Economic Growth and Energy

DAVID I. STERN
Rensselaer Polytechnic Institute
Troy, New York, United States

Glossary

capital There are two meanings in economics. One is the monetary value of assets owned by an individual firm, etc. or the amount that is available to invest. The second meaning, used herein, refers to a human-made input into the production process, including things such as buildings and machinery, and infrastructure (telecommunications systems and the like). Capital in these forms is also described as reproducible capital, manufactured capital, or human-made capital. Some writers use the term "capital" more widely, to include human capital and/or environmental or natural capital. Human capital refers to expertise or know-how embodied in people through processes of education and training. The term "natural capital" has been used to refer to the stock of environmental resources.

demand elasticity The percentage response in the quantity of a good demanded for a 1% increase in its price.

diminishing returns If use of an input to production is increased, holding the amounts of the other inputs constant, the law of diminishing returns states that eventually the marginal product of an input will decline, though not necessarily becoming negative.

ecological economics A transdisciplinary field including ideas from economics, natural sciences, and other social sciences, in which scholars and practitioners are interested in the broad question of the use of resources by society and in understanding the nature, behavior, and dynamics of the economy–environment system. Many, but not all, ecological economists are highly critical of the mainstream (neoclassical) school of economics.

econometrics The specialized branch of mathematical statistics applied to the analysis of economic data.

economic growth An increase in economy-wide economic production, usually measured by an increase in gross domestic product (GDP); also, the process of the economy growing over time.

elasticity of substitution A measure of the responsiveness of the ratio of the quantities of two productive inputs to a change in the relative prices of the quantities. For example, if the price of energy increases 1% relative to the price of capital and the amount of energy used per dollar of capital decreases by 1%, then the elasticity of substitution is 1. If the decrease in energy use is less, then the elasticity of substitution is less than 1, and vice versa.

embodied energy Used directly in the production of goods and services, energy is also used indirectly through its involvement in producing the other inputs to production. The sum of the two uses is called the embodied energy of the good or service.

externality A negative externality, or external effect, occurs when a production or consumption activity has unintended damaging effects, such as pollution, on other firms or individuals, and no compensation is paid by the producer or consumer to the affected parties. A positive externality occurs when activities have beneficial effects for others who do not pay the generator of the effect, such as freely available research results.

institutions Commonly accepted rules, practices, etc. that govern human social and economic behavior, such as laws, customs, markets, etc.

marginal product If use of an input to production is increased, holding the amounts of the other inputs constant, the increase in output produced is called the marginal product of that input.

neoclassical economics The mainstream school or paradigm of economics characterized by mathematical models based on optimization by individual consumers and producers. A paradigm that arose in the late 19th century and emerged to its dominant position after the World War II; other paradigms include Keynesian, Marxist, Institutional, and neo-Ricardian, and post-Keynesian models.

nonrenewable resource A natural resource such as petroleum, which has zero natural growth over ordinary human timescales; though it may be replenished over geological time. Renewable resources such as forests grow or regenerate at rates that are relatively fast on human timescales.

production function A mathematical representation of the quantitative relationship between productive inputs and the level of output of some good or service, expressing the maximum quantity of output that can be obtained from any particular amounts of inputs.

productivity An index of economic output relative to input. Increases in productivity are often seen as being due to improvements in technology but may also be due to other factors. The most general productivity index is known as total factor productivity. Labor productivity is just output per worker hour and is the measure usually discussed in the financial media.

public goods The consumption of a public good by one person, which does not reduce the amount available for others and once provided to some people, other people cannot be excluded from consuming them.

substitution Shifting the mix of inputs used to produce a given quantity of output. Substitution occurs among already known production methods and does not imply that innovations are introduced.

sustainability As used herein, nondeclining individual consumption or utility over time; many other definitions exist.

technological change The invention and introduction of new methods of production as well as new products; the main focus in economics and herein is on the former.

utility The level of satisfaction (or happiness) of an individual. In a more aggregate context, we refer to the social welfare of a group of people, which is an indicator related to the utilities of the individuals in society.

Resource and ecological economists have criticized the theory that energy plays a minor role in economic growth on a number of grounds, especially the implications of thermodynamics for economic production and the long-term prospects of the economy. Although a fully worked out alternative model of the growth process does not seem to exist, extensive empirical work has examined the role of energy in the growth process. The principal finding is that energy used per unit of economic output has declined, but that this is to a large extent due to a shift in energy use from direct use of fossil fuels such as coal to the use of higher quality fuels, and especially electricity. When this shift in the composition of final energy use is accounted for, energy use and the level of economic activity are found to be tightly coupled. When these and other trends are taken into account, the prospects for further large reductions in the energy intensity of economic activity seem limited. These findings have important implications for environmental quality and economic and environmental policy.

1. INTRODUCTION

This article surveys the relation between energy and economic growth, and more generally the role of energy in economic production. Business and financial economists pay significant attention to the impact of oil and other energy prices on economic activity, but the mainstream theory of economic growth pays little or no attention to the role of energy or other natural resources in promoting or enabling economic growth. The extensive discussions concerning the “productivity slowdown” following the 1970s oil crises were an exception.

This article is structured to cover these key points in a systematic fashion. In the first section there is a review of the background theory of production and growth from different points of view—those based in economics and those based in the natural sciences. The starting premise is that gaining an understanding of the role of energy in economic growth cannot be achieved without first understanding the role of energy in production. The scientific basis of the role of energy in production, and hence also in the increasing scale of production involved in economic growth, is considered first. However, institutional phenomena also affect how this role plays out and therefore the economic view of growth and production and of the potential role of energy is necessarily more complex than just this scientific understanding. The mainstream theory of economic growth is, therefore, reviewed next. The limitations of its consideration of energy and other resource issues have been the subject of strong criticism grounded in the biophysical theory of the role of energy. A review of these alternative viewpoints completes this first section of the article.

The mainstream economics concept of the production function is next used to examine the key factors that could reduce or strengthen the linkage between energy use and economic activity over time. This production theory is very general and is less subject to criticism than are the specific models of economic growth. These key factors are (1) substitution between energy and other inputs within an existing technology, (2) technological change, (3) shifts in the composition of the energy input, and (4) shifts in the composition of economic output. Each of these themes will be discussed in detail.

Numerous ideas and views exist about the potential linkages between energy and economic growth. Choice between these theories has to be on the basis of both inherent plausibility and consistency and, perhaps more crucially, empirical evidence. Therefore, studies that investigate the strength of the linkage between energy and growth are reviewed. To be useful, such studies must not be grounded in a single theory, potential mechanism, or school of thought. Therefore, the studies reviewed here do not specify structural linkages between energy and output. Because correlation and regression analysis does not imply causality from one variable to another, most of these studies employ the econometric notions of Granger causality and cointegration to test the presence of and direction of causality between the variables.

Finally, the implications of the theory for environmental quality are examined and the recent empirical literature on that topic is discussed.

2. THEORY OF PRODUCTION AND GROWTH

2.1 Energy in Production: Physical Theory and Economic Models

Reproducibility is a key concept in the economics of production. Some inputs to production are nonreproducible, whereas others can be manufactured at a cost within the economic production system. Primary factors of production are inputs that exist at the beginning of the period under consideration and are not directly used up in production (though they can be degraded and can be added to), whereas intermediate inputs are created during the production period under consideration and are used up entirely in production. Mainstream economists usually think of capital, labor, and land as the primary factors of production, and goods such fuels and materials as the intermediate inputs. The prices paid for all the different inputs are seen as eventually being payments to the owners of the primary inputs for the services provided directly or embodied in the produced intermediate inputs. In the theory of growth, this approach has led to a focus on the primary inputs, in particular on capital and land, and a much lesser and somewhat indirect treatment of the role of energy in the growth process. The primary energy inputs are stock resources such as oil deposits. But these are not given an explicit role in the standard growth theories, which focus on labor and capital. However, capital, labor, and, in the longer term, even natural resources are reproducible factors of production, whereas energy is a nonreproducible factor of production, although, of course, energy vectors (fuels) are reproducible factors. Therefore, natural scientists and some ecological economists have placed a very heavy emphasis on the role of energy and its availability in the economic production and growth processes.

The first law of thermodynamics (the conservation law) implies the mass-balance principle. In order to obtain a given material output, greater or equal quantities of matter must enter the production process as inputs, with the residual as a pollutant or waste product. Therefore, there are minimum material input requirements for any production process producing material outputs. The second law of thermodynamics (the efficiency law) implies that a minimum quantity of energy is required to carry out the transformation of matter. All production involves the transformation or movement of matter in some way, and all such transformations require energy. Therefore, there must be limits to the substitution of other factors of production for energy. Energy is also an essential factor of production. Though all economic processes require energy, some service activities may not require the direct processing of materials. However, this is only true at the micro level. At the macro level (economy-wide level), all economic processes require the indirect use of materials, either in the maintenance of labor or in the production of capital goods.

Some aspects of organized matter—that is, information—might also be considered to be nonreproducible inputs. Several analysts argue that information is a fundamentally nonreproducible factor of production in the same way as energy. Energy is necessary to extract information from the environment, but energy cannot be made active use of without information and possibly accumulated knowledge. Unlike energy, information and knowledge cannot be easily quantified. But these latter factors of production must be incorporated into machines, workers, and materials in order to be made useful. This provides a biophysical justification for treating capital, labor, etc. as factors of production. Though capital and labor are easier to measure than information and knowledge, their measurement is, still, very imperfect compared to that of energy.

In the mainstream neoclassical economics approach (discussed later), the quantity of energy available to the economy in any period is endogenous, though restricted by biophysical constraints (such as the pressure in oil reservoirs) and economic constraints (such as the amount of installed extraction, refining, and generating capacity), and the possible speeds and efficiencies with which these processes can proceed. Nevertheless, this analytical approach leads to a downplaying of the role of energy as a driver of economic growth and production. Some alternative, biophysical models of the economy propose that energy is the only primary factor of production. This could be understood as there being a given stock of energy that is degraded (but, due to the law of the conservation of energy, not used up) in the process of providing services to the economy. But this means that the available energy in each period needs to be exogenously determined. In some biophysical models, geological constraints fix the rate of energy extraction. Capital and labor are treated as flows of capital consumption and labor services rather than as stocks. These flows are computed in terms of the embodied energy use associated with them and the entire value added in the economy is regarded as the rent accruing to the energy used in the economy. The actual distribution of the surplus depends on the relative bargaining power of the different social classes and foreign suppliers of fuel. Therefore, the owners of labor, capital, and land appropriate the energy surplus. The production process in the economy as a whole can be represented as an input–output model as originally developed by Nobel Laureate Wassily Leontief and adapted in the energy context by Bruce Hannon and others. The input–output model specifies the required quantities of each input needed to produce each output, with each output potentially an input to another process. It is also possible to trace all final production back to its use of a single primary factor of production, in this case energy.

These ecological economists argue that the energy used to produce intermediate resource inputs such as fuels increases as the quality of resources (such as oil reservoirs) declines. This is like negative productivity growth or technological change. Thus, changing resource quality is represented by changes in the embodied energy of the resources rather than by changes in the input–output coefficients. If resource stocks were explicitly represented, energy would no longer be the only primary factor of production. The neo-Ricardian models developed by Charles Perrings and Martin O'Connor, like all other neo-Ricardian models, have a fixed-proportions technology in terms of capital stocks instead of the flows in the Leontief model. They do not distinguish between primary and intermediate factors of production. Yet that approach can still take biophysical constraints such as mass balance and energy conservation into account.

If the economy can be represented as an input–output model where there is no substitution between factors of production, the embodied knowledge in factors of production can be ignored. An accurate accounting for all energy used to support final production is important. But in the real world the contribution of knowledge to production cannot be assumed to be proportional to its embodied energy. Though thermodynamics places constraints on substitution, the actual degree of substitutability among capital stocks embodying knowledge and energy is an empirical question. Neither the Leontief nor the neo-Ricardian model allows substitution between inputs. The neoclassical production models that are considered next, however, do.

2.2 The Mainstream Theory of Growth

As already explained, there is an inbuilt bias in mainstream production and growth theory to downplay the role of resources in the economy, though there is nothing inherent in economics that restricts the potential role of resources in the economy. The basic model of economic growth is the Nobel prize-winning work by Robert Solow that does not include resources at all. This model subsequently was extended with nonrenewable resources, renewable resources, and some waste assimilation services. These extended models, however, have only been applied in the context of debates about environmental sustainability, not in standard macroeconomic applications.

2.2.1 The Basic Growth Model

Economic growth models examine the evolution of a hypothetical economy over time as the quantities and/or the qualities of various inputs into the production process and the methods of using those inputs change. Here, the simplest model based on the work of Solow is described. In this model, a constant-sized labor force using manufactured capital produces output in terms of gross domestic product (GDP). The model assumes that output increases at a decreasing rate as the amount of capital employed rises. The uppermost curve in Fig. 1 shows this relationship between output (Y) and capital (K). This is termed diminishing returns to capital. It is assumed that a constant proportion, s, of the output is saved and invested in the capital stock. A constant proportion of the existing capital stock depreciates (becomes productively useless) in each period of time. The capital stock is in equilibrium (and so unchanging in size) when saving equals depreciation. This is also shown in Fig. 1. The dynamics implied by Fig. 1 are very simple. To the left of K^*, where capital per worker is scarce, capital investment generates a relatively large increase in future income, and so will

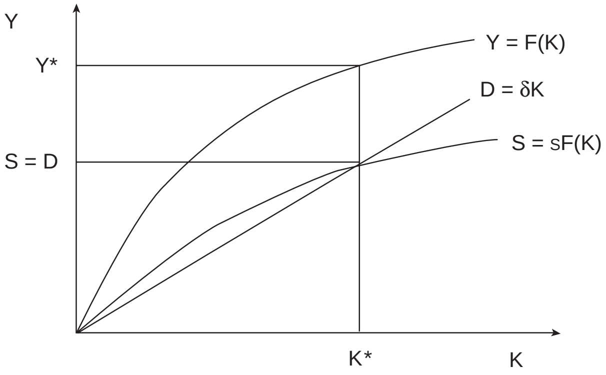

FIGURE 1 The neoclassical growth model.

offer a high rate of return. Moreover, it is clear from the relative positions of the *S* and *D* curves to the left of *K** that the addition to the capital stock (*S*) is greater than depreciation (*D*) and so capital rises. When the capital stock has reached *K**, it will be at a stationary, or equilibrium, state. Additions to capital due to saving are exactly offset by reductions in capital from depreciation.

This simple economy must sooner or later reach a stationary state in which there is no net (additional) investment and economic growth must eventually halt. In a transition process, while a country is moving toward this stationary state, growth can and will occur. An underdeveloped economy, with a small capital stock per worker, can achieve fast growth while it is building up its capital stock. But all economies will eventually settle into a zero growth equilibrium if the savings rate remains constant. No country can grow in perpetuity merely by accumulating capital, due to the diminishing returns to capital accumulation. If the savings rate is increased, growth will occur for a while until a new equilibrium is reached. Of course, the higher the savings the rate, the lower the current standard of living of the population.

If the labor force grows at a fixed rate over time, the total capital stock and the total quantity of output will rise, but capital per worker and output per worker will remain constant once an economy has developed to its equilibrium. The only necessary adjustment to Fig. 1 is that all units are now measured in per capita terms.

According to neoclassical growth theory, the only cause of continuing economic growth is technological progress. As the level of technological knowledge rises, the functional relationship between productive inputs and output changes. Greater quantities or better qualities of output can be produced from the same quantity of inputs. In the simple model being examined, technological progress continually shifts the output function upward, and so raises the equilibrium per capita capital stock and output levels. Intuitively, increases in the state of technological knowledge raise the rate of return to capital, thereby offsetting the diminishing returns to capital that would otherwise apply a brake to growth.

2.2.2 Endogenous Technological Change

The simple model just described does not explain how improvements in technology come about. They are just assumed to happen exogenously, so that this model is said to have exogenous technological change. More recent models attempt to endogenize technological change, explaining technological progress within the growth model as the outcome of decisions taken by firms and individuals. In endogenous growth models, the relationship between capital and output can be written in the form $Y = AK$. Capital, *K*, is defined more broadly than in the neoclassical model. It is a composite of manufactured and knowledge-based capital. Endogenous growth theorists have been able to show that, under reasonable assumptions, the *A* term in the preceding expression is a constant, and so growth can continue indefinitely as capital is accumulated.

The key point is that technological knowledge can be thought of as a form of capital. It is accumulated through research and development (R&D) and other knowledge-creating processes. Technological knowledge has two special properties. First, it is a public good: the stock of this form of capital is not depleted with use. This is important because it implies that the knowledge stock can be stored over time, even when it is being used. Second, it generates positive externalities in production: although the firm doing R&D obtains benefits from the knowledge acquired, others benefit too—the benefits that the firm accrues when it learns and innovates are only partly appropriated by itself. There are beneficial spillovers to the economy from the R&D process so that the social benefits of innovation exceed the private benefits to the original innovator. These externalities create momentum in the growth process. As firms install new capital, this tends to be associated with process and product innovations. The incentive to devote resources to innovation comes from the prospect of temporary monopoly profits for successful innovations. The growth of *K* thus means the growth of a composite stock of capital and disembodied technological knowledge. Therefore, output is able to rise as a constant proportion (*A*) of the composite capital stock, and is not subject to the diminishing returns shown in Fig. 1.

So, in an endogenous growth model, the economy can sustain a constant growth rate in which the diminishing returns to manufactured capital are exactly offset by the technological growth external effect just described. The growth rate is permanently influenced by the savings rate; a higher savings rate increases the economy's growth rate, not merely its equilibrium level of income.

2.2.3 Growth Models with Natural Resources

The growth models examined so far do not include any natural resources. All natural resources exist in finite stocks or flows, though some, such as sunlight or deuterium, are available in very large quantities. Some environmental resources are nonrenewable, and many renewable resources are potentially exhaustible. Finiteness and exhaustibility of resources make the notion of indefinite economic growth problematic. Even maintaining the current level of output indefinitely—so-called sustainable development—may be difficult.

When there is more than one input (both capital and natural resources), there are many alternative paths that economic growth can take. The path taken is determined by the institutional arrangements that are assumed to exist. Analysts have looked at both optimal growth models that attempt to either maximize the sum of discounted social welfare over some relevant time horizon (often an infinite horizon) or achieve sustainability (nondeclining social welfare) and models intended to represent real economies with perfectly competitive markets or other arrangements.

The neoclassical literature on growth and resources centers on what conditions permit continuing growth, or at least nondeclining consumption or utility. Technical and institutional conditions determine whether such sustainability is possible. Technical conditions refer to things such as the mix of renewable and nonrenewable resources, the initial endowments of capital and natural resources, and the ease of substitution among inputs. The institutional setting includes things such as market structure (competition versus central planning), the system of property rights (private versus common property), and the system of values regarding the welfare of future generations.

In a 1974 paper published in a special issue of *Review of Economic Studies*, Solow showed that sustainability was achievable in a model with a finite and nonrenewable natural resource with no extraction costs and nondepreciating capital, which was produced using capital and the natural resource. However, the same model economy under competition results in exhaustion of the resource and consumption and social welfare eventually falling to zero. In another paper in the same special journal issue, Dasgupta and Heal showed that with any constant discount rate the so-called optimal growth path also leads to eventual depletion of the natural resource and the collapse of the economy. A common interpretation of this theory is that substitution and technical change can effectively decouple economic growth from resources and environmental services. Depleted resources or degraded environmental services can be replaced by more abundant substitutes, or by "equivalent" forms of human-made capital (people, machines, factories, etc.). But this is a misinterpretation. Neoclassical economists are primarily interested in what institutional arrangements, and not what technical arrangements, will lead to sustainability, so that they typically assume a *priori* that sustainability is technically feasible, and then investigate what institutional arrangements might lead to sustainability if it is technically feasible. However, there is a tendency among mainstream economists to assume that sustainability is technically feasible unless proved otherwise.

The elasticity of substitution (σ) between capital (factories, machines, etc.) and inputs from the environment (natural resources, waste assimilation, ecosystem services) is a critical technical term that indicates by how much one of the inputs must be increased to maintain the same level of production, when the use of the other input is reduced. A large value of σ implies that this substitution is easy, and vice versa. Figure 2 shows the different combinations of two inputs that can produce a given level of output for different values of σ. Different levels of output were chosen for the three values of σ to make the figure clearer.

The marginal product is the additional contribution to production of using one more unit of an input, holding the levels of the other inputs constant. A unitary elasticity of substitution ($\sigma = 1$), referred to as "perfect substitutability," means that as the ratio of the two inputs is changed by a given percentage, holding output constant; the ratio of their marginal products changes by the same percentage (in the opposite direction). This relation is shown by the curve (known as an isoquant) in Fig. 2, which is asymptotic to both axes. As resource use decreases toward zero, production can be maintained by increasing capital use toward infinity. Additionally,

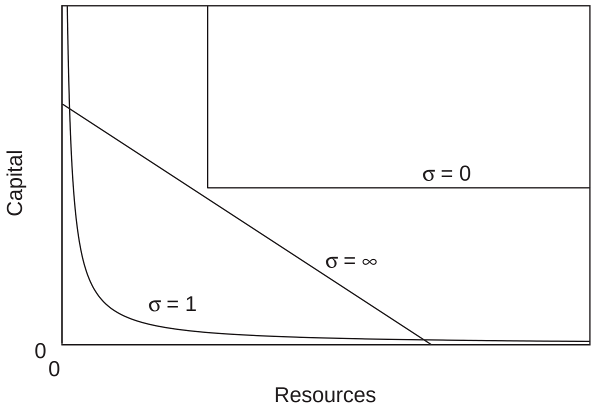

FIGURE 2 The elasticity of substitution.

the total cost of production is constant along the isoquant. Perfect substitutability does not mean that resources and capital are equivalently useful—that case is termed "infinite substitutability" ($\sigma = \infty$). In the latter case, producers see no difference between the two inputs and use the cheapest one. Whenever the elasticity of substitution is greater than 1, the isoquants cross the axes and inputs are nonessential for production. The figure also illustrates the case where no substitution is possible ($\sigma = 0$). In this case, the two inputs must be used in a fixed ratio. As is discussed later, perfect substitutability is an unrealistic assumption from a biophysical perspective, at least if it is assumed to apply at all possible ratios of capital and resources. Demand elasticities for energy, which in theory are related to the elasticity of substitution, also indicate that the elasticities of substitution between energy and other inputs and among different fuels may be between 0 and 1.

Economists such as Solow explicitly dispose of cases where σ for nonrenewable resources and capital is greater or less than unity. In the former case, substitution possibilities are large and therefore the possibility of nonsustainability is not an issue. In the latter case, sustainability is not feasible if an economy uses only nonrenewable resources. Of course, where there are renewable resources, sustainability is technically feasible, at least in the absence of population growth. Neoclassical economists argue that the class of growth models that includes resources can account for mass balance and thermodynamic constraints with the "essentiality condition." If σ is greater than 1, then resources are "nonessential." If σ is less than or equal to 1, than resources are "essential." Essential in this case means that, given positive nonresource inputs, output is only zero when the resource input is zero, and strictly positive otherwise. The Cobb–Douglas production function, a form frequently used in growth models, has the essentiality condition. Economists argue that this at least accounts for the fact that some amount of energy and materials is required to produce goods and services. But when the elasticity of substitution is unity, this "essential" amount can be infinitesimal if sufficient manufactured capital is applied. Economists also note that resources and capital are interdependent in the neoclassical models in that some positive quantity of resources is required to produce capital assets. Thus, the capital stock cannot be increased without depleting the resource stock.

Substitution that is technically possible will not occur unless society invests in sufficient capital over time to replace the depleted natural resources and ecosystem services. How much investment does take place depends on the institutional setting of the economy. For example, in an economy in which sustainability is just technically feasible ($\sigma = 1$) and there are only nonrenewable resources, sustainability will not occur in either a competitive or centrally planned economy where the decision rule is the maximization of the discounted flow of utility of future generations using a constant and positive discount rate. Consumption per capita will eventually decline to zero after an initial period of economic growth because resources and ecosystem services are depleted faster than capital can be accumulated to replace them. However, if the utility of individuals is given equal weight without regard to when they happen to live, and the aim is to maximize the sum of utilities over time, then growth in consumption can occur indefinitely. This is equivalent to maximizing net present value with a zero discount rate. The Hartwick rule is an investment rule that can, in theory, achieve sustainability if sustainability is technically feasible. A constant level of consumption can be achieved by reinvesting resource rents in other forms of capital, which, in turn, can substitute for resources.

The other key factor permitting growth in the face of a limited resource base is technological change. A technological improvement is defined as a gain in total factor productivity, which implies that output increases while a weighted sum of the inputs to production is held constant. Therefore, growth would be possible in the face of finite resources. Studies that examine the roles of resources in growth models with endogenous technological change have so far been less general in their assumptions, compared to research using the exogenous technological change assumption, and do not yet provide necessary or sufficient conditions for the achievement of sustainability. For example, Sjak Smulders and Michiel de Nooij assume in their model that energy use per capita continues to grow over time and that conservation policies consist of either a reduction in the growth rate or a one-time cut in the level of energy use. This set of assumptions is very different from those in the models just discussed. The results depend on whether the growth rate of energy use is reduced or the level of energy use is cut and on the value of the elasticity of substitution. In these models, energy scarcity or conservation policies can induce technological change. Though this can mitigate the effects of scarcity, non-energy-related research and development may be crowded out. This means that attempts to conserve energy can depress the long-run economic growth rate, even when they would not in a model with exogenous technological change. Integrating these ideas into research on sustainability policy will clearly be a future area of active research.

2.3 Critique and Alternative Views

Many ecological economists have a “preanalytic vision” of the economic process that is fundamentally different from that presented in neoclassical economics. Mainstream growth theory focuses on institutional limits to growth. Ecological economists tend instead to focus on the material basis of the economy. The criticism of growth theory focuses on the limits to substitution and the limits to technological progress as ways of mitigating the scarcity of resources. Substitution of manufactured capital for resources and technological change could potentially get more production out of a limited resource input and circumvent the limited capacity of natural environments to absorb the impacts of energy and resource use by reducing those impacts. But if these two processes are limited, then limited resources or excessive environmental impacts may restrict growth (empirical evidence on these questions is reviewed in Sections 4 and 5).

2.3.1 Limits to Substitution

There is more than one type of substitution between inputs and, therefore, there is more than one reason why substitution may be limited. There can be substitution within a category of similar production inputs (for example, between different fuels) and between different categories of inputs (for example, between energy and machines). There is also a distinction to be made between substitution at the micro level (for example, in a single engineering process or in a single firm) and at the macro level (in the economy as a whole). Additionally, some types of substitution that are possible in a single country are not possible globally.

The first type of substitution (within category), and, in particular, the substitution of renewable for nonrenewable resources, is undoubtedly important. The long-run pattern of energy use in industrial economies has been dominated by substitutions: from wood to coal, to oil, to natural gas, and to primary electricity. It is possible that the elasticity of substitution for within-category types of substitution exceeds unity. This would imply that some particular inputs are nonessential.

Ecological economists emphasize the importance of limits to the other type of substitution, and, in particular, the substitution of manufactured capital for natural capital. Natural capital is needed both for energy capture and for absorbing the impacts of energy and resource use—the sink function. Even if smaller amounts of energy were needed for production, all productive activity, being a work process transforming the materials using energy, will disrupt the natural environment. Often one form of environmental disruption (for example, pollution) is replaced by another form of environmental disruption (for example, hydroelectric dams). Potential reasons for limited substitutability are discussed in the following section.

2.3.1.1 Thermodynamic Limits to Substitution

Thermodynamic limits to substitution are easily identified for individual processes by an energy–materials analysis that defines both the fundamental limitations of transforming materials into different thermodynamic states and the use of energy to achieve that transformation. These types of analyses have shown where technological improvements exhibit strong diminishing returns due to thermodynamic limits, and where there is substantial room for improvements in the efficiency of energy and material use. For example, the thermal efficiency of power plants has been relatively constant for many years, reflecting the fact that power plant design is approaching the thermodynamic limit.

2.3.1.2 Complementarity Limits Substitution

Production is a work process that uses energy to transform materials into goods and services. Nicholas Georgescu-Roegen’s fund-flow model describes production as a transformation process in which a flow of materials, energy, and information is transformed by two agents of transformation: human labor and manufactured capital. The flow of energy, materials, and services from natural capital is what is being transformed, while manufactured capital effects the transformation. Thus, some ecological economists argue that, for example, adding to the stock of pulp mills does not produce an increase in pulp unless there also is the wood fiber to feed them. The latter is essential an argument about material balance.

Mainstream economists think about this question differently. First, they argue that though additional capital cannot conjure wood fibers out of a vacuum, more capital can be used with each amount of wood fibers to produce more sophisticated and valuable products from them and that this is the relevant substitution between capital and resources. In the energy industries, more capital can extract more oil from a petroleum reservoir and extract more useful work downstream in cleaner ways, only subject to thermodynamic limits. Even thermodynamic limits apply only to production of physical product. There is no limit in this view to the potential value of product created through sophisticated manipulation using larger amounts of capital.

2.3.1.3 Physical Interdependence and Macroeconomic Limits to Substitution The construction, operation, and maintenance of tools, machines, and factories require a flow of materials, i.e., energy from natural capital. Similarly, the humans that direct manufactured capital consume energy and materials (i.e., food and water). Thus, producing more of the "substitute," i.e., manufactured capital, requires more of the thing for which it is supposed to substitute. Ecological economists argue that production functions used in growth models do not account for this interdependence, and thus assume a degree of substitutability that does not exist. But both environmental and ecological economics have erred by not distinguishing among the micro- and macro-applications of production functions. Substitution is fundamentally more constrained at the macro level of analysis than at the micro level. For example, home insulation directly substitutes for heating fuel within the household sector. But interdependence means that insulation requires fuel to manufacture, so, for the economy as a whole, the net substitution of insulation for fuel is less than that indicated by an analysis of the household sector, in isolation from the rest of the economy.

In Fig. 3, the curve $E = f(M)$ is a neoclassical isoquant for a constant level of output, where E is energy and M is materials. The indirect energy costs of materials are represented by $g(M)$. For simplicity, the diagram unrealistically assumes that no materials are required in the extraction or capture of energy. Addition of direct and indirect energy costs results in the "net" isoquant $E = h(M)$. Generalizing for material costs to energy extraction appears to indicate that there are eventually decreasing returns to all factors at the macro level, and therefore the socially efficient region of the aggregate production function does not include areas with extreme factor ratios.

At a global level, a country such as Kuwait or Nauru can deplete its natural resources and invest in manufactured capital offshore through the financial markets. But this route to substituting manufactured capital for natural capital is clearly not possible for the world as a whole.

2.3.1.4 Critical Natural Capital Limits Substitution Ecological economists have also argued that at the macro level some forms of "natural capital" are not replaceable by produced capital, at least beyond certain minimum stock sizes. These stocks may provide life-support services to the economy or represent pools of irreplaceable genetic information or "biodiversity." The limited substitutability argument has also been extended to incorporate nonlinear dynamics and irreversible changes. The fear is that excessive substitution of human-made capitals for natural capital will cause the system to approach a threshold beyond which natural systems will lose resilience and suffer catastrophic collapse. Though we cannot demonstrate these forms of nonsubstitutability from basic physical laws, they may be just as important as thermodynamics in constraining actual production functions. In the energy context, this argument is most relevant regarding the sink func-

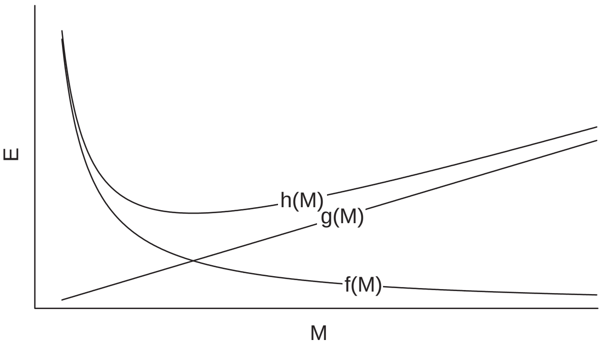

FIGURE 3 Macro-level limits to substitution.

tion of the environment. Using more and more of the environment as a sink for pollution means less and less of it is available for other life-support uses. Compared to current fossil fuel technologies, alternative energy sources may require larger areas of the environment for energy capture and may also generate wastes, etc. in the production of the energy capture and transmission capitals.

2.3.2 Limits to Technological Change

Even if substitution possibilities are limited, sustainability is possible if technological change is natural capital augmenting and unlimited in scope. The arguments for technological change as a solution would be more convincing if technological change were really something different from substitution. The neoclassical approach assumes that an infinite number of efficient techniques coexist at any one point in time. Substitution occurs among these techniques. Changes in technology occur when new, more efficient techniques are developed. However, in a sense, these new techniques represent the substitution of knowledge for the other factors of production. The knowledge is embodied in improved capital goods and more skilled workers and managers, all of which require energy, materials, and ecosystem services to produce and maintain. Thus, however sophisticated the workers and machinery become, there are still thermodynamic restrictions on the extent to which energy and material flows can be reduced.

Another question is whether technology will follow the "right" direction. If natural resources are not priced correctly due to market failure—a common and pervasive phenomenon that is the main topic of study of mainstream environmental economics—then there will be insufficient incentives to develop technologies that reduce resource and energy use. Instead, technological change would result in more resource use, not less.

3. FACTORS AFFECTING LINKAGE BETWEEN ENERGY AND GROWTH

There has been extensive debate concerning the trend in energy efficiency in the developed economies, especially since the two oil price shocks of the 1970s. Taking the example of the U.S. economy, energy consumption hardly changed in the period 1973 to 1991 (Fig. 4). This was despite a significant increase in GDP. These facts are indisputable. The reasons for the break in the trend have been the subject of argument. It is commonly asserted that there has been a decoupling of economic output and resources, which implies that the limits to growth are no longer as restricting as in the past.

The discussion here starts from the neoclassical perspective of the production function to examine the factors that could reduce or strengthen the linkage between energy use and economic activity over time. A general production function can be represented as follows:

$$(Q_1, \ldots, Q_m)' = f(A, X_1, \ldots, X_n, E_1, \ldots, E_p), \quad (1)$$

where the Q_i are various outputs (such as manufactured goods and services), the X_j are various inputs (such as capital, labor, etc.), the E_k are different energy inputs (such as coal, oil, etc.), and A is the state of technology as defined by the total factor productivity indicator. The relationship between energy and an aggregate of output such as gross domestic product can then be affected by substitution between energy and other inputs, technological change (a change in A), shifts in the composition of the energy input, and shifts in the composition of output. Also, shifts in the mix of the other inputs—for example, to a more capital-intensive economy from a more labor-intensive economy—can affect the relationship between energy and output, but this issue has not been extensively discussed in the literature and so will not be pursued further here. It is also possible for the input variables to affect total factor productivity, though in models that invoke exogenous technological change, this is assumed not to occur.

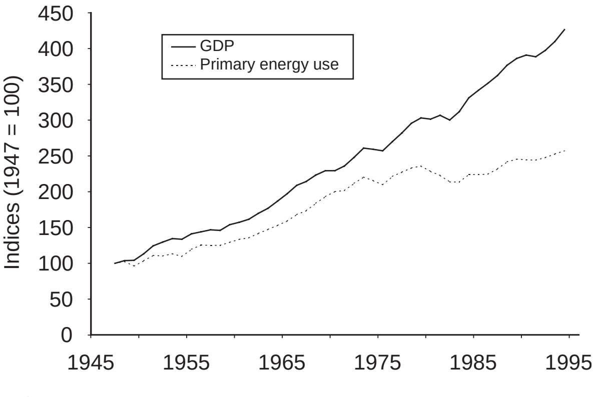

FIGURE 4 United States gross domestic product (GDP) and total primary energy use. GDP is in constant dollars, i.e., adjusted for inflation. Both variables are indexed to 100 in 1947.

3.1 Energy and Capital: Substitution and Complementarity

Econometric studies employing the translog and other functional forms have come to varying conclusions regarding whether capital and energy are complements or substitutes. These studies all estimate elasticities at the industry level. It seems that capital and energy are at best weak substitutes and possibly are complements. The degree of complementarity likely varies across industries, the level of aggregation considered, and the time frame considered.

There are few studies that look at macroeconomic substitution possibilities. In a 1991 paper, Robert Kaufmann and Irisita Azary-Lee demonstrated the importance of accounting for the physical interdependency between manufactured and natural capital. They used a standard production function to account for the indirect energy used elsewhere in the economy to produce the capital substituted for fuel in the U.S. forest products sector. They found that from 1958 to 1984, the indirect energy costs of capital offset a significant fraction of the direct fuel savings. In some years, the indirect energy costs of capital are greater than the direct fuel savings. The results of Kaufmann and Azary-Lee's analysis are consistent with the arguments made previously that scale is critical in assessing substitution possibilities. In this case, the assessment of substitution at one scale (the individual sector) overestimates the energy savings at a larger scale (the entire economy).

3.2 Innovation and Energy Efficiency

Substitution between different techniques of production using different ratios of inputs occurs due to changes in the relative prices of the inputs. Changes in the energy/GDP ratio that are not related to changes in the relative price of energy are called changes in the autonomous energy efficiency index. These could be due to any of the determinants of the relationship between energy and output listed at the beginning of this section, not just technological change. Actually, when there is endogenous technological change, changes in prices induce technological changes. In reality, technological change is at least in part endogenous. As a result, an increase in energy prices does tend to accelerate the development of energy-saving technologies. Periods of decreasing energy prices may result in technological development tending toward more intense use of energy.

Estimates of the trend in autonomous energy efficiency or the related energy augmentation index are mixed. This is likely because the direction of change has not been constant and varies across different sectors of the economy. Jorgenson and Wilcoxen estimated that autonomous energy efficiency in the United States is declining. Jorgenson also found that technological change tended to result in an increase in the share of production costs allocated to energy use—so-called biased technological change. Berndt *et al.* used a model in which the energy augmentation index is assumed to change at a constant rate. They estimate that in the U.S. manufacturing industry between 1965 and 1987, the energy augmentation index was increasing between 1.75 and 13.09% per annum, depending on the assumptions made. Judson *et al.* estimate separate environmental Kuznets curve relations for energy consumption in each of a number of energy-consuming sectors for a large panel of data using spline regression. The sectors are industry and construction, transportation, households and others, energy sector, nonenergy uses, and total apparent consumption, as well as households and agriculture, which are subsets of households and others. Judson *et al.* estimate time effects that show rising energy consumption over time in the household and other sectors but flat to declining time effects in industry and construction. Technical innovations tend to introduce more energy-using appliances to households and energy-saving techniques to industry.

Figure 5 presents the author's estimate of the autonomous energy efficiency index using the Kalman filter, an advanced time series technique that can extract unobserved stochastic series from time series data given a theoretical model that defines the trend of interest. This approach is more sophisticated than that of the studies cited previously, which assume that the index follows a simple linear or quadratic time path. Relative to an overall upward trend, the index shows large fluctuations. Until the mid-1960s, autonomous energy efficiency is increasing and then it starts a sharp decline. However, the results show that the first oil shock in 1973 does not disrupt the overall downward trend in energy efficiency. Only after the second oil shock does the trend reverse and energy efficiency increase.

The Khazzoom–Brookes postulate argues that energy-saving innovations can end up causing even more energy to be used, because the money saved is spent on other goods and services that require energy in their production. Energy services are demanded by the producer or consumer and are produced using

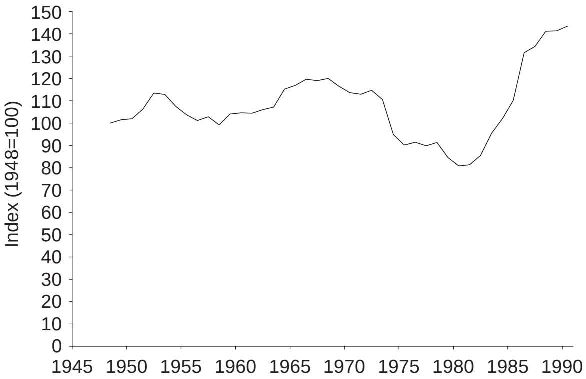

FIGURE 5 United States autonomous energy efficiency index.

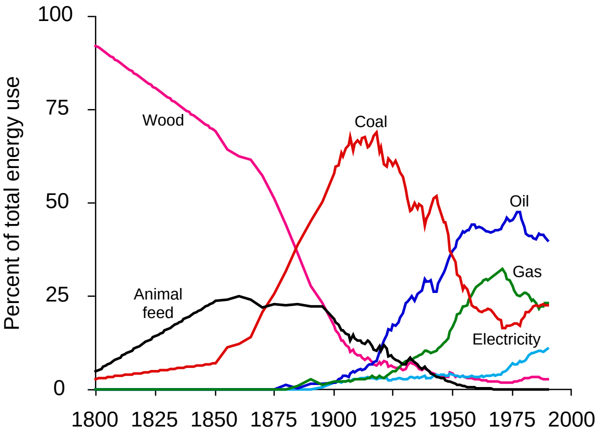

FIGURE 6 Composition of U.S. energy input.

energy. An innovation that reduces the amount of energy required to produce a unit of energy services lowers the effective price of energy services. This results in an increase in demand for energy services and therefore for energy. The lower price of energy also results in an income effect that increases demand for all goods in the economy and therefore for the energy required to produce them. There may also be adjustments in capital stocks that result in an even further increased long-run demand response for energy. This adjustment in capital stocks is termed a "macroeconomic feedback." In a 1997 paper, Richard Howarth argued persuasively that the rebound effect is less than the initial innovation-induced reduction in energy use, so improvements in energy efficiency do, in fact, reduce total energy demand.

3.3 Energy Quality and Shifts in Composition of Energy Input

Energy quality is the relative economic usefulness per heat equivalent unit of different fuels and electricity. One way of measuring energy quality is the marginal product of the fuel, which is the marginal increase in the quantity of a good or service produced by the use of one additional heat unit of fuel. These services also include services received directly from energy by consumers. Some fuels can be used for a larger number of activities and/or for more valuable activities. For example, coal cannot be used directly to power a computer whereas electricity can. The marginal product of a fuel is determined in part by a complex set of attributes unique to each fuel: physical scarcity, capacity to do useful work, energy density, cleanliness, amenability to storage, safety, flexibility of use, cost of conversion, and so on. But also the marginal product is not uniquely fixed by these attributes. Rather the energy vector's marginal product varies according to what activities it is used in, how much and what form of capital, labor, and materials it is used in conjunction with, and how much energy is used in each application. Therefore, energy qualities are not fixed over time. However, it is generally believed that electricity is the highest quality type of energy, followed by natural gas, oil, coal, and wood and other biofuels in descending order of quality. This is supported by the typical prices of these fuels per unit of energy. According to economic theory, the price paid for a fuel should be proportional to its marginal product.

Samuel Schurr was among the first to recognize the economic importance of energy quality. Noting that the composition of energy use has changed significantly over time (Fig. 6), Schurr argued that the general shift to higher quality fuels reduces the amount of energy required to produce a dollar's worth of GDP. If this is ignored, apparent total factor productivity (TFP) growth is greater than is really the case. Researchers such as Cutler Cleveland, Robert Kaufmann, and the Office of Technology Assessment, have presented analyses that explain much of the decline in U.S. energy intensity in terms of structural shifts in the economy and shifts from lower quality fuels to higher quality fuels. In a recent paper, Kaufmann estimated a vector autoregressive model of the energy/GDP ratio, household energy expenditures, fuel mix variables, and energy price variables for the United States. He found that shifting away from coal use and in particular shifting toward the use of oil reduced energy intensity. This shift away from coal contributed to declining energy intensity over the entire 1929–1999 time period.

Figure 7 illustrates the increase in the second half of the 20th century in U.S. GDP and a quality-adjusted index of energy use computed by the author. The index accounts for differences in the productivity of different fuels by weighting them by their prices. There is clearly less evidence of decoupling of energy use and GDP in these data than in those in Fig. 4. If decoupling is mainly due to the shift to higher quality fuels, then there appear to be limits to that substitution. In particular, exhaustion of low-cost oil supplies could mean that economies have to revert to lower quality fuels such as coal.

3.4 Shifts in the Composition of Output

Typically, over the course of economic development, the output mix changes. In the earlier phases of development there is a shift away from agriculture towards heavy industry, and in the later stages of development there is a shift from the more resource-intensive extractive and heavy industrial sectors toward services and lighter manufacturing. Different industries have different energy intensities. It is often argued that this will result in an increase in energy used per unit of output in the early stages of economic development and a reduction in energy used per unit output in the later stages of economic development. Pollution and environmental disruption would be expected to follow a similar path.

Service industries still need large energy and resource inputs. The product being sold may be immaterial but the office towers, shopping malls, warehouses, rental apartment complexes, etc. where the activity is conducted are very material intensive and energy is used in their functioning as well as in their construction and maintenance. Other service industries such as transport clearly heavily draw on resources and energy. Furthermore, consumers use large amounts of energy and resources in commuting to work, to shop, etc. Therefore, a complete decoupling of energy and growth as a result of shifting to the service sector seems unlikely. When the indirect energy use embodied in manufactured products and services is taken into account, the U.S. service and household sectors are not much less energy intensive than are the other sectors of the economy, and there is little evidence that the shift in output mix that has occurred in the past few decades has significantly lowered the energy/GDP ratio. Rather, changes in the mix of energy used are primarily responsible. Furthermore, on a global scale, there may be limits to the extent to which developing countries can replicate the structural shift that has occurred in the developed economies and, additionally, it is not clear that the developed world can continue to shift in that direction indefinitely.

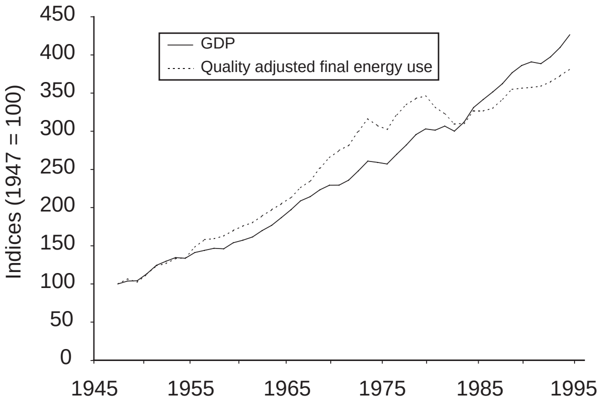

FIGURE 7 United States gross domestic product (GDP) and quality-adjusted final energy use GDP is in constant dollars, i.e., adjusted for inflation. Energy use is a Divisia index (the accumulation of weighted sums of growth rates) of the principal final energy use categories (oil, natural gas, coal, electricity, biofuels, etc.). The different fuels are weighted according to their average prices.

4. *EMPIRICAL TESTING*

Here the focus is on the empirical evidence for the tightness of coupling between energy use and economic output. Ordinary linear regression or correlation methods cannot be used to establish a casual relation among variables. In particular, it is well known that when two or more totally unrelated variables are trending over time they will appear to be correlated simply because of the shared directionality. Even after removing any trends by appropriate means, the correlations among variables could be due to causality between them or due to their relations with other variables not included in the analysis. Two methods for testing for causality among time series variables are Granger causality tests and cointegration analysis.

Granger causality tests whether one variable in a linear relation can be meaningfully described as a dependent variable and the other variable as an independent variable, whether the relation is bidirectional, or whether no functional relation exists at all. This is usually done by testing whether lagged values of one of the variables significantly add to the explanatory power of a model that already includes lagged values of the dependent variable, and perhaps also lagged values of other variables.

Whereas Granger causality can be applied to both stationary and integrated time series (time series that follow a random walk), cointegration applies only to linear models of integrated time series. The irregular trend in integrated series is known as a stochastic trend, as opposed to a simple linear deterministic time trend. Time series of GDP and energy use are integrated. Cointegration analysis aims to uncover causal relations among variables by determining if the stochastic trends in a group of variables are shared by the series. If these trends are shared, either one variable causes the other or they are both driven by a third variable. It can also be used to test if there are residual stochastic trends that are not shared by any other variables. This may be an indication that important variables have been omitted from the regression model or that the variable with the residual trend does not have long-run interactions with the other variables. The presence of cointegration can also be interpreted as the presence of a long-run equilibrium relationship between the variables in question.

Both cointegration and Granger causality tests are usually carried out within the context of vector autoregression models. These models consist of group of regression equations in which each dependent variable is regressed on lagged values of itself and of all the other variables in the system.

Many analysts have used Granger causality tests or the related test developed by Christopher Sims to test whether energy use causes economic growth or whether energy use is determined by the level of output in the context of a bivariate vector autoregression. The results have been generally inconclusive. When nominally significant results were obtained, the majority indicated that causality runs from output to energy use. In a 1993 study, the author tested for Granger causality in a multivariate setting using a vector autoregression (VAR) model of GDP, energy use, capital, and labor inputs in the United States over a the post-World War II period. He also used a quality-adjusted index of energy input in place of gross energy use. The multivariate methodology is important because changes in energy use are frequently countered by the substitution of other factors of production, resulting in an insignificant overall impact on output. When both of these innovations are employed, energy is found to "Granger-cause" GDP. These results are supported by other studies that found that changes in oil prices Granger-cause changes in GDP and unemployment in VAR models, whereas oil prices are exogenous to the system.

Yu and Jin were the first to test whether energy and output cointegrate. They found that no such relationship exists between energy use and either employment or an index of industrial production in the United States. However, the lack of a long-run equilibrium relationship between gross energy use and output alone does not necessarily imply that there is no relation between the variables. Few analysts believe that capital, labor, and technical change play no significant role in determining output. If these variables are integrated, then there will be no cointegration between energy and output, no matter whether there is a relationship between the latter two variables. Also, decreasing energy intensity (due to increased energy efficiency), shifts in the composition of the energy input, and structural change in the economy mean that energy and output will drift apart. Similar comments apply to the bivariate energy-employment relationship. Further, using total energy use in the economy as a whole but measuring output as industrial output alone may compound the insensitivity of the test.

It would seem that if a multivariate approach helps in uncovering the Granger causality relations between energy and GDP, a multivariate approach should be used to investigate the cointegration relations among the variables. The author investigated the time series properties of GDP, quality weighted energy, labor, and capital series, estimating a dynamic cointegration model using the Johansen methodology. The cointegration analysis showed that energy is significant in explaining GDP. It also showed that there is cointegration in a relationship including GDP, capital, labor, and energy. Other analysts have found that energy, GDP, and energy prices cointegrate and that when all three variables are included there is mutual causation between energy and GDP. The inconclusive results of the earlier tests of Granger causality are probably due to the omission of necessary variables—either the quantities of other inputs (and quality adjustment of the energy input) or energy prices.

5. ENVIRONMENTAL IMPLICATIONS

5.1 Theory

The argument has been made here that there is a very strong link between energy use and both the levels of economic activity and economic growth. What are the implications of this linkage for the quality of the environment? The focus here is now on immediate environmental impacts (issues of sustainability were discussed in Section 2).

Energy use has a variety of impacts. Energy extraction and processing always involve some forms of environmental disruption, including both geomorphological and ecological disruption as well as pollution. Energy use involves both pollution and other impacts, such as noise from transport, and land-use impacts, such as the construction of roads, etc. As all human activities require energy use; in fact, all human impacts on the environment can be seen as the consequences of energy use. Energy use is sometimes seen as a proxy for the environmental impact of human activity in general. Creating order in the economic system always implies creating disorder in nature, though this disorder could be in the sun or space rather than on Earth. The factors that reduce the total amount of energy needed to produce a dollar's worth of GDP, discussed in Section 3, therefore, also act to reduce the environmental impact of economic growth in exactly the same way as they reduce energy consumption. However, not all impacts of energy use are equally harmful to the environment and human health.

A shift from lower to higher quality energy sources not only reduces the total energy required to produce a unit of GDP, but also may reduce the environmental impact of the remaining energy use. An obvious example would be a shift from coal use to natural gas use. Natural gas is cleaner burning and produces less carbon dioxide per unit of energy derived. However, we need to be careful here. Nuclear-generated electricity is a higher quality fuel compared to coal, at least as measured by current prices, but its long-term environmental impacts are not necessarily lower. Incorporating the cost of currently unpriced environmental externalities into the prices of fuels would, though, raise the apparent quality of those fuels. This is not a contradiction, because a higher market price would mean that those fuels would only be used in more productive activities.

The environmental impact of energy use may also change over time due to technological innovation that reduces the emissions of various pollutants or other environmental impacts associated with each energy source. This is in addition to general energy-conserving technological changes that reduce the energy requirements of production, already discussed. Therefore, despite the strong connections between energy use and economic growth, there are several pathways through which the environmental impact of growth can be reduced. Again, however, if there are limits to substitution and technological change, then the potential reduction in the environmental intensity of economic production is eventually limited. Innovations that reduce one type of emission (for example, flue gas desulfurization) often produce a different type of waste that must be disposed of (in this case, gypsum that is possibly contaminated with heavy metals) as well as other disruptions required to implement the technology (in this example, limestone mining).

5.2 Empirical Evidence: The Environmental Kuznets Curve

In the 1970s, the debate on the relation between growth and the environment focused on the report, *The Limits to Growth*, by the Club of Rome. The common wisdom was that economic growth meant greater environmental impacts and the main way to environmental gains was reduced population and consumption. Economists and others argued that substitution and innovation could reduce environmental impacts (as was described in Section 2). But these were minority views. The mainstream view was that environment and economy were conflicting goals. By the late 1980s, however, with the emergence of the idea of sustainable development, the conventional wisdom shifted to one of "too poor to be green." Less developed countries lacked the resources for environmental protection and growth, and development was needed to provide for environmental protection. This idea was embodied in the empirical models that became known as environmental Kuznets curves (EKCs). The hypothesis states that there is an inverted U-shape relation between various indicators of environmental degradation and income per capita, with pollution or other forms of degradation rising in the early stages of economic development and falling in the later stages. The EKC is named for the economist Kuznets, who hypothesized that the relationship between a measure of inequality in the distribution of income and the level of income takes the form of an inverted U-shape curve. However, Kuznets had no part in developing the EKC concept.

The idea has become one of the "stylized facts" of environmental and resource economics. This is despite considerable criticism on both theoretical and empirical grounds. The EKC has been interpreted by many as indicating that no effort should be made to adopt environmental policies in developing countries, i.e., when those countries become rich, the current environmental problems will be addressed by policy changes adopted at that later time. As a corollary, it is implied that little in the way of environmental clean-up activity is being conducted in developing countries. These views are challenged by recent evidence that, in fact, pollution problems are being addressed and remedied in developing economies. In addition to the data and case studies provided by Susmita Dasgupta *et al.*, the author shows that for sulfur (widely believed to show the inverted U-shape relation between emission and income per capita), emissions in fact rise with increasing income at all levels of income, but that there are strong time effects reducing emissions in all countries across all income levels. In the author's opinion, this new evidence supersedes the debate about whether some pollutants show an inverted U-shaped curve and others—for example, carbon dioxide and "new toxics"—show a monotonic relationship. All pollutants show a monotonic relation with income, but over time pollution has been reduced at all income levels, *ceteris paribus*. Similarly, the debate about whether the downward sloping portion of the EKC is an illusion resulting from the movement of polluting industries to offshore locations is also now moot. This phenomenon might lower the income elasticity of pollution in developed economies relative to developing economies, but it does not seem sufficient to make it negative. The true form of the emissions–income relationship is a mix of two of the scenarios proposed by Dasgupta *et al.,* illustrated in Fig. 8. The overall shape is that of their "new toxics" EKC—a monotonic increase of emissions in income. But over time this curve shifts down. This is analogous to their "revised EKC" scenario, which is intended to indicate that over time the conventional EKC curve shifts down.

Although the environmental Kuznets curve is clearly not a "stylized fact" and is unlikely to be a useful model, this does not mean that it is not possible to reduce emissions of sulfur and other pollutants. The time effects from an EKC estimated in first differences and from an econometric emissions decomposition model both show that considerable time-related reductions in sulfur emissions have been achieved in countries at many different levels of income. Dasgupta *et al.* provide data and case studies that illustrate the progress already made in developing countries to reduce pollution. In addition, the income elasticity of emissions is likely to be less than 1—but not negative in wealthy countries, as suggested by the EKC hypothesis. In slower growing economies, emissions-reducing technological change can overcome the scale effect of rising income per capita on emissions. As a result, substantial reductions in sulfur emissions per capita have been observed in many Organization for Economic Cooperation and Development (OECD) countries in the past few decades. In faster growing middle income economies, the effects of rising income overwhelm the contribution of technological in reducing emissions.

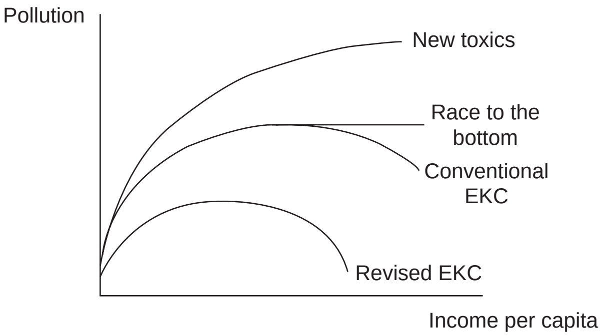

FIGURE 8 Environmental Kuznets curve (EKC): different scenarios. From Dasgupta *et al.* (2002).

This picture relates very closely to the theoretical model described in the previous Section 5.1. Overall, there is a strong link between rising energy use, economic growth, and pollution. However, the linkages between these three can be mitigated by a number of factors, including shifting to higher quality fuels and technological change aimed both at general increases in economic productivity and, specifically, at reducing pollution. However, given the skepticism regarding the potential for unlimited substitution or technological progress, there may be limits to the extent that these linkages can continue to be loosened in the future.

SEE ALSO THE FOLLOWING ARTICLES

Development and Energy, Overview • Economic Growth, Liberalization, and the Environment • Economics of Energy Demand • Economics of Energy Supply • Energy Ladder in Developing Nations • Environmental Kuznets Curve • Population Growth and Energy • Sustainable Development:

Basic Concepts and Application to Energy • Thermodynamics and Economics, Overview

Further Reading

Berndt, E. R., Kolstad, C., and Lee, J-K. (1993). Measuring the energy efficiency and productivity impacts of embodied technical change. *Energy J.* **14**, 33–55.

Brookes, L. (1990). The greenhouse effect: The fallacies in the energy efficiency solution. *Energy Policy* **18**, 199–201.

Cleveland, C. J., Costanza, R., Hall, C. A. S., and Kaufmann, R. K. (1984). Energy and the U.S. economy: A biophysical perspective. *Science* **225**, 890–897.

Dasgupta, P., Heal, G. (1974). The optimal depletion of exhaustible resources. Review of economic studies. Symp. Econ. Exhaustible Resources, pp. 3–28.

Dasgupta, S., Laplante, B., Wang, H., and Wheeler, D. (2002). Confronting the environmental Kuznets curve. *J. Econ. Perspect.* **16**, 147–168.

Georgescu-Roegen, N. (1976). "Energy and Economic Myths." Pergamon, New York.

Hannon, B. (1973). The structure of the ecosystem. *J. Theoret. Biol.* **41**, 535–546.

Hartwick, J. M. (1977). Intergenerational equity and the investing of rents from exhaustible resources. *Am. Econ. Rev.* **66**, 972–974.

Howarth, R. B. (1997). Energy efficiency and economic growth. *Contemp. Econ. Policy* **25**, 1–9.

Jorgenson, D. W. (1984). The role of energy in productivity growth. *Energy J.* **5**(3), 11–26.

Jorgenson, D. W., and Wilcoxen, P. J. (1993). Reducing U.S. carbon emissions: An econometric general equilibrium assessment. *Resour. Energy Econ.* **15**, 7–25.

Judson, R. A., Schmalensee, R., and Stoker, T. M. (1999). Economic development and the structure of demand for commercial energy. *Energy J.* **20**(2), 29–57.

Kaufmann, R. K. (2003). The mechanisms for autonomous energy efficiency increases: A cointegration analysis of the US energy/GDP ratio. *Energy J.* (in press).

Kaufmann, R.K., Azary-Lee, I.G. (1991). A biophysical analysis of substitution: Does substitution save energy in the U.S. forest products industry? Proc. Ecol. Econ.: Implications Forest Manage. Practice, 1991.

Khazzoom, D. J. (1980). Economic implications of mandated efficiency standards for household appliances. *Energy J.* **1**(4), 21–39.

O'Connor, M. P. (1993). Entropic irreversibility and uncontrolled technological change in the economy and environment. *J. Evolut. Econ.* **34**, 285–315.

Perrings, C. A. (1987). "Economy and Environment: A Theoretical Essay on the Interdependence of Economic and Environmental Systems." Cambridge University Press, Cambridge.

Schurr, S., and Netschert, B. (1960). "Energy and the American Economy, 1850–1975." Johns Hopkins University Press, Baltimore.

Sims, C. A. (1972). Money, income and causality. *Am. Econ. Rev.* **62**, 540–552.

Smulders, S., and de Nooij, M. (2003). The impact of energy conservation on technology and economic growth. *Resour. Energy Econ.* **25**, 59–79.

Solow, R. M. (1956). A contribution to the theory of economic growth. *Q. J. Econ.* **70**, 65–94.

Solow, R.M. (1974). Intergenerational equity and exhaustible resources. Review of economic studies. Symp. Econ. Exhaustible Resources, pp. 29–46.

Solow, R. M. (1997). Reply: Georgescu-Roegen versus Solow/Stiglitz. *Ecol. Econ.* **22**, 267–268.

Stern, D. I. (1993). Energy use and economic growth in the USA. A multivariate approach. *Energy Econ.* **15**, 137–150.

Stern, D. I. (2000). A multivariate cointegration analysis of the role of energy in the U.S. macroeconomy. *Energy Econ.* **22**, 267–283.

Stern, D. I. (2002). Explaining changes in global sulfur emissions: An econometric decomposition approach. *Ecol. Econ.* **42**, 201–220.

Stern, D. I., and Common, M. S. (2001). Is there an environmental Kuznets curve for sulfur? *J. Environ. Econ. Manage.* **41**, 162–178.

U.S. Congress, Office of Technology Assessment (1990). "Energy Use and the U.S. Economy." OTA-BP-E-57. U.S. Government Printing Office, Washington, D.C.

Yu, E. S. H., and Jin, J. C. (1992). Cointegration tests of energy consumption, income, and employment. *Resour. Energy* **14**, 259–266.