8. The legacy of dualism in new growth theory

Salvatore Capasso and Maria Rosaria Carillo

8.1. INTRODUCTION

The most recent advances in growth theory have shown renewed interest in issues that had long been marginalised, at least by the mainstream literature on growth. Structural changes in production systems, increasing urbanization during development, unemployment persistence, the possibilities of market failures and the emerging of poverty traps are some of these issues. The study of such issues requires complex theoretical approaches and the assumption of frictions and impediments to the proper functioning of markets. Hence balanced growth cannot always provide a proper framework for such analysis. However, as amply shown by empirical evidence, development[1] very rarely occurs in a smooth pattern. Almost always, development involves deep restructuring of the economy, with the rescaling of some sectors and the emergence and growth of others. It entails changes in health standards, education, and in institutional settings. Inevitably, an analysis of growth and development which does not fully take such considerations into account risks being partial and limitative. Although these same issues have long occupied much of the attention of development literature; the latter has been more interested in the empirical description of the process of development as this unfolds, rather than in growth. On these grounds, growth theory has recently attempted to modify the standard balanced growth framework to allow for some empirical regularities which were hitherto not taken into account. As a result, this literature has been able to provide an innovative rationale for poverty traps, the emergence of multiple equilibria, the importance of institutions and market structure.

Notwithstanding such efforts, a central feature of development has long been grossly neglected by standard growth literature, namely the issue of structural change. And one can easily understand the reasons. Since structural change is the main empirical feature of development (Chenery and Syrquin, 1975; Chenery et al. 1986) it has been at the core of more

traditional development analysis. However, structural change cannot be analysed, by definition, in a standard one-sector model. Neither can it be simply described by a multisectoral framework in which markets can freely and immediately adjust (Uzawa, 1961, 1963). It appears to be a complex and multifaceted process by which production and consumption systems undergo slow and non-monotonic transformation, in which sectors change strongly, interacting with each other, and which might often not lead to balanced growth and equilibrium. These kinds of features are extremely difficult to capture in a fully working general equilibrium model of balanced growth.

Yet these are particular aspects of development which need to be accounted for. Since the seminal work by Lewis (1954), a large number of development studies have focused on the issue of dualism. Dualism is a representation of structural change as it empirically appears to be. This framework can provide a rationale for the shift of employment and production from agriculture to manufacturing, in the first stage of development, and then from manufacturing to services, later on. It attempts to explain the urban unemployment which follows early industrialisation as opposed to the reduction of employment in agriculture. Standard dual economy models are built not simply on the assumption of an existing contraposition of two sectors or markets, but more importantly, they are designed to take into account the dynamic linkages between sectors and the consequent development path of the economy. It is labour reallocation between sectors that in most models drives economy transformation. The migration flow of excess labour supply from agriculture to the more advanced manufacture sector makes capital accumulation and technological improvements possible in the manufacturing sector. In turn, growth and development of the manufacturing sector allows the growth and development of the economy and the adoption of new production standards in agriculture as well.

Recently, many economists have attempted to engineer the traditional theory of dualism into new growth theory frameworks.[2] The objective is to reconcile two, up to now, alternative strands of literature and to construct a unified model of growth and development. While moving in this direction, the literature has followed different routes. The common denominator of these works is the two-sector Uzawa-type growth model, aptly modified, in which a technologically developed sector, industry or manufacturing, coexists with a less developed sector, agriculture. The link between sectors feeds on different assumptions. For example, if one assumes that consumers have a lower income elasticity of demand for agriculture goods – the so-called Engel's law, then as the economy develops, the relative demand for these goods as compared with manufacturing goods will decrease. Similarly,

if the production function in each sector displays different elasticity for each accumulable factor, then as factors accumulate at different rates, there will be sector reallocation of factors and output. Finally, labour migration flows from one sector to another could explain, as in traditional dualism, the relative development of one sector and the restructuring of the economy.

The descriptive potential of this new framework of analysis is clear. It is able to display dynamics that standard growth theory cannot properly predict. Indeed, not necessarily do such models end up displaying balanced growth. Multiple equilibria, development traps and endogenous dualism are often a possible result. Also, the issue of convergence is now completely different. Most of the time, structural change is associated with the idea of sectoral convergence, that is a dynamic process which leads to the progressive disappearance (or severe decline) of one sector. Technologically mature and outdated sectors give way to more technologically advanced sectors. This kind of convergence, however, does not necessarily coincide with regional convergence, intended as progressive homogenisation in growth rates and in per capita income. Two economies might be involved in the same convergent structural change (disappearance of more mature sectors), but still be further apart in the growth rate. This framework can thus explain facts which are still puzzling. Why are some economies still lagging behind, despite deep structural change? Is the observed dualism between and within some economies a transitory phenomenon or a permanent condition? Is dualism endogenous?

Starting with Lewis's (1954) seminal contribution, a large number of studies have attempted to describe the process of development by means of dual economy models. Our objective is not to survey such literature, which has been already extensively reviewed in a number of studies (Dixit, 1973; Kanbur and MacIntosh, 1988, amongst others), but rather to determine the possible links of this literature with the more recent new growth theory. Indeed, given that these two theoretical frameworks allow us to capture different aspects of the process of economic development, their merging might give birth to a new theoretical framework which is potentially able to shed new light on growth and economic development. Very recently new studies have started to investigate development and growth within the endogenous growth framework. They have done so by reconsidering and drawing on typical features of dualism. The resulting dynamics are completely new, as are the results these studies obtain. To understand the real novelty of this new framework, one needs to take a step back and have a look at the traditional literature on dualism.

The chapter is structured as follows. In Section 8.2 we determine the main features of dualism in traditional development theory. In Section 8.3 we review the advances of new growth theory on the back of traditional

dualism. Section 8.4 contains the most recent developments of growth theory and dualism through migration and human capital accumulation. Section 8.5 concludes.

8.2. TRADITIONAL THEORIES OF DUALISM

During development, complex multifaceted changes take place in technologies, institutional settings and markets. Among such changes one appears to be a major player on the stage of economic development, requiring particular attention. This is the process of structural change (Chenery and Syrquin, 1975; Chenery et al., 1986). Typically, during the process of development the share of the agricultural sector in GDP decreases considerably. And this occurs, at least during the first stages of development, in favour of the manufacturing sector whose share in GDP increases at the same time. The data show that normally the share of agricultural sector tends to decrease from 60 to less than 5 per cent of GDP. Subsequently, as growth occurs, also manufacturing declines and the services sector starts to develop and become predominant in the economy. Of course, with sector transformation there follows a substantial change in factor allocation in the economy. The share of labour employed in agriculture decreases while its productivity increases. At the same time, the manufacturing and industry sectors allow human capital accumulation by absorbing skilled labour and becoming more capital intensive. These changes have not only strict economic, but also social and demographic effects. As such, structural change is clearly a complex phenomenon which cannot be integrated in a simple one-sector model of economic growth, by definition, but which cannot either be explained by a multisectoral model with no frictions or obstacles to factor allocation.

By pivoting on specific non-standard assumptions such as asymmetries in the mechanisms of market clearing and non-convexities in production and utility functions, traditional dual models (Lewis, 1954; Jorgenson, 1961, 1967; Sen, 1966; Taylor, 1969; Harris and Todaro, 1970, amongst others) can depict specific features of development which standard neoclassical multisectoral model cannot explain. Because of the specific aspects of development they describe, and because of the high level of diversity in the theoretical frameworks, it is difficult to group and classify the literature on dualism. Yet, some common factors can be identified

- the economy is dual since it consists of two sectors: a traditional sector (agriculture) and a modern or advanced sector (industry). The distinction between sectors hinges not on the specific features of the sectors and on

the goods they produce, but rather on the initial technological differences. A sector displaying a low productive technology is opposed to a sector displaying a more advanced productive technology.

- Dualism is determined by the specific assumptions on the structure of markets and sectors. Differences in the production functions, in the preferences, in the wage setting mechanism, and in factor allocation, define the sectors and the economic structure.
- Simple differences among sectors are not sufficient to define a dual economy. Dualism requires a dynamic functional linkage between the advanced sector and the traditional one. The excess labour supply in the agricultural sector which allows the development of the industrial sector and capital accumulation is a typical example (Lewis, 1954). Such linkage determines the dynamics of the economy and the degree of persistence of dualism.
- In order to reproduce slow and progressive adjustments, a low (or even absent) mobility of factors among sectors is required. While in Uzawa's two-sector model factors freely flow from one sector to another, equating marginal productivities, in dual economy models these adjustment are hindered.[3]

The seminal work on dualism is the model developed by Lewis (1954) which describes the transition during development from an agricultural based economy to an industry based one. Lewis's economy initially consists of an underdeveloped agricultural sector, which does not employ capital in production, but only labour and land, and an advanced industrial sector, which employs labour and capital. An important assumption is that there is an excess supply of labour in the agricultural sector. This in turn implies that labour can migrate from the agricultural sector to the industrial one without affecting productivity and output in the agricultural sector. As labour supply moves out of the agricultural sector, the industrial sector and the economy accumulate capital and develop. Under these circumstances, it is the advanced sector that develops on the back of the less developed one. As the economy moves on and accumulates capital, the industrial sector grows and it is able to absorb more and more labour, wages being constant. Of course, the growth of manufactured output, the decrease in overall unemployment and increasing investment result in the overall improvement in standard of life in the economy. The migration of labour from agriculture to industry is the functional link behind Lewis's dualism. Clearly, the flow of labour migration is the major determinant of the dynamics of the economy and the development path. Yet Lewis and most of the subsequent literature do not fully clarify what the long-term dynamics should be in this kind of model. The focus is the transitional dynamics rather than the long-run steady state

equilibrium. If one sticks to the previous assumptions, it may be argued that dualism will persist as long as labour flows from the agricultural sector into the industrial sector. Since the link is unidirectional, the economy will grow until the industrial sector accumulates capital and employs increasing labour. This might lead to the hypothetical disappearance of the agricultural sector, or to a condition in which two sectors coexist with no linkages between them. In the first case the economy will became virtually a one-sector one; in the second, the two sectors will continue to coexist, but dualism, as we have defined it, might disappear as well.

By following a different route, Jorgenson (1961, 1967), Fei and Ranis (1964, 1966) and Dixit (1973) assume that while the wage equates its marginal productivity in the industry sector, it equates, instead, its average productivity in the agricultural sector. The link between the agricultural sector and the industry sector is essentially one of physical dependence: the industrial sector depends on the agricultural sector for its food supply. Without the agricultural sector and the supply of food, the advanced sector cannot exist and develop. These models usually assume constant returns in the advanced sector and decreasing returns in the agricultural one. This assumption results in a development process in which the advanced sector will continuously grow and the relative size of the agricultural sector will decrease: a process that leads to a virtual one-sector economy.

Further compelling evidence that plays a key role in economic development is the change in the unemployment rate. Todaro (1969) and Harris and Todaro (1970) focus on the role of unemployment and, specifically, of urban unemployment in the process of development. Unlike previous works, they assume that capital is employed not only in the advanced sector but also in agriculture. This implies that progress and capital accumulation can potentially occur both in the advanced and in the underdeveloped sector. The peculiarity of these models lies in the workings of the labour market. Indeed, it is assumed that while wage equates labour marginal productivity in the agriculture sector, in industry, wage is artificially kept above labour marginal productivity and above the level that equates demand and supply. The latter assumption implies the existence of unemployment in the advanced sector. In turn, expectations of higher wages lead to the migration of labour from agriculture to industry and the increase in unemployment. It is the wage differential, then, as happens in Lewis's model, which leads to the reallocation of labour among sectors and, ultimately, the dynamics of capital accumulation in the two sectors and in the economy. The paths of capital accumulation and growth strongly depend on the assumptions governing the production functions and the degree of factor mobility. As a consequence, in these models the dynamics is not uniquely determined. Development might, indeed, involve both the

expansion of industry and agriculture, through capital accumulation and labour reallocation, or it can lead to the disappearance of the agriculture sector. Long-run convergence, among other exogenous frictions, depends on the return to scale on each production function. In whatever case, the functional link between sectors, that is labour reallocation, tends to disappear.

In order to avoid possible confusion, it is necessary to stress that we define dualism as not simply the contraposition of two different sectors (or economies) at different stages of development. To have dualism the sectors must be strictly interconnected. Labour migration from one sector to another is one example of such a linkage. If the link dies off, dualism should end as well. Therefore, dualism can change since this linkage changes over time or because the differences among the sectors vary. In a dynamic context, therefore, we should strictly not define an economy as a dual one if there is no such linkage. This specification is necessary when analysing the long-term dynamics of these models. The growth path and convergence of a dual economy is governed by the inter-sectoral speed of adjustment of variables and by the rate of capital accumulation. Therefore, because of development the economy might display the disappearance of the underdeveloped sector and become a one-sector economy. This, by definition, results in the end of dualism. Alternatively, the economy might tend to preserve both sectors but not the linkage between the sectors. Even in this case, under a strict definition, we should speak of a two-sector economy, but not of a dual economy. Finally, there is the possibility that the economy might preserve both the differentiated sectors and the link between them. In the latter case, we are in the presence of a persistent disequilibrium.

In reality, the multiplicity of development paths in dual economies is strictly connected to stability conditions: only persistent instability allows us to observe a relatively slow adjustment in sectoral composition through a relatively slow reallocation of production factors. The large number of traditional dual models are, in fact, models of disequilibrium. It is possible, however, to think of a framework in which dualism persists over time, even under the conditions of a long-term equilibrium (Bartlett, 1983).

Clearly, while instability and dynamic disequilibrium make it very difficult to handle the theory and analytics of the models, it is this condition of persistent disequilibrium that constitutes the core of dualism and allows magnified explanatory powers which are able to take the multifaceted aspects of development into account.

Models of dualism are characterized by asymmetry in resource allocation and by imperfect mobility of factors of production. The long-run dynamics and convergence towards a one-sector economy in which sector contraposition disappears or, instead, the permanence of dualism and the

coexistence of an advanced sector together with an underdeveloped one, depends on the mechanisms of adjustment and on the speed at which resources tend to flow from one sector to another. In such models, asymmetries work similarly to the difference in potential in electric systems: dualism generate the 'potential energy' for the resources to flow where they get the highest remuneration. Such adjustments tend to fill the gaps and to reduce the differential of potential between sectors. These short-term changes in the stocks of resources tend to reduce the degree of dualism in the economy only if it occurs at a slower speed than that at which capital accumulates in the advanced sector. The latter ultimately determines the rate of development in the economy. When factors and resources tend not to move or to move at a slow pace, dualism tends to persist over time. In the latter case, the dynamics certainly become more difficult to determine. This is the reason why these models of traditional dualism focus on adjustments outside the equilibrium, and very few (Jorgenson, 1961, 1967) analyse long-term dynamics and convergence. However, such issues are the focus of the most recent models of growth and dualism.

8.3.   DUALISM AND STRUCTURAL CHANGE IN THE NEW GROWTH THEORY

There is common agreement on the fact that the new growth theory has brought about significant progress in understanding the processes behind growth and economic development. By endogenising technical innovations and highlighting the role of human capital and externalities, this new strand of research has been able to shed new light on economic development and overcome the limitation of Solow's model. Most importantly, the new growth theory has attempted, at least up to a point, to reunite two seemingly irreconcilable disciplines, the classical theory of development and the growth theory, which have long moved along different paths, both in terms of methodology and results. The classical theory of development has attempted to explain the differences in the growth rates of economies by focusing on the importance of savings, educational attainments, institutional frameworks and initial conditions. Yet classical theorists of development have mainly done so by paying little attention to the analytical framework.

Though based on more solid analytical models, growth theory has been unable to capture important aspects of development, such as the role of government and institutions, the role of financial markets, and the role of human capital accumulation. Indeed, these aspects of development, which had been previously and rather extensively analysed by the traditional theory of development, have been the main focus of endogenous growth literature

in recent decades. The literature on endogenous growth, in this sense, has been particularly prolific and it has produced a large number of studies. Yet certain major issues, which lay at the centre of previous studies on development, have been largely neglected by the new growth theory, namely the role of dualism and structural change in the growth process. The reasons for marginalising these issues are merely technical. As already noted, structural change and dualism require restrictions on the proper functioning of market adjustment processes and, often, out-of-equilibrium analysis. Very recently, the literature on growth has attempted to recover these aspects of economic development. Starting with the seminal work by Matsuyama (1991, 1992), a substantial number of studies have engineered the assumption of a dual economy and structural transformation in models of endogenous growth. These studies are not only able to shed new light on the issue of convergence, clarifying the distinction between sectoral and regional convergence, but they are also able to understand the complex dynamics which often appear to characterise the development path of many economies. At the end of the day, questions such as whether dualism is a transitory or long-term phenomenon, whether dualism can explain the growth trap, and if structural change affects growth and convergence, are all important issues which have important policy implications and which have not yet been fully investigated.

The crucial step one has to make in order to introduce structural change in general equilibrium dynamic models is to consider the possibilities of non-balanced growth paths.

These new studies on dualism and endogenous growth can be grouped following different routes. One can focus on the mechanism implemented to introduce dualism, the major frictions and the channels of interaction between variables. Or one can focus, instead, on the phenomenon these studies attempt to depict. Every classification, in any case, would be partial or limitative, since one feature is mixed with the other and often overlaps. Bearing this in mind, we choose to label these studies according to a weak criterion which nonetheless allows us to capture the most interesting features of this literature. The classification is the following. The first group comprises models which assume differentiated outputs and considers as a source of dualism and structural change an exogenous innovation in the advanced sector and Engel's law; the second comprises models which base dualism on technological differences between sectors and considers structural change as generated by exogenous innovation in the less developed sector; the third still considers dualism as defined by technological differences between sectors, but in which technological change is endogenous and is caused by the initial technological differences between sectors; in a fourth group we include models which mainly focus on

unemployment as a source of dualism. Finally, we describe separately the workings of models on dualism and human capital accumulation, since these models can offer the greatest opportunities for research and insights into development.

8.3.1.   Engel's Law as the Engine of Dualism

A central assumption of dualism is the contraposition of a developed sector (industry) to an underdeveloped sector (typically agriculture). Indeed, this is often regarded as the essence of dualism. However, a strand of recent literature has reconsidered dualism in quite a different manner: not focusing on the technology gap between sectors, but on the nature of goods produced. These works still consider the presence in the economy of two sectors; however, now the sectors do not differ in the level of technology advancement, but on the demand elasticity in the goods produced. It may be sufficient to assume, for example, a lower income elasticity of demand for agriculture goods than that for manufactured goods to obtain that, as the economy develops and income increases, the relative weight in the economy of the agriculture sector is decreasing over time. This assumption on the preferences, which is based on well-known observed pattern behaviour of consumers, constitutes the so-called Engel's Law. These preferences, often referred to as Stone-Geary preferences, ensure that as income increases the demand for primary goods increases less than the demand for other goods. As a result, as development occurs, we will observe a restructuring in the economy with a shrinking agriculture sector. The reason, this time, is not the production shift to a more advanced sector, but a change in consumer demand. This result does not require, as in previous studies, a delay in agriculture technology. Technological improvements, wherever they occur, will boost overall growth and development. Technological improvements in the agricultural sector, for example, will lead to an increase in income and an increase in the demand for primary goods which is lower than the increase in the demand for manufactured goods. As a result, a higher proportion of resources can be reallocated to the production of manufactured goods which will expand as well. The impact on the economy and on growth will depend on the production functions in both sectors. Under the assumption of constant returns to scale in the manufacturing sector, and decreasing returns in the agricultural sector, such resource reallocation will increase total factor productivity in the economy and have a positive impact on growth. To illustrate the working of this channel of structural change, we consider a simple two-sector model adapted from Matsuyama (1992).

The j-th (j=1,2) sector economy produces its output with $Y_j = A_j F_j(n_j)$, where A_j is total factor productivity and F_j is an increasing, concave

production function, and n_j is the employment rate, with $n_1+n_2=1$. Consumers have Stone-Geary preferences, $U(c_1,c_2)=\beta\log(c_1-\gamma)+\log(c_2)$ with $\gamma>0$. In competitive equilibrium, the marginal values of labour in the two sectors are equal:

$$A_1F_1'(n)=pA_2F_2'(1-n) \tag{8.1}$$

where p is the relative price of good 2 and $n=n_1=1-n_2$ is the first sector employment share. Since consumer demand satisfies $C_1=\gamma+(\beta p)C_2$, the goods market equilibrium is given by: $A_1F_1=\gamma+A_2F_2(1-n)(\beta p)$. Combining this condition with Equation (8.1) yields:

$$F_1(n)-\frac{\beta F_2(1-n)F_1'(n)}{F_2'(1-n)}=\frac{\gamma}{A_1} \tag{8.2}$$

which implicitly defines n as a decreasing function of A_1, $n=N(A_1)$. By interpreting the first sector as agriculture, and the second as industry, Engel's law offers an explanation for the decline of agriculture and the rise in industry.

Following the seminal works by Murphy et al. (1989) and Matsuyama (1992), other economists – including Echevarria (1997), Laitner (2000), Zweimuller (2000), Kongsamut et al. (2001), Gollin et al. (2002), Greenwood and Uysal, (2005) and Foellmi and Zweimuller (2006) – have developed this idea of a dualism based on Engel's law.

Laitner (2000) develops a model of growth and structural change by relying on a satiation level of agricultural consumption. An increase in agricultural consumption provides no more utility at a certain point. At this stage individuals start to consume manufacturing goods and structural change occurs. Foellmi and Zweimuller (2006) and Zweimuller (2000), by assuming non-homothetic preferences, analyse the effect on growth of the introduction of new goods, which can be consumed only at higher levels of income. This implies that in the first stage of development only a small group of rich people can purchase the production of the advanced sector. However, the innovation increases incomes and this also increases the market size for new goods, as less wealthy people also become willing to buy. As the process of development goes on, the goods produced by the modern sector replace the old goods produced by the traditional sector, structural change occurs and the traditional sector shrinks. Along a similar line of arguments are Greenwood and Uysal (2005) who assume three sectors: agriculture, manufacturing and the new goods sector, a separable utility function in the three types of goods, a fixed cost associated to the purchasing of new goods and technological progress which occurs only in the new goods sector. Under these assumptions they demonstrate that along

the development process there is continuous replacing of new goods with agricultural and manufacturing goods, with a strong decline in the agricultural sector.

Finally, also Gollin et al. (2002) assume hierarchic preferences, but their work differs from the previous papers in a crucial aspect: technological progress occurs, exogenously, in the agricultural sector rather than in the manufacturing one. They specifically assume that the introduction of new, more efficient technologies in agriculture make existing ones obsolete, and reduce the amount of labour required per unit of capital. The excess of labour can thus be employed in the manufacturing sector. The demand for, and production of, manufacturing goods increase under the assumption of non-homothetic preferences. This captures a version of 'the staple theory of growth' (Watkins, 1963), which argues that a productivity primary sector triggers the growth of the industry. Hence, the inability of the agriculture system to provide enough goods to sustain an increasing population employed in the manufacturing sector can explain the delay in development of many economies.

8.3.2.   Models with Technological Gaps Between Sectors

The traditional theory of dualism distinguishes between a technologically advanced sector and a more mature sector employing less advanced technology. The standard approach identifies the former sector with the industrial or the manufacturing sector, the latter with agriculture. However, as already stressed, such a distinction can be generalised and does not necessarily require the contraposition of agriculture to industry. Dualism will still be a useful framework of analysis if one considers any other two separate sectors employing technologies at different stages of development. Following this route, recent studies have generalised the dualism to enclose any framework which sees the contraposition of two interacting sectors with different technologies. The two sectors might still produce the same, or similar, goods, but they have to display different technologies. Coexistence of sectors is ensured by trade barriers, market imperfections or other frictions. The engine of dynamics in these models is the technological innovation which occurs in the advanced sector.

To illustrate this view and the differences with the Engel's law mechanism, let us change the consumer's preferences to $U_1 = C_1 + pC_2$. One interpretation is that there are two goods, the first one is produced in a traditional, land-based sector, and the second one is produced in modern, manufacturing sector. In this case Equation (8.1) determines on its own the equilibrium allocation, which means that n increases with A_1/A_2. Thus

faster productivity growth in the modern sector induces more workers to abandon the traditional sector.

Galor and Weil (2000), Hansen and Prescott (2002) and Temple and Woessman (2006) provide significant examples of such dual economy models in which sectors differ in their prevalent technology. Other features, such as the structure of the labour market or the endogeneity of population growth rates, contribute to shape the economy. In Galor and Weil (2000) as well as in Hansen and Prescott (2002), while the advanced sector is capital-intensive and can benefit from technological innovation, the underdeveloped sector employs predominantly non-accumulable factors (land or other) and is not subject to technical advancements. Given the predominant use of non-accumulable factors, the backward sector will reach decreasing returns more rapidly. This might occur later, or not occur at all, in the advanced sector which instead operates through accumulable factors. Furthermore, the advanced sector benefits from exogenous innovation technology which, combined with an adequate rate of capital accumulation, can bring about constant returns to scale. If this is the case, labour moves from the backward sector to the advanced one. In many instances, as populations grow, the process will only end when the complete reallocation of labour asymptotically reduces to zero the weight of the backward sector in the economy. Hence, in terms of convergence, the dual economy tends to turn into a one-sector economy. For the mechanism involved and the short-term dynamics, these models appear to be very close to traditional dual economy models. Labour reallocation is the channel by which capital can freely accumulate and, together with the assumption of constant returns to scale, it fosters growth and development.

Yet this type of model disregards the institutional framework, which, instead was playing a central role in more traditional dual models. Dualism, and its consequences as the functional connections between sectors, is completely exogenous. It is only determined by initial technological differences and by the exogenous subsequent technical innovations. To some extent, Temple and Woessman (2006) get closer to traditional approach by introducing institutional aspects such as the process of wage setting. Indeed, they assume that the wage is differently determined in the two sectors. While in the advanced sector wage reduction finds a floor in the minimum wage requirement, this limit does not exist in the underdeveloped sector in which the wage can be reduced without boundaries. Several gains surface from such a novelty in modelling dualism. It is not only able to describe the nature of development dynamics focusing on a different channel, that is the speed of inter-sector labour readjustment, but it also introduces the more subtle idea of the importance of the institutional setting

and market structure in determining the nature of dualism and the corresponding dynamics.

In all these studies dualism essentially remains an exogenous phenomenon. A dual economy is assumed at the outset and the analysis is focused on the long-run implications of dualism on growth and development. Is it really fully satisfactory to consider dualism as exogenously given?

8.3.3.   Endogeneity and the Dynamics of Dualism

One of the most interesting questions concerning dualism is whether or not it can be considered an endogenous phenomenon. Is economic dualism an exogenous predetermined condition which is determined by merely historical and geographical factors or, rather, is it a time-varying condition which can be strongly affected by other economic and political factors? This is not merely an academic question since it opens the door to important related policy issues.

An attempt to answer such questions is made by Matsuyama (1991) and Graham and Temple (2001). By assuming increasing returns in the advanced sector due to learning by doing and by introducing Marshallian externalities, these authors provide a dynamic framework which is able to shed new light on the dynamics of dualism itself and on its possible endogeneity. Of course, this also leads to innovative dynamics for the overall economy and it shapes different development and growth paths. Under the assumption of increasing returns in the advanced sector, labour reallocation between sectors will not reduce productivity gaps, but may well increase them by fostering the development of the advanced sector. Indeed, by freeing excess labour resources, the underdeveloped sector allows technological innovation in the advanced sector and its consequent development. This connecting wire which transfers impulses boosting development from the underdeveloped sector to the advanced sector determines the nature of dualism itself in such a framework.

Yet the most interesting consequences of the assumption of increasing returns to scale is the possibility of multiple equilibria. Indeed, both Matsuyama (1991) and Graham and Temple (2001) show that two possible stable equilibria might emerge in such a framework: a low equilibrium with permanent dualism in which there is a predominant underdeveloped sector in the economy; and a high equilibrium in which dualism is overcome by the development of the advanced sector and the disappearance of the underdeveloped sector. The occurrence of one of the two equilibria depends on expectations and initial conditions. As a matter of fact, cumulative causation between growth and technological change establishes a dynamic

interrelationship between technological dualism and structural change. A given technological innovation has the potential to initiate structural change, which in turn spurs technological innovation. On the other hand, a slow increase in productivity and an insufficiently rapid structural change (persistent dualism) can reinforce each other and bring about a vicious circle of underdevelopment. In this case dualism turns into an endogenous phenomenon whose persistence and duration depends on the degree of dualism itself, that is on the technological gap between sectors.

To illustrate this point, let us go back to a version of the above model where the economy has the traditional and modern sectors which produce two perfectly substitutable goods, such that the equilibrium condition is given by Equation (8.1). And let us modify this by assuming that the modern sector is subject to economies of scale, such that its technology is an increasing function of its employment share as: $A_2 = A(1-n)$. For simplicity, let us assume that a higher productivity is due to external economies such that firms take A_2 as given. The equilibrium is then given by:

$$A_1 F_1'(n) = pA(1-n)F_2'(1-n).$$

This could generate multiple equilibria, each of which corresponds to a different level of development, as shown in the Figure 8.1. One of them, *E*, can be considered a state of persistent dualism, where the advanced sector cannot attract many workers because of excessively low productivity. Hence it cannot take advantage of scale economies and the undeveloped sector remains of a considerable size. The other equilibrium, *E'*, is characterized by high productivity and a high employment share in the modern sector. Shifting from *E* to *E'* implies productivity growth and a change in sectoral composition that can be considered as an overcoming of dualism since the undeveloped sector is strongly reduced or even disappears.
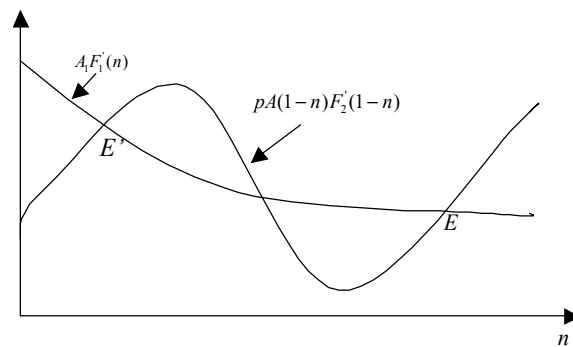

Figure 8.1. Possibility of multiple equilibria

8.3.4.   Unemployment, Migration and Human Capital Accumulation as a Source of Dualism

The functional linkages between sectors, and the resulting dynamics of the economy, represent necessary and qualifying features of dualism. Indeed, factor reallocation between sectors can affect productivities and influence technological change. Labour reallocation is typically what theorists have considered the major engine of dynamics in dual economies. Starting with Lewis, a number of economists have considered labour market segmentation as the most plausible candidate to capture the nature of dualism. Lewis assumes considerable institutional differences in the labour market within the economy: while the labour market obeys market clearing conditions only in the advanced sector, it is unable to clear and function freely in the underdeveloped sector. The reason is the presence of exogenous forces and strong frictions. In Harris and Todaro (1970), although market clearing forces are at work, labour migration is unable to equate wages to labour marginal productivity. The presence of frictions in the adjustment process, such as migration costs or the presence of minimum wages, impedes labour reallocation between sectors and is an obstacle to the elimination of productivity differentials between sectors. This can explain urban unemployment and persistent dualism. The new literature on growth has reintroduced labour market failure and frictions in fully working general equilibrium models in order to explain growth through the lens of structural change in the economy.

Just to quote one of the most recent works showing considerable empirical insight, Caselli and Coleman (2001) attempt to explain the United States development process focusing on the initial dualism between Southern regions, characterized by prevailing agriculture, and Northern ones in which the manufacturing sector is predominant. In this dual economy model, the authors assume that excess of labour supply in the agricultural sector can migrate to the advanced sector only if workers can acquire human capital: the advanced sector can only employ skilled workers. Because of the decreasing cost of education, labour can move from Southern to Northern regions without affecting wages and production in both sectors. Given decreasing returns in the agricultural sector and constant returns in the manufacturing sector, this reallocation of labour leads to an increase in total factor productivity and in growth. As the economy develops, the weight of agriculture in the economy shrinks, there is increasing regional specialization and increasing human capital accumulation.

The same idea that labour reallocation is driven by human capital accumulation is also at the core of Lucas's (2004) work. Lucas assumes that the necessary skills to work in the advanced sector can be acquired only by

migrating to the city: it is only by actually settling in the city that workers can acquire the necessary skills in advanced technology which allow them to be employed in the developed sector. Therefore, in the first stage of development workers will move from the countryside to the city. These workers cannot be immediately employed in production in the advanced sector since they lack the necessary skills. The result is an increase in urban unemployment. However, as the economy develops further and workers acquire more skills, the level of urban unemployment decreases through time. During this last stage of development there will also be convergence between wages in the two sectors. Not necessarily, however, is unskilled labour destined to die off in the city. If one assumes that above a threshold level on the skilled/unskilled labour supply, there is complementarity between skilled and unskilled labour (Lucas, 2004), then above a certain degree of development there could well be positive feedback on the migration flow from the agricultural sector, out of which more workers can move.

8.4.   THE NEW PATH TO DUALISM: HUMAN CAPITAL ACCUMULATION AND MIGRATION

Since Lewis's seminal work, dualism has fed on the idea that labour mobility among sectors can drive the dynamics of capital accumulation and development. This occurs 'because of' structural change and it is a 'cause' of structural change. Put simply, the key engine of dualism and development is labour mobility among sectors. Generalising this idea, one can argue that strong differences among different regions or economies, and the issue of convergence, can also be analysed by means of dualism and through long-term effects of labour migration.

   Indeed the most recent literature on migration, human capital accumulation and development employs the typical features of dual economy to explain the effects of the interplay between migration and human capital accumulation on possible development paths (see for example, Bencivenga and Smith, 1991; Bourguignon and Morrison, 1998; Masson, 2001; Carillo and Marselli, 2003). Particularly interesting in this light is the paper by Masson (2001), which presents a model where rural–urban migration is associated with skills acquisition, but the possibility of being unemployed and credit market constraints may lead to a poverty trap where wealth is no longer sufficient to educate one's children and give them access to higher-skilled jobs. This model can explain persistent urban poverty and rural poverty among those not migrating to the cities.

Vidal's model (1998) also follows this route of explaining structural change and the persistence of dualism by considering the interplay between migration decision and human capital accumulation, stating that incentives to accumulate human capital might come from migration. Higher return to human capital in other economies is an incentive to invest in human capital and to migrate. In turn, human capital accumulation may spark a virtuous cycle and spur technological advancement also in the country or sector of origin. In short, labour migration affects productivity and human capital accumulation and, in turn, return on human capital influences the decision to migrate. Factors that may affect the cost of education and capital accumulation can thus influence development and growth. For instance, credit constraints, income distribution and the level of GDP per capita are all factors that can influence the migration decision and resource allocation.

We think that this line of research is particularly interesting and it is useful, at this stage, to illustrate the new possibilities for analysing development and dualism that comes through the interaction between migration, structural change and human capital accumulation. In order to do so, we will use a simple model in which the decision to migrate is endogenous. This decision influences the level of human capital accumulation and development. In turn, growth and development affect the return on human capital and hence migration flows. To keep things simple, let us consider a two-period overlapping generation economy which comprises two sectors (we can alternatively consider two economies at different stages of development): an underdeveloped sector, *A*, and a developed sector *B*. Production in each sector requires only human capital according to the following:

$$Y_t^A = A(H_t^A)^\alpha \tag{8.3}$$

$$Y_t^B = B(H_t^B)^\alpha \tag{8.4}$$

where H_t^i, $i = A$, B is human capital employed in each sector. Agents are at birth assigned to one of the two sectors (economies) and can move to the other only when old. We assume that each period a population of mass one of agents is assigned to sector-economy *A* and a population of one of agents is assigned to sector-economy *B*. Moreover, in order to keep things as simple as possible, we assume that *B* is the developed sector-economy and that agents might migrate only from *A* to *B*. Agents go to school in the first period of their life and supply labour inelastically in the second. At the beginning of their second period of life they also decide in which sector to work and how much to buy of consumption and education for their children. Human capital accumulation depends on their education, e_{it}, on parents' human capital, h_{it}, and on personal abilities, φ,

$$h_{it+1} = \varphi e_{it}^{\theta} h_{it}^{1-\theta} \tag{8.5}$$

The initial level of human capital is given, $h_{i0} > 0$. Personal abilities are distributed according to the probability distribution function $\Phi(i)$, and hence,

$$\bar{h}_t = \int \varphi_i e_{it-1}^{\theta} h_{it-1}^{1-\theta} d\Phi(i) \tag{8.6}$$

The timing of events is the following. By assumption the decision to migrate concerns only agents in sector-economy *A*. This decision, which depends on the expected income under the two alternatives, is taken at the beginning of their second period of life, before supplying labour. We solve the agents' problem backwardly by first determining the utility they would get under the two alternatives and then by comparing them. Agents derive utility from consumption. They also care about the education and human capital their children will get. Hence, an agent *i* born at time $t-1$ in sector-economy *A* who decides not to move will solve the following problem:

$$Max \ln c_{it} + \gamma \ln h_{it+1}$$

s.t.

$$w_t^A h_{it} = c_{it} + e_{it}$$

$$h_{it+1} = \varphi e_{it}^{\theta} h_{it}^{1-\theta}$$

$$w_t^A, h_{it} \text{ given}$$

where w_t^A is the wage per unit of human capital in sector *A*.

The solution to the above problem entails

$$c_{it}^A = \frac{w_t^A h_{it}}{1+\gamma\theta} \quad ; \quad e_{it}^A = \frac{\gamma\theta}{1+\gamma\theta} w_t^A h_{it} \tag{8.7}$$

By analogy, if he/she moves to sector *B*, agent *i* would solve

$$Max \ln c_{it} + \gamma \ln h_{it+1}$$

s.t.

$$w_t^B h_{it} = c_{it} + e_{it}$$

$$h_{it+1} = \varphi e_{it}^{\theta} h_{it}^{1-\theta}$$

$$w_t^B, h_{it} \text{ given}$$

and obtain

$$c_{it}^B = \frac{w_t^B h_{it}}{1+\gamma\theta} \;\; ; \;\; e_{it}^B = \frac{\gamma\theta}{1+\gamma\theta} w_t^B h_{it} \tag{8.8}$$

The level of consumption and education under the two alternatives depends, as one would have guessed, on the expected wage rate per unit of human capital in the two sectors, w_t^A and w_t^B. In turn, the wage rate in each sector-economy depends on the number of workers employed. Assuming competitive markets, we can write

$$w_t^A = \alpha A (\bar{h}_t N_t^A)^{\alpha-1} N_t^A \tag{9.9}$$

and

$$w_t^B = \alpha B (\bar{h}_t N_t^B)^{\alpha-1} N_t^B \tag{8.10}$$

where N_t^A and N_t^B are the number of workers in each sector. These, of course, depend on the number of agents deciding to move sector. If π_t is the fraction of agents in *A* that decides to move, and recalling that the number of workers in each period per sector is 1, then $N_t^A = 1 - \pi_t$ and $N_t^B = 1 + \pi_t$. It is not so straightforward to determine the equilibrium level of wages and migration flows. Indeed, the decision to move depends on the expected income. In turn, the expected income depends on the number of agents who decide to move. By comparing expected income we can write the following proposition.

Proposition 1
If $w_t^A h_{it} > w_t^B h_{it}$ then no agent will move and therefore $\pi_t = 0$.
If $w_t^A h_{it} < w_t^B h_{it}$ then all agents will move and therefore $\pi_t = 1$.
Only if $w_t^A h_{it} = w_t^B h_{it}$ agents will be indifferent on moving or not and $0 < \pi_t < 1$.

Proposition 1 implies that an equilibrium flow of migration, $\hat{\pi}_t$, requires that $w_t^A(\hat{\pi}_t) = w_t^B(\hat{\pi}_t)$. At each point in time each individual will move with probability $\hat{\pi}_t$. Recalling that there is a unit mass of agents, this is also the fraction of people that move sector-economy. In order to be consistent with the equilibrium, this fraction of people must be such that ex post the wages in the two sectors are the same. If this is not the case, and, for example, $w_t^A h_{it} < w_t^B h_{it}$, then π_t will increase. The opposite happens if $w_t^A h_{it} > w_t^B h_{it}$. Only if the ex ante decision is consistent with the equilibrium wages is the level of agents leaving sector *A* (the probability of moving each agent face) at its equilibrium value. By equating wages we get

$$X = \left[\frac{B}{A} \right]^{\frac{1}{1-\alpha}} = \frac{1+\hat{\pi}_t}{1-\hat{\pi}_t} \tag{8.11}$$

It is clear that if B and A are constant, the migration flow will remain constant over time, that is $\hat{\pi}_t = \hat{\pi}$. This is equivalent to implying persistent dualism. In each period a constant fraction of agents move from the less developed sector to the most developed one. By substituting for e_{it} and averaging human capital, Equation (8.5), across all individuals we obtain the rate of human capital accumulation and growth in the economy

$$\frac{\overline{h}_{t+1}}{\overline{h}_t} = \varphi \left[\frac{\gamma\theta}{1+\gamma\theta} w_t \right]^{\theta} \tag{8.12}$$

where, $w_t^A = w_t^B = w_t$.

The rate of growth evidently depends on $\lim_{t\to\infty} \varphi[\gamma\theta/(1+\gamma\theta)w_t]^{\theta} \lesseqgtr 1$. If $\lim_{t\to\infty} \varphi[\gamma\theta/(1+\gamma\theta)w_t]^{\theta} < 1$ there will be an equilibrium level of average human capital, $\overline{h}_t = \overline{h}$. The system will display a positive rate of growth only if $\lim_{t\to\infty} \varphi[\gamma\theta/(1+\gamma\theta)w_t]^{\theta} > 1$. Although this is a simple model, it is potentially able to display different and interesting dynamics for the economy. For example, if one assumes only externalities in the production functions of the advanced sector, that is $B = B(\overline{h})$, as the literature usually does, then the coefficient ratio X will increase with capital accumulation, and it is easy to show that this will entail an increase in the fraction of people moving sector, $\hat{\pi}_t$. Given that $0 < \pi_t < 1$, this is equivalent to stating that the underdeveloped sector tends to disappear in the long run. If one introduce externalities in sector A as well, that is $A = A(\overline{h})$, the dynamics of the economy will be driven by the dynamics of X. Thus, the model is able to display a degree of dualism which is endogenous in the sense that the migration flow which constitutes the link between sectors closely depends on the interaction between capital accumulation, agents' decisions and the production possibilities set.

8.5. CONCLUSIONS

Structural change and dualism are very often the most important features of development. As economies develop, some sectors tend to expand, others shrink. Technological progress makes some production obsolete and spurs change in consumer preferences and lifestyles. Institutions and market structures are deeply transformed. A comprehensive theory of growth and development needs to take such changes into account. Hence, oversimplified models of growth, albeit very elegant and compact, cannot be fully

explicative of these facts. With the intent to depict these features of development, the most recent literature on growth has attempted to introduce typical aspects of the traditional theory of development on structural change and dualism within a new growth theory framework. The result is an innovative unified model which is able to account for major aspects of development and explain rich and complex growth dynamics.

NOTES

1. In the chapter we often use the term 'growth' interchangeably with the term 'development' well aware that these are usually meant to intend two conceptually different subjects. Yet since we are attempting to bridge the distance between very different strands of literature, which very often also use different terminology, we allow ourselves the licence of adopting these two terms in their wider concept and almost as a synonym.
2. Usually the term 'new growth theory' refers only to models of endogenous growth. More extensively, for our objective, with this term we refer to all recent models of growth which entail both endogenous or exogenous processes of capital accumulation. Indeed, since we focus on dualism and structural change, our interest is in describing the features of the dynamics involving growth traps and steady state per capita capital, as well as endogenous growth.
3. Note that this is not a necessary assumption to have structural change. It only affects the dynamics.

REFERENCES

Bartlett, W. (1983), 'On the Dynamic Instability of Induced-migration Unemployment in a Dual Economy', *Journal of Development Economics*, **22**(13), 85–96.

Bencivenga, V. and B. Smith (1991), 'Financial Intermediation and Endogenous Growth', *Review of Economic Studies*, **58**(2), 195–209.

Bourguignon, F. and C. Morrison (1998), 'Inequality and Development: the Role of Dualism', *Journal of Development Economics*, **57**(2), 223–57.

Carillo, M.R. and R. Marselli (2003), 'Migrazione interna e costi di ricerca: il ruolo della struttura produttiva regionale', *Studi Economici*, **80**(2), 31–51.

Caselli, F. and W.J. Coleman, (2001), 'The U.S. Structural Transformation and Regional Convergence: a Reinterpretation', *Journal of Political Economy*, **109**(3), 584–616.

Chenery, H. and M. Syrquin (1975), *Patterns of Development, 1950–1970*, London: Oxford University Press.

Chenery, H., Robinson, S. and Syrquin, M. (1986), *Industrialization and Growth: a Comparative Study*, New York: Oxford University Press.

Dixit, A. K. (1973), 'Models of Dual Economies', in J.A. Mirrlees and N.H. Stern (eds), *Models of Economic Growth*, New York: Wiley, pp. 325–52.

Echevarria, C. (1997), 'Changes in Sectoral Composition Associated with Economic Growth', *International Economic Review*, **38**(2), 431–52.

Fei, J.C.H. and G. Ranis (1964), *Development of the Labor Surplus Economy: Theory and Policy*, Homewood, IL: Richard D. Irwin.

Fei, J.C.H. and G. Ranis (1966), 'Agrarianism, Dualism and Economic Development', in I. Adelman and E. Thoerbecke (eds), *The Theory and Design of Economic Development*, Baltimore: John Hopkins Press, pp. 3–44.

Foellmi, R. and J. Zweimuller (2006), 'Structural Change, Engel's Consumption Cycles and Kaldor Facts of Economic Growth', University of Zurich, mimeo.

Galor, O. and D. Weil (2000), 'Population, Technology, and Growth: from Malthusian Stagnation to the Demographic Transition and Beyond', *American Economic Review*, **90**(4), 806–28.

Gollin D., S. Parente and R. Rogerson (2002), 'The Role of Agriculture in Development', *American Economic Review*, **92**(2), 160–64.

Graham B.S. and J. Temple (2001), 'Rich Nations, Poor Nations: How much can Multiple Equilibria Explain?', *CEPR Discussion Paper*, no. 3046.

Greenwood J. and G. Uysal (2005), 'New Goods and the Transition to a New Economy', *Journal of Economic Growth,* **10**(2), 99–134.

Hansen, G. and E. Prescott (2002), 'Malthus to Solow', *American Economic Review,* **92**(4), 1205–17.

Harris, J.R. and M.P. Todaro (1970), 'Migration, Unemployment and Development: a Two-sector Analysis', *American Economic Review*, **60**(1), 126–42.

Jorgenson, D.W. (1961), 'The Development of a Dual Economy', *Economic Journal*, **71**(282), 309–334.

Jorgenson, D.W. (1967). 'Surplus Agricultural Labour and the Development of a Dual Economy', *Oxford Economic Papers*, **19**(3), 288–312.

Kanbur, S.M.R. and J. McIntosh (1988), 'Dual Economy Models: Retrospect and Prospect', *Bulletin of Economic Research*, **40**(2), 83–114.

Kongsamut P., S. Rebelo and D. Xie (2001), 'Beyond Balanced Growth', *Review of Economic Studies*, **68**(4), 869–82.

Laitner, J. (2000), 'Structural Change and Economic Growth', *Review of Economic Studies*, **67**(3), 545–61.

Lewis, W.A. (1954), 'Economic Development with Unlimited Supplies of Labour', *The Manchester School*, **22**(2), 139–91.

Lucas R. (2004), 'Life Earnings and Rural–Urban Migration', *Journal of Political Economy*, **112**(1), 29–59.

Masson P.R. (2001), 'Migration, Human Capital, and Poverty in a Dual-Economy Model of a Developing Country', *International Monetary Fund Working Paper 01/28*.

Matsuyama K. (1991), 'Increasing Returns, Industrialization, and Indeterminacy of Equilibrium', *Quarterly Journal of Economics*, **106**(2), 617–50.

Matsuyama, K. (1992), 'Agriculture Productivity, Comparative Advantage, and Economic Growth', *Journal of Economic Theory*, **58**(2), 317–34.

Murphy K., A. Shleifer and R. Vishny (1989), 'Industrialization and the Big Push', *Journal of Political Economy*, **97**(5), 1003–26.

Sen, A.K. (1966), 'Peasants and Dualism with and without Surplus Labor', *Journal of Political Economy*, **74**(5), 425–50.

Taylor, M.P. (1969), *Structuralist Macroeconomics*, New York: Basic Books.

Temple J. and L. Woessmann (2006), 'Dualism and Cross-Country Growth Regressions', *Journal of Economic Growth*, **11**, 187–228.

Todaro, M.P. (1969), 'A Model of Labour Migration and Urban Unemployment in Less Developed Countries', *American Economic Review*, **59**(1), 138–48.

Uzawa, H. (1961), 'On a Two-sector Model of Economic Growth: I', *Review of Economic Studies*, **29**(1), 40–47.

Uzawa, H. (1963), 'On a Two-sector Model of Economic Growth: II', *Review of Economic Studies*, **30**(2), 115–18.

Vidal, J.P. (1998), 'The Effects of Migration on Human Capital Formation', *Journal of Population Economics*, **11**(4), 589–600.

Watkins, M. (1963), 'A Staple Theory of Economic Growth', *Canadian Journal of Economics and Political Science*, **29**(2), 141–58.

Zweimuller, J. (2000), 'Schumpeterian Entrepreneurs Meet Engel's Law: The Impact of Inequality on Innovation-Driven Growth', *Journal of Economic Growth*, **5**(2), 185–206.