture and quality, and hence productivity, both *within* and *across* the study villages. In Aurepalle, for example, farmers divide all village soils into five broad groups, and at the finest level of classification, they recognize twenty-one different soil categories. In contrast, consider Shirapur and Kalman, both villages in the Sholapur district of Maharashtra. Shirapur has a rich endowment of deep black soil, whereas Kalman has more upland area with shallower soils that do not retain enough moisture for good cropping in the postmonsoon season (although in this regard it does better than some other villages). The difference in land fertility across the two villages is borne out by the fact that owner–cultivated plots in Shirapur had an average plot value of Rs 29.68 per acre, whereas the corresponding figure for Kalman was Rs 17.55 (Shaban [1987]).

Rainfall and cropping patterns

Although the villages are in reasonably close proximity, rainfall patterns differ widely across them, and this is also true of a single village *across* years. Farmers react sharply to these differences: they can adopt very different cropping strategies across villages and also respond to changing climatic conditions. For example, in the villages in Mahbubnagar and Akola districts, more than 90% of the area cultivated is planted during the *kharif* or rainy season because these villages receive more copious and less variable rainfall, and are marked by shallower soils, with less moisture retention capacity.

The Sholapur villages, on the other hand, have (relatively) deep and rich soils that can hold moisture for longer periods. Rainfall is relatively low and far more erratic in terms of arrival time. Farmers in these villages, therefore, wait until the end of the monsoon and rely more heavily on *rabi* or post-monsoon season planting. Planting on dry soil at the onset of the monsoon surely would have increased the risk of crop failure in these villages, because of the uncertain arrival time of the rains.

At the same time, in Sholapur, the relative importance of monsoon and postmonsoon season cropping can vary considerably from year to year. During "normal" years of rainfall, rainy season crops account for about 40% of gross cropped area, whereas during an unusually dry season (e.g., in 1977–78), this share may drop below 10%. Sometimes, farmers also react to adverse weather conditions by altering the crop planted. An instance of this occurred in Aurepalle in 1976–77, when the southeast monsoon arrived late, and most farmers planted castor, a hardy cash crop, instead of the

traditional sorghum,[4] which runs a greater risk of pest damage in case of initially scanty rainfall.

Irrigation

The uncertainty introduced by the erratic pattern of rainfall can, of course, be reduced by irrigation. However, because the farmers are poor and capital is in short supply, irrigation is not very widespread. In the late 1970s, the proportion of gross cropped area that was irrigated varied from a high of 32% in Dokur to less than 1% in Kinkheda. (The average across six villages, excluding Boriya and Rampura, was around 12% in 1975–76.)

However, thanks to increased investment by the government in agricultural infrastructure, irrigation is on the rise. In 1983–84, 20% of the land was irrigated on average in the six villages. The technology of irrigation is also changing: groundwater irrigation is becoming more popular, replacing surface irrigation from small catchment reservoirs. Furthermore, in response to rapid rural electrification, cheaper institutional credit, and technological advances in groundwater pumping, electric pump sets have rapidly replaced traditional animal-drawn and diesel-powered lifting devices. For example, in Aurepalle, agricultural wells increased by 25% from 1974 to 1984; the number was around 190 in 1984. During the same time period, electric pump sets grew in number from 75 to 136. Except in the Akola villages, joint ownership of wells is fairly common. Among the sample households, the average number of owners per well in Shirapur, Aurepalle and Kanzara were 4.8, 2.4 and 1.0 respectively.

The ownership of wells is to be contrasted with the ownership of land, which is privately owned and very intensively cultivated (for more on land ownership and operations, see Chapter 12). Public and fallow land is rare. The quantity and quality of the "village commons"—jointly owned open-access land used for animal grazing—has declined over the years, from around 20% of the total area in the study villages in the early 1950s to about 10% today.

Fertilizer use

The successful application of fertilizers requires plentiful drainage. It is not surprising, therefore, that the use of fertilizers and the availability of irrigation facilities go hand in hand. Where irrigation facilities enjoy a limited spread, the use of fertilizers is also thin. In the Mahbubnagar villages, 40% of the land is irrigated and 98% of the total volume of fertilizers used was applied to such land. Likewise in Akola, only 5% of gross cropped area had access to irrigation, but these accounted for 37% of the fertilizer usage. On the whole, the mean nutrient consumption (i.e., fertilizer use) in kilograms per hectare of gross cropped area has varied from a meager 2 in the drought-prone Sholapur villages to a somewhat healthy 25 in the better irrigated Mahbubnagar villages.

Draft power

Many important agricultural operations, such as plowing, harrowing, and tilling, require draft power. In all villages, most farmers rely on traditional animal draft power, mainly bullocks. Mechanized draft power such as that provided by tractors and harvesting machines is still out of reach for the majority of farmers, due to the large capital expenses involved. Indeed, there is an acute scarcity of

draft power, even of bullocks. As in the rest of India, many households that own small amounts of land do not own bullocks. The scarcity is most pronounced in the Sholapur villages, where less than one landowning household in three owns a bullock. Sharply fluctuating fodder prices make the bullock not only an expensive but also a risky asset.

The problem of scarcity of draft animals, is somewhat mitigated if there is a well functioning market for *hiring* bullocks, but this is not the case for two reasons. First, most cropping activities must be executed within a very narrow time frame, so that all farmers across the village feel the need for bullock power more or less simultaneously: their use cannot be phased across farmers. Second, there is a fundamental incentive problem in hiring bullock power: the leasing party usually drives the animal too hard, which extracts better service, but at the cost of depleting its health and stock value (which the lessee does not internalize). The scarcity of bullocks and the incomplete market for hired bullock power drives some of the widely observed features of traditional agriculture, for example, the pattern of land-lease contracts (see Chapter 12).

Technical change

Technical change in the study villages has been most pronounced, not in the development of infrastructure or greater capital intensity of cultivation, but in the adoption of new and improved inputs, particularly seeds. High-yielding varieties (HYV) of seeds for many popular crops were introduced in India in the mid-1960s. Some of these improved varieties were adopted on a wide scale and at a remarkably fast rate in many villages, whereas others were just as quickly rejected. Among the HYVs that have enjoyed successful adoption are hybrid pearl millet[5] and modern castor varieties in Aurepalle, improved paddy varieties in Aurepalle and Dokur, and sorghum hybrids and improved upland cotton cultivars in Kanzara and Kinkheda. For these varieties, the adoption rates (proportion of farmers cultivating that crop who are using the HYV) stood at more than 70%, sometimes close to 100%.

To be sure, technical and economic problems with some of the HYVs do occur. For example, sorghum hybrids do not perform well in the red-soil Mahbubnagar villages, because they are afflicted with numerous diseases and pests. Cotton hybrids, though promising in terms of yield, require intensive plant protection, high soil fertility, and a copious supply of water for success: their adoption in the dry upland Akola villages has therefore been limited. These examples illustrate the need to mold the development and introduction of seeds to suit local conditions, as well as the need to provide complementary inputs and economic services such as irrigation and credit.

The ICRISAT villages will reappear more than once in the chapters to follow, as we study the structure of the rural sector in some detail. Not only will we look at how productive activity is carried out, we will also study the background lubricants of that activity, such as credit markets, land rental arrangements, insurance schemes, and labor contracts.

Before we do that, however, it is useful to study the overall interaction between the rural and urban sectors. That is what this chapter is all about.

10.2. Rural–urban interaction

10.2.1. Two fundamental resource flows

The most important of many rural–urban interactions is the synergistic role that agriculture plays in the development of the nonagricultural sector. From agriculture comes the supply of labor to industry and the surplus of food that allows a nonagricultural labor force to survive. These are the two fundamental resource flows from agriculture, and they lie at the heart of the structural transformation that occurs in most developing countries.

There are other connections as well. Industry supplies inputs to agriculture: tractors, pump sets, chemicals of various kinds, and so on. With a large population in the rural sector, agriculture is often a major source of demand for the products of industry, which include not just durables, but final consumption goods as well. Agrarian exports can serve as the source of vital foreign exchange, which permits the import of inputs to industrial production. While these links are important, the flow of labor from agriculture to industry and the parallel flow of agricultural surplus to nurture workers in industry are often basic to the development process.[6]

10.2.2. The Lewis model

The dual economy

Lewis [1954] outlined a view of development that was based on the foregoing fundamental resource flows. This approach, which views economic development as the progressive transformation of a "traditional" sector into a "modern" sector, goes beyond the narrower picture of agriculture-to-industry transformation, but essentially builds on it. The starting point of the Lewis model is the idea of a *dual economy*.[7] In a sentence, dualism is the coexistence of "traditional" and "modern," where the words in quotes can have several shades of meaning. The traditional sector is often equated to the agricultural sector, which after all produces the traditional output of all societies. In contrast, the modern sector is the industrial sector, which produces manufactured commodities. At the same time, "traditional" can mean the use of older techniques of production that are labor-intensive and employ simple instruments. In contrast, "modern" might refer to the use of new technology, which is intensive in the use of capital. Finally, and perhaps most important at a conceptual level, "traditional" refers to traditional forms of economic *organization*, based on family as opposed to wage labor, with overall output distributed not in the form of wages and profits, but in the form of shares that accrue to each family member.[8] In contrast, "modern" describes production organized on capitalist principles, which relies on the use of wage labor and is carried out for economic profit.

At one level, these distinctions are all a bit vague. Agricultural activity can be commercial, highly capital-intensive, and employ wage labor, just like any other "modern" economic organization. The terms labor-intensive versus capital-intensive are certainly not related one for one with traditional versus modern. Similarly, it is unclear what "traditional" modes of organization mean: the form of organization may simply depend on the particular environment (the presence of uncertainty, the lack of a capital market, or limited resources). At the same time, even if we cannot furnish a perfectly logical distinction between the two concepts, they have general usefulness and help us to organize our thoughts.

Essentially, the dual economy consists of two sectors that can be characterized in a number of ways; each way has suggestive advantages, but each carries with it the possibilities of error as well. We label the two sectors "agriculture" and "industry," but recognize that these are provisional labels and subject to change when the particular issue under discussion needs a more precise description.

For instance, it may be useful in some cases to view the urban informal sector as part of the "traditional" sector.

Surplus labor

Arthur Lewis proposed a framework of economic development that put the movement of labor from traditional to modern sectors on the center stage. The traditional sector, in this theory, is viewed as a supplier of labor, whereas the role of the modern sector is to soak up this supply. Why isn't all the supply immediately absorbed? The answer is that the scale of the modern sector is limited by the supply of capital. Thus capital accumulation in the modern sector becomes the engine of development. The fundamental assumption, then, is that labor is virtually unlimited in supply, being drawn from a vast traditional sector, whereas the rate of savings and investment limits the pace of development. In this latter sense Lewis is in agreement with the Harrod–Domar view of economic growth (see Lewis [1954]):

> The central problem in the theory of economic development is to understand the process by which a community which was previously investing and saving 4 or 5 per cent of its national income or less, converts itself into an economy where voluntary saving is running at about 12 to 15 per cent of national income or more."

We have seen the Harrod–Domar theory and its many extensions in Chapters 3 and 4, and in the description of modern sector expansion due to Arthur Lewis and many other writers on the dual economy, there is very little that adds to what we have already studied. We concentrate instead on the assumption that the supply of labor is "unlimited" and on the associated problem of an adequate agricultural surplus. To understand these features, we focus on the traditional sector of the economy.

The main idea of the Lewis model is that there is a large surplus of labor in the traditional sector of the economy, that can be removed at little or no potential cost. By cost, we refer to *opportunity cost*: the loss of traditional sector output as labor supply is reduced. Figure 10.1 explains this concept in one particular context.

Figure 10.1 plots the production function on a family farm. Quantities of labor are on the horizontal axis and output is on the vertical axis. In the background is a fixed plot of land, on which this labor is applied. Because land is fixed, there are diminishing returns to the labor input. In keeping with our view that the family farm uses "traditional" techniques of production, we neglect the use of capital inputs.

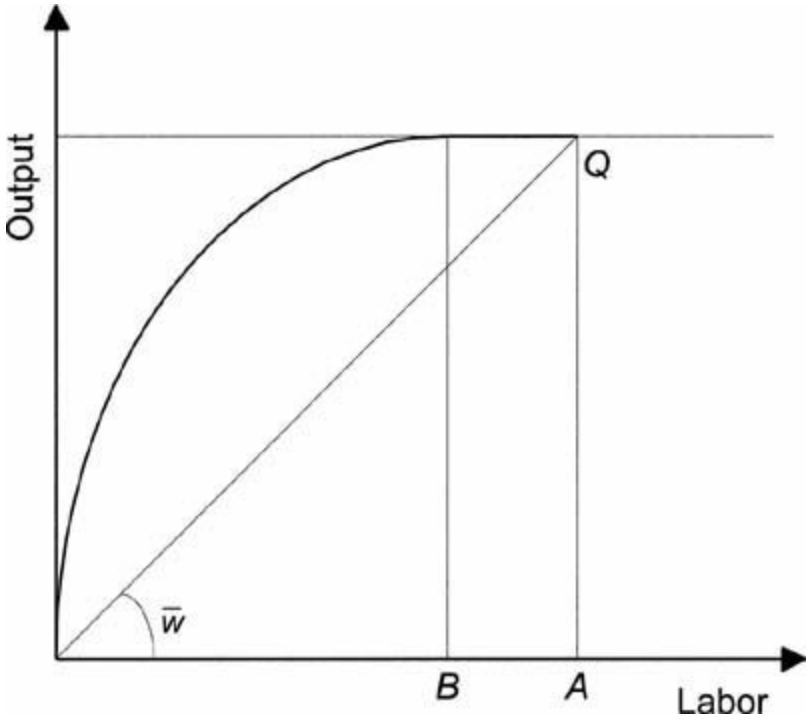

Figure 10.1. Surplus labor on the family farm.

The production function is drawn so that after a certain level of labor input, there is no significant effect on output. After all, there is only so much intensity at which a given plot of land can be cultivated, and after a point additional inputs of labor may have no effect at all. Thus the marginal product of labor at points such as *A* is zero or close to zero.

Now consider a reduction in the amount of labor from *A* to *B*. Because the marginal product of labor is assumed to be close to zero, total output practically stays constant when the reduction occurs. Because the family farm has so much labor relative to land, labor is in *surplus*.

When might such a situation occur? First, it might occur in economies where there is high population pressure, so that there are large numbers of people per acre of arable land. According to Lewis [1954]:

> "[This state of affairs] is obviously not true of the United Kingdom, or of Northwest Europe. It is not true either of some of the countries usually lumped together as underdeveloped; for example there is an acute shortage of male labour in some parts of Africa and of Latin America. On the other hand it is obviously the relevant assumption for the economies of Egypt, of India, or of Jamaica . . . an unlimited supply of labour may be said to exist in those countries where population is so large relatively to capital and natural resources, that there are large sectors of the economy where the marginal productivity of labour is negligible, zero, or even negative."

But the phenomenon is not just limited to agriculture. Again, Lewis [1954] puts it best:

> "Another large sector to which it applies is the whole range of casual jobs—the workers on the docks, the young men who rush forward asking to carry your bag as you appear, the jobbing gardener, and the like. These occupations usually have a multiple of the number they need, each of them earning very small sums from occasional employment; frequently their number could be halved without reducing output in this sector. Petty retail trading is also exactly of this type; it is enormously expanded in overpopulated economies . . ."

If you remember our statistical description of the proportion of the labor force in "services" (see Chapter 2), this account will strike a responsive chord.

Third, a simplistic aggregate view of population density might miss the point in many instances. Latin America has a relatively large per capita endowment of land, but at the same time land is very

unequally distributed, so there are a large number of family farms in the state described by Figure 10.1. In such a case we might ask why the extra labor is not hired out to those with a surfeit of land. This is indeed possible and it does happen, but at the same time it might also be the case that the large *haciendas* use capital-intensive methods of production and do not hire labor in proportion to their size. In a word, agriculture itself may be dualistic. We will have more to say on this and related phenomena in Chapter 12.

Income sharing and surplus labor

At this stage, a natural question arises: if the marginal product of labor is close to zero, how is it that such labor is hired, unless wages themselves are close to zero as well? We all know from introductory microeconomics that an entrepreneur hires labor only to the point where marginal product equals the wage. With more labor than this, gains can be realized by cutting back on employed labor and saving on the wage bill. In other words, how can we reconcile the observation that the wage is positive (and provides for minimum subsistence at least) with the parallel observation that the marginal product is close to zero?

This brings to the forefront a second asymmetry between traditional and modern sectors. One asymmetry already was used in *production methods*. We conceived of the traditional sector as an activity intensive in labor (and in land, in the case of agriculture), but not requiring significant quantities of capital. The second asymmetry is in *organization*. A profit-maximizing firm regards wage payments to employees as a cost of production, that is subtracted from revenues in order to arrive at final profits. In contrast, a family farm values the incomes received by each of its members. For instance, the output of the farm may be shared equally among its members.[9] Thus a family farm might employ labor *beyond* the point where the marginal product equals the "wage", because the wage in this case is not really a wage at all, but the average output of the farm (which is what each member receives as compensation). In Figure 10.1, if the total labor input is A and the total output is AQ, then the average income is simply the output AQ divided by input A, which is represented by the angle marked as $\bar{w}$. Contrast this with the marginal product, which is the slope of the (almost) flat tangent to the production function at the point Q. Sharing rules of this kind shelter family members from the difficulties of finding employment elsewhere.

Income sharing is not just an agricultural phenomenon and occurs not just among families. It is not uncommon to see this in the urban informal sector as well. Thus the neighborhood store may be run by a joint family, with revenues divided among siblings. A cab driver might share his driving with a friend. The bus conductor of a crowded bus might sublease part of his ticket-collecting duties to a teen aged nephew. There are aspects of mutual insurance in these relationships that also have value. To these Lewis adds social prestige and charity:

> Social prestige requires people to have servants, and the grand seigneur may have to keep a whole army of retainers who are really little more than a burden upon his purse. This is found not only in domestic service, but in every sector of employment. Most businesses in underdeveloped countries employ a large number of "messengers," whose contribution is almost negligible; you see them sitting outside office doors, or hanging around in the courtyard. And even in the severest slump the agricultural or commercial employer is expected to keep his labour force somehow or other—it would be immoral to turn them out, for how would they eat, in countries where the only form of unemployment assistance is the charity of relatives? So it comes about that even in the sectors where people are working for wages, and above all the domestic sector, marginal productivity may be negligible or even zero.

Lewis was not alone in asserting the existence of surplus labor. Already in the 1940s there were

claims that large numbers of able-bodied people were in surplus in the agricultural sectors of eastern and southeastern Europe and in the Soviet Union. Rosenstein-Rodan [1943] and Nurkse [1953] were among those who held this view.[10] These writers realized that the presence of redundant labor in the agricultural sector meant that the population surplus could be transferred out of the agricultural sector with no loss in agricultural output. Surplus labor is, therefore, a supply of labor that, given the preponderance of the agricultural sector in less developed economies, is likely to be of major quantitative importance in the development process of less developed economies. This is the classical tradition that Lewis inherited.

Surplus Labor: A Natural Experiment

Economic development with unlimited supplies of labora–Arthur Lewis's phrase was provocative enough to instigate a flood of research on the existence of disguised unemployment in agriculture. We have seen that disguised unemployment refers to a situation where marginal product is less than the going wage. However, Lewis had something stronger in mind: the possibility that "there are large sectors of the economy where the marginal productivity of labour is negligible, zero, or even negative." Although this assertion is not strictly necessary for the Lewis framework, it suggests the existence of a free resource in agriculture: labor.

One of the most interesting early studies on surplus labor was that of Schultz [1964], who studied the effect of the influenza pandemic in India (1918–19). The epidemic was sudden; the death rate reached a peak within weeks and then diminished rapidly There were a large number of deaths. Schultz chose two years, one before 1916–17 and one after 1919–20, when weather conditions were approximately equal. He then estimated the existence of surplus labor by comparing the reduction in acreage sown with the reduction in the labor force.

His findings were that, as a result of the epidemic, the agricultural population fell by 8.3% over these two years. He made the following observation [1964, p. 67]:

> The area sown in 1919–20 was, however, 10 million acres below, or 3.8% less than that of the base year 1916–17. In general, the provinces of India that had the highest death rates attributed to the epidemic also had the largest percentage decline in acreage sown to crops. It would be hard to find any support in these data for the doctrine that a part of the labor force in agriculture in India at the time of the epidemic had a marginal product of zero.

According to Schultz, therefore, surplus labor did not exist in India at the time of the epidemic.

This study is interesting in its use of a "natural experiment" to address an economic question. However, was the experiment "natural" enough? Consider the *pattern* of population decrease. Influenza epidemics attack entire households and the epidemic of 1918–19 was no exception. Thus entire plots of land were left uncultivated during this period of time. As Sen [1967] pointed out in his comment on the Schultz study, if land is not redistributed following the labor removal, it is not surprising that the sown acreage decreases. In this short span of time, this redistribution could not have taken place. Contrast this with the view implicit in the theories of Lewis and others: that is, in each family unit there is a surplus of labor. The pattern of labor removal from agriculture critically affects the fate of agrarian output.

Two extensions of the surplus labor concept.

Two extensions of the surplus labor concept are of some interest. First, note that surplus labor as defined in the previous section is purely a *technological* concept: there is simply too much labor relative to land, or more generally, too many people relative to other inputs of production, so that individuals are in surplus relative to production possibilities: remove them to other activities and output will not change because the additional labor power is of no use at all: the marginal product of labor is literally zero.

The inability of labor to add *anything* to output was criticized by several economists as an unrealistic phenomenon (see, for example, the box on the influenza epidemic in India). For instance,

Viner [1957] writes:

> I find it impossible to conceive a farm of any kind on which, other factors of production being held constant in quantity, and even in form as well, it would not be possible, by known methods, to obtain some addition to the crop by using additional labor in more careful selection and planting of the seed, more intensive weeding, cultivation, thinning, and mulching, more painstaking harvesting, gleaning and cleaning of the crop.

Thus, the narrow technological concept of surplus labor may be inapplicable except in special situations. Is there a broader yet still useful specification? This raises the question of just why we are interested in the concept of surplus labor. This question can be answered from two viewpoints, each of which leads to a useful extension of the concept.

(1) *Disguised unemployment*. First of all, there is the question of efficient *allocation*. If marginal product is zero in some activity and positive in some other activity, there are efficiency gains to be had in switching resources away from the former activity to the latter. Why doesn't the market, left to its own devices, spontaneously accomplish this switch? The reason is that the zero marginal-product activity is usually characterized by a payment system that is not based (and cannot be based) on marginal product. As we saw in the previous section, it is often based on income sharing, which means that people in such activities receive the *average* product, which is surely positive (see Figure 10.1 to verify this). As long as average product in this activity is equal to marginal product in activities elsewhere, no individual would be interested in making the switch (although see the problem on family-based migration at the end of this chapter).

This line of reasoning indicates that if efficient allocation of resources is the underlying objective that motivates the concept of surplus labor, the concept is surely way too strong. It isn't necessary that the marginal product in the traditional activity be *exactly* zero. As long as the marginal product is *lower* than in activities elsewhere, there are gains to be had from a reallocation of (labor) resources. If we suppose that there is a capitalist sector elsewhere that does pay according to marginal product, then the economy will exhibit a wage rate (for unskilled labor) that is a true measure of the marginal product elsewhere, and there will be efficiency gains available as long as the marginal product on the traditional activity is *less than the wage*, whether it is zero or not. This extended concept is known as *disguised unemployment*. The amount of disguised unemployment may be measured roughly by the difference between the existing labor input in the traditional activity and the labor input that sets marginal product equal to the wage.[11]

Now surplus labor may be viewed as a special case of disguised unemployment, but the generalization greatly increases the value of the narrower concept.

(2) *Surplus labor versus surplus laborers*. Our next extension takes us back to the narrower concept of surplus labor and again starts from the Viner criticism. This extension is motivated by a second possible answer to the question, Why are we interested in surplus labor? The answer is that once labor is removed from agricultural pursuits, the issue of maintaining an adequate surplus of food in the economy becomes very important. Recall that in economies with limited international trade in food, an internally produced agricultural surplus is necessary if an industrial sector is to be supported.[12] After all, workers in the industrial sector demand food in the marketplace, and if such food isn't forthcoming, the resulting inflationary spiral can destroy the prospects of industrialization.[13]

So in this view, the question of *maintaining* agricultural output (or at least not letting it fall by too much) is of independent interest, quite apart from the efficiency calculus that underlies marginal

product comparisons.[14]

This raises a new point. We remove *laborers*, not labor. The meaning of this cryptic sentence is that the *remaining* laborers in the traditional activity typically adjust *their* labor input once some laborers are removed (say, through rural–urban migration). If there is an increase in work effort on the part of the remaining laborers, total output may not fall even though the marginal product of labor is zero. This argument was originally made by Sen [1966] (see also Takagi [1978]).

Why would the members of a family farm raise their work hours to compensate for the departure of some of their compatriots? The answer depends on the alternative uses of labor applied to the farm. Such uses may involve leisure or working elsewhere part time. If the marginal product of such alternatives (which is exactly the marginal cost of working on the farm) rises as more and more labor is drawn away, then indeed there will not be full compensation for the lost workers, but there will be *some*. In the extreme case where the marginal cost of labor is constant, there will be full compensation for the lost workers. Even though the marginal product of labor is not zero, the farm will exhibit a surplus of *laborers*, in the sense that as laborers are removed from the farm, output will not fall.[15]

The reason for all this is very simple. Efficient resource allocation on the family farm requires that the value of marginal product of effort be equal to marginal cost. The first panel of Figure 10.2 shows how this familiar calculation yields a total labor input for the family. Now the point is just this: *if the marginal cost of family labor is constant*, then the total cost is just a straight line as shown in the figure, and total family input is determined *independently* of family size! This is just another way of saying that the removal of some members has no effect on total output (but note that the marginal product of labor *is* positive). The insight is that the removal of family labor has no effect on marginal cost if the marginal cost of labor for each family member is constant (and identical across members).

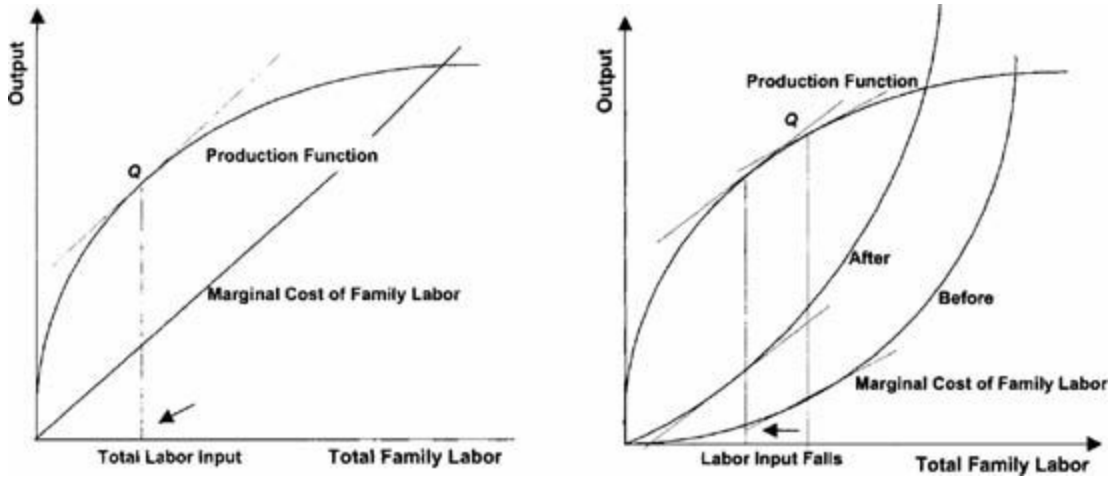

Figure 10.2. Surplus labor and surplus laborers.

Of course this state of affairs is altered if marginal cost increases with effort. The second panel of Figure 10.2 shows this scenario. Then the total cost curve of the family is shifted upward as laborers are removed (provision of the same level of *family* effort as before now involves a higher marginal cost). Output will fall, but even so, the distinction between labor *effort* and the *number* of laborers is a point that is worth appreciating.

Economic development and the agricultural surplus

Armed with these forgoing concepts, we can describe the interplay between rural and urban sectors as envisaged by Lewis and later extended by Ranis and Fei [1961].

In the traditional agricultural sector there is disguised unemployment, perhaps even a core of surplus labor, and the wage rate is given by income sharing. The industrial sector is capitalistic. Economic development proceeds by the transfer of labor from agriculture to industry and *the simultaneous transfer of surplus food-grain production, which sustains that part of the labor force engaged in nonagricultural activity*.

Figure 10.3, which is based on Ranis and Fei [1961], provides a schematic description of how the labor force and the corresponding agricultural surplus is transferred in the process of development. In each panel of the diagram, the industrial labor force is read from left to right, whereas the agricultural labor force is read from right to left. Assume for simplicity that the total labor force is divided between agriculture and industry. Then the width of the panels corresponds to the entire labor force in the economy.

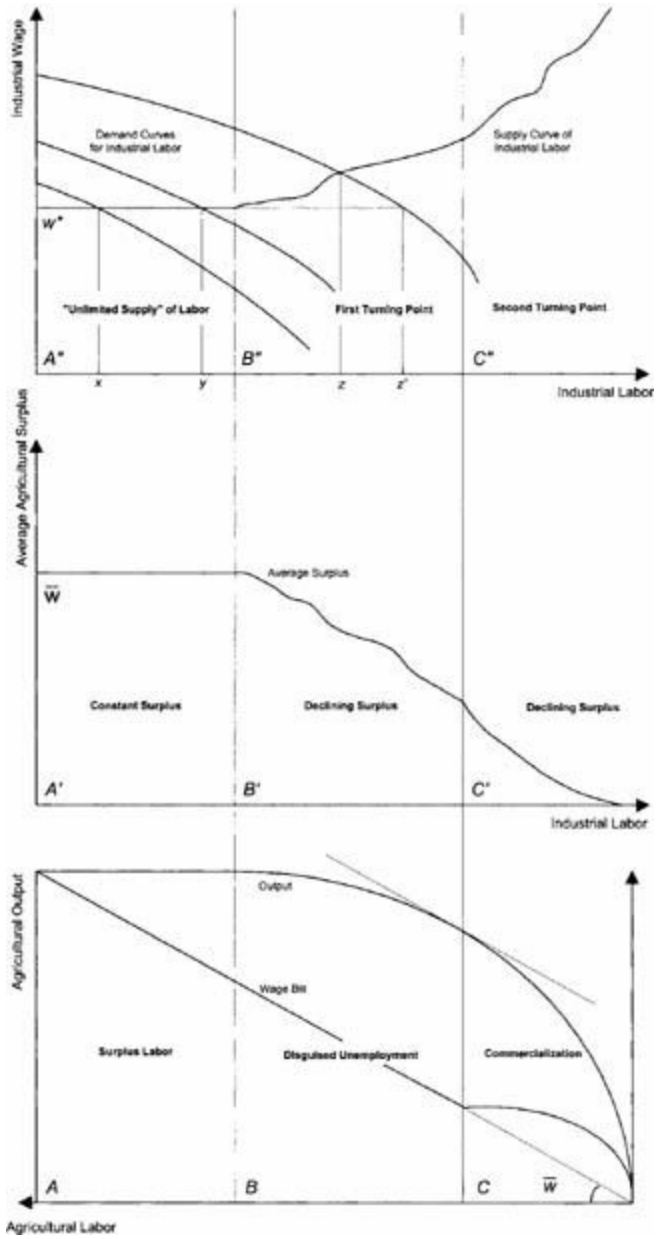

Figure 10.3. The Lewis–Ranis–Fei model.

It is best to read this figure from the bottom up. In the lowest panel, we have drawn a typical agricultural production function, except that it is drawn from right to left to reflect the way we read

agricultural labor on the panels. The production function levels off just as in Figure 10.1, and there is a phase of surplus labor provided that the entire labor force is in agriculture. This is shown by the segment *AB* on the diagram. Moreover, if "wages" in this sector are decided by income sharing, then the average wage is just $\bar{w}$, which is proportional to the angle shown in this panel. This turns out to be the wage in the nascent industrial sector, as we will presently see. Thus the segment *BC* has no surplus labor, but does exhibit disguised unemployment, because the marginal product of labor in agriculture is less than the wage $\bar{w}$ for labor inputs in this segment. To the right of *C*, the phase of disguised unemployment ends.

What we will do in the next few paragraphs is carry out a thought experiment. Starting with a situation where the entire labor force is in agriculture, we will trace the consequences of transferring labor to the industrial sector. In particular, we will describe the minimum cost of hiring transferred labor in industry. This will give us something like a "supply curve" of labor to industry.[16] The topmost panel of Figure 10.3 shows this supply curve.

Begin with the entire labor force in agriculture. Suppose we decrease this by a small amount, so that we are still in the surplus labor phase. Then the total wage bill in agriculture falls along the diagonal straight line in the lowest panel, *provided that the wage in agriculture does not rise*. At the same time output does not fall, because we are in the surplus labor phase. An *agricultural surplus* therefore opens up; this is given by the vertical gap between the production function and the wage bill line. If we divide this surplus by the number of transferred workers, then we obtain the *average* agricultural surplus, where we are taking the average or per capita surplus amount relative to the transferred workers. It is easy to see that the average agricultural surplus in the surplus labor phase must be exactly $\bar{w}$. In the imagery of Nurkse [1953], it is as if each laborer simply moved out of agriculture with his food parcel on his back.

The average surplus is depicted in the middle panel of Figure 10.3. As discussed in the previous paragraph, the average surplus is unchanging in the surplus labor phase and is equal to $\bar{w}$. This is depicted as a horizontal line of height $\bar{w}$.

Now, observe that each transferred laborer in industry must be able to buy back his food parcel, because he is no longer employed in the agricultural sector but in industry. The minimum industrial wage required to do this is depicted in the topmost panel. Because the industrial wage is described in units of industrial goods, we must multiply $\bar{w}$ by the relative price, or the *terms of trade*, between agriculture and industry to arrive at the required minimum industrial wage. This is shown by the value w^* in the topmost panel. In the surplus labor phase, the minimum industrial wage required for compensation does not change, because the average agricultural surplus is not changing.

This creates a perfectly elastic supply of labor in the surplus labor phase, which is depicted as a horizontal line emanating from the point w^* in the topmost panel. This is the zone where it is possible to have economic development with "unlimited supplies" of labor: an expansion in the industrial sector does not drive up the wage rate.

As we now move into the phase of disguised unemployment, the average agricultural surplus begins to decline. This is because total output in the agricultural sector begins to fall, while those who are still there continue to consume the same amount per capita. This is shown by the decreasing line in the zone $B'C'$ of the middle panel.

Now what is the minimum wage in industry? Well, if the wage is still w^* as before, transferred workers will *not* be able to compensate themselves for the move, because it is physically not possible for each of them to buy $\bar{w}$ units of food. This is because the average agricultural surplus has

fallen below $\bar{w}$. The immediate effect of this is that food prices start to rise: the terms of trade between rural and urban sectors begin to move against industry. To compensate for this price effect, the industrial wage must rise.

However, rising wages that cannot solve the problem. No matter how much the industrial wage rises, it is not possible for workers to buy their old food parcel back, because there simply is not enough to go around. The only way that compensation can be achieved, then, is for industrial workers to consume a mix of agricultural and industrial products, the latter compensating them for the loss of the former.

Is such compensation possible? It depends on how close the traditional wage is to minimum subsistence. The closer is $\bar{w}$ to the minimum subsistence level, the larger is the compensation required and the steeper is the increase in the required industrial wage. Conversely, the easier it is to substitute industrial consumption for agricultural consumption, the softer is the necessary increase in the compensatory industrial wage.

Ranis and Fei [1961] referred to this phase, where the supply wage of labor tilts upward, as the "first turning point."

Continue the transfer of labor until we reach the point *C*, where the disguised unemployment phase comes to an end. At this point, the marginal product of labor begins to exceed the traditionally given wage rate. It then becomes profitable to actively bid for labor, because the additional contribution of labor in agricultural production exceeds the cost of hiring labor. This situation means that the wage in agriculture rises. One implication is that the wage bill falls more slowly than it did before along the diagonal line of the lowest panel. It now traces the curve after *C*, because wages rise as the agricultural labor force decreases.

This phenomenon, which we may think of as the *commercialization of agriculture*, is associated with an even sharper decrease in the average agricultural surplus. In terms of the topmost panel, this phenomenon induces a second turning point in the industrial wage. Not only must the wage compensate for a declining agricultural surplus and a movement of the terms of trade against industry, it must now compensate workers for a higher income foregone in the agricultural sector, and this creates a still sharper upward movement in the industrial wage rate. This completes the construction of the supply curve.

We are now ready to see how the model works. The industrial sector demands labor for production. The topmost panel of Figure 10.3 shows a family of such demand curves; begin with the lowest one. This demand for labor induces a situation where the amount of industrial labor is *x*, hired at a wage of *w**. With industrial production, profits are realized, parts of which are plowed back as extra capital in the industrial sector. This part of the story is very similar to the Harrod–Domar and Solow paradigms. The expansion of capital means that the demand for labor rises (shift to the second demand curve in the topmost panel). Because the economy is in the surplus labor phase, this labor is forthcoming from the traditional sector with no increase in the wage, as we already discussed. Industrial employment is now at point *y*. However, with further investment, the demand curve for labor shifts to a point where the compensatory wage must rise. Employment rises to *z*. However, it would have risen even further (to point *z*′) had the turning point not occurred. The fall in the agricultural surplus chokes off industrial employment to some extent, because it raises the costs of hiring industrial labor.

Our account is now complete. Capital accumulation in the industrial sector is the engine of growth. More capital means a greater demand for labor, which in turn induces greater rural–urban

migration. As development proceeds, the terms of trade gradually turn against industry: food prices rise because a smaller number of farmers must support a greater number of nonagricultural workers. The rise in the price of food causes an increase in the industrial wage rate. The pace of development is driven by the accumulation of capital, but is limited by the ability of the economy to produce a surplus of food.

Despite all the qualifications and imperfections that we subsequently note, this is the heart of the Lewis–Ranis–Fei story: development proceeds via the *joint* transfer of labor and agricultural surplus from the "traditional" agricultural sector to the "modern" industrial sector. The ability to expand the industrial sector is determined in part by production conditions in agriculture. In particular, without the existence of a surplus in the latter sector, it is difficult to create growth in the former sector.

Policy issues

As we will see in this section, the Lewis model of economic development needs several qualifications: the details of the framework should not be taken literally, but the model helps us organize our thoughts along certain lines and throws light on different policies that can be adopted in the course of development.

(1) *Agricultural taxation*. The assumption that the wage rate in agriculture is fixed until the phase of commercialization is reached is strong. Take another look at Figure 10.3. As labor is progressively withdrawn from the agricultural sector, there is more income left for the *remaining* workers to share. Why don't they share it and raise the wage upward from $\bar{w}$? If they do, then there are two effects: (i) the agricultural surplus available to industry is reduced and (ii) the compensating wage paid to transferred workers must rise *even* in the phase of surplus labor. Even if farmers willingly market the freed surplus (if the price is right), effect (ii) remains, and the supply curve of labor to industry fails to be perfectly elastic.

This observation uncovers a problematic issue in the Lewis–Ranis–Fei theory: industry has a vested interest in taxing agriculture, because it is only through taxation that the incomes of family farmers stay low as labor is withdrawn (as shown in Figure 10.3). Indeed, the model implicitly assumes that family farms are being taxed as labor is withdrawn, thus keeping per capita income constant in agriculture and allowing the supply curve of labor to industry to remain perfectly elastic. In contrast, if taxes are not imposed, agricultural incomes will rise—surplus labor or not—and industrial wages must rise to keep migration incentives alive. The rise in industrial wages chokes off industrial profit, and this is the source of the tension between agriculture and industry.

Who would support agricultural taxation? Industrialists would: such taxation keeps agricultural incomes down and this reduces industrial wages. Industrial workers would not support such taxes, not necessarily out of solidarity for farmers, but because it would raise migration and increase the competition for jobs. Small farmers certainly oppose the policy: they are the ones who are taxed! As for large landowners, the situation is more complicated. Their response depends on the intensity of the taxation and the ease with which some of it can be evaded. Certainly they prefer no taxation to taxation, but taxation also has the effect of driving down rural wages. To the extent that large landowners are significant employers of labor, this effect may be beneficial.

Thus we see that a policy of agrarian taxation may run into severe political problems, even though it may have a beneficial effect on industrial growth. This is especially true of countries with a large fraction of the population in agriculture: the governments of such countries often draw on farmers for

political support.

There is a postscript to all this. We've assumed in the preceding discussion that agricultural taxation does assist in industrial development, but this may be a shortsighted view. There are longer-term considerations. If farmers believe that greater output will be systematically taxed away, then they will lose all incentives to create, improve, or maintain productive inputs such as irrigation facilities or soil quality. There will be underinvestment in agriculture, and this underinvestment will have repercussions for the availability of *future* surplus.

The tension between a static or short-term view of agriculture as a sector to be squeezed for *current* surplus and a dynamic or long-term view of agriculture as a sector to be invested in and encouraged for the generation of *future* surplus, represents an economic issue of great importance. Walking this tightrope is no easy task and can have enormous political connotations.

Agriculture versus Industry in the New Soviet Union

Of the newly created Soviet Union of the 1920s, Dobb [1966, p. 208] wrote: "The rate at which agricultural production could expand and afford a growing supply of raw materials for industry *and foodstuffs for industrial workers* appeared as the crucial question in economic discussion in the second half of the decade: an issue upon which all other hopes and possibilities rested" (Italics added for emphasis). It wasn't that agricultural investment was nonexistent: large imports of tractors occurred in the 1920s, the majority of these going to the collective or state farms. However, the *marketed* surplus of food grain continued to be abysmally low: in 1925–26, while the total agricultural land area under cultivation was close to that of pre-war sown acreage, the surplus on the market was only around 70% of the pre-war amount. With the land reform of 1917, land was now more equally distributed and it was clear that the newly endowed peasantry were eating more and selling less.

There was no end of fretting and fuming about the problem of agriculture. Oddly enough, some intellectual hardliners centered around Trotsky in the mid-1920s continued to view "further development as only possible in the existing situation in Russia if industry were to expand at the *expense* of the peasantry" (Dobb [1966, p. 183]). This view contrasted with the relatively moderate earlier views of Lenin, who regarded agriculture as a sector to be (at least provisionally) treated as complementary to the development process, creating and retaining a *smytchka*, or bond, between the peasant and the industrial worker. In the mid-1920s the government embarked on a program of food price stabilization that limited competition among purchasers of grain and required all private traders to register with the government. State collecting organizations, which set price limits on purchases, came to occupy a larger share of the grain trade. The policy of price stabilization was successful: food prices rose only by 2% between October 1926 and March 1927, but grain purchases collapsed. Thus grain to the towns was sharply limited, as were grain exports in exchange for needed imports of industrial inputs.

In 1928, Stalin described the situation thus[17]:

> On January 1st of this year there was a deficit of 128 million poods of grain as compared with last year . . . . What was to be done to make up the lost ground? It was necessary first of all to strike hard at the *kulaks* (rich peasants) and the speculators . . . . Secondly, it was necessary to pour the maximum amount of goods into the grain regions . . . the measures taken were effective, and by the end of March we had collected 275 million poods of grain . . . . [But] from April to June we were unable to collect even 100 million poods . . . . Hence the second relapse into emergency measures, administrative arbitrariness, violation of revolutionary laws, raids on peasant houses, illegal searches, and so forth, which affected the political conditions of the country and created a menace to the *smytchka* between the workers and the peasants.

This was a two-edged sword: grain was wanted, but despite the intended pouring of the "maximum amount of goods into the grain regions," incentives to farmers were absolutely minimal. Short-run gains in food collection were met only with longer-run resistance. Finally, the historic decision was taken to embark on a massive state collectivization of agriculture, a story in itself. Other governments, for whom large-scale collectivization is not a desirable option, will have to solve this problem differently, but the problem *is* there all the same.

---

(2) *Agricultural pricing policy*. Agricultural taxation is not the only way to extract a food surplus. As we have seen, this policy has several problems: informational (can the government verify how

much is produced or, indeed, how much land a farmer owns?), political (farmers are a powerful voting bloc), and economic (taxation creates long-run disincentives to invest in agriculture, which lowers future surplus). Food can be coaxed instead of coerced into the market by lucrative prices for output or subsidies to agricultural inputs. To be sure, a higher output price is a more costly route as far as industrial capitalists are concerned: to them, every concession to agriculture has repercussions for the industrial wage, either directly (as agrarian incomes rise) or indirectly (via a higher relative price of food).

The typical price support program consists in offering guaranteed procurement prices at which the government stands ready to buy food grain. The idea, of course, is to increase the marketed surplus of grain. At the same time, governments often are unwilling to pass on these prices to urban consumers, partly because these consumers are typically incensed by higher prices (never mind if their salaries are adjusted as a result) and partly because of the effect on the industrial wage. Thus price support programs are usually accompanied by a subsidy to urban consumers: the procured food is sold at or below market prices by the government. Of course, someone has to pay for this subsidy, and it usually comes out of the government budget.

An alternative to high procurement prices is the policy of keeping *input* prices low. Water, electricity, and fertilizer may be supplied free or at reduced prices. In India, the fertilizer subsidy *alone* accounted for an enormous chunk of the government budget: in the late 1980s, the size of the subsidy exceeded total revenues from noncorporate income taxes!

Yet another option, which has the dubious advantage of being less transparent, is to maintain an overvalued exchange rate (see Chapter 17 for much more on this). The overvaluation is kept in place by tariff or quota-based restrictions on imports. To be sure, an overvalued exchange rate serves many purposes, and this is not the place to discuss them, but one implication is that the prices of exports are kept artificially low (in terms of the domestic currency). If the country is a food exporter, this policy has the effect of discouraging food exports and shifting food sales into the domestic market. The policy has sufficient opacity about it—farmers may not be aware that the exchange rate is overvalued and reduces their export earnings—and it has the desired impact of making food available to urban consumers without a price rise.

But it is important to note that policies such as export restrictions, opaque though they may well be, do involve serious efficiency losses and can run the country into severe balance of payments problems. Moreover, if there is a sudden liberalization of the exchange rate, the opacity of the policy may just as quickly disappear and be replaced with a (now more transparent) policy of an outright ban on food exports. The Indian government has faced this dilemma in the 1990s.

Perhaps there is no way out but to endure the short-run rise in food prices, and the inherent shift in national income in favor of agriculture, with the understanding that this will be good for the entire economy in the longer run. Certainly in countries where (to begin with) there were artificial controls to selling on the market, the simple policy of allowing market-determined sales had significant effects. Contrast Russian and Chinese reforms. Russian agriculture never really recovered from the collectivization programs of the 1930s: it is a sector that neither Gorbachev nor Yeltsin seriously touched in any way. Bureaucratic collectives meant low productivity and low output, and Russians have had to import food, particularly from the rest of the Soviet bloc (in exchange for armaments and other heavy-industry products). The collapse of the Soviet bloc, the lowered demand for armaments, and the need to now pay in hard currency all led to a food shortage and high inflation.

China stands in sharp contrast. They began post-1978 reforms with agriculture. Land was given to

farmers (on long leases) under the new "household responsibility system" and collective farms were disbanded. Farmers were allowed to sell on the market and market prices were unregulated. This plan actually implied two conceptual departures from earlier policy: the introduction of price incentives and the abandonment of a regional self- sufficiency-in-grain program, which had been in place since the Great Leap Forward in 1958. This policy actually required that each region plant crops for self-sufficiency, regardless of whether or not they were suited to grow those crops.

The resulting gain in agricultural productivity (and output) was impressive. In the 1970s, TFP in agriculture was 20–30% lower than in 1952, the year before collectivization. Just a few years after the reforms, TFP was back to the 1952 level and continued to grow steadily through the 1980s (Wen [1993]). Agricultural output expanded by over 40% between 1978 and 1984 (see McMillan, Whalley, and Zhu [1989]).

How much of the productivity growth can be traced to the new surge in price incentives and how much to the abandonment of the self-sufficiency policy? It appears that the former accounts for almost all of the productivity gains according to a decomposition analysis carried out by Lin and Wen [1995]. The self-sufficiency policy, macabre though it was, apparently was not the main culprit in the earlier productivity stagnation.

Thus the initial price increase promoted much of the output gains. In the longer run, this meant that China was able to avoid the problem of *continuously* rising food prices or food shortages, which in turn kept industrial wages competitive and fuelled industrial growth.

The question is whether the *initial* rise in food prices that comes from a pro-farmer pricing policy can be politically tolerated. In the case of China led by Deng Xiao Ping, the base of the Chinese Communist party was formed by the farmers (who supported the party through the Tiananmen episode).

However, there are limits to where price policy can take us unless we start from an unusually repressed base. Although Lipton's [1968] point that there is often "urban bias in rural planning" is certainly well taken, over- inflated food prices can only retard industrial development. There are other gaps to fill: as we shall see in the chapters to come, agricultural growth is often limited by access barriers to capital and credit, as well as intrinsic disincentives that arise from agrarian contracts (which occur primarily because such contracts need to play the double role of assuring *some* incentives and *some* insurance). Land reform, credit expansion, and infrastructural investment all go a long way to assuring agricultural and industrial growth. There would be no need to do this if the markets to ensure such developments functioned smoothly, but as we will see, they do not. A fuller consideration of these issues is postponed to Chapters 12–15.

10.3. Rural–urban migration

10.3.1. Introduction

The Lewis model tells us that agricultural surpluses and labor must be transferred in tandem for industrial development to begin. But as we have already noted, labor moves from one sector to another in obedience to its own wishes and objectives. To the extent that these objectives may be out of line with social goals or policies, we might have over- or undermigration to the cities. The purpose of this section is to discuss patterns of rural–urban migration.

The classic theory of rural–urban migration is based on Harris and Todaro [1970]. We start by talking about the basic theory, and then extend the framework in a number of different directions.

The main idea of the Harris–Todaro model is that the formal urban sector pays a high wage to workers and it is this high wage that creates urban unemployment (the mechanism will be examined in what follows). Many reasons might be provided for the phenomenon of an overly high urban wage. The sector may be unionized and subject to collective bargaining over wages, whereas other sectors of the economy are not remotely as organized, so that wages are more flexible in those sectors. In addition, the urban formal sector is often treated as the showcase of government policy, so that minimum wage laws, pension schemes, unemployment benefits, day care, and other facilities may be required by law. These provisions may not raise the wage directly, but amounts to the same thing, because such forms of compensation raise worker utility.

Finally, it may well be the case that firms in the urban formal sector *deliberately* pay wages that exceed levels found elsewhere so they can hire workers of the best quality and fire inferior workers after their quality is revealed. Even if there are no quality differences across workers, "supermarket" wages may still be paid if firms wish to elicit effort from their workers. The idea is that if such effort is not forthcoming, then workers are fired and returned to the informal or rural labor market. The threat of being fired induces higher effort. Of course, being fired can carry no threat if the wage package is no different from what the worker can get elsewhere; in other words, to make being fired a serious punishment, the firm must "buy the threat" by paying a higher-than-normal wage. We will see more of this sort of contract in Chapter 13.

In contrast to the high wages paid in the formal urban sector, the informal urban sector and the rural sector have low wages that fluctuate according to supply and demand considerations. There is no unionization here and government policy is difficult to implement. Moreover, if the bulk of labor is family labor (as it is in much of the urban informal businesses, as well as in rural family farms) or if the bulk of labor effort is readily monitorable (as in harvest labor), then there will be little incentive for employers in these sectors to pay higher wages as a potential threat. Even if there were such an incentive, the net effect is unlikely to dominate the huge premiums that are paid in the urban formal sector.

Migration in the Harris–Todaro model is then viewed as a response to the significant wage gap that prevails between the two sectors. Of course, not everyone can be absorbed into the formal sector at these high wages: some people are unlucky and fail to find a job, in which case they turn to the urban informal sector for some meager sustenance. Thus the migration decision is akin to leaving behind a relatively sure thing (employment as an agricultural labor or on the family farm) for the great uncertainty of employment as a formal laborer. Those who fail in this quest join the queue of the unemployed, perhaps in disguised form in the informal sector. Thus the urban informal sector (in the Harris–Todaro view) contains the failed aspirants to the formal sector dream—the lottery tickets that didn't win.

10.3.2. The basic model

We begin by assuming that there are only two sectors in the economy: a rural sector and a formal urban sector. Solely for the purpose of setting a benchmark, we assume that wages in *both* sectors are fully flexible. Later, we introduce rigidity in the urban formal wage.

Figure 10.4 captures the basic story. The width of the horizontal axis is the entire labor force in

the economy. The labor force is divided between the agricultural sector, which we denote by A, and the formal urban sector, which we denote by F. The left axis of the figure records various formal wages in the urban sector, whereas the right axis records agricultural wages. The curve AB may be thought of as a demand curve for labor in the urban formal sector: like most demand curves, it is downward sloping, so that more labor can be absorbed in the sector only at a lower wage. Likewise, the curve CD captures the absorption of labor in agriculture (you can think of it as a demand curve as well, but there are other interpretations that we will discuss presently). Just as in the urban sector, more agricultural labor typically can be absorbed only at a lower wage.

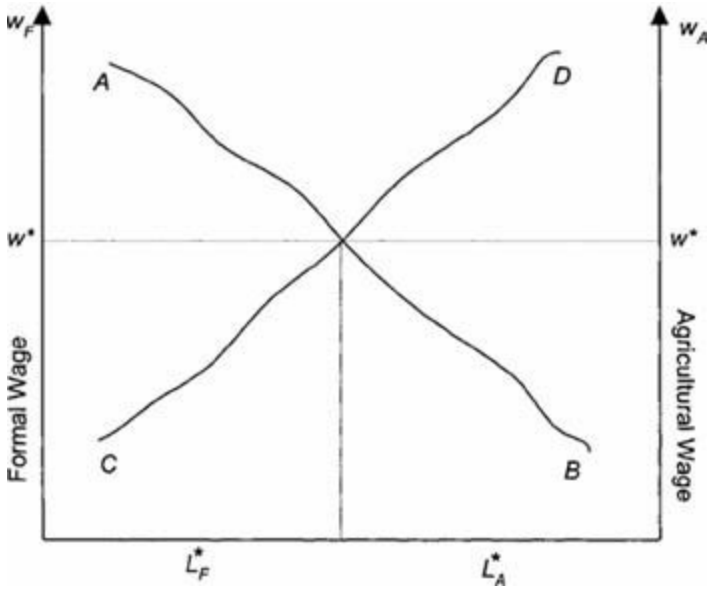

Figure 10.4. Market equilibrium with flexible wages.

It is now easy enough to combine these two "absorption curves" to analyze the equilibrium of this simple economy. To alleviate persistent migration between one sector and the other, the wages in the two sectors must be equalized.[18] This equalization occurs at the intersection of the curves AB and CD, and we can read the equilibrium wage rate and intersectoral allocation of labor from this intersection. Figure 10.4 records the equilibrium wage rate in this case as w^*, with L^*_A individuals in the agrarian sector and L^*_F individuals in the urban sector.

10.3.3. Floors on formal wages and the Harris–Todaro equilibrium

What is wrong with the preceding argument? Not much, it would appear: what we have in Figure 10.4 and the accompanying discussion is a textbook case of competitive equilibrium. The problem is that it assumes that the urban wage rate is perfectly flexible. We have already seen that this is not the case. Indeed, it is not at all unreasonable to argue that the formal urban wage is too high for market clearing to occur as described by Figure 10.4. We have provided several reasons for this. Now let us see what the implications are. In terms of our simple model, then, imagine that the wage rate in the formal sector is fixed at too high a level for market equilibrium w^* to occur. Figure 10.5 captures this situation by drawing the minimum formal wage, $\bar{w}$, at a level that lies above the intersection of the two absorption curves. It follows that private-sector formal firms will hire no more than the amount $\bar{L}_F$ of labor at this wage. Where do the remainder go?

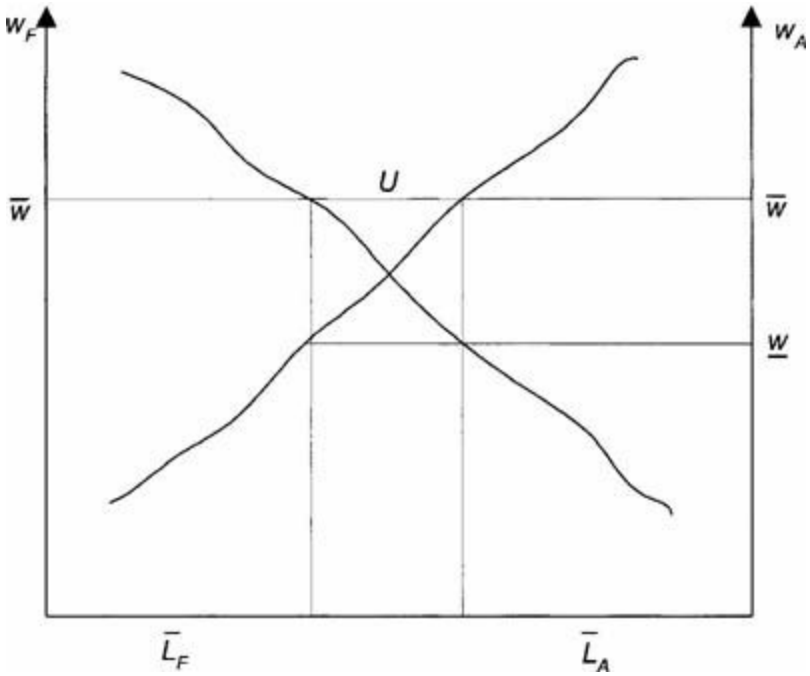

Figure 10.5. A floor on the formal wage.

One possibility is that all the remaining individuals are employed in the agricultural sector. In that case, Figure 10.5 tells us that the wage in the agricultural sector must drop to $\underline{w}$. Now step back and look at the final outcome. In both sectors we have full employment, so that no *individual* job seeker needs to fear unemployment if she looks for a job in either sector. Nonetheless, the wages, $\bar{w}$ and $\underline{w}$, are different. This *cannot* be an equilibrium state for the economy, because with full employment in both sectors, workers will wish to migrate to the sector with the higher wage.

On the other hand, simply imposing the equality of wages across the two sectors is problematic as well. Try it. Figure 10.5 then reveals that only an amount $\bar{L}_A$ can be soaked up in the agricultural sector. If the formal and the agricultural sectors are the only two sectors in the economy, we must have a pool of unemployed people. (In the figure, U denotes the size of the resulting unemployed pool.) This cannot be an equilibrium state either. Given that agriculture has flexible wages, the unemployed workers cannot be physically located in agriculture. If they were, they would simply pour into that labor market and consequently drive the wages down. Therefore, they must be located in the urban sector. Now we have a situation in which these workers rationally migrate to the urban formal sector, even though the wages there are the same as those in agriculture *and* there is significant risk of unemployment. Under no stretch of the imagination can such a state of affairs be described as an equilibrium, even from an ex ante point of view.

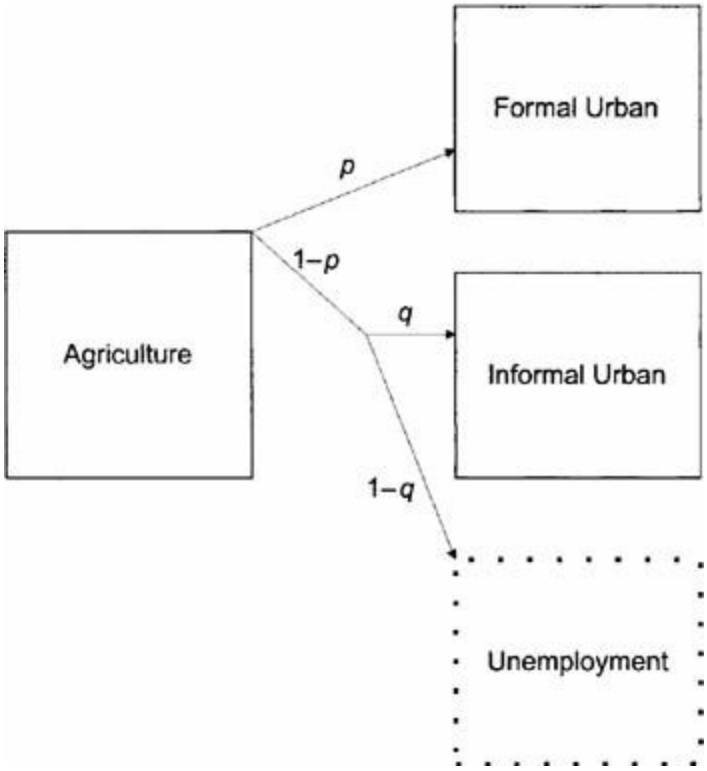

Figure 10.6. Options open to a potential migrant.

Although these alternatives do not work as descriptions of the final outcome, they are suggestive of what the equilibrium might look like. The main idea is that potential migrants choose between a relatively safe (though possibly unpleasant) option, which is to stay in the agricultural sector, and the gamble of moving to the urban sector, where a high-paying formal job may or may not be attainable. In turn, the probability of getting such a job is determined by the ratio of formal job seekers to available formal jobs. Those who do not get a job might be referred to as the unemployed, but this description is not entirely accurate (and here is where the informal sector comes in). Frustrated formal job seekers may enter the informal sector, where jobs or businesses are easy enough to find but pay (relatively speaking) a pittance. Figure 10.6 schematically captures the gamble that is involved.

In this diagram, there are two sets of boxes. The left set is a single box: agriculture, with its wage w_A.[19] The right set describes the various options open in the urban sector, together with the probabilities of access. First, there is the formal sector at some high wage $\bar{w}$. The probability of obtaining such a job depends on the ratio of vacancies to job seekers. Denote this by p. Next, there is the informal urban sector, in which our migrant can get absorbed in the event that no formal job is forthcoming. Denote the wage rate in the informal sector by w_I and assume that it is fixed regardless of the number of people in that sector.

What is needed is a calculation of the *expected value* of these two risky options. The expected value is calculated in the usual way: weigh each outcome by its probability of occurrence and add up over all outcomes. Thus the expected wage in the urban sector is neither $\bar{w}$ nor w_I, but the combination $p\bar{w} + (1 - p)\, w_I$. It is this *expected* wage that is compared to the wage in the agricultural sector.

In the preceding calculation, we implicitly assumed that there are only two options in the urban sector: formal or informal employment. However, once we understand how the calculation is carried out, it is easy enough to expand the urban sector to include more possibilities. For instance, it is reasonable to suppose that not everyone is guaranteed to receive even the lower income w_I in the

informal sector. It may be that some individuals do not get any employment at all, so that they are "openly" unemployed. This additional option is displayed by the dotted box in Figure 10.6, with associated wages equal to zero.

How can we now compute expected values? We need to know the probability of getting an informal sector job, conditional on having been turned away from the formal sector: denote this by q. Thus after being turned away from the formal sector, the migrant manages to join the informal sector with probability q and remains openly unemployed with probability $1 - q$. The expected value of this latter set of possibilities is $qw_I + (1 - q)0 = qw_I$. Thus the overall expected wage is now $p\bar{w} + (1 - p)qw_I$.

With this small digression completed, let us return to the simpler case of just two urban outcomes: employment in the formal sector or employment in the informal sector. Suppose that we use L_I to denote informal employment. Then we can see that the ratio

$$\frac{L_F}{L_F + L_I}$$

captures the probability of getting a job in the formal sector. The number of employed people L_F tells us how many jobs there are, whereas the number $L_F + L_I$ is the measure of the total number of potential job seekers. The ratio of the two thus gives us the chances that an urban dweller will get a job in the formal or informal sector.[20]

Now we can work toward the important equilibrium concept first introduced by Harris and Todaro [1970]. Migration from the rural sector may be thought of as an irreversible decision, at least for the proximate future. Because the fate of a potential migrant is not known, we must consider the *expected income* from migration and compare it with the actual income received in agriculture. Thus we may conclude that if

$$\frac{\bar{L}_F}{\bar{L}_F + L_I}\bar{w} + \frac{L_I}{\bar{L}_F + L_I}w_I = w_A, \tag{10.1}$$

we are at an equilibrium where no person wishes to migrate from one sector to the other. This is the *Harris–Todaro equilibrium condition*.

Some remarks are in order. First, the equilibrium condition represents a situation where ex ante people are indifferent between migrating and not migrating; ex post, they will not be indifferent. The lucky subgroup who land a job in the formal sector will be very pleased that they did migrate, whereas those who seek solace in the informal sector will regret that they made the move.

Second, observe that the equilibrium concept implies a *particular* allocation of labor between the three sectors of the economy. This is because it is the allocation of labor that affects the perceived probabilities of getting a job. If it is known, for instance, that the formal sector accounts for a smaller proportion of total urban employment, individuals will think harder before they hope for a job in the formal sector. Their expected wage calculation will yield a lower wage. This prospect will lower the size of the urban labor force, but increase the size of the formal sector as a proportion of total urban employment, which in turn, feeds back on the probability of getting the formal job.

Third, the equilibrium concept in no way requires that we stick to merely *two* subsectors of the

urban sector (formal and informal) or that we have only one sector in agriculture. The fundamental requirement is that *expected* wages are equalized over the two sectors for a migration equilibrium to be obtained, but these expectations may be the outcome of wages in three or more urban sectors (e.g., open unemployment may be thought of simply as another sector in which wages happen to be zero) or in several sectors in agriculture.

The Harris–Todaro equilibrium may be depicted on the sort of the diagram we have been using so far, but not with the greatest degree of clarity.[21] Recall Figure 10.5 and note that the agricultural wage of $\bar{w}$ was too high to be an equilibrium and that $\underline{w}$ was too low. It stands to reason that the equilibrium agricultural wage is somewhere between these two extremes. Note that there is no necessary relationship between this equilibrium wage and the wage rate w^* that arose in the flexible market case. Figure 10.7 denotes a typical Harris–Todaro equilibrium condition.

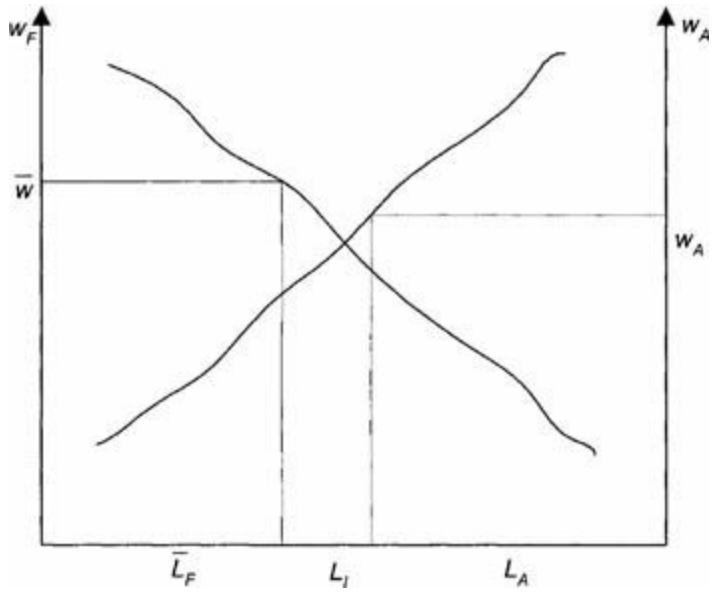

Figure 10.7. Harris–Todaro equilibrium.

In this figure, the equilibrium agricultural wage is given by w_A. L_A people are employed in agriculture, $\bar{L}_F$ people are in the formal urban sector, and the remainder, L_I, take refuge in the informal sector where they obtain an income of w_I. The allocation is such that (10.1) holds.

10.3.4. Government policy

The paradox of urban job creation

In this view, then, the informal sector is an outgrowth of the fact that the formal sector has wages that are too high, so that not everyone is capable of obtaining employment in this sector. At the same time, not everyone else can stay in agriculture as well, for that would make the formal sector look too attractive and induce a great deal of migration. The informal sector is a result of this migration. In the Harris–Todaro view, the informal sector acts as a necessary counterweight to the attractiveness of the formal sector and slows the pace of rural–urban migration.

To the urban observer and to the government official, however, the informal sector is an eyesore with not very pleasing properties. Unregulated economic activity in this sector is often responsible for congestion, pollution, and a high crime rate. There are several ways in we can deal with this

problem, and in the rest of this section we will study some of them.

The most obvious policy that comes to mind is to somehow accelerate the rate of absorption of labor in the formal sector. Even though wages are fixed at $\bar{w}$, it is possible to generate additional demand for formal labor by offering urban businesses various setup incentives (such as tax holidays) or ongoing investment incentives (such as better treatment in the credit market). The government might itself expand the demand for formal labor by expanding the employment of public sector enterprises.

No doubt, such policies initially reduce the size of the informal sector, by simply channeling people from this sector into the greater number of formal jobs that are now available, but matters do not end there. The size of the urban sector is endogenous, and migration will rise in response to this policy. The Harris–Todaro equilibrium concept helps us to explore the nature of the final outcome, *after* fresh rural–urban migration is taken into account.

To see this, trace the effect on the Harris–Todaro equilibrium condition. Imagine that the formal labor demand curve shifts out and to the right, so that, in particular, labor demand at the wage rate $\bar{w}$ rises from $\bar{L}_F$ to $\bar{L}'_F$. In the short run, all this extra labor simply comes from the informal pool. This means that relative to the initial outcome, L_F rises and L_I falls. This raises the probability of getting a formal job. Consequently, *the expected urban wage must initially rise*.

But the initial increase cannot be fully persistent. Rural–urban migration picks up. More migrants enter the urban sector. Of course they add to the informal sector, which after its initial decline, now begins to increase once again. This phenomenon sets in motion two related forces. First, as the labor force in agriculture falls, the agricultural wage tends to rise (by how much it rises will depend on the slope or elasticity of the agricultural absorption curve). Second, as migration continues, the expected urban wage once again begins to fall (relative to the initial sharp rise). One glance at (10.1) tells you why. The fraction $L_F/(L_F + L_I)$ begins to move down as migration continues, and this brings down the probability of getting a formal job (relative to what prevailed just after the institution of the policy) and the expected urban wage drops with it.

With the agricultural wage climbing up and the expected urban wage creeping down, the two are bound to come into line once again. In the process, we have a fresh allocation of labor in the three sectors: $(\bar{L}'_F, \bar{L}'_A, \bar{L}'_I)$. The new allocation must satisfy the new Harris–Todaro equilibrium condition

$$\frac{\bar{L}'_F}{\bar{L}'_F + L'_I}\bar{w} + \frac{L'_I}{\bar{L}'_F + L'_I}w_I = w'_A, \tag{10.2}$$

where w'_A denotes the new agricultural wage after the policy.

How do we compare the magnitudes in (10.1) and (10.2)? Recalling that the agricultural wage rises (or at least does not fall) after the introduction of the policy, it must be the case that the new expected wage in the urban sector exceeds the old expected wage. The only way in which this can happen in the model is if

$$\frac{\bar{L}'_F}{\bar{L}'_F + L'_I} > \frac{\bar{L}_F}{\bar{L}_F + L_I};$$

in other words, if the share of the formal sector in total urban sector employment goes up. This is a beneficial implication of the policy: the informal sector does shrink, *measured as a fraction of the*

total urban sector.

However, there is another way to look at the outcome, and that is to study the resulting size of the informal sector as a fraction of the *total* labor force. Is it possible that this increases even though the assertion in the last paragraph is true? Interestingly enough, the answer is yes. Although it may be true that the informal sector shrinks as a fraction of the urban labor force, it is also true that the size of the urban labor force has expanded. If the latter effect dominates the former, the informal sector may well expand—an implication of a policy that was directly aimed at *reducing* the size of that sector!

To see this, imagine that the agricultural sector is in the surplus-labor phase of Lewis, so that the wage in the agricultural sector adjusts very little or not at all as fresh migrants move out of agriculture. In that case, the share of the informal sector in the total urban sector is practically unaltered [simply look at (10.1) and (10.2), and use the fact that w'_A is very close to w_A], At the same time, it is certainly the case that the entire urban sector has grown from the first equilibrium to the second. This must mean that the informal sector has expanded as well.

What accounts for this seeming paradox? How is it that a policy designed to absorb people from the informal sector ends up enlarging its size? As a matter of fact, there is no paradox at all, but an observation that we see repeated in one developing country after another. Attempts to increase the demand for labor in the formal sector may enlarge the size of the informal sector, as migrants respond to the better job conditions that are available. The migration effect may dominate the initial "soak-up effect."

This observation is not limited to cases in which the demand for formal- sector labor is increased. A similar application can be carried out with regard to urban congestion, pollution, or the provision of health facilities. In each of these cases, policies aimed at directly reducing urban congestion (say, by building more roads), reducing pollution (say, by building a subway), or increasing the provision of health (say, by building new public hospitals) might all have the paradoxical effect of finally worsening these indicators. In each case, the ultimate worsening occurs because fresh migration in response to the improved conditions ends up exacerbating the very conditions that the initial policy attempted to ameliorate.[22]

Efficient allocation and migration policy

Recall for a moment the case of fully flexible wages with which we began our analysis of migration. The equilibrium there has an interesting efficiency property provided we think of the two absorption curves *AB* and *CD* (see Figure 10.4) as the competitive demand curves for labor. From elementary economic theory, we know that the competitive demand curve for labor is nothing but the value of the marginal product curve, where the price of the final output is used to compute value. In this situation, the intersection of these two curves corresponds to the case in which the values of marginal products in the two sectors are *equalized*. In all other allocations, there is a discrepancy between the two marginal products. In addition, the informal sector has a still lower value of marginal product (given by w_I), so that these allocations cannot maximize the value of total national product. The reason is simple. As long as marginal products are not equal, a small transfer of labor from the sector in which the marginal product is lower to the sector in which it is higher increases the total value of national income.

This observation should make us think twice about what it is, exactly, that government policy is

trying to achieve. It is not that the informal sector, per se, is something to eliminate. Indeed, in the policies that we will now examine, getting rid of the informal sector is not the main problem at all. Getting as close to the efficient allocation of labor resources, epitomized by the crossing of the two demand curves, is what policy should be all about.

I want to reiterate that this prescription is valid only in the case where both demand curves arise through competitive profit maximization, so that they correspond to the value of marginal product. When this is not the case, it is unclear whether the intersection of the two absorption curves possesses any efficiency meaning and should somehow represent a target. However, let us ignore this qualification for now.

Consider two policies that reduce or remove the informal sector. One policy is to physically *restrict migration*. Figure 10.8 illustrates this. All individuals who do not have formal sector jobs are prevented from entering the cities. If this policy can be enforced (and this is not a trivial issue), then migration restrictions certainly get rid of the informal urban sector. The number of people in the urban sector is now just $\bar{L}_F$; the remainder, L_A^M, stay in agriculture.

Note, however, that simply getting rid of the informal sector does not ensure that we have an efficient outcome. Compare the allocation achieved in this way to that achieved in the case of fully flexible wages. It is clear that under a policy of migration restrictions, we have too few people in the cities relative to the efficient allocation.

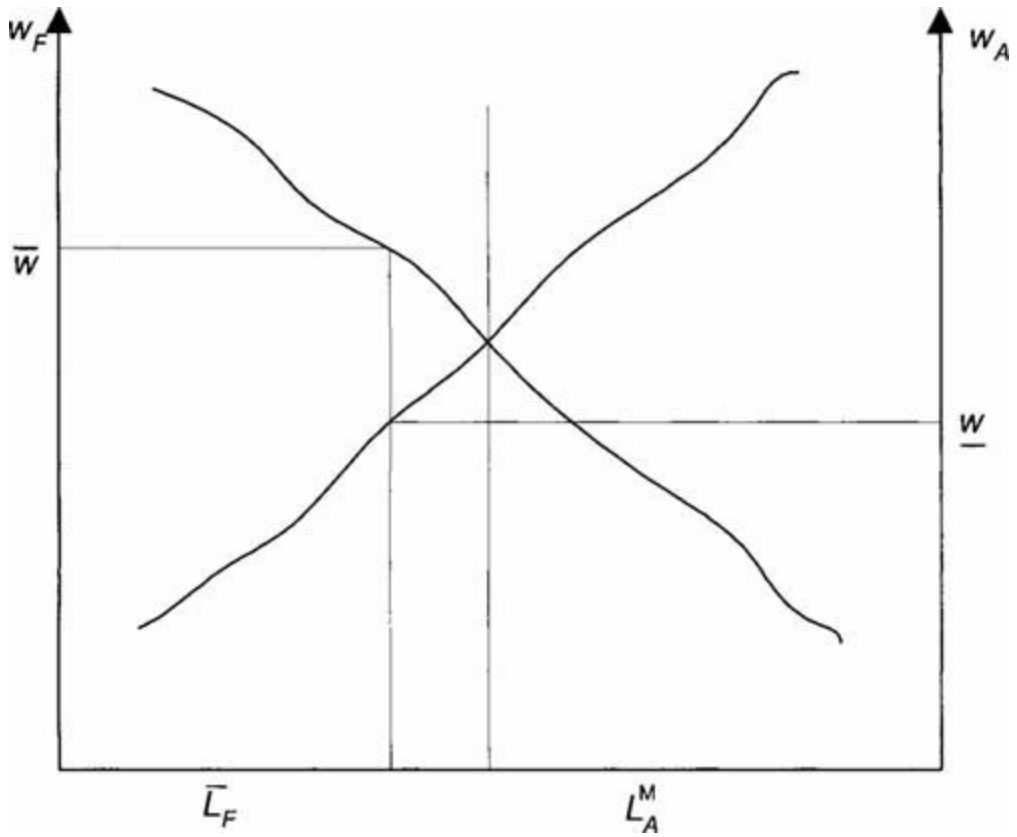

Figure 10.8. Migration restrictions.

The second policy is to offer a subsidy to employers in the formal sector for every unit of labor that they hire. Let us suppose that the subsidy involves financing s dollars of the formal wage for every extra labor hour that is hired by a formal-sector employer. In this case the wage that is paid by the employer is effectively $\bar{w} - s$, but the worker receives the full wage $\bar{w}$. Thus the effect is to push out the demand for labor at the formal wage $\bar{w}$, as Figure 10.9 illustrates.

Note that as the subsidy increases, formal labor demand increases. There comes a point when the formal-sector labor demand increases so much that agricultural wages are pushed up to equal $\bar{w}$, the formal-sector wage. At this point there is no urban informal sector and no incentive for anyone to migrate. This situation is illustrated in Figure 10.9, where formal employment is now at the level L_F^S and agricultural employment is at the level L_A^S under the subsidy

I am going to return later to the issue of how the subsidy is financed and whether the subsidy can be enforced. Leaving these questions aside, we see that although the urban informal sector has been removed, there is now *too much* labor in the urban sector relative to the efficient allocation. Thus in a way, wage subsidies achieve exactly the opposite of migration restrictions, even though both serve to eliminate the urban informal sector.

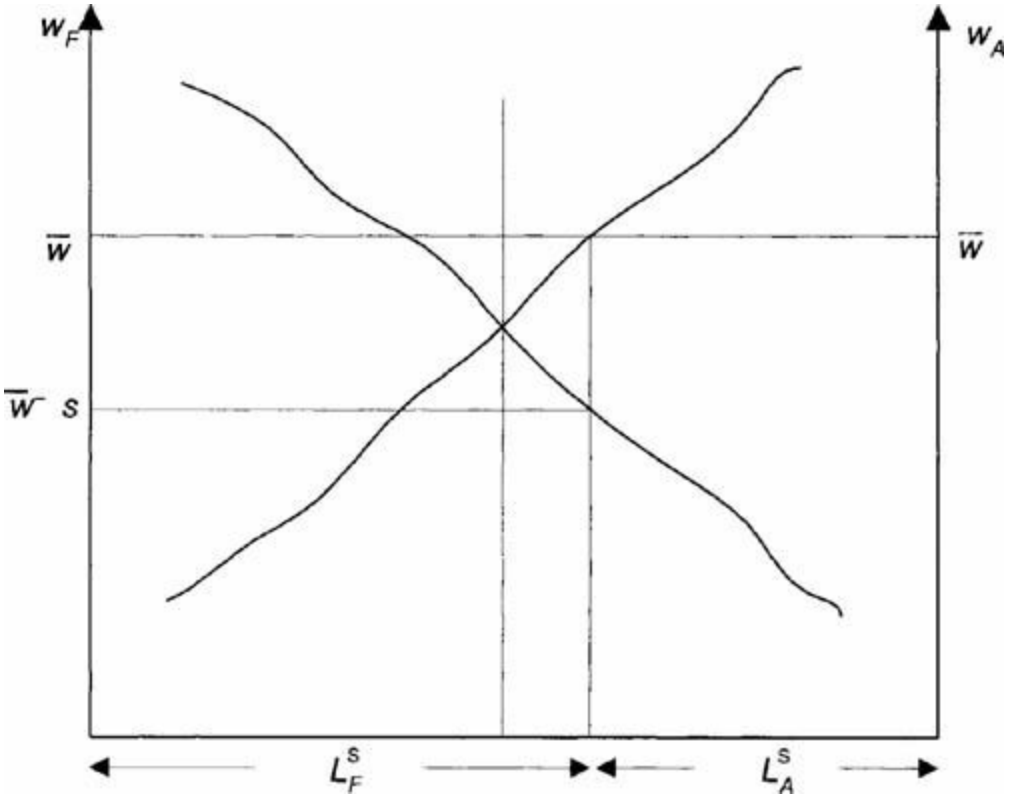

Figure 10.9. A formal-sector wage subsidy.

It is in this sense that we entertain a mixed policy that combines migration restrictions with wage subsidies in the formal sector. Figure 10.10 illustrates this. The subsidy is carefully chosen so that formal labor demand is hiked up to precisely the flexible equilibrium level, which is L_F^*. Note that workers still receive wage $\bar{w}$ in the formal sector, so that the flexible agricultural wage in this situation is still smaller than the formal wage if all remaining laborers stay in agriculture. Thus migration restrictions are still needed under this policy to make sure that the remaining labor (L_A^*) stays in the agricultural sector.

Are there policies that can make do without migration restrictions? The answer is that there are and, oddly enough, they involve the subsidization of employment in agriculture *even though* agriculture has perfectly flexible wages to begin with! Consider a *uniform* subsidy of s dollars per worker-hour to both agriculture and industry.[23] The first panel of Figure 10.11 shows how such a policy works, starting from a Harris–Todaro equilibrium. The demand for labor in both agriculture and the formal sector increases because the wage payouts in these sectors, *from the viewpoint of the*

employer, are $\bar{w} - s$ and $w_A - s$ (instead of $\bar{w}$ and w_A). The *worker*, on the other hand, continues to compare the wages $\bar{w}$ and w_A. Because the informal sector must have shrunk, the urban wage must have gone up (even though none of the separate wages has changed). To restore equilibrium, the agricultural wage must rise.

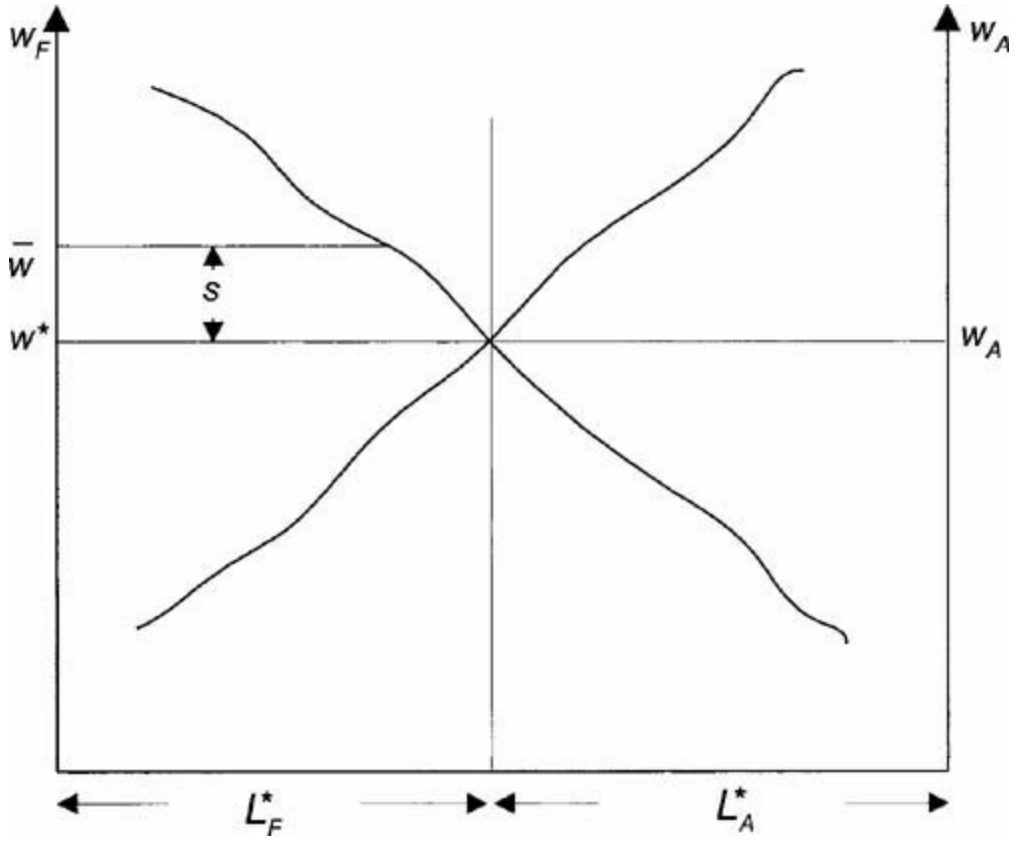

Figure 10.10. A combination policy of migration restrictions and wage subsidies.

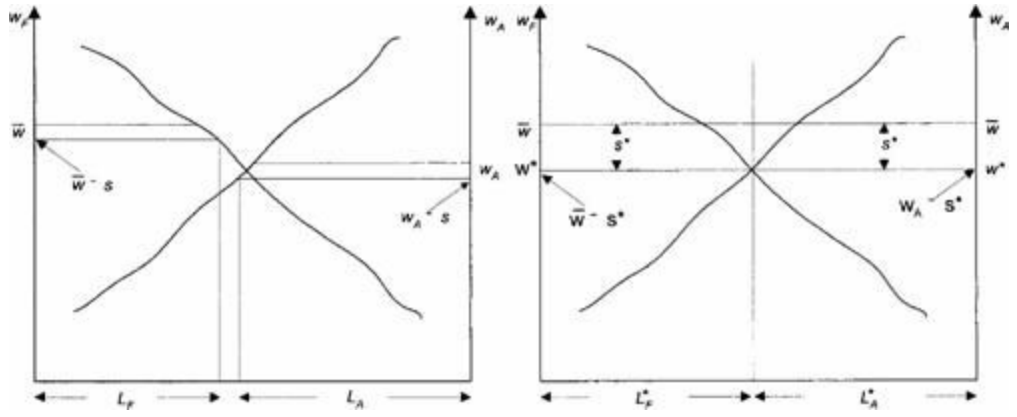

Figure 10.11. A uniform wage subsidy.

This process continues until the uniform subsidy reaches the level $\bar{w} - w^*$, where w^* is the old flexible-wage equilibrium. At this point, the agricultural wage must have risen to precisely $\bar{w}$ as well! This is shown in the second panel of Figure 10.11. Now there is full employment in both agriculture and the formal urban sector, there is no informal sector, *and* there is no need for migration restrictions, because the wages in the two sectors are perfectly equalized!

10.3.5. Comments and extensions

Some remarks on policy

Is there, then, a magical government policy that solves the migration problem and restores the efficient outcome without coercive restrictions on labor movement? To answer this question necessitates that we step outside the rigid confines of this simple model and examine the robustness of its predictions. Several objections may be raised. We consider them one at a time.

First, it can be argued that getting the subsidy *exactly* right might require too precise a knowledge of the parameters of the economy; specifically, the position and shape of the two absorption curves. Note that in Figure 10.11, the subsidy is chosen to be equal to $\bar{w} - w^*$. This requires that the government or planner *know* what the flexible equilibrium wage should be. This may be a tall order.

Fortunately, this criticism can be countered quite easily.[24] To see this, suppose that the subsidy was not chosen at the "correct" level s^* in the second panel of Figure 10.11, but at some point exceeding it. Then it can be seen easily that the only effect is to push up the wages *gross* of the subsidy in a uniform way, so that the net wage (paid by employers) still settles at w^* automatically. Figure 10.12 illustrates this. In this diagram, the subsidy was chosen so that it is larger than s^*. This has the effect of pushing the formal sector wage that is actually paid to workers *above* the institutional minimum $\bar{w}$. The very same thing happens in agriculture as well. It is exactly as if the greater subsidy is outweighed, dollar for dollar, by an increase in gross wage costs, so that the net effect is zero.

What this implies, then, is that the uniform subsidy does not require a precise knowledge of the parametric structure of the model. Mistakes in the size of the subsidy, as long as they do not involve too *small* a subsidy, tend to be washed out.

The second criticism is that the subsidy is not free. It has to be financed from some source. Shouldn't the costs of financing be taken into account in determining the efficiency of the overall outcome? The answer is that it certainly should, but, in principle, there is a way to make the entire policy self-financing: simply impose a profits tax on firms and return the proceeds as a wage subsidy. The profit tax will form the revenue to implement the subsidy.

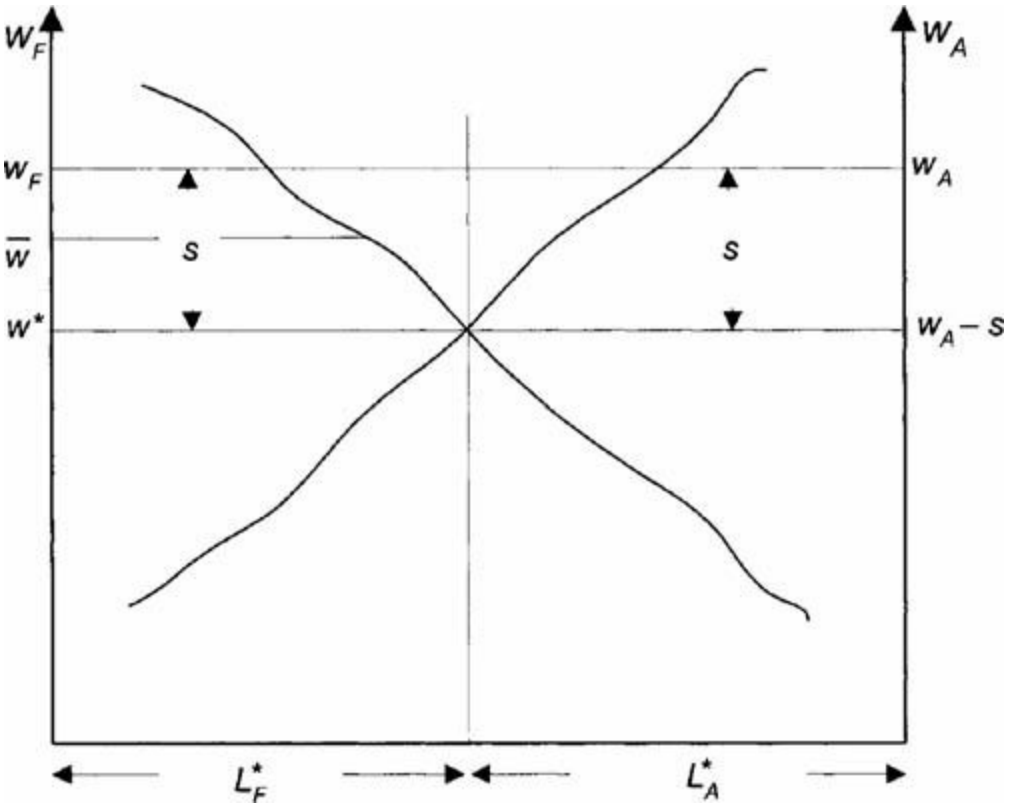

Figure 10.12. An overly generous uniform wage subsidy.

At first sight, this appears absurd. Note that in the end, firms would receive a net subsidy or pay a net tax. How is it that a financial policy that has no net effect on the balance sheets of firms nevertheless affects firm behavior? The answer is simple. The tax and the subsidy are conditioned on two *different* indicators of firm performance: profits and employment. If the firm does well on profits, but does not expand employment, it would pay a net tax. But the point is that it *will* want to expand employment to take advantage of the subsidy. Thus even though the two indicators cancel out when all is said and done, firm behavior can be altered substantially.

The third criticism is damaging and probably represents the single most important reason why a policy of wage subsidies is difficult, if not impossible, to implement. It is that employment figures are often very difficult to verify. Without adequate verification, it is not possible to pay a subsidy based on employment. Of course, in principle, it is perhaps possible to adequately verify that each employer is telling the truth about the number of laborers in his employ, but this would cost an immense amount of resources, which then need to be factored into an assessment of the overall efficiency of the migration policy.

Verification is a particularly serious problem in agriculture, which is an "informal sector" *par excellance*. Perhaps we could devise a scheme in which employee tax returns, for instance, are matched against their place of work, so that no fictitious employees can be invented,[25] but to do this we need a sector where tax returns are filed by all eligible employees. Such is generally not the case with agriculture, so the usual expedient of relying on alternative, complementary sources of information is not available.

We now turn to some extensions of the basic migration model.

A digression on risk and risk aversion

In this subsection, we introduce the concept of *risk aversion*. We give it plenty of importance because the concept plays a role not just here, but in the chapters that follow.

Many economic ventures and production processes are marked by substantial uncertainty regarding the final outcome. At the same time, investment decisions may need to be made well in advance of the resolution of this uncertainty. In the context of agricultural production, a farmer may have to decide how much fertilizer to use or how intensively to cultivate his land before knowing whether the weather will be favorable or adverse for a good harvest. He might even opt for the production of an entirely new crop without being sure of the return that it will bring. Similarly, in manufacturing, a firm may need to make its production decisions without knowing exactly what market conditions are going to be. For instance, when Coca Cola decided to open plants and marketing divisions in China, they made an investment commitment while being less than certain about their business prospects there. Almost all economic endeavors are marked by risk regarding the exact returns. In turn, such risk fundamentally affects the way in which people contract with one another. The migration decision is no exception, as we have already seen.

Consider a simple example to illustrate some of the issues. Suppose you are Nazim, a Turkish entrepreneur about to participate in an investment project to produce silk hats. This project is going to produce one of two possible sums of money. If your silk hats are a hit with the Turkish bourgeoisie, you are going to make a tidy profit of $10,000. If your hat factory is sabotaged, however, you will

make a profit of only $2,000. Because you are a bit of a worrier, you think that there is a 50–50 chance of sabotage. Now put yourself wholeheartedly into this situation and ask yourself the following question: if the money is all that you care about, what is the *minimum* compensation for which you would be willing to surrender the rights to the proceeds of this venture? After a little bit of introspection, you may want to write down a figure on a piece of paper and then think about the significance of your choice after the following discussion.

We begin by reviewing a term very popular with statisticians, called the mathematical expectation. It's something that we have already used implicitly to calculate the expected wage in the urban sector. In the preceding example, the mathematical expectation (expectation for short) of the returns to the project (before the actual outcome is known) is $\frac{1}{2}\$10{,}000 + \frac{1}{2}\$2{,}000 = \$6{,}000$—simply a weighted average of the various *possible* outcomes, where the weight on each outcome is the probability of its occurrence. More formally, if a project has n possible outcomes indexed by i (i.e., $i = 1, 2, \ldots, n$) and the ith outcome has a monetary value of x_i with a probability of occurrence p_i, then the expectation of the project is given by

$$E = \sum_{i=1}^{n} p_i x_i. \tag{10.3}$$

Going back to the previous example, check whether the minimum acceptable compensation you wrote down is more or less than the expected value of the project, namely, $6,000. If you have given it a little thought and if you are psychologically similar to most people, then the amount you wrote should be less than the expected value $6,000. This is because people usually dislike risk; they prefer to have the expected value of a project *for sure* rather than go into the uncertain prospect where the return can be either more or less than that expected value with fairly even chance. Thus people will generally be willing to receive somewhat less than the expected return in guaranteed compensation, in order to give up their claim to the proceeds. This attitude is known as *risk aversion*. Of course, the more risk averse a person is, the lower will be the minimum compensation he will need to be paid. In contrast, a person who is indifferent between enjoying the uncertain returns to a project and its expected value as guaranteed compensation is said to be *risk-neutral*.

One way to capture the attitudes of individuals toward risk is to think of them as having a utility function of money.[26] The idea is that individuals act as if they are maximizing the expected value of this utility under various uncertain circumstances.

What would be the utility function of a risk-neutral person? Recall that such a person acts to maximize the expected value of her *monetary* return. This is the same as postulating that her utility function coincides with the amount of money that she makes; in other words, that her utility function of money can be represented by a straight line. Put another way, the marginal utility of money for such a person is independent of the amount of money in her possession. From this angle, it turns out that risk aversion can be equated with the notion of a *diminishing* marginal utility of money. Figure 10.13 shows us the connection. In this diagram, the utility function is drawn so that it displays diminishing marginal utility: it is strictly concave.

Suppose that this utility function represents Nazim's preferences. The point A shows Nazim's utility when his profit is $2,000 and point B shows the utility of a profit of $10,000. How do we calculate Nazim's *expected utility* under these circumstances, assuming that the probability of each

occurrence is 1/2? This is not a deep question: we calculate the expected value of utility just as we calculate the expected value of a monetary gamble or the expected number of eggs a hen might lay on any given day. We take the value *A*, multiply it by 1/2 and add it to the value *B* multiplied by 1/2 as well. There is an easy way to do this on the diagram. Simply connect the points *A* and *B* by a line and find the point *C* between the two that lies at distances proportional to the probabilities of occurrence, in this case halfway between *A* and *B*. The (height of the) point *C* represents the expected utility of the gamble that Nazim is contemplating.

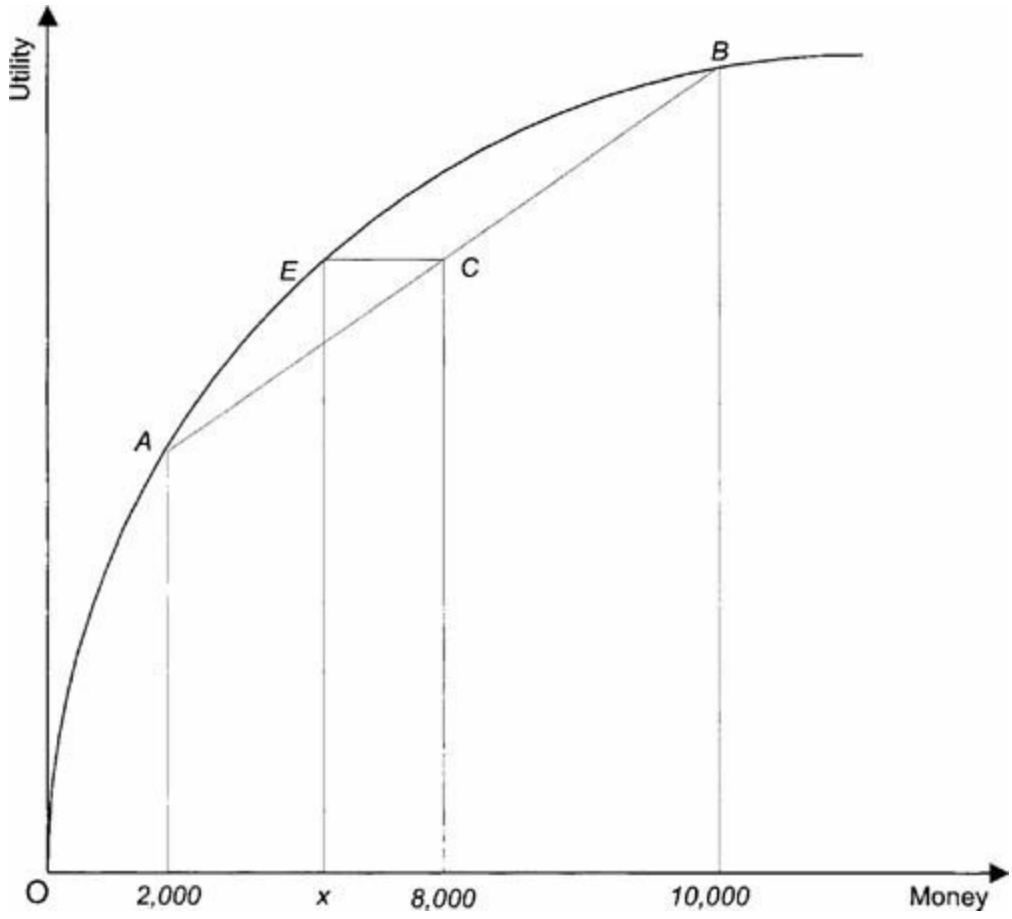

Figure 10.13. Risk aversion.

Note that if the utility function is curved in the way shown in Figure 10.13, then the point *C* must actually lie *below* the utility function. The meaning of this graphical observation is that the *expected utility of the gamble is lower than the utility of the expected value of the gamble*. This is just a way of saying that Nazim is risk-averse; he likes the gamble less than he likes receiving the expected value of the gamble for sure. This means that an amount of money smaller than the expected value will be enough to compensate him for foregoing the risky venture. To find this amount on the graph, we only need to find the sum of money whose utility equals the expected utility of the gamble. This is easy to do: simply draw a horizontal segment leftward from *C* until it meets the utility function (at *E*). The sum of money corresponding to this utility is the required amount. As we have already noted, if the utility function is shaped the way it is in Figure 10.13, this amount, marked *x* on the diagram, will fall short of the expected value. Thus this shape captures the idea of risk aversion.

With diminishing marginal utility, a dollar lost is always dearer than a dollar gained. That is why a risk-averse person faced with the prospect of gain or loss of a dollar (or any higher amount) with equal probability will be willing to pay a premium (say to an insurance agent) to avoid that risk. A risk-neutral person's utility function of money, on the other hand, can be represented by a straight line.

Such a person will in fact be perfectly willing to assume the risk upon himself (if a small fee is paid to him). In a risky economy, therefore, optimal arrangements are those in which risk-neutral agents bear all the risk, while risk-averse persons settle for stable income levels, implicitly or explicitly paying the risk-averse agents premiums for bearing risk. Insurance is a straightforward example of such arrangements: an insurance company promises to cover the risk of accidental damage and loss on behalf of their clients and charges a premium in return during normal times. Many other contractual arrangements (e.g., that between firms and employees, landlords and tenants) also have a dimension of insurance to them, as we shall see in the chapters to follow.

Finally, what determines an agent's attitude toward risk? Partly, it is a matter of individual taste: some people simply tend to be more cautious than others. To a large extent, however, it is also shaped by the economic circumstances of a person. In particular, two features of the economic environment play an important role: wealth and diversification.

It is plausible to argue that wealthier people have a better ability and willingness to bear a given (absolute) amount of risk compared with relatively poor people. A given loss can be ruinous for someone living off a modest saving, whereas a millionaire would be able to shrug it off by saying you can't win 'em all. Think of Nazim's silk hats and picture Nazim as an enormously rich billionaire. In that case, it would be reasonable to expect that Nazim would evaluate this business venture at its expected value; no "risk premium" would be involved. On the other hand, if we look at this venture against a background in which Nazim's own wealth is a few thousand dollars as well, the riskiness of the venture assumes its old significance.[27]

A second crucial question is how diversified an agent's source of income is. A person who earns his income from many different sources, each subject to *independent* risks (e.g., income from a diversified portfolio of stocks) will be less sensitive to the risks at each source, because on the whole, their effects tend to cancel out. It is highly unlikely that all sources will suffer bad outcomes at once. This explains why large insurance companies can comfortably bear the risks of their clients and act as if they are risk-neutral—at any time, only a small fraction of the clientele is likely to have accidents. On the other hand, people with only one or two major sources of income (e.g., those who rely exclusively on a wage income or self-employed individuals with a single occupation or crop) are likely to be much more sensitive to the associated risks. Hence, wealth and diversification are two key factors that positively affect the ability and willingness to bear risk.

Risk aversion and migration

It is easy enough to apply the theory of risk aversion to migration decisions. Recall that the original Harris–Todaro equilibrium condition [see equation (10.1)] equated the expected wage in the urban sector to the wage in agriculture. Thus we implicitly assumed that all individuals were *risk-neutral*. The fact that their expected income might be coming from an intrinsically uncertain lottery was of no consequence to them. Expected income from a lottery is the same as a guaranteed income, as long as the two have the same value. We now see that if individuals are risk-averse, this equivalence fails to hold.

Let us try to understand what a risk-averse potential migrant would feel if the Harris–Todaro equilibrium condition (10.1) were to hold economy wide. If you have read and absorbed the discussion in the previous subsection, the answer should be very easy. The potential migrant will not migrate. The reason is that he is comparing a *certain* wage in agriculture to an *uncertain* prospect in

the urban sector, both of which have the same expected value. However, expected value alone is not good enough for the risk-averse person. He craves insurance as well. Because the urban sector is risky, he will need an expected wage that is *strictly* higher than the wage he receives in agriculture in order to be induced to migrate. Consider Figure 10.13 once again. Note that the risky prospect has the same expected value as the safe prospect, but the expected *utility* of the risky prospect is lower.

Now let us see what effect this has on the allocation of employment across the sectors. Clearly, an economy of risk-averse individuals will exhibit *less* migration than an economy of risk-neutral agents, so in such an economy, the urban informal sector will be somewhat smaller and the agricultural sector will be somewhat larger than that predicted by the Harris–Todaro model. In this equilibrium, the expected wage in the urban sector will exceed that in the informal sector, and the extent of the difference will mirror the degree of risk aversion in the economy.

Social capital and migration

The wage in the agricultural sector is often not a full measure of the payoff to being in agriculture. The rural sector may be relatively attractive in its ability to provide forms of social support, insurance, emergency credit, and use of common property that the anonymity of the urban sector might fail to deliver.[28] The majority of these forms of support rely on two features that a traditional sector might possess in relative abundance: information and low mobility In the chapters to follow, we will see how these two characteristics assist in the provision of various forms of support, but a quick overview is not out of place here.

Consider insurance: the provision of help in cash, kind, or services if a fellow farmer or laborer is in economic difficulties. Such difficulties may arise, for instance, if a farmer's harvest fails for some "localized" or idiosyncratic reason connected to that farmer, such as pest damage. (A generalized harvest failure means that everyone is in trouble, so the question of insuring one farmer by his compatriots is not relevant.) Now, the damage may have occurred because the farmer was willfully negligent in his application of pesticides or it may have happened because of simple carelessness or events beyond the farmer's control. This is where the role of information comes in. If people know a lot about the daily lives of their neighbors, there is a greater chance that individuals will know just *why* the damage occurred. It is very important that they know this and that the *farmer* knows that they know this, because it is only in this case that genuine insurance is possible. If bad luck cannot be separated from deliberate behavior, insurance will only create incentives for the farmer to cut back on his effort and other inputs, and the system will break down. Thus information plays a role in the provision of social support.

Low mobility plays a different but equally important role. Much of the social support that we observe is predicated on some notion of *reciprocity*. I help you out today (with a loan, for instance) because I know that you will be around tomorrow to pay back the loan and perhaps help me out if I am in trouble. If this link between present and future is missing, it is very unlikely that my loan or assistance will be forthcoming in the first place. Thus low mobility justifies the expectation that true reciprocity is possible.

Low mobility serves a related function. It enables the imposition of social sanctions on a deviant. Again, consider insurance. If I have been informally participating in an insurance arrangement with my fellow farmers (where we help one another in times of trouble) and I suddenly decide not to offer any help, I can be punished by being denied insurance in the future. On the other hand, if I can just as

easily migrate to another village without great personal loss, the threat of this sanction carries little punitive value, and insurance schemes will fall apart.

You can think of low mobility and high information as forms of *social capital*: they permit productivity-enhancing arrangements that would not be possible otherwise. To the extent that the urban sector lacks this social capital, the pace of rural–urban migration will be reduced, because a potential migrant will now factor in the cost of this loss. At the same time, social capital is itself eroded by the act of migration (which serves to increase mobility and lower information). Thus an initial spurt of migration from a region may balloon as social capital is thereby eroded, making it now profitable for others (and still others) to join the exodus.

Migration and family structure

Just as risk aversion, or the existence of social capital in agriculture, reduces the amount of migration compared to the benchmark Harris–Todaro model, there are other variants that predict increased migration. One particularly interesting example requires us to think about the labor absorption curve in agriculture. In the policy options discussed in the preceding text, we thought of this curve as the competitive demand curve for labor in agriculture that was induced by profit maximization, but we already know of situations where this is not the case. If agriculture is largely composed of family farms that share their income, then this curve may be regarded as the *average* product curve. To be sure, it is still downward sloping as long as there is diminishing returns to labor in agriculture.

In this case, the intersection of the two labor absorption curves has a different interpretation. If the industrial sector is capitalistic, then the labor demand curve is the marginal product curve. However, the labor absorption curve in agriculture is the average product curve, as we've already observed, and so the intersection of the two curves does *not* correspond to the point of efficient resource allocation (which is the allocation for which marginal products are equalized). The policy analysis then has to be redone with this change in mind (we omit the extension here).

I mention in passing that the same outcome is also true in cases where there is uncertainty and individuals are risk-averse, or when we take into account the presence of social capital in agriculture (as in the previous section). In each of these cases, the location of the efficient allocation must take these additional features into account.

It is interesting, however, to emphasize a somewhat different aspect of the family farm model. Suppose that potential migrants make their migration decisions in order to maximize *family income*, as opposed to individual income. Let us walk ourselves through the steps of the argument. If all family income is shared, whether earned in the rural or the urban sector, then a potential migrant who seeks to maximize family income will see the expected income in the urban sector as the gain from migration, just as before. However, the loss from migration is different. This individual's contribution to the farm, his *marginal* product, is *not* measured by his stated income on the farm, which as we've seen is equal to the average product. Because average product exceeds marginal product (consult Figure 10.1 once again), a family-income maximizer will migrate *even if expected urban income is less than the agricultural income per person*, as long as the former exceeds the marginal product on the farm. This will yield excessive migration relative to the prediction of the Harris–Todaro equation (see the problem on this subject at the end of the chapter).

10.4. Summary

This chapter contains a detailed study of *intersectoral interaction* in the development process, notably between the agricultural and industrial sectors. The basis for this study is the view that development is rarely distributed evenly across sectors; rather, it typically manifests itself in *structural transformation*—resources move out of one sector to fuel the growth of another.

By far the most important structural transformation that a developing economy goes through is the change from a predominantly rural economy to an industrial economy. This intersectoral movement is typically accompanied by a move from traditional forms to modern forms of organization: an economy in which such forms coexist is often referred to as a *dual economy*.

We began with a description of the urban sector and introduced the notions of *formal sector* and *informal sector*. We then described the basic features of the agricultural sector, and supplemented this with an introduction to the ICRISAT villages, which will reappear at several points in the book. We noted that the transformation from rural to urban is marked by two massive resource flows: the move of labor, and a parallel move of food, to support the basic needs of those individuals no longer engaged in farming. The study of balance in these resource flows is often critical to our understanding of economic development.

A theoretical framework that studies structural transformation is the *Lewis model*, which we turned to next. Development is characterized by an ongoing move of labor and resources from a "traditional sector" to a "modern sector." Ongoing capital accumulation in the modern sector provides the fuel for sustained transfers. Lewis argued that the traditional sector is characterized by *surplus labor* (a situation in which labor can be removed without loss in output). In principle, this permits, industrial development *with unlimited supplies of labor*, at least until the surplus-labor phase comes to an end.

The part of Lewis's model that deals with industrial accumulation is pretty standard, so we focused on the traditional sector. We studied, first, what surplus labor means and what forms of economic organization permit it to exist. A narrow definition of a surplus labor situation is simply one in which the physical marginal product of labor is equal to zero. However, this state of affairs cannot persist under a capitalist organization that pays a positive wage, so the concept of surplus labor naturally led to a discussion of economic organization in the traditional sector. Typically, traditional forms of organization are characterized by income sharing (or payment according to *average* product), and this allows family farms to pay positive "wages" even when the marginal product of labor is close to zero. This form of organization is of interest in itself, with or without surplus labor.

We then returned to surplus labor per se and introduced two extensions, one of which is *disguised unemployment*, a situation in which the marginal product of labor is positive but smaller than the going wage. We argued that this is just as relevant a concept as surplus labor (and a generalization of it): both concepts signal inefficiency in the intersectoral allocation of labor resources. The second extension carefully distinguished between the notions of surplus *laborers* and surplus labor.

We then integrated the traditional and the modern sectors into one interactive model. It turned out that the supply of labor to industry was perfectly elastic in the surplus-labor phase, but began to rise thereafter as the available food surplus per capita began to shrink and the *terms of trade* between agriculture and industry turned against industry. The model brings out a fundamental tension between agricultural and industrial development: industrialists like to keep agricultural prices low, because

that ensures a low wage bill. On the other hand, a policy that keeps agricultural prices low has disincentive effects on agriculture and can strangle industrial development in the longer run.

This discussion led to the theme of *agricultural taxation*: how should a government that is interested in promoting industry deal with agriculture? Should it tax agriculture as much as possible, thus ensuring a large food surplus in the short to medium run, or should it invest in agriculture and look to possible long-run gains? The great debate on this subject that took place in the Soviet Union in the 1920s is still relevant today, and we discussed this briefly.

Related to agricultural taxation is the question of agricultural *pricing policy*: price support programs, subsidies to inputs such as water, electricity, and fertilizer, and (international) exchange rate policy. There is evidence from China and other countries that agricultural output reacts strongly to price incentives, and this suggests that a more liberal attitude to food prices in the short run may pay off in the longer run.

To be sure, all of these arguments rely on the premise that for economic or political reasons, free trade in food grains is problematic. Food is a special commodity: governments often desire self-sufficiency in food grains for political reasons. To the extent that an economy seeks to produce its own food supply, the considerations here are of great relevance. As economists we might *wish* that countries did not behave in this way, but pragmatic considerations of reality force us to treat the question of internal food supply as an issue of great importance.

Finally, we turned to an explicit consideration of the other resource flow involved in structural transformation: the movement of labor. We studied rural–urban migration in the framework of the *Harris–Todaro model*—a theoretical framework in which formal-sector wages have lower bounds or floors, whereas informal and agricultural wages are flexible. This thesis leads to a view of migration equilibrium in which the formal sector is characterized by an excess supply of labor, with the excess spilling over into the informal urban sector or manifesting itself in the form of open unemployment. Thus it is not wages that are equalized across sectors, but the *expectation* of wages: in the Harris–Todaro equilibrium, the average of various urban wages *weighted by the probability of employment in formal and informal sectors* is equal to the agricultural wage. This model endogenously delivers a prediction for the size of the urban informal sector and allows us to examine how different policies affect this sector. Of special interest is the so-called *Todaro paradox*, in which an expansion of formal employment leads to an enlargement of the informal sector as fresh migrants from the rural sector swarm into the urban sector response to the policy.

Policies that move the economy toward an efficient labor allocation were also considered: among them are *migration restrictions* and *wage subsidies*. Finally, we studied various extensions of the Harris–Todaro model. The extensions include a treatment of *risk aversion* (indeed, this chapter introduces this important concept which we will use later), a discussion of *social capital* in the traditional sector, and finally, the role of family structure in rural–urban migration.

Exercises

■ (1) Review the concepts of formal and informal sectors. Explain why labor might receive better treatment (in pay or conditions of work) in the formal sector. How do you think such differentials persist? Why can't informal sector workers simply enter the formal sector by offering to work on terms that undercut existing formal-sector workers?

■ (2) Describe the two fundamental resource flows that link the agricultural sector with the industrial sector. Discuss the market forces that are relevant to the magnitudes of these flows.

■ (3) Review the Lewis model of economic development. In particular, discuss the following concepts: agricultural surplus, average agricultural surplus, surplus labor, disguised unemployment, family farming, capitalist farming, and the three phases of development in the Lewis model.

■ (4) (a) Consider a family farm that is in the surplus labor phase. Now suppose that some members migrate to work elsewhere. Describe what happens to the average income of the family farm.

(b) Reconcile your observation with the assertion that the supply curve is perfectly elastic (or flat) in the surplus labor phase of development. In other words, describe when the observation in (a) is consistent with the observation in this paragraph.

■ (5) Present arguments why, all other things being the same, the industrial supply curve of labor is steeper, when the economy is closer to minimum subsistence to begin with. If food can be freely traded on the world market, do you anticipate that the supply curve will be flatter or steeper? Justify your answer.

■ (6) Taxes on industrial profits will leave less room for capital accumulation and slow down the rate of growth. Do workers already employed in the industrial sector have an incentive to lobby for such taxes, the proceeds of which are transferred to them in the form of additional benefits? Show that the answer to this question depends, among other things, on the slope of the labor supply curve to industry. Argue, in particular, that if the supply curve is horizontal (as in phase 1 of the Lewis model), the tendency for already-employed workers to vote for industrial taxation will be higher.

■ (7) Consider a labor surplus economy producing a single output, which can be consumed (as food) or invested (as capital). Labor, once employed, must be paid a fixed wage of w which is fully consumed. All surpluses from production are reinvested.

(a) Draw a diagram showing how output is distributed between consumption and reinvestment, for some chosen level of employment. Show the profit-maximizing employment level: call it L^*.

(b) Reinvested surplus raises consumption tomorrow. Suppose that a social planner cares about consumption today *and* consumption tomorrow. Show that she will always wish to choose an employment level not less than L^*.

(c) Suppose that the planner wishes to employ $\hat{L}$ units of labor, where $\hat{L} > L^*$. Carefully describe a subsidy scheme to profit maximizing employers that will make them choose this level of employment.

■ (8) Pim and her three sisters own a small farm in the agricultural sector of the land of Grim. They work equally hard, and the value of their output measured in the local currency, *nice*, is 4,000 nice, which they divide equally. The urban sector of Grim has two kinds of jobs. There are informal jobs which *anybody can get*, which pay 500 nice, and there are formal jobs which pay 1,200 nice. The probability of getting these jobs depends on the proportion of such jobs to the urban labor force, exactly as in the Harris–Todaro model.

(a) Assume that Pim compares her *own* expected returns in the two sectors and there are no costs of

migration. Calculate the threshold proportion of formal jobs to urban labor force that will *just* deter Pim from migrating.

(b) The full production function on Pirn's farm is given in the following table.

Number working on farm	*Output (in nice)*
One sister	1,500
Two sisters	2,500
Three sisters	3,300
Four sisters	4,000

Suppose that Pim and her sisters seek to maximize their *total family income*, instead of Pim simply acting to maximize her own. Assume that the threshold proportion that you derived in (a) *actually does* prevails in the urban sector. Now prove that Pim will migrate.

(c) Will any of Pirn's sisters *also* wish to migrate?

(d) Provide a brief description that uses your economic intuition to contrast cases (a) and (b).

■ (9) A farm household in rural Mexico consists of five adult brothers and no other dependents. Total annual income depends on the number of brothers working on the farm through the year and is given by the following schedule:

Number of brothers	*1*	*2*	*3*	*4*	*5*
Total farm output (in $)	1,000	1,800	2,400	2,800	3,000

Each brother, at the beginning of the year, can decide to migrate to Mexico City, where a typical job, commensurate with his skills, pays $1,300 per annum, but the unemployment rate is as high as 50%. A person who migrates to the city cannot come back and work on the farm that year. Furthermore, all city jobs are temporary one-year appointments. Also assume, for all questions that follow, that the brothers are risk-neutral and there is no difference in the cost of living between the city and the countryside. Now consider three different scenarios.

(a) Suppose the family is completely individualistic: those brothers who work on the farm share the farm income equally among themselves. There are no remittances to or from any family member who goes to the city. Find the number of brothers who will migrate.

(b) Now suppose the family is completely altruistic: the total family income, whether from a city job or from the farm, is pooled and shared equally. How many brothers will this family send off to Mexico City to look for jobs?

(c) Here is a third possibility. Those brothers who migrate to the city become selfish and never send home any remittances, even if they are employed (maybe those in the countryside have no way to verify whether their brothers in the city have a job). However, the family sits down and makes the following arrangement: those brothers who try out their luck in the city will each be sent $200 a year from the farm income, to insure against possible joblessness (assume that this contract is always

honored). Find how many brothers will decide to migrate. (Assume that if a brother is indifferent between migrating and not migrating, he does migrate.)

(d) Compare the numbers in parts (a)–(c). In light of the comparison, discuss the following assertion. "The extent of rural urban migration depends, *ceteris paribus*, on the nature and degree of altruistic links within families."

■ (10) Are the following statements true, false or uncertain? Provide a brief explanation to back up your answer.

(a) In the dual economy model, the phase of disguised unemployment must be associated with a horizontal supply curve of industrial labor.

(b) A low or moderate inequality of land holdings should slow down the pace of rural–urban migration.

(c) Migration restrictions alone lead to too many people in the informal sector.

(d) In the Harris–Todaro model, an increase in the formal sector labor demand at a fixed wage rate *must* lower the percentage of people in the informal sector, as a fraction of the urban labor force.

(e) If governments cannot tax agriculture, the supply curve of labor to industry in the Lewis model is always upward sloping.

■ (11) In the 1950s, facing massive unemployment in the cities (much of it disguised in the informal sector), the Kenyan government embarked on a "Keynesian" policy of creating new urban jobs through public investment. By many accounts, the size of the informal sector in Kenya went up instead of dropping in the months that followed. Give an economic explanation of this phenomenon, using the Harris–Todaro model.

■ (12) Carefully review the different migration policies studied in this chapter. Explain under what circumstances the flexible equilibrium allocation is the efficient allocation and how different policies situate themselves relative to the efficient allocation.

■ (13) (a) Calculate the expected values of the following lotteries: (i) 100 with probability 0.4 and 200 with probability 0.5; (ii) 100 with probability p and 200 with probability $1 - p$ (evaluate the amount as p varies between 0 and 1: does this make sense?); (iii) 100 with probability p and if this does not happen (which is the case with probability $1 - p$), then another lottery where you get 50 with probability q and 200 with probability $1 - q$; (iv) 100 with probability p, 200 with probability q, 300 with probability r, and nothing with probability $1 - p - q - r$.

(b) Suppose that you are asked to participate in a lottery where you get 1,000 with probability 0.1 and 200 otherwise. If you are risk-neutral, what is the maximum you would pay to enter the lottery? Would you be willing to pay more if you were risk-averse? Now suppose that the probability of winning is unknown. If you are risk-averse and willing to pay 600 to enter the lottery, what must be the *minimum* probability of winning 1,000?

■ (14) Suppose that two individuals A and B meet and undertake a joint project in which the returns are 1,000 with probability 0.5, and 2,000 otherwise. They are negotiating an agreement regarding the

division of the returns. That is they decide on a division rule *before* the project comes to fruition and they know what the outcome is. An example is: if the outcome is 1,000, then *A* pays 20 to *B*, and *B* gets 1,200. If it is 2,000, then they split it 1,000–1,000.

Prove that if *A* is risk-averse and *B* is risk-neutral, then any efficient division rule will give the *same* amount to *A* irrespective of project outcome and *B* will bear all the risk. (Note: The division rule is efficient in the sense that no other rule exists in which *both* parties enjoy higher ex ante expected utility.)

[1] These observations are drawn from material prepared by the Federal Research Division of the United States Library of Congress under the Country Studies/Area Handbook Program sponsored by the Department of the Army.

[2] The villages are Aurepalle, Dokur (in the Mahbubnagar district of Maharashtra state), Shirapur, line Kalman (in the Sholapur district of Maharashtra), Kanzara, Kinkheda (in the Akola district of Andhra Pradesh), Boriya, and Rampura.

[3] This material is largely based on Walker and Ryan [1990].

[4] Sorghum is used as both food and fodder and is cultivated throughout the semi-arid tropics. Sorghum originated in the northeastern quadrant of Africa and was distributed along trade and shipping routes throughout Africa, and through the Middle East to India at least 3,000 years ago. Sorghum is now widely found in the drier areas of Africa, Asia, the Americas, and Australia. The ICRISAT home page at http://www.cgiar.org/icrisat/, from which this account is drawn, contains more on sorghum.

[5] Probably the world's hardiest crop, pearl millet is a food staple in the semi-arid tropics. Pearl millet has been used as a cereal for thousands of years in Africa and parts of the Near East, and is cultivated for both forage and grain. It is grown today in many African countries and in some Asian countries, particularly India. See ICRISAT's home page at http://www.cgiar.org/icrisat/ for more information.

[6] There are exceptions to every rule, of course. In some countries, the creation of an agricultural surplus was not fundamental to economic development. These countries relied on the export of manufactures to fund their import of food items. With international trade in food grain, it is possible in principle to have *nobody* in the agricultural sector. A country can rely entirely on imported food. Such countries do not follow the general rule that we have outlined, but why are such situations the exception rather than the rule? The answer to this question must ultimately lie with the notion of *self-sufficiency* in food: food is so basic and so much the foundation of all activity that most governments cannot bear to think of a contingency in which their nations must depend on others for this most basic of wants. Such attitudes go some way in explaining why agriculture enjoys extraordinary protection in many developed countries, such as the United States, Japan, or in the European Union.

[7] See also Nurkse [1953], Jorgenson [1961], Ranis and Fei [1961], Sen [1966], Dixit [1970], Amano [1980], and Rakshit [1982].

[8] On the so-called peasant mode of production, based on the notion of traditionally organized family farms, see Georgescu-Roegen [1960] and Chayanov [1991].

[9] Equal sharing is only a simplifying assumption and may not hold in all situations. For instance, in Chapter 8 we studied the possibility of unequal division of output among family members in very poor family farms, because of nutritional considerations.

[10] See the survey by Kao, Anschel, and Eicher [1964].

[11] This is actually an overestimate. As labor is released from traditional activities and put to work elsewhere, the marginal product in these other activities typically falls. But as a micro estimate for one particular farm, this measure isn't a bad one.

[12] I reiterate that the assumption of limited food-grain trade, although realistic, is crucial to the argument. Without it only the efficiency argument that we just examined makes any sense.

[13] For theories of development traps and inflationary spirals based on a limited surplus, see Rao [1952], Kalecki [1976], and Rakshit [1982].

[14] Of course, simply *maintaining* output isn't enough. The excess also must be released to the market for consumption by nonagricultural workers. We return to this issue after a discussion of the Lewis–Ranis–Fei model.

[15] To be sure, this cannot happen if many family members are removed, because the few that are left will have to put in immense effort to compensate and the marginal cost of such effort will surely be higher than before.

[16] It is not exactly a supply curve in the traditional sense because in its construction we account for movements in the relative price of agricultural to industrial output, as well as for changes in the agricultural wage rate induced by the transfer. See the description that

follows for details.

[17] Stalin's speech was made to the Leningrad organization of the Party and is quoted in Dobb [1966].

[18] We neglect here the costs of migration, which can easily be incorporated into the model.

[19] To be sure, this is a simplification. Agriculture may have its own variability of wages, depending the form of the contract and the nature of employment, but for simplicity, we ignore this variability here.

[20] The careful reader will see that this statement is only correct if there is a rapid enough rate of turnover in the formal sector, so that the current level of *employment* can roughly be equated to the number of available *vacancies*. With a smaller rate of turnover, the number of vacancies is not L_F, but some number less than that, and likewise, the number of job seekers is smaller than $L_F + L_I$. The mode of analysis is very similar.

[21] The main problem is that the informal sector cannot be depicted explicitly on this diagram.

[22] This phenomenon is often referred to as the Todaro paradox.

[23] Bhagwati and srinisavan [1974] discussed the uniform subsidy that we study in this text.

[24] On this point and its resolution, see Basu [1980].

[25] This is not to say that such a scheme is entirely foolproof.

[26] Alough the expositon here is unashamedly biased toward simplicity, it must be noted that the postulate of a utility function for money is a bit misleading. What the literature does is begin with preferences of individuals over various risky gambles, which is just an extension of the usual preferences in consumer theory over goods and services. The well-developed theory of decision making under uncertainty then shows that these preferences can be represented (under some conditions) by a utility function for money and the behavioral postulate that an individual acts to maximize the expected value of this utility. See Arrow [1971] for an exposition of this theory.

[27] This is often captured by assuming that individuals have *decreasing absolute risk aversion*, where the qualifier "absolute" refers to the fact that we are talking about the same gamble against the background of alternative incomes or wealth. However, matters are more complicated when we refer to a gamble whose amounts bear some given *proportion* to wealth or income. For instance, we could consider a variation on Nazim's story: he stands to gain an amount that equals his wealth if he "wins" and amount equal to that of one-fifth of his wealth if he "loses." Now we are not talking about the same gamble (as wealth changes), but about the same *relative* amount of risk. Whether relative risk aversion increases or decreases in wealth is a subtle issue. An assumption of constant relative risk aversion is certainly not out of place.

[28] On these matters, see Das Gupta [1987]. For a model of rural–urban migration that includes these features, see Banerjee and Newman [1997].