ECONOMIC GROWTH

Economic growth is the increase in a country’s standard of living over time. Growth economists study how living standards differ across countries as well as across time. This article discusses some of the broad facts of economic growth and some of the main approaches to its study.

Economic growth is typically measured as the change in per-capita Gross Domestic Product (GDP). Sustained long-term economic growth at a positive rate is a fairly recent phenomenon in human history, most of it having occurred in the last 200 years. According to Maddison’s (2001) estimates, per-capita GDP in the world economy was no higher in the year 1000 than in the year 1, and only 53% higher in 1820 than in 1000, implying an average annual growth rate of only one nineteenth of one percent over the latter 820 year period. Some time around 1820, the world growth rate started to rise, averaging just over one half of one percent per year from 1820 to 1870, and peaking during what Maddison calls the “golden age”, the period from 1950 to 1973, when it averaged 2.93% per year. By 2000, world per-capita GDP had risen to more than 8 ½ times its 1820 value.

Growth has been uneven not only across time but also across countries. Since 1820, living standards in Western Europe and its offshoots in North America and the Antipodes have raced ahead of the rest of the world, with the exception of Japan, in what is often referred to as the “Great Divergence”. As shown in Figure 1 below, the proportional gap in per-capita GDP between the richest group of countries and the poorest group (as classified by Maddison) grew from 3 in 1820 to 19 in 1998. Pritchett (1997) tells a similar story, estimating that the proportional gap between the richest and poorest countries grew more than five-fold from 1870 to 1990.

[Insert Figure 1 Here]
This widening of the cross-country income distribution seems to have slowed during the second half of the twentieth century, at least among a large group of nations. Indeed, Figure 1, which is drawn on a proportional scale, shows that with the acceleration of growth in Asia there has been a narrowing of the spread between the richest and the second poorest group since 1950. Evans (1996) shows a narrowing of the top end of the distribution (that is, among OECD countries) over the period. However, not all countries have taken part in this convergence process, as the gap between the leading countries as a whole and the very poorest countries has continued to widen. In Figure 1 the gap between the Western Offshoots and Africa grew by a factor of 1.75 between 1950 and 1998. Likewise the proportional income gap between Mayer-Foulkes's (2002) richest and poorest convergence groups grew by a factor of 2.6 between 1960 and 1995.

Jones (1997) argues that continuing divergence of the poorest countries from the rest of the world does not imply rising income inequality among the world’s population, mainly because China and India, which contain about 40% of that population, are rising rapidly from near the bottom of the distribution. Indeed Sala-i-Martin (2006) shows, using data on within-country income distributions, that the cross-individual distribution of world income narrowed considerably between 1970 and 2000, even as the cross-country distribution continued to widen somewhat. But between-country inequality is still extremely important; in 1992 it explained 60% of overall world inequality (Bourguignon and Morrison, 2002). Another reason that growth economists are typically more concerned with the cross-country than the cross-individual distribution is that many of the determinants of economic growth vary across countries but not across individuals within countries.

The Production Function Approach

The main task of growth theory is to explain this variation of living standards across time and countries. One way to organize one’s thinking about the sources of growth is in terms of an aggregate production function, which indicates how a country’s output per worker \( y \) depends on the (per worker) stocks of physical, human, and natural capital, represented by the vector \( k \), according to

\[
y = f(k, A),
\]
where $A$ is a productivity parameter. Economic growth, as measured by the growth rate of $y$, depends therefore on the rate of capital accumulation and the rate of productivity growth. Similarly, countries can differ in their levels of GDP per capita either because of differences in capital or because of differences in productivity. Much recent work on the economics of growth has focused on trying to identify the relative contributions of these two fundamental factors to differences in growth rates or income levels among countries.

Modern growth theory started with the neoclassical model of Solow (1956) and Swan (1956), who showed that in the long run growth cannot be sustained by capital accumulation alone. In their formulation, the diminishing marginal product of capital (augmented by an Inada condition that makes the marginal product asymptote to zero as capital grows) will always terminate any temporary burst of growth in excess of the growth rate of labor-augmenting productivity. But this perspective has been challenged by the more recent endogenous growth theory. In the $AK$ theory of Frankel (1962) and Romer (1986), growth in productivity is functionally dependent on growth in capital, through learning by doing and technology spillovers, so that an increase in investment rates in physical capital can also sustain a permanent increase in productivity growth and hence in the rate of economic growth. In the innovation-based theory that followed $AK$ theory, the Solow model has been combined with a Schumpeterian theory of productivity growth, in which capital accumulation is one of the factors that can lead to a permanently higher rate of productivity growth (Howitt and Aghion, 1998).

**Capital**

Having introduced the production function in a general sense, we now examine the accumulation of different types of capital in more detail, and then turn to an assessment of the relative importance of factor accumulation and productivity in explaining income differences among countries and growth over time.

*Physical Capital*

Physical capital is made up of tools, machines, buildings, and infrastructure such as roads and ports. Its key characteristics are, first, that it is produced (via investment), and second that it is in turn used in producing output. Physical capital differs importantly
from technology (which, as is discussed below, is also both produced and productive) in that physical capital is rival in its use: only a limited number of workers can use a single piece of physical capital at a time.

Differences in physical capital between rich are poor countries are very large. In the year 2000, for example, physical capital per worker was $148,091 in the United States, $42,991 in Mexico, and $6,270 in India. These large differences in physical capital are clearly contributors to income differences among countries in a proximate sense. That is, if the US had India’s level of capital it would be a poorer country. The magnitude of this proximate effect can be calculated by using the production function. For example, using a value for capital’s share of national income of 1/3 (which is consistent with the findings of Gollin (2002) for a cross section of countries), the ratio of capital per worker in the US to that in India would by itself explain a ratio of income per capita in the two countries of 7.9 (=($148,091/6,270)^{1/3}).

Differences in physical capital among countries can result from several factors. First, countries may differ in their levels of investment in physical capital relative to output. In an economy closed to external capital flows, the investment rate will equal the national saving rate. Saving rates can differ among countries because of differences in the security of property rights, due to the availability of a financial system to bring together savers and investors, because of government policies like budget deficits or PAYGO old age pensions, differences in cultural attitudes toward present vs. future consumption, or simply because deferring consumption to the future is a luxury that very poor people cannot afford.

A second factor that drives differences in investment rates among countries is the relative price of capital. The price of investment goods relative to consumption goods is two to three times as high in poor countries as in rich countries. Measuring both output and investment at international prices, investment as a fraction of GDP is strongly correlated with GDP per capita, (correlation of 0.50) and poor countries have on average between ½ and ¼ the investment rate of rich countries. When investment rates are expressed in domestic prices, the correlation between investment rates and GDP per capita falls to 0.05 (Hsieh and Klenow, forthcoming).
But levels of capital can also differ among countries for reasons that have nothing to do with the rate of accumulation. Differences in productivity (the $A$ term in equation 1) will produce different levels of capital even in countries with the same rates of physical capital investment. Similarly, differences in the accumulation of other factors of production will produce differences in the level of physical capital per worker.

**Human capital**

Human capital refers to the qualities of such as education and health that allow a worker to produce more output and which themselves are the result of past investment. Like physical capital, human capital can earn an economic return for its owner. However, the two types of capital differ in several important respects. Most significantly, human capital is “installed” in a person. This makes it very difficult for one person to own human capital that is used by someone else. Human capital investment is a significant expense. In the United States in the year 2000, spending by governments and families on education amounted to 6.2% of GDP; forgone wages by students were of a similar magnitude.

Information on the productivity of human capital can be derived from comparing wages of workers with different levels of education. So called “Mincer regressions” of log wage on years of education, controlling by various means for bias due to the endogeneity of schooling, yield estimated returns to schooling in the neighborhood of 10% per year. In the year 2000, average schooling of workers in advanced countries was 9.8 years and among workers in developing countries was 5.1 years. Applying a rate of return of 10% implies that the average worker in the advanced countries supplied 56% more labor input because of this education difference. If labor’s share in a Cobb-Douglas production function is 2/3, this would imply that education differences would explain a factor of 1.35 difference in income between the advanced and developed countries, which is very small relative to the observed gap in income. Allowing for differences in school quality increases somewhat the income differences explained by human capital in the form of schooling.

A second form of human capital is health. The importance of health as an input into production can be estimated by looking at microeconomic data on how health affects
individual wages. Health differences between rich and poor countries are large, and in wealthy countries worker health has improved significantly over the last 200 years (Fogel, 1997). Weil (forthcoming) using the Adult Survival Rate as a proxy for worker health, estimates that eliminating gaps in worker health among countries would reduce the log variance of GDP per worker by 9.9%.

Natural Capital

Natural capital is the value of a country’s agricultural and pasture lands, forests, and subsoil resources. Like physical and human capital, natural capital is an input into production of goods and services. Unlike other forms of capital, however, it is not itself produced.

Natural capital per worker and GDP per worker are positively correlated, but the link is much weaker than for the other measures of capital discussed above. The poor performance of many resource rich countries has led many observers to identify a “resource curse” by which the availability of natural capital undermines other forms of capital accumulation or reduces productivity. Among the suggested channels by which this happens are that resource booms lead countries to raise consumption to unsustainable levels, thus depressing saving and investment (Rodriguez and Sachs, 1999); that exploitation of natural resources suppresses the development of a local manufacturing sector, which holds back growth because manufacturing is inherently more technologically dynamic than other parts of the economy (this is the so called Dutch disease); and that economic inefficiencies are associated with political competition or even civil war to appropriate the rents generated by natural resources.

Population and Economic Growth

Population affects the accumulation of all three forms of capital discussed above, and through them the level of output per worker. Rapid population growth dilutes the quantities of physical and human capital per worker, raising the rates of investment and school expenditure required to maintain output per worker. The interaction of natural capital with population growth is at the center of the model of Malthus (1798). For a
fixed stock of natural capital, higher population lowers output per capita. Combined with a positive feedback from the level of income to population growth, this resource constraint produces a stable steady state level of output per capita and, with technology fixed, a stable level of population as well. This Malthusian feedback is the explanation for the long period of nearly constant living standards that preceded the Industrial Revolution (Galor and Weil, 2000). Because of resource-saving technological progress, as well as expansion of international trade, which allows countries evade resource constraints, the interaction of population and natural capital is much less important today than in the past, with the exception of very poor countries that are reliant on subsistence agriculture.

In addition to its effect on the level of factors of production per worker, population also matters for economic growth because demographic change produces important changes in the age structure of the population. A reduction in fertility, for example, will produce a long period of reduced dependency, in which the ratio of children and the elderly, on the one hand, to working age adults, on the other, is temporarily below its sustainable steady state level. This is the so-called “demographic dividend.” (see Weil, this volume).

In addition to these effects of population on the level of income per capital, there is also causality that runs from the economic to the demographic. Over the course of economic development, countries generally move through a demographic transition in which first mortality rates and then fertility rates fall. While the decline in mortality is easily explained as a consequence of higher income and technological progress, the decline in fertility is not fully understood. Among the factors thought to contribute to the decline in fertility are falling mortality, a shift along a quality-quantity tradeoff due to rising returns to human capital, the rise of women’s relative wages, the reduced importance of children as a means of old age support, and improvements in the availability of contraception.

**Growth Accounting and Development Accounting**

The discussion above makes clear that stocks of different forms of capital are positively correlated with GDP per capita. Similarly, as countries grow, levels of capital
per worker grow as well. It is natural to ask whether these variations in capital are sufficiently large to explain the matching variations in growth. The techniques of growth accounting (Solow, 1957) and development accounting (Klenow and Rodriguez-Clare, 1997; Hall and Jones, 1999) attempt to give quantitative answers to this question. Using a parameterized production function and measures of the quantities of human and physical capital, one can back out relative levels of productivity among countries and rates of productivity growth within a country.

Caselli (2005) presents a review of development accounting along with his own thorough estimates. His finding is that if human and physical capital per worker were equalized across countries, the variance of log GDP per worker would fall by only 39%. In other words the majority of variation in income is due to differences in productivity, not factor accumulation. Differences in productivity growth, rather than differences in the growth of physical and human capital, are also the dominant determinants of differences in income growth rates among countries (Weil, 2005, chapter 7; Klenow and Rodriguez-Clare, 1997); Differences in productivity levels among countries are striking. For example, comparing the countries at the 90th and 10th percentiles of the income distribution (which differ in income by a factor of 21), the former would produce seven times as much output as the latter with equal quantities of human and physical capital.

**Productivity, Technology, and Efficiency**

Development accounting shows that productivity differences among countries are the dominant explanation for income differences. Similarly, differences in productivity growth are the most important explanation for differences in income growth rates among countries. And as a theoretical matter, the Solow model shows that as long as there are decreasing returns to capital per worker, productivity growth can be the only source of long-term growth. The question is: what explains these changes over time and differences in the level of productivity? Over the long-term it is natural to associate productivity growth with technological change. However, especially as an explanation for differences in productivity at a point in time, a second possibility is that productivity differences reflect differences not in technology, in the sense of inventions, blueprints,
and so on, but rather differences in how economies are organized and use available technology and inputs. We label this second contributor to productivity as “efficiency.”

**Technology**

Technology consists of the knowledge of how to transform basic inputs into final utility. This knowledge can be thought of as another form of capital, an intangible intellectual capital. What distinguishes technology from human or physical capital is its non-rival character. For example, the knowledge that a particular kind of corn will be immune to caterpillars, or the knowledge of how to produce a 3 GHz CPU for a portable computer, can be used any number of times by any number of people without diminishing anyone’s ability to use it again. By contrast, if you drive a truck for an hour, or if you employ the skills of a doctor for an hour, then that truck or those skills are not available to anyone else during that hour.

Different growth theories have different approaches to modeling the accumulation of technology, i.e. technical progress. According to neoclassical theory, for example, the relationship between technology and the economy is a one-way street, with all of the causation running from technology to the economy. It portrays technical progress as emanating from a scientific progress that operates outside the realm of economics, and thus takes the rate of technical progress as being given exogenously.

This neoclassical view has never been accepted universally. Specialists in economic history and the economics of technology have generally believed that technical progress comes in the form of new products, new techniques and new markets, which do not spring directly from the scientific laboratory; instead they come from discoveries made by private business enterprises, operating in competitive markets, and motivated by the search for profits. For example, the transistor, which underlies so much recent technological progress, was discovered by scientists working for AT&T on the practical problem of how to improve the performance of switch boxes that were using vacuum tubes. Rosenberg (1981) describes many other examples of scientific and technological breakthroughs that originated in profit-oriented economic activity.

What kept this view of endogenous technology from entering the mainstream of economics until recently was the difficulty of incorporating increasing returns to scale
into dynamic general equilibrium theory. Increasing returns arise once one considers technology as a kind of capital that can be accumulated, because of its non-rival nature. That is, the cost of developing a technology for producing a particular product is a fixed setup cost, which does not have to be repeated when more of the product is produced. Once the technology has been developed then there should be at least constant returns to scale in the factors that use that technology, on the grounds that if you can do something once then you can do it twice. But this means that there are increasing returns in the broad set of factors that includes the technology itself. Increasing returns creates a problem because it generally implies that a competitive equilibrium will not exist, at least not without externalities.

These technical difficulties were overcome by the new “endogenous growth theory” that was introduced by Romer (1986) and Lucas (1988), which incorporated techniques that had been developed for dealing with increasing returns in the theories of industrial organization and international trade. The first generation of endogenous growth theory to enter the mainstream was the “AK theory” according to which technological progress takes place as a result of externalities in learning to produce capital goods more efficiently. The second generation was the innovation-based theory of Romer (1990) and Aghion and Howitt (1992), which emphasizes the distinction between technological knowledge and other forms of capital, and analyzes technological innovation as a separate activity from saving and schooling.

Historically, technical progress has engendered much social conflict, because it involves what Schumpeter (1942) called “creative destruction.” That is, new technologies render old technologies obsolete. As a result, technical progress is a game with losers as well as winners. From the handloom weavers of early 19th century Britain to the former giants of mainframe computing in the late 20th century, many people’s skills, capital equipment and technological knowledge have been devalued and their livelihoods imperiled by the same innovations that have created fortunes for others.

The destructive side of technical progress shows up most clearly during periods when a new “General Purpose Technology” (GPT) is being introduced. A GPT is a basic enabling technology that is used in many sectors of the economy, such as the steam engine, the electric dynamo, the laser or the computer. As Lipsey, Carlaw and Bekar
(2005) have emphasized, a GPT typically arrives only partially formed, creates
technological complementarities and opens a window on new technological possibilities.
Thus it is typically associated with a wave of new innovations. Moreover, the period in
which the new GPT is diffusing through the economy is typically a period of rapid
obsolescence, costly learning and wrenching adjustment. Greenwood and Yorukoglu
(1997) argue that the productivity slowdown of the 1970s is attributable to the arrival of
the computer, and Howitt (1998) argues that the rapid obsolescence generated by a new
GPT can cause per capita income to fall for many years before eventually paying off in a
much higher standard of living.

New technologies are often opposed by those who would lose from their
introduction. Some of this opposition takes place within the economic sphere, where
workers threaten action against firms that adopt labor-saving technologies and firms try
to pre-empt innovations by rivals. But much of it also takes place within the political
sphere, where governments protect favored firms from more technically advanced foreign
competitors, and where people sometimes vote for politicians promising to preserve
traditional ways of life by blocking the adoption of new technologies.

The leading industrial nations of the world spend large amounts on R&D aimed at
generating innovations. In the United States, for example, R&D expenditures constituted
between 2.2 and 2.9 percent of GDP every year from 1957 through 2004. But not much
cutting-edge R&D takes place outside these leading countries. In 1996, for example, over
90% of the world’s R&D expenditure, as measured by UNESCO, was done in just 8
countries. In the majority of countries that do very little measured R&D, technology
advances not so much by making frontier innovations as by implementing technologies
that have already been developed elsewhere. But the process of implementation is not
costless, because technologies tend to be context-dependent and technological knowledge
tends to be tacit. So implementation requires an up-front investment to adapt the
technology to a new environment (See, for example, Evenson and Westphal, 1995) This
investment plays the same role analytically in the implementing country as R&D does in
the original innovating country.

Implementation is important in accounting for the patterns of cross-country
convergence and divergence noted above. This is because a country in which firms are
induced to spend on implementation have what Gerschenkron (1952) called an “advantage of backwardness”. That is, the further they fall behind the world’s technology frontier the faster they will grow with any given level of implementation expenditures, because the bigger will be the improvement in productivity when they implement any given foreign technology. In the long run, as Howitt (2000) has shown, this force can cause all countries that engage in R&D or implementation to grow at the same rate, while countries in which firms are not induced to make such investments will stagnate. But technology transfer through implementation expenditures is no guarantee of convergence, because the technologies that are being developed in the rich R&D-performing countries are not necessarily appropriate for conditions in poor implementing countries (Basu and Weil, 1998; Acemoglu and Zilibotti, 2001) and because financial constraints may prevent poor countries from spending at a level needed to keep pace with the frontier (Aghion, Howitt and Mayer-Foulkes, 2005).

*Efficiency*

The efficiency with which a technology is used is not likely to play a major role in accounting for long-run growth rates, because there is a finite limit to how high you can raise living standards simply by using the same technologies more efficiently. But there is good reason to believe that differences in efficiency account for much of the cross-country variation in the level of productivity.

Inefficiencies take several different forms. Economic resources are sometimes allocated to unproductive uses, or even unused, as when union featherbedding agreements kick in. Resources can be misallocated as the result of taxes, subsidies and imperfect competition, all of which create discrepancies between marginal rates of substitution. Technologies can be blocked by those who would lose from their implementation and have more market power or political influence than those who would win.

The distinction between differences in technology and differences in efficiency is often unclear. Suppose firms in country A are using the same machinery and the same number of workers per machine as in country B, but output per worker is higher in A than B. This may appear to be an obvious case of inefficiency, since the technology embodied
in the machines used by workers in the two countries is the same. But maybe it is just that people in country B lack the knowledge of how best to use the machines, in which case it may actually be a case of differences in technology. As an example, General Motors has has little success in their attempts to emulate the manufacturing methods that Toyota has deployed successfully even in their U.S. operations for many years.

Moreover, identical technologies will have different effects in different countries, because of differences in language, raw materials, consumer preferences, workers expectations and the like. Euro Disney, for example, was plagued initially with labor disputes when it first opened its park in the outskirts of Paris in 1987. It took the American managers several years to realize that the problem was not recalcitrant workers but rather that French workers consider it an intolerable indignity to be forced to wear items such as mouse ears when serving the public. A minor adjustment in amusement park technology was needed to eliminate make Disney’s amusement park technology as productive in France as it had been in the United States.

*Deeper determinants of growth*

Even if we knew how much of the cross-country variation in growth rates or income levels to attribute to different kinds of capital or to technology or efficiency, we would still be faced with the deeper question of why these differences in capital and productivity arise. A large number of candidate explanations have been offered in the literature. These candidates can be classified into four broad categories: geography, institutions, policy and culture.

Geographical differences are perhaps the most obvious. As Sachs (2003) has emphasized, countries that are landlocked, that suffer from a hazardous disease environment and that have difficult obstacles in the way of internal transportation, will almost certainly produce at a lower level than countries without these problems, even if they use the same technology and the same array of capital. In addition, the lower productivity of these countries will serve to reduce the rate of return to accumulating capital and to generating new technologies,

Institutions matter because of the way they affect private contracts and also because of the way they affect the extent to which the returns to different kinds of
investments can be appropriated by the government. The origin of a country’s legal system has been shown by La Porta et al (1998) to have an important effect on private contracts. In particular they show that countries with British legal origins tend to offer greater protection of investor and creditor rights, which in turn is likely to affect both capital accumulation and investment in technology by making outside finance more easily available.

Because long-term productivity growth requires technical progress, it depends on political, institutional and regulatory factors that affect the way the conflict between the winners and losers of technical progress will be resolved, and hence affect the incentives to create and adopt new technologies. For example, the way intellectual property is protected will affect the incentive to innovate, because on the one hand no one will want to spend resources creating new technologies that his or her rivals can easily copy, while on the other hand a firm that is protected from competition by patent laws that make it difficult for rivals to innovate in the same product lines will be under less pressure to innovate. Likewise, a populist political regime may erect barriers to labor-saving innovation, resulting in slower technical progress.

Economic policies matter not only because of the way they affect the return to investing in capital and technology but also because of the inefficiencies that can be created by taxes and subsidies. But how these policies affect economic growth can vary from one country to another. In particular, Aghion and Howitt (2006) have argued that policies that are growth promoting in technologically advanced countries are not necessarily growth promoting in poorer countries, because innovation and implementation are affected differently by the same variables. For example, tighter competition policy in a relatively backward country might retard technology development by local firms that will be discouraged by the threat of foreign entry, whereas in more advanced countries firms will be spurred to make even greater R&D investments when threatened by competition.

As this example suggests, international trade is one of the policy domains most likely to matter for growth and income differences, because of the huge productivity advantage that is squandered by policies that run counter to comparative advantage, because protected firms tend to become technologically backward firms, and because for
many countries international trade is the only way for firms to gain a market large enough to cover the expense of developing leading edge technologies. So it is probably no accident that export-promotion has been a prominent feature of all the East Asian countries that started escaping from the lower end of the world income distribution towards the end of the 20th Century, whereas import-substitution was a prominent feature of several Latin American countries that fell from the upper end of the distribution early in the 20th Century.

Culture is a difficult factor to measure. In principle, however, it is capable of explaining a great deal of cross-country variation in growth, because a society in which people are socialized to trust each other, to work hard, to value technical expertise and to respect law and order is certainly going to be thriftier and more productive. Recent work has begun to quantify the role of culture using measures of social capital, social capability, ethno-linguistic fractionalization, religious belief, the spread of Anglo-Saxon culture and many other variables.

References


