ANNALS OF ECONOMICS AND FINANCE **4**, 219–229 (2003)

Reflections on Growth and Development*

Robert M.Solow

1. INTRODUCTION

Economists share an intuitive grasp of the distinction between growth and development. When I write the words and you read them, we understand roughly the same thing. It is much harder to be precise about the relation between the two. This is probably because most dimensions of economic growth are quantifiable, while the phrase "economic development" seems to refer to more complex qualitative aspects of social and economic organization that are hardly measurable and in fact only imperfectly definable. The same pattern occurs elsewhere, for example in the growth and development of children. We measure their growth directly, through height and weight or even vocabulary. But child development involves the emergence of capabilities and behavior patterns whose very description may be problematic and disputed.

This contrast probably explains why there is such a well-developed body of evolving theory about growth, well enough established to be written up in textbooks and taught to undergraduates. The study of economic development, on the other hand, has produced a much smaller amount of theory; and the theory that exists is much less systematic,and certainly less cumulative. Descriptions and discussions of economic development tend to be more politicized (for want of a more accurate word). This is not an aberration; it seems to come with the territory.

2. SOCIAL INSTITUTIONS AND ECONOMIC GROWTH

There is obviously a two-way connection between growth and development. Growth clearly encourages and facilitates development, if only because the costs of transition and possible dislocation are less burdensome when aggregate income is rising anyway. Conversely, development in certain directions may be a necessary foundation for durable, stable growth. In any case, the institutions and attitudes that arise in the course of de-

* This is the introduction written for the China Development Report, 2004.

1529-7373/2002

velopment certainly affect the growth path of the economy. An important question is:what are those favored directions? That is where complication and what I called politicization come into the picture.

One disciplined way to elucidate the connections between development and growth is to start with a simple, clearly stated model of growth, and look for the ports of entry through which the facts of development can influence the growth path generated by the model.

The standard model of economic growth surely has its limitations, the most obvious one being its high degree of aggregation. There must be aspects of economic growth that can not be adequately captured in a model with one, two or three sectors. Growth theory is limited to those that can. More specifically, the standard model is essentially an elaboration of the aggregative balance equation that says: output (as limited by technology and available inputs of labor services, capital services, and perhaps natural resources) must provide for (private and public) consumption, (private and public) net investment, and depreciation.

All the other assumptions must find their place in that framework .They and the discipline imposed by the balance equation determine the growth path.

So where do the capabilities, attitudes and institutions that emerge from economic development fit in to this picture? The answer appears to be : almost everywhere. The fact that current output is limited by known technology and available inputs is usually expressed in an aggregate production function, $Y = F(S_k, S_l, R; T)$, according to which total output or value added (Y) depends on the services of capital and labor, the use of natural resources (R), and the level of "total factor productivity" (indexed by T). Business-cycle effects are conventionally ignored; they belong to a different branch of economics.

The production function is tacitly taken as fact of technology, and this is harmless if the model is applied to a society whose institutions are changing slowly or not at all. But it can be a distortion in a process of rapid development. For example:

(a) Distortions due to monopoly or restrictions imposed by convention, religion, caste rules, gended roles, and other institutions can keep the level of output below the purely technological limit.

(b) The same sorts of norms and rules can limit the services obtainable from a given supply of capital and number of workers, by placing restrictions on hours and days of work, seasonal activities, degrees of effort, willingness to cooperate and make decisions, and so on.

(c) The use of natural resources may be especially affected by social convention, especially in societies still close to the land.

(d) Attitudes toward instrumental rationality can affect the use made of "known" technology, and thus total factor productivity.

(e) It is worth separate emphasis that **changes** in total factor productivity, one of the main sources of economic growth, can be strongly affected by institutional factors, not only through the willingness of a society to adopt new technology, but also through its ability and willingness to create and adopt organizational innovations; of the two, organizational innovation may be more directly affected by social institutions , attitudes and norms, especially because new technology can be acquired simply by imitation, whereas many organizational and management innovations may need to be home-grown in light of local conditions.

Clearly, then, the speed and direction of economic development can have a strong influence on the level and growth of aggregate output. The story does not end there, however. I have been talking about the relation between development and the **source** of output; there are also connections to the **uses** of output. According to the basic balance equation, current output has to be allocated among consumption, investment, and governmental uses (and also net exports). Much of economics is devoted to analyzing just this process. It is obvious that the direction of economic development is a major causal factor governing the uses of output. Attitudes toward the future, including perceptions of risk, will influence the savings decisions of household. The scope provided for enterprise, along with those same perceptions of risk, will help determine the volume of private investment. In both respects, the capacity of government to tax, and the nature of the tax system, will also figure importantly in private decisions. And what the government does with its revenues will also change as the development process unfolds.

At this point the mechanism of growth theory takes over. The allocation decisions taken this year will affect the supply side of the economy next year. Most obviously, private and public investment, in conjunction with depreciation, determine the amount and kind of industrial capital and social infrastructure available for next period. Some part of government expenditure is devoted to education at various levels. In the short run, education may draw people out of the labor force; in the longer run, the quantity and quality of education is an important determinant of the skill level of the labor force and thus the quality and quantity of the productive services delivered by workers.

These various considerations tend to strain the highly aggregative character of the underlying model of growth. For example, ongoing technological change will alter the relative costs of consumption goods and investment goods. Rising income per person will change the particular goods consumed. Both such changes will affect the mix of skills needed from workers, and perhaps the location of production and thus the need for transportation. All such compositional affects are ignored in a completely aggregated model. It is possible to disaggregate slightly without compromising the

transparency and comprehensibility of the model. One can distinguish, say, consumption goods and capital goods, and assign them to separate productive sectors. Or one can do the same for agricultural and industrial goods.

It is also natural to wonder whether an economy with the immense area and population of China can be analyzed without an explicit geographical dimension. To put the point more picturesquely: can the same modelling straty work both with Uruguay and with China? There is not enough experience with regionally differentiated models to suggest an answer. The U.S. economy has been successfully studied without any explicit geographical dimension. The U.S. is bigger than Uruguay, but not as big as China. Perhaps some experimentation is required. Soon enough experiments in growth modelling will be carried out for the European Union, and we will see what happens when once separate national economies are aggregated.

Just to mention one researchable question: whether a geographically large economy can usefully be analyzed as an integrated whole must depend on the extent to which relative factor prices differ from one area to another. We know that there is some regional variation in the U.S., even with an excellent transportation and communication infrastructure. Regional variation within the European Union is presumably diminishing with the end of trade barriers and obstacles to the mobility of capital and labor, and the beginning of a common currency. It would be interesting to know the corresponding facts for China.

At this level of generality, the nature of the growth-development connection is not very controversial. The hard questions and controversial answers arise when one asks: how exactly do the various choices that can be made in the course of development affect the path of economic growth? Are some choices distinctly superior from the point of view of growth, and are there other overriding considerations? That is what I had in mind when I described discussions of development as almost inevitably "politicized".

It will certainly be a matter of great scientific interest, to say the least, when it will be possible to do serious growth modelling of the Chinese economy in this spirit. I do not know if it will ever be possible to make model-based comparisons of pre-reform and post-reform China from the point of view of growth theory. For economists, that would be a little like acquiring some photographs taken through a new and more powerful telescope. Second-best would be model-based comparisons among China, the U.S. and the European Union: this is a second-best alternative because it will not be so straightforward to infer the significance of institutional difference brought about by reform.

3. ECONOMIES OF SCALE AND TOTAL FACTOR PRODUCTIVITY

The main obstacle to achieving these goals is probably the unavailability of long time series, and perhaps also the quality of the underlying data where they exist. I am not acquainted with the repertoire of research on the Chinese economy, but in one relevant study that I have seen, the time series of outputs and inputs for pre-reform China covers only the period from 1953 to 1977. That is long enough for macroeconomic purposes, but probably not for growth-modelling. The interval beginning with the reform of 1978 is of equal length, but of course it gets longer every year.

Would it be possible to combine the two periods to try to detect the effects of reform on the growth path? That depends mainly on the comparability of data before and after reform, and on that question I can have no opinion.

Merely to illustrate the importance of these questions, and without any intention of substituting my judgment for that of experts, I will mention one example that I happen to have seen: the study "Natural Decomposition of Total Factor Productivity Growth," by Zhou Fang of the Institute of Quantitative and Technical Economics of the Chinese Academy of Social Sciences (Chapter 6, pages 291-306, in *Econometric Modelling of China,* edited by Lawrence R. Klein and Shinichi Ichimura, World Scientific Publishing, Singapore, 2000).

Zhou Fang studies both the pre-reform (1953-77) and post-reform (1978-96) periods, but entirely separately, using a Cobb-Douglas production function with only capital and labor as inputs. (The post-reform data are said to be for the "total economy;" if agriculture is included, the omission of land may be significant. The pre-reform data are for the "industrial economy".) Zhou Feng does not impose constant returns to scale. The result is that for 1978-96 the sum of the capital and labor elasticities is 1.6, a finding of very strong increasing returns to scale. (For the earlier period the sum is 1.07, thus close to constant returns to scale. This sharp difference between the two periods is worth careful investigation.)

Zhou Fang does indeed make the "natural" decomposition. If the estimated elasticities of capital and labor are a and b with $a+b>1$, then the constant-returns-to-scale part of the production function is assigned elasticities $a/(a+b)$ and $b/(a+b)$, and the increasing-returns part is assigned the remaining $a(a+b-1)/(a+b)$ and $b(a+b-1)/(a+b)$. This procedure leads to a remarkable conclusion. The observed growth in output between 1978 and 1996, namely 9.34 percent per year, is partitioned between growth in inputs *as if there were constant returns to scale* (5.70 percent per year) and increasing returns to scale (3.64 percent per year), leaving nothing to be accounted for by technological progress.

The analogous exercise for 1953-77 leads to the same conclusion: no growth imputed to pure technological progress, and essentially everything imputed to economies of scale.(Zhou Fang identifies economies of scale with "embodied technological progress", but that is not the way that concept is usually understood.) I think that what has happened here is that the growth of measured inputs of capital and labor was so strong that they dominate the underlying regressions. (Capital multiplied by 16 times between 1953 and 1977, and by another five times between 1978 and 1996; labor input multiplied by more than five times in the first period and grew by another 60 percent in the second.) If a time trend had been inserted into those regressions, the collinearity with the input series would no doubt have resulted in very poor estimates of coefficients. But leaving out what might be a "true" time trend will cause it to be absorbed by the fast and smoothly growing inputs, with the appearance of increasing returns to scale.

The other obvious alternative would have been to impose constant returns to scale, and then include a time trend. The result would certainly have been a major imputation to technological progress or total factor productivity, and the fit would have been just about as good. I do not want to argue for either alternative, but rather to point out that the difficulty inheres in analysis of a few strongly trend-affected time series, especially short ones. As time goes on, one can hope that a little more independent variation will show itself. As the data are now, they contain little or no usable information about the relation between institutional development and aggregative growth.

The point can be made more sharply with a little algebra. The basic finding is that if the production function $Y = K^a L^b$ is estimated, $a+b$ turns out to exceed one. Let $c = a+b$. Then always $K^a L^b = [K^{\frac{a}{c}} L^{\frac{b}{c}}][K^{\frac{a}{c}} L^{\frac{b}{c}}]^{c-1}$. The left-hand bracket is the constant-returns part; the right-hand bracket is the same expression raised to a positive power. Since $c-1 > 0$, the right-hand bracket, evaluated for China or for any rapidly-growing economy, will be a strongly increasing function of time. If we write the product as MN, where M is what Zhou Fang calls "extensive growth of output" and N is "economies of scale", then the growth rate of N will be $c-1$ times the growth rate of M, and the two growth rates will add up to the growth rate of output. That is exactly what Zhou Fang finds.

The finding is that, while output grew at about the same rate in the two sub-periods, about 93.5 percent of it was extensive growth pre-reform and 60 percent of it was extensive growth post-reform, with the rest coming from economies of scale. This corresponds exactly to the different estimates of c for the two periods. I think it is odd to suppose that the effect of reform was primarily to shift the source of growth sharply in the direction of increasing returns to scale. I would be more inclined either to question the data, or at least to try a decomposition into constant-returns-to-scale

growth plus a pure time-effect, and then to decompose the latter into effects of increasing returns and a remainder. It is precisely in the remainder that the footprints of reform — or of institutional change in general — might be found.

This is the procedure actually followed in the chapter entitled "ICSEAD's Econometric Model of the Chinese Economy(1997 Version", by Yoshihisa Inada, in the volume cited earlier. This model estimates Cobb-Douglas production functions for three sectors of the Chinese economy: primary industry, state-owned manufacturing, non-state-owned manufacturing, and tertiary industry. In each case constant returns to scale is imposed; and a time-trend is allowed for the primary and tertiary sectors. But the model is estimated only for the post-reform period.

4. LEVELS AND GROWTH RATES

Even in comparisons among the older industrial economies, there has been a tendency to focus too much on differences in growth rates of output, and to focus too little on differences in levels of output. Comparisons of growth rates inevitably carry overtones suggesting implicitly that the growth rates are almost indefinitely sustainable. In fact steady-state growth rates have to be inferred, uncertainly, from long time series and specific models. I suspect that comparisons of growth rates between countries at different stages of development are even more likely to be problematic.

Just to be clear, I will state very baldly what I have in mind. Imagine a country with neither population growth nor growth in total factor productivity. Its trend growth rate is zero. Now suppose it experiences a fairly large but one-time agricultural or industrial innovation. Aggregate output may grow rapidly for quite a long time, until the new technology has been fully exploited. But then the zero-growth trend will reassert itself. Attempts to analyze this episode in terms of growth rates will probably go wrong. The situation is no different if the country has a positive long-term growth rate. A one-time improvement in technology or organization will allow and encourage a faster growth rate for some time, but only for that time.

For this kind of reason, I think it may be more fruitful to identify and analyze the sources of differences in productivity itself, and not its rate of change, among countries at different stages of development, and often among countries at the same stage of development. This is as true for comparisons between France and Germany on one hand and Greece and Portugal on the other as it is for comparisons between the U.S. and the U.K. or between China and Japan or China and the U.S. And it is as true for comparisons industry by industry as for comparisons on the scale

of the economy as a whole. (In the latter case, differences in industrial composition will be one factor to be considered.)

The most important reason for paying attention to levels is that countries relatively new to modern industry can and do imitate what is already available at the technological frontier elsewhere. The implication is that growth, say from one five or ten year period to the next, is dependent primarily on the ability to absorb already known technology and to provide the physical and human capital needed to translate it into productive capacity. The same can be said for organizational innovation (where human and cultural factors may be even more important). These possibilities are of course ultimately limited by production possibilities in the most advanced economies, but the limit may be far away. This is rather different from the situation facing an economy already at the technological frontier. The important comparisons are fundamentally between two situations, not between growth rates.

This general point has even broader scope. Microeconomic research indicates that there are often substantial and persistent differences in productivity within a given industry as between two advanced economies, like the U.S. and France or Germany. These difference can not be ascribed either to differences in technological know-how or to differences in the availability of capital, tangible or intangible. Sometimes the lower-productivity industry is more capital intensive than the other. The inference is that there is simply **slack** or inefficiency: known technology and existing capital are not being fully exploited. One possible reason is the absence of powerful competitive pressure, domestic or foreign, from rivals or from the capital markets or from other agencies. Owners and managers, unless they are pushed, may not get the most out of their productive capacity. There are other possible explanations, for example the sorts of legal or customary restrictions that were mentioned earlier as stemming from different paths of "development".

Persistent differences in productive efficiency will usually be classified by standard methods as differences in total factor productivity. But then they ought to be analyzed as such. The search for international differences in the level of total factor productivity could lead to explanations in terms of organizational and institutional characteristics, and these in turn might suggest either corrective policies or perhaps lead to the conclusion that they are not after all differences in efficiency but instead reflect local situational values that are not reflected in conventional measurement of output. I do not remember having seen such an argument carefully made, but it is clearly a logical possibility.

5. EXOGENOUS AND ENDOGENOUS IN GROWTH THEORY

In applications of the simple neoclassical growth model, the custom has been to treat the population and labor force along with the state of technology as exogenous variables. The endogenous variables are then the demand-side division of output between consumption and saving-investment, the stock of capital, and the level of output itself.(Often government uses of output are ignored altogether; otherwise they are classified as exogenous.)

It is generally taken as a sign of intellectual progress to endogenize the previously exogenous. One would think that the most obvious candidate for endogenization would be the size of the population; and the next step would be to deal explicitly with the supply of labor. Of course there are clear precedents for this, going back to Malthus in the case of demography itself and to the traditional theory of labor-supply to be found in every textbook. Instead, it is the level of technology that attracted most of the attention of endogenizers. The hardest problem has been tackled first. Perhaps this because the Malthusian mechanism is not really relevant in the advanced industrial world, and that is the main context in which growth theory has been elaborated.

When the topic is the interplay between growth and development, however, it seems appropriate to interchange the position of cart and horse. Endogenous technological change matters much less for the reason already mentioned: the relevant technology is already in existence to be learned and adopted. In some cases foreign direct investment is the natural vehicle; in others, direct imitation is suitable. The limiting factor for growth is more likely to be the need for capital equipment designed for modern techniques, and for managers, engineers, technicians and skilled operators of the new equipment, and maybe marketers. Whether or not birth and death rates are to be treated as endogenous or exogenous, the mechanisms and incentives for producing and reproducing human capital are certainly an endogenous part of the economic system to be modelled.

This puts the question of incentives squarely in the center of the picture. It is not always the case that private entrepreneurs or public bureaucrats are motivated to assume the risks involved in the development process. I mentioned earlier the finding that, even in rich, advanced industrial economies, competitive forces are sometimes too weak to generate economic efficiency. What begins as regulation sometimes ends as protection. (It sometimes begins as protection too.) State bureaucracies are subject to the same temptation-the best of all monopoly profits is a quiet life, as John Hicks remarked-especially since the rewards for success may be — **may** be — smaller.

So I would think that the natural endogenous mechanisms to explore are those that govern which pieces of advanced technology are adopted, and how fast they are adopted. The answer to the first question should be fairly routine, as soon as we understand the goals of the decision-makers. The answer to the second question will presumably turn on the availability of funds to be directed into suitable investment , and the availability of workers, technicians and managers with the right skills. In both cases, the full set of incentives is obviously important, as is the set of institutional arrangements, habits, attitudes and expectations that I have been lumping together under the general heading of alternative "development paths".

6. CROSS-COUNTRY ANALYSIS

In recent years, the notion of an "empirical growth model" has come to mean, not an econometric investigation of the long-term path of a single economy,but an investigation in which the basic observations come from a cross-section of many national economies over a common time-interval. Typically the thing to be explained is the average rate of growth of real GDP per person over that interval, and the explanatory variables are a list of geographical, demographical, political and socio-cultural indexes. (Each country's initial level of income is also included to allow for the fact that a country is likely to move more quickly toward its steady state level if starts far below it.) The estimated coefficients in a multiple regression can then be interpreted as the causal effect of each explanatory variable on the country's (steady-state) growth rate.

There is now an enormous literature of this kind. Robert Barro, one of the pioneers, discusses it in his contribution to this volume. The question I want to ask is: what do these regressions mean? The fact that there may be an element of reverse causality is well understood by the authors. If a measure of political stability is positively associated with the growth rate, comparing one country with another, it is not hard to believe both that political stability favors growth and that growth favors political stability. Even if one accepts that possibility, it is difficult to separate those two causal chains empirically, and in fact I am very doubtful that it has been done successfully. But that is not the main point I want to emphasize here.

Perhaps I can state the issue clearly by posing a different question: if you wanted to predict the growth rate for China, would you use such a regression equation? That is, would you collect the appropriate values for China of the list of explanatory variables and simply apply the regression coefficients estimated (without China) from a collection of countries? I doubt it, and I hope that is not simply because I already know something about recent growth rates in China.

I would be more inclined to believe that something could be learned from looking at the deviation of China (or any other country) from such a regression; but my uneasiness goes further than that.

Suppose we understood such a regression as a purely descriptive statement. It would then be telling us that the joint distribution (across countries) of growth rates and the various explanatory variables tends to cluster around the plane described by the regression. Is it sensible to imagine the countries we observe as some kind of sample from a universe of "possible" countries? That seems very odd to me. If not, then we have to look for an even simpler "merely" descriptive interpretation.

After all, it is useful to know — if it is true — that more open economies have tended to grow faster than roughly similar countries that are less open to trade, or that the same can be said of more politically stable of less regulated economies, and so on down list. I have earlier emphasized that these same questions might better be asked about difference in levels of productivity or total factor productivity than about differences in growth rates. I am not sure that the cross-country regression is the best way, or even a reasonable way, to understand those differences. This is because the notion of a "development path" seems more complex in my view. But almost any approach to exploration can be useful. If one merely expects the cross-country regression to suggest questions rather than answers, that would be more acceptable. And it would then on longer be an interesting exercise to worry about the "right" formulation of the regression.