Rensselaer
Working Papers in Economics

Department of Economics, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180-3590, USA. Tel: +1-518-276-6387; Fax: +1-518-276-2235; URL: http://www.rpi.edu/dept/economics/; E-Mail: keenak@rpi.edu

Energy and Economic Growth

David I. Stern
Rensselaer Polytechnic Institute

Cutler J. Cleveland
Boston University

Number 0410

March 2004

Energy and Economic Growth

March 2004

David I. Stern
Department of Economics, Sage 3208, Rensselaer Polytechnic Institute, 110 8th Street Troy, NY, 12180-3590, USA. E-mail: sternd@rpi.edu, Phone: 518-276-6386, Fax: 518-276-2235.

Cutler J. Cleveland
Department of Geography and Center for Energy and Environmental Studies, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA. E-mail: cutler@bu.edu, Phone: 617-353-3083, Fax: 617-353-5986.

Abstract

Physical theory shows that energy is necessary for economic production and therefore growth but the mainstream theory of economic growth, except for specialized resource economics models, pays no attention to the role of energy. This paper reviews the relevant biophysical theory, mainstream and resource economics models of growth, the critiques of mainstream models, and the various mechanisms that can weaken the links between energy and growth. Finally we review the empirical literature that finds that energy used per unit of economic output has declined, but that this is to a large extent due to a shift from poorer quality fuels such as coal to the use of higher quality fuels, and especially electricity. Furthermore, time series analysis shows that energy and GDP cointegrate and energy use Granger causes GDP when additional variables such as energy prices or other production inputs are included. As a result, prospects for further large reductions in energy intensity seem limited.

Acknowledgements

We thank Jeroen van den Bergh, Jack Pezzey, and Mark Diesendorf for useful comments and discussions.

JEL Codes: O40, Q43

1. Introduction

In a recent paper in the *Energy Journal*, Toman and Jemelkova (2003) argue that most of the literature on energy and economic development discusses how development affects energy use rather than vice versa. This paper surveys the literature on the effect of changes in energy supply on economic growth in general in both developing and developed countries. As Toman and Jemelkova (2003) state, the mainstream economics literature on this issue is somewhat limited. Business and financial economists do pay significant attention to the impact of oil and other energy prices on economic activity in the short-run, 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. An exception was the extensive discussions concerning the "productivity slowdown" following the 1970s oil crises. Much of the relevant literature is outside the mainstream in what has come to be known as ecological economics.

Resource economists have developed models that incorporate the role of resources including energy in the growth process but these ideas remain segregated in the resource economics field. The mainstream theory of growth has been criticized on a number of grounds, especially on the basis of the implications of thermodynamics for economic production and the long-term prospects of the economy. Extensive empirical work has examined the role of energy in the growth process. The principal findings are 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 remain fairly tightly coupled. Furthermore, time series analysis shows that energy and GDP cointegrate and energy use Granger causes GDP when additional variables such as energy prices or other production inputs are included. When theory and empirical results are taken into account the prospects for further large reductions in the energy intensity of economic activity seem limited. This has important implications for environmental quality and both economic and environmental policy.

This paper is structured to cover these key points in a systematic fashion. The first section of the paper reviews the background theory of production and growth from different points of view –

those based in economics and those based in the natural sciences - in an attempt to assess to what degree energy availability enables and constrains or limits economic growth. We focus on the long-run prospects for economic growth. Our 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 section starts by reviewing the scientific basis of the role of energy in production and hence also in the increasing scale of production involved in economic growth. However, institutional phenomena also affect how this role plays out and therefore the economics of growth and production and 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 paper.

The next section uses the concept of the production function to examine the factors that could reduce or strengthen the linkage between energy use and economic activity over time. These key factors are:

- substitution between energy and other inputs within an existing technology,
- technological change,
- shifts in the composition of the energy input, and
- shifts in the composition of economic output.

Each of these themes has a subsection dedicated to its discussion.

Choice between these competing theories and models has to be on the basis of both inherent plausibility and consistency and, perhaps more crucially, empirical evidence. Therefore, the next section of the paper moves on to review studies that investigate the strength of the linkage between energy and growth. To be useful, such studies must not be grounded in a single theory, potential mechanism, or school of thought. Therefore, the studies reviewed here are reduced form time series models that do not specify structural linkages between energy and output. As 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. The final section of the paper summarizes our conclusions and points to some implications for environmental policy.

2. Theory of Production and Growth

A. Energy in Production: Physical Theory and Economic Models

i. Basic Principles

Reproducibility is a key concept in the economics of production. Some inputs to production are non-reproducible, while others can be manufactured at a cost within the economic production system. Capital, labor, and in the longer term even natural resources, are reproducible factors of production, while energy is a nonreproducible factor of production, though of course energy vectors - fuels - are reproducible factors (Stern, 1999). 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 (e.g. Hall et al., 2001, 2003). In the extreme, energy use rather than output of goods is used as an indicator of the state of economic development (e.g. Kardashev, 1964).

The first law of thermodynamics (the conservation law) implies the mass-balance principle (Ayres and Kneese, 1969). In order to obtain a given material output greater or equal quantities of matter must be used as inputs with the residual a pollutant or waste product. Therefore, there are minimal 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. Some form of matter must be moved or transformed though particular elements and chemicals may be substitutable. Therefore there must be limits to the substitution of other factors of production for energy. All economic processes must, therefore, require energy, so that energy is always an essential factor of production (Stern, 1997a).

Some aspects of organized matter - that is information - might also be considered to be non-reproducible inputs. Several analysts (e.g. Spreng, 1993; Chen, 1994; Stern, 1994; Ruth, 1995) argue that information is a fundamentally nonreproducible factor of production in the same way as energy, and that economics must pay as much consideration to information and its accumulation as knowledge as it pays to energy. Energy is necessary to extract information from the environment while energy cannot be made active use of without information and possibly accumulated knowledge. Obviously energy can provide uncontrolled heating, lighting etc. without any activity on the part of economic agents. But even non-intelligent organisms need to use information to make controlled use of energy. For example, when plants use some sunlight for photosynthesis rather than just heating and lighting their leaves they are using the information in their genetic code to produce chlorophyll, construct chloroplasts, and generate sugar. Unlike energy, information and knowledge cannot be easily quantified. However, the fact that they must be incorporated into machines, workers, and materials in order to be made useful 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 (Stern, 1999).

A second key concept is the notion of primary and intermediate factors of production. 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), while intermediate inputs are those 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, while goods such fuels and materials are 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 (Stern, 1999).

This approach has led to a focus in mainstream growth theory on the primary inputs, and in particular, capital and labor, and the attribution of a lesser and somewhat indirect role to energy. The primary energy inputs are stock resources such as oil deposits. Therefore, 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 (Stern, 1999). But these are not given an explicit role in the standard macroeconomic growth theories that focus on labor and capital. Therefore, understanding the role of energy in the mainstream theory of growth is not so straightforward and the role of energy as a driver of economic growth and production is downplayed.

ii. Biophysical Models of the Economy

Some alternative, ecological economic 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 (Stern, 1999). In some biophysical economic models (e.g. Gever *et al.*, 1986) geological constraints fix the rate of energy extraction. On the other hand, 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 (Costanza, 1980; Hall *et al.*, 1986; Gever *et al.*, 1986; or Kaufmann, 1987). Prices of commodities should then be determined by embodied energy cost (Hannon, 1973b) – a normative energy theory of value - or are actually correlated with energy cost (Costanza, 1980) - a positive energy theory of value (Common, 1995). This theory – like the Marxian paradigm - must then explain how labor, capital etc. end up receiving part of the surplus. Energy surplus must be appropriated by the owners of labor, capital, and land with the actual distribution of the surplus depending on the relative bargaining power of the different social classes and foreign suppliers of fuel (Kaufmann, 1987). If we assume constant returns to scale, the production process of the economy as a whole can be represented by a Leontief input-output model with a single primary factor of production (Hannon, 1973a; Stern, 1999).

However, these ecological economists also argue that the energy required to produce fuels and other intermediate resources increases as the quality of resources such as oil reservoirs declines over time. Thus changing resource quality can be represented by changes in the embodied energy of the intermediate inputs. However, this implies that it is not energy per se that is the primary input or the only primary input. Rather the level of organization and information embodied in energy sources appear to play a role with energy sources being referred to as low or high entropy. However, declining resource quality can also be represented as negative productivity growth or technological change (Cleveland and Stern, 1999) and therefore can be formally modeled in the input-output model as changes in the input-output coefficients over time with only a single primary input. On the other hand, in the approach developed by Costanza (1980) and Odum's emergy approach (see Brown and Herendeen, 1996) resources are represented by their embodied solar and geological energy. Thus changing resource quality is represented by changes in the embodied energy of the resources rather than by changes in the input-output coefficients. The concept of energy quality – the notion that different energy vectors have different productivities - also features in some of these biophysical models. Again this is explained in terms of entropy.

If resource stocks were explicitly represented, energy would clearly no longer be the only primary factor of production. The neo-Ricardian models developed by Perrings (1987) and O'Connor (1993), 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 this approach can take the biophysical constraints of mass balance and energy conservation into account (Stern, 1999).

If the economy could actually be represented as an input-output model where there is no substitution between factors of production, and a single source of uniform quality energy, the embodied knowledge in the factors of production can itself be ignored, though its embodied energy content is of course counted. But 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 we consider next do.

B. The Mainstream Theory of Growth

i. Growth Models without Resources

In Solow's (1956) original growth model - known as the neoclassical growth model - the economy must reach a stationary state in which there is no net (additional) investment. Growth is a transitional phase, where a country is moving towards the stationary state, An underdeveloped economy, with a small capital stock per worker, can achieve fast growth while it is building up its capital stock. But if the savings rate remains constant all economies will eventually reach a zero growth equilibrium. No country can grow in perpetuity merely by accumulating capital. If the savings rate is increased growth will occur for a while until a new equilibrium is reached, though, the higher the savings rate, the lower the current standard of living of the population. According to this basic neoclassical growth theory, the only cause of continuing economic growth is technological progress. 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.

The original models did not explain how improvements in technology come about. They are just assumed to happen exogenously, so that these models are 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.

There are a few different types of endogenous growth models (Aghion and Howitt, 1998). Early endogenous technological change models allowed the state of technology to respond to changes in one of the variables in the model but do not explicitly model an optimizing process. In learning-by-doing models the state of technology is a function of cumulative production. In the original Arrow (1962) model, the productivity of capital goods improves over time as more of them are cumulatively produced. In other versions, the learning curve implies rising productivity in the production of a good as more of it is cumulatively produced. In induced technological change models, originated by Hicks (1932), innovation increases when the price of an input such as energy increases.

In the second class of endogenous growth model, the relationship between capital and output can be written in the form $Y = AK$, where A is constant and K is a composite of manufactured capital and disembodied technological knowledge thought of as a form of capital. Therefore, economic growth can continue indefinitely as this very broadly defined capital is accumulated, as output is not subject to diminishing returns. In AK models saving is directed to either manufactured capital accumulation or the increase of knowledge. However, the models do not explicitly model research and development activities (R&D). Technological knowledge has two special properties. First it is a non-rival good - the stock of this form of capital is not depleted with use. Second, it generates positive externalities in production. While the firm doing R&D obtains benefits from the knowledge acquired, 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. As some of the benefit of knowledge generation is external to those producing it, the growth rate of the economy is below the socially optimal level. However, the economy can sustain a constant growth rate in which the diminishing returns to manufactured capital are exactly offset by the external effect of knowledge creation. 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 (Perman and Stern, 2001).

The incentive to devote resources to innovation comes from the prospect of temporary monopoly profits for successful innovations. Schumpterian growth models are a third class of endogenous technology models (Aghion and Howitt, 1998) that explicitly model this incentive structure. Firms invest in R&D in order to receive monopoly profits. Innovations appear stochastically and are embodied in new generations of capital goods and there is imperfect competition in the capital goods industry. The average growth rate may be too high or too low to maximize welfare, as there are both positive and negative externalities. There are positive externalities to consumers who benefit from innovation and to future researchers who benefit from past ideas. There are negative externalities due to new innovations making old vintages of capital obsolete. Both capital accumulation and innovation determine the long-run growth rate. Capital accumulation raises the returns to innovation activity. However, if there are diminishing returns in the innovation sector as technology becomes more complex the economy could have a constant growth rate (Aghion and Howitt, 1998).

ii. Growth Models with Natural Resources and No Technological Change

All natural resources exist in finite quantities though some such as sunlight or deuterium are available in very large quantities. Some environmental resources are non-renewable; and many renewable resources are potentially exhaustible. Finiteness and exhaustibility of resources make the notion of indefinite economic growth problematic. Even sustainable development - non-declining consumption - may not be feasible.

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 assumed institutional arrangements. Analysts have looked at both optimal growth models which attempt to either maximization the sum of discounted social welfare over some relevant time horizon (often an infinite horizon) or achieve sustainability (non-declining 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 non-declining consumption or utility. Technical and institutional conditions determine whether sustainability – defined as non-declining consumption - is possible. Technical conditions include 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 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.

Solow (1974) showed that sustainability is achievable in a model with a finite and nonrenewable natural resource with no extraction costs and non-depreciating capital, which is produced using capital and the natural resource when the elasticity of substitution between the two inputs is unity, and certain other technical conditions are met. Sustainability occurs when the utility of individuals was given equal weight without regard to when they happen to live and the aim was to maximize the sum of utilities over time. In fact, growth in consumption can occur indefinitely. However, the same model economy under competition results in exhaustion of the resource and consumption and social welfare eventually fall to zero (Stiglitz, 1974). Dasgupta and Heal (1979) show that with any constant discount rate the efficient growth path also leads to eventual depletion of the natural resource and the collapse of the economy. Sustainability occurs when society invests in sufficient capital over time to replace the depleted natural resources. The

Hartwick rule (Hartwick, 1977) shows that if sustainability is technically feasible, a constant level of consumption can be achieved by reinvesting the resource rents in other forms of capital, which in turn can substitute for resources. Dixit *et al.* (1980) extended the rule to multiple capital stocks while Hartwick (1995) extended the rule to open economies. It is difficult to apply this rule in practice as the rents and capital must be valued at sustainability compatible prices (Asheim, 1994; Asheim *et al.*, 2003; Stern, 1997b) and not simply the competitive prices that would hold were all ordinary market failures (externalities, intertemporal inefficiency etc.) corrected, which are hard enough to determine as it is. Rather these are the prices that would emerge if the sustainability constraint were imposed.

A common interpretation of this body of work is that substitution and technical change can effectively de-couple economic growth from energy and other resources. Depleted resources can be replaced by more abundant substitutes, or by "equivalent" forms of human-made capital (people, machines, factories, etc.). But this is a misinterpretation. As explained above, 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. Solow (1974) and others explicitly dispose of cases where the elasticity of substitution between non-renewable resources and capital is greater or less than unity. In the former case substitution possibilities are large and therefore the possibility of non-sustainability is not an issue. In the latter case, sustainability is not feasible if an economy uses only non-renewable resources. Of course, where there are renewable resources sustainability is technically feasible, at least in the absence of population growth. However, there is a tendency among mainstream economists to assume that sustainability is technically feasible unless proved otherwise (Solow, 1993, 1997).

iii. Growth Models with Natural Resources and Technological Change

In addition to substitution of capital for resources, technological change might permit growth or at least constant consumption in the face of a finite resource base. Growing total factor productivity makes sustainability technically easier to achieve and sustainability may be possible even with an elasticity of substitution of less than one. However, again technical feasibility does not imply that sustainability will occur. Technological improvements imply that production per unit resource will be higher in the future. Depending on preferences for current versus future consumption, current depletion may as a result be faster (Smulders, in press). This result is related to the Khazzoom-Brookes postulate or rebound effect discussed below. As noted above, due to externalities in knowledge production there is too little innovation in an endogenous growth world. As a result, depletion of a non-renewable resource is nonoptimal. But as explained here this rate could be either too fast or too slow.

Studies that examine the roles of resources in growth models with endogenous technological change have been less general in their assumptions than research using the exogenous technological change or no technological change assumptions and do not yet provide necessary conditions for the achievement of sustainability. In particular, research on growth with non-renewable resources in combination with endogenous technological change has been somewhat limited (Smulders, 1999). Some models make very specific assumptions. For example, Smulders and de Nooij (2003) assume that energy use has a positive growth rate apart from a possible one-time reduction in the level of energy use. Smulders (1999) provides a survey of earlier endogenous growth work and Smulders and de Nooij (2003) provide references to the more recent literature.

The most general results are provided by Aghion and Howitt (1998) who analyze four different models, determining which allow for sustained growth and which not. Two of the models involve environmental pollution (with quality of the environment is a renewable resource) and two a nonrenewable resource. Each set of two models includes models using the AK and Schumpeterian frameworks. The renewable resource models require resources to be diverted from final goods production to reduce pollution, while environmental quality is an argument in utility. In the AK version, the long-run growth rate cannot be positive, which stands in contrast to the results for the AK model without resources. The Schumpetarian model can allow unlimited growth, but only under certain assumptions about model parameters that seem extreme to Aghion and Howitt (1998). The nonrenewable resource models assume that the nonrenewable resource is essential in production. The AK model again cannot have a positive long-run growth rate of consumption. The Schumpeterian model with nonrenewable resources allows unbounded growth in consumption under weaker conditions than the model with renewable resources. This might seem counterintuitive but would seem to stem from the fact that in the latter case consumers only care about consumption. With renewable resources that do not affect utility, continued growth would seem to be even easier.

Tahvonen and Salo (2001) develop a model economy with both renewable and non-renewable energy resources that is both very general and more realistic than the earlier "neoclassical" literature (Solow, 1974 etc.). Like Stiglitz (1974) they intend to see how the growth process would actually work. The models have extraction costs for fossil fuels and production costs for renewable energy resources, which also rise as cheaper sources are exploited first. The model can also deal with no technological change or technical change of exogenous and to some degree endogenous – learning by doing - varieties as well. It is assumed that technical knowledge in extraction increases proportionally to extraction and that technical knowledge in final production is proportional to the capital stock. The optimal development of such an economy appears to mimic history much more effectively than the "neoclassical models". The economy passes through pre-industrial, industrial, and post-industrial eras as the use of fossil fuels first rises and then falls and capital is accumulated. The price of non-renewables first falls and then rises. This seems to be a very promising platform for the future investigation of growth and sustainability issues especially if it was integrated with some of the more general endogenous technological change models discussed above.

C. Critique and Alternative Views

i. Ecological Economics and Mainstream Views on Growth

Many ecological economists have a fundamentally different "pre-analytic vision" of the economic process than that presented in neoclassical economics. As explained above, mainstream growth theory focuses on institutional limits to growth. When mainstream economists address the technical limits to growth they tend to not take these possible constraints very seriously (e.g. Solow, 1978, 1993, 1997). Ecological economists tend instead to focus on the material basis of the economy. The criticism of mainstream growth theory focuses on limits to substitution and limits to technological progress as ways of mitigating the scarcity of resources. If these two processes are limited then limited resources or excessive environmental impacts may restrict growth.

ii. 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. Finally, some types of substitution that are possible in a single country are not possible globally.

Solow (1997) argues that within category substitution and in particular the substitution of renewable for nonrenewable resources, is most important and seems to assume that new substitutes will always be found. It is possible that the elasticity of substitution for within category types of substitution exceeds unity. The long run pattern of energy use in industrial economies has been dominated by the substitutions from wood and waterpower to coal, oil, natural gas and primary electricity (Hall *et al.*, 1986). In large part the industrial revolution was enabled by the use of fossil fuels that freed economic activity from reliance on low power and variable but renewable solar energy. When fossil fuels are economically exhausted the next stage of energy development may see a return to solar energy, albeit captured in a more sophisticated way, rather than a move to anew substitute.

Ecological economists emphasize the importance of limits to the between category type of substitution, and in particular, the substitution of manufactured capital for resources including energy. A number of arguments for limited substitutability have been put forward, with the main ones that are relevant to the energy case being:

Thermodynamic limits to substitution Thermodynamic limits to substitution are easily identified for individual processes by an energy-materials analysis that defines the fundamental limitations of transforming materials into different thermodynamic states and on the use of energy to achieve that transformation (Ruth 1993; Islam, 1985). 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 it is approaching the thermodynamic limit. It might be argued that production functions can account for mass balance and thermodynamic constraints with the “essentiality condition.” If the elasticity of substitution is greater than one, then energy is "non-essential." If the elasticity of substitution is less than or equal to one, then energy is "essential." Given positive non-energy inputs, output is only zero when the energy input is zero, and strictly positive otherwise. The Cobb-Douglas production function has the essentiality condition. This at least accounts for the fact that *some* amount of energy and materials are 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. Therefore, this condition does not satisfy thermodynamic considerations throughout the domain of the function (Dasgupta and Heal, p211).

Material cause and efficient cause Georgescu-Roegen’s (1976) fund-flow model describes production as a transformation process in which a flow of materials, energy, and information – the material cause - is transformed by two agents of transformation, human labor and manufactured capital – the efficient cause. 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 (Daly, 1991). 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 downstream it can extract more useful work in cleaner ways, only subject to thermodynamic limits. Even thermodynamic limits only apply to production of physical product. There is no limit in their view to the potential value of product created through sophisticated manipulation using larger amounts of capital (van den Bergh, 1999).

Physical interdependence and macroeconomic and global limits to substitution The construction, operation, and maintenance of tools, machines, and factories require a flow of materials and energy. 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 that it is supposed to substitute for.

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 (Georgescu-Roegen, 1979; Cleveland *et al.*, 1984; Ayres and Nair, 1984; Kaufmann, 1992; Daly, 1997, Stern, 1997a). But we must distinguish between micro-and macro-applications of production functions. Substitution is fundamentally more constrained at the macro- level of analysis than at the micro-level (Stern, 1997a). 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. Put another way, the aggregate of potential energy savings at the macroeconomic level is less than the sum of the savings one would calculate by adding the savings from sectoral-level analyses that do not account for the indirect costs.

In Figure 1 the curve $E = f(M)$ is a neoclassical isoquant for a constant level of output, where E is energy, and M materials. The indirect energy costs of materials are represented by $g(K)$. 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 $E = h(K)$. Generalizing for material costs to energy extraction suggests 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 the 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.

iii. Limits to Technological Change

Even if substitution possibilities are limited, sustainability is possible if technological change is resource augmenting and unlimited in scope. Of course technological change might be biased to be energy consuming rather than saving. In any case, this arguments would be more convincing if technological change were really something different from substitution. This is not the case. 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, these new techniques really 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.

The difference between knowledge and other forms of capital is that knowledge is non-rival in use – in other words the same idea can be used simultaneously in different locations and production processes without any reduction in the productivity of the knowledge in the different locations and processes. This means that there are constant returns to the application of knowledge in production while other inputs experience diminishing returns. But the knowledge must be used in conjunction with the other inputs. The marginal productivity of knowledge is still determined by the available quantities of those inputs.

3. Factors Affecting the Linkage between Energy and Growth

A. Introduction

The previous section established that energy is an essential input and that in theory in the long-run energy availability could constrain economic growth. However, there has been extensive debate concerning the trend in energy intensity in the developed economies, especially since the two oil price shocks of the 1970s. 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. This was one of the messages of the 1992 World Development Report that addressed environmental issues prior to the 1992 Earth Summit in Rio de Janeiro. (IBRD, 1992). See also the discussions in de Bruyn and Opschoor (1997) and Bohi (1989). Taking the example of the US economy, energy consumption hardly changed in the period 1973 to 1991 (Figure 2). This was despite a significant increase in GDP. These facts are indisputable. What has been the subject of argument is what were the reasons for the break in the trend

This section of the paper 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 and summarize the empirical evidence on each of these mechanisms. A general production function can be represented as:

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

where the Q_i are various outputs, such as manufactured goods and services, the X_i are various inputs such as capital, labor etc., the E_i are different energy inputs: 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.
– 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 X to affect total factor productivity. This possibility is discussed in the subsection on technological change below.

B. Energy and Capital: Substitution and Complementarity

Econometric studies have come to varying conclusions regarding whether capital and energy are complements or substitutes (Berndt and Wood, 1979; Apostolakis, 1990). Based on the differences between time series and cross-sectional results, Apostolakis (1990) concluded that capital and energy act more as substitutes in the long run and more as complements in the short run. However, in the light of the cointegration literature it is now dubious that we can assert that time-series regressions in levels represent short-run results. Frondel and Schmidt (2002) revisit the studies reviewed by Apostolakis and additional data from Germany and find that evidence of complementarity only occurs in cases where the cost share of energy is small. When materials are included the cost shares of capital and energy are smaller and a finding of complementarity is more likely. More time series studies than cross-sectional studies have data on materials use. Obviously the cost of materials should be included if possible and econometric results that exclude this variable are likely to be biased. Similarly, Berndt and Wood (1979) found that econometric studies using the KLE specification (i.e. not including materials) and engineering studies indicate substitution, while cost functions with the KLEM specification indicate complementarity.

Thompson and Taylor (1995) argue that when the Morishima elasticity of substitution is used is place of the more common Allen-Uzawa elasticities capital and energy are universally found to be substitutes. However, the Morishima elasticity rarely finds any inputs to be complements (Frondel and Schmidt, 2002; Thompson, 1997). Blackorby and Russell (1989) state that “the elasticity of substitution concept, as originally conceived by Hicks, has nothing to do with the substitute/complement taxonomy” (885). That discrimination should be made according to the sign of the cross-price elasticity, which is necessarily the same as the sign of the Allen-Uzawa substitution elasticity. It is worth noting that neither the Morishima nor Allen-Uzawa elasticities represent movement along an isoquant of a production function with the quantities of the other inputs held constant. Instead the prices of the other inputs are held constant so that their quantities adjust optimally to minimize cost. Only the original Hicks elasticity of substitution measures movement along an isoquant (Stern, 2004). With only two inputs all the elasticities of substitution are the same. But with multiple inputs there is no longer a single elasticity of substitution (see Stern, 2004 for more details).

Most of these studies estimate elasticities at the industry level – usually in manufacturing industry - rather than at the economy-wide level that is perhaps most relevant to economic growth. Using a VAR analysis of the US macro-economy Stern (1993) finds energy and capital to be neither substitutes nor complements. Kaufmann and Azary-Lee (1991) demonstrate the importance of accounting for the physical interdependency between manufactured and natural capital. They use 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 above that substitution possibilities are different at macro and micro levels.

It seems that, in conclusion, capital and energy are at best weak substitutes and possibly are complements. The degree of complementarity likely varies across industries and the level of aggregation considered. However, if the cost share of energy is small relative to that of capital, only small percentage increases in capital will be needed for large percentage reductions in energy use.

C. Innovation and Energy Efficiency

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 (AEEI). These could be due to any of the determinants of the relationship between energy and output listed at the beginning of this section and not just technological change. A more specific indicator is an index of energy augmenting technical change (Stern, 1999). This involves a reformulation of the production function (1):

$$Q = f(A_1 X_1, \ldots, A_n X_n, A_E E) \qquad (2)$$

so that each input is multiplied by its own technology factor A_i that converts crude units of the input into “effective units”. A_E is the index of energy augmenting technical change, which holds the use of all other inputs and their augmentation indices constant.

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. Jorgensen and Wilcoxen (1993) estimated that autonomous energy efficiency is declining. Berndt *et al.* (1993) use a model in which t index is assumed to change at a constant rate. They estimate that in US manufacturing industry between 1965 and 1987 the energy augmentation index was increasing at between 1.75% and 13.09% per annum depending on the assumptions made. Judson *et al.* (1999) estimate separate EKC 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, non-energy uses, and total apparent consumption, as well as households and agriculture which are subsets of households and others. They estimate time effects that show rising energy consumption over time in the household and other sector 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 (Stern, 2002).

The Khazzoom-Brookes Postulate (Brookes, 1990; Khazzoom, 1980) or “rebound effect” argues that energy saving innovations can end up causing even more energy to be used as the money saved is spent on other goods and services which themselves require energy in their production. Energy services are demanded by the producer or consumer and are produced using energy itself. 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 (Binswanger, 2001). The lower price of energy also results in an income effect (Lovins, 1988) 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 (Howarth, 1997). This adjustment in capital stocks is termed a "macro-economic feedback". Howarth

(1997) argues 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.

Actually, when there is endogenous technological change, changes in prices may induce technological changes. As a result, an increase in energy prices does tend to accelerate the development of energy saving technologies, while periods of falling energy prices may result in energy-using technological change. There can also be an effect on the general rate of TFP growth (Berndt, 1990). Jorgenson (1984) found that technical change was biased and tended to be energy using. If this is the case, lower energy prices tend to accelerate TFP growth and vice versa. More recent results may contradict this conclusion (e.g. Judson *et al.*, 1999). The Schurr hypothesis (Schurr and Netschert, 1960) argued that innovations that allowed the use of energy sources such as electricity were embodied in capital equipment that then subsequently allowed the organization of workplaces along more efficient and productive lines inducing further productivity gains. Toman and Jemelkova (2003) argue that this is the main way in which there could be apparent increasing returns to energy use so that energy has a disproportionate effect on economic development.

Newell *et al.* (1999) provide some information on the degree to which energy price increases induce improvements in the energy efficiency of consumer products. They decompose the changes in cost and energy efficiency of various energy using appliances using the concept of a transformation frontier of possible cost and efficiency combinations. For room air conditioners, large reductions in cost holding efficiency and cooling capacity occurred from 1960 to 1980 in the US. Also the cost of high efficiency air conditioners relative to inefficient ones was reduced. From 1980 to 1990 the former trend ended but the mix of air conditioners offered from those that were feasible to manufacture shifted sharply in favor of higher efficiency. Only about one quarter of the gain in energy efficiency since 1973 was induced by higher energy prices. Another quarter was found to be due to raised government standards and labeling. For gas water heaters the induced improvements were close to one half of the total. Much less cost reducing technical change occurred though. Popp (2002) similarly finds that increased energy prices have a significant though quantitatively small effect on the rate of patenting in the energy sector.

D. 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. Some fuels can be used for a larger number of activities and/or for more valuable activities. For example coal cannot be used to directly power a computer while 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 but also 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 biofuels in descending order of quality. This is supported by the typical prices of these fuels per unit of energy, which should be proportional to its marginal product.

Schurr and Netschert (1960) were among the first to recognize the economic importance of energy quality. Noting that the composition of energy use has changed significantly over time (Figure 3), Schurr and Netschert argued that the general shift to higher quality fuels reduces the amount of energy required to produce a dollar's worth of GDP. Berndt (1990) also notes the key role played by the shifting composition of energy use towards higher quality energy inputs. If this is ignored, apparent TFP growth is greater than is really the case.

Cleveland *et al.* (1984), Kaufmann (1992) and OTA (US Congress, 1990) presented analyses that explain much of the decline in US energy intensity in terms of structural shifts in the economy and shifts from lower quality fuels to higher quality fuels. Kaufmann (2004) estimates a vector autoregressive model of the energy/GDP ratio, household energy expenditures, energy mix variables, and energy price variables for the US. He finds that shifting away from coal use and in particular shifting towards the use of oil reduces energy intensity. This shift away from coal made contributes to declining energy intensity over the entire 1929-99 time period.

Figure 4 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 Stern (1993). 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 Figure 2. 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 (Kaufmann, 1992).

E. 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, while in the later stages of development there is a shift from the more resource intensive extractive and heavy industrial sectors towards 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 (Panayotou, 1993).

However, service industries still need large energy and resource inputs. The service being sold may be intangible but the office towers, shopping malls, warehouses, rental apartment complexes etc. where the activity is conducted are very tangible and energy is used in their functioning as well as in their construction and maintenance. Other service industries such as transport are clearly heavily resource and energy using. Furthermore, consumers use large amounts of energy and resources in commuting to work, 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 US service and household sectors are not much less energy intensive than the other sectors of the economy and there is little evidence that the shift in output mix that has occurred in the last few decades has significantly lowered the energy/GDP ratio. Rather, changes in the mix of energy used are primarily responsible (Cleveland *et al.* 1984). There may also be a tendency for consumers to use more energy directly over time as their consumption of the services appliances, housing, transport etc. increases. Judson *et al.* (1999) find that the consumer sector sees rising energy intensity over time, *ceteris paribus*, while the manufacturing sector sees decreasing energy intensity.

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 (Stern *et al.*, 1996) to the extent that this has occurred by outsourcing manufacturing overseas rather than simply from an expansion in service activities. Eventually developing economies will find no countries remaining to outsource those activities to. Additionally, if the service sector does require substantial material support, it is not clear whether the developed world can continue to shift in the direction of a growing service share of GDP indefinitely. An alternative view is that, as manufacturing prices have fallen relative to the prices of services, even the relative decline of manufacturing in developed countries is exaggerated when the relative sizes of the sectors are computed in current prices (Kander, 2002).

4. Empirical Testing

The previous section reviewed empirical evidence on specific mechanisms, which might weaken the coupling between energy use and economic output. In this section we take a step back and look at the empirical evidence on the overall relation between the two variables and studies that are not grounded in a single theory, potential mechanism, or school of thought. The studies that we review here are, therefore, reduced form time series models that do not specify structural linkages between energy and 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 (Granger, 1969) and cointegration analysis (Engle and Granger, 1987). See Enders (1995) for an accessible introduction to these methods.

Many analysts (e.g. Kraft and Kraft, 1978; Akarca and Long, 1980; Yu and Hwang, 1984; Abosedra and Baghestani, 1991; Yu and Choi, 1985; Erol and Yu, 1987; Ammah-Tagoe, 1990) used Granger (1969) causality tests or the related test developed by Sims (1972) 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 were generally inconclusive. Where nominally significant results were obtained they indicated that causality runs from output to energy. However, in many cases results differed depending on the samples used, the countries investigated etc. Stern (1993) tested for Granger causality in a multivariate setting using a vector autoregression (VAR) model of GDP, capital and labor inputs, and a quality-adjusted index of energy input in place of gross energy use. The multivariate methodology is important because reductions in energy use are frequently countered by the substitution of other factors of production for energy and vice versa, resulting in an insignificant overall impact on output. When both the multivariate approach and the quality adjusted energy index were employed, energy was found to Granger cause GDP. These results are supported by Hamilton (1983) and Burbridge and Harrison (1984), who found that changes in oil prices Granger-cause changes in GNP and unemployment in VAR models whereas oil prices are exogenous to the system.

Ohanian (1988) and Toda and Phillips (1993) showed that the distribution of the test statistic for Granger causality a VAR with non-stationary variables is not the standard chi-square distribution. This means that the significance levels reported in previous studies of the Granger-causality relationship between energy and GDP may be incorrect, as both variables are generally integrated series. If there is no cointegration between the variables then the causality test should be carried out on a VAR in differenced data, while if there is cointegration standard chi-square distributions apply when the cointegrating restrictions are imposed. Thus testing for cointegration is a necessary prerequisite to causality testing.

This may explain the problematic nature of this earlier literature. Few analysts believe that capital, labor, and technical change play no significant role in determining output. If these variables, or perhaps energy prices instead, are omitted from the model, there will be no cointegration and a spurious regression will result. Results are frequently sample dependent in the face of omitted variables and non-cointegration (e.g. Stern and Common, 2001).

Yu and Jin (1992) 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. 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 as other variables, mentioned above, maybe should be included in the model. If these variables are integrated, then there will be no cointegration between energy and output whether there is a relationship between the latter two variables or not. 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. Yang (2000), Soytas and Sari (2003), and Cheng and Lai (1997) all fail to find bivariate cointegration in the countries and time periods that they examine. However, Glasure and Lee (1997) claim to find cointegration in South Korea and Singapore.

Masih and Masih (1996) found cointegration between energy and GDP in India, Pakistan, and Indonesia, but no cointegration in Malaysia, Singapore, or the Philippines. Granger causality runs from energy to GDP in India but in the opposite direction in the other two countries. Again, bivariate methods yield indeterminate results.

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. Stern (2000) 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. The multivariate analysis shows that energy Granger causes GDP either unidirectionally or possibly through a mutually causative relationship depending on which version of the model is used. Oh and Lee (2004) apply Stern’s (1993, 2000) methodology to Korea and comes to exactly the same conclusions, extending the validity of Stern’s results beyond the United States.

Glasure (2002) also investigates the role of omitted variables in the energy income relation in Korea though the variables he investigates reflect fiscal and monetary policy – real money and real government expenditure. There is weak evidence of cointegration and bidirectional causality between energy and income in this model. Other analysts (Hondroyiannis *et al.*, 2002; Masih and Masih, 1997) 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.

An alternative view is that the relation between GDP and energy prices is asymmetric. Rising energy prices have a greater impact on GDP than falling energy prices and when energy price rises are merely correcting previous declines they have less effect than when they come after periods of stable prices (Hamilton, 2003). This nonlinear relation is said to explain why the relation between oil prices and GDP appeared to weaken after the mid-1980s. The leading explanation for the asymmetry is that adjustment costs occur whether prices rise or fall and thus they blunt the boom that would occur when oil prices decrease (Brown and Yücel, 2002). Modeling the GDP-oil price relation nonlinearly really just allows some variables that are not explicitly modeled to adjust endogenously to their optimal values. We have argued above that the changes in the quantities of the other inputs can account for this supposed nonlinearity.

5. Conclusions and Implications

The introduction to this paper summarized our principal conclusions, that the theoretical and empirical evidence indicates that energy use and output are tightly coupled with energy availability playing a key role in enabling growth. Furthermore, we argued that a large part of reductions in energy intensity was explained by a shift to higher quality fuels. Gains in autonomous energy efficiency are possible and have occurred but we suggest that ultimately technological change must obey the same thermodynamic constraints as substitution. As a result, we suggest that prospects for further large reductions in energy intensity seem limited.

Recent results in the environmental Kuznets curve literature show that pollution emissions also tend to rise with increasing income rather than follow an inverted U shaped curve (Stern, in press). Pollution concentrations, however, may follow an inverted U shape due to urban and industrial decentralization, taller smoke stacks etc. As emissions of these pollutants are generally linked to energy use this provides further evidence of the strength of linkages between energy use and GDP. Again, there is scope for technological change to reduce emissions. For any single pollutant the scope is much greater than that for reducing energy use as different types of impacts and emissions can be substituted for one another. As we explained in section 2, energy is used to transform or move matter and therefore any energy use causes some environmental disruption. Often one form of environmental disruption - for example pollution – is replaced by another form of environmental disruption – for example hydroelectric dams. Additionally, the second law of thermodynamics implies that creating order in one part of the environment-economy system creates a greater degree of disorder elsewhere – creating order in the economic system always implies creating disorder in nature though this could increased disorder could be partly in the sun or space. For these reasons, energy use is sometimes seen as a proxy for environmental impact of human activity in general (Common, 1995).

Therefore, even if future growth can leave behind its ties to finite supplies of fossil fuels and even go beyond the non-exhaustible but limited, variable, and diffuse supply of solar power, the environmental impacts of growth will remain critical.

References

Abosedra, S. and H. Baghestani (1991). “New evidence on the causal relationship between United States energy consumption and gross national product.” *Journal of Energy and Development* 14: 285-292.

Aghion, P. and P. Howitt (1998). *Endogenous Growth Theory*. : Cambridge, MA: MIT Press.

Akarca, A. and T. Long (1980). “On the relationship between energy and GNP: A reexamination.” *Journal of Energy and Development* 5: 326-331.

Ammah-Tagoe, F. A. (1990). *On Woodfuel, Total Energy Consumption and GDP in Ghana: A Study of Trends and Causal Relations*. Center for Energy and Environmental Studies, Boston University, Boston, MA.

Apostolakis, B. E. (1990). “Energy-capital substitutability / complementarity: the dichotomy.” *Energy Economics* 12: 48-58.

Arrow, K. J. (1962). “The economic implications of learning-by-doing.” *Review of Economic Studies* 29: 155-173.

Asheim, G. B. (1994). “Net national product as an indicator of sustainability.” *Scandinavian Journal of Economics* 96: 257–265.

Asheim, G. B., W. Buchholz, and C. Withagen (2003). “The Hartwick rule: myths and facts.” *Environmental and Resource Economics* 25: 129–150.

Ayres, R.U., and A.V. Kneese (1969). “Production, consumption and externalities.” *American Economic Review* 59: 282-97.

Ayres, R. and I. Nair (1984). “Thermodynamics and economics.” *Physics Today* 35: 62-71.

Berndt, E. R. (1990). “Energy use, technical progress and productivity growth: a survey of economic issues.” *The Journal of Productivity Analysis* 2: 67-83.

Berndt, E. R. and D. O. Wood (1979). “Engineering and econometric interpretations of energy-capital complementarity.” *American Economic Review* 69: 342-354.

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

Binswanger, M. (2001). “Technological progress and sustainable development: what about the rebound effect?” *Ecological Economics* 36: 119-132.

Blackorby, C. and R. R. Russell (1989). “Will the real elasticity of substitution please stand up? (A comparison of the Allen/Uzawa and Morishima elasticities).” *American Economic Review* 79: 882-888.

Bohi, D. (1989). *Energy Price Shocks and Macroeconomic Performance*. Washington, DC: Resources for the Future.

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

Brown, M. T. and R. A. Herendeen (1996). “Embodied energy analysis and emergy analysis: a comparative view.” *Ecological Economics* 19: 219-236.

Brown, S. P. A. and M. K. Yücel (2002). “Energy prices and aggregate economic activity: an interpretative survey.” *The Quarterly Review of Economics and Finance* 42: 193-208.

Burbridge, J. and A. Harrison (1984). “Testing for the effects of oil prices rises using vector autoregressions.” *International Economic Review* 25: 459-484.

Chen, X. (1994). “Substitution of information for energy: conceptual background, realities and limits.” *Energy Policy* 22: 15-28.

Cheng, B. S. and T. W. Lai (1997). “An investigation of co-integration and causality between energy consumption and economic activity in Taiwan.” *Energy Economics* 19: 435-444.

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

Cleveland, C. J. and D. I. Stern (1999). “Indicators of natural resource scarcity: a review and synthesis.” In: J. C. J. M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*. Cheltenham: Edward Elgar.

Common, M. S. (1995). *Sustainability and Policy: Limits to Economics*. Melbourne: Cambridge University Press.

Costanza, R. (1980). “Embodied energy and economic valuation.” *Science* 210: 1219-1224.

Costanza, R. and Daly, H. E. (1992). “Natural capital and sustainable development.” *Conservation Biology*, 6: 37-46.

Daly, H. E. (1991). “Elements of an environmental macroeconomics.” In: R. Costanza (ed.), *Ecological Economics* New York: Oxford University Press. pp. 32-46.

Daly, H. E. (1997). “Georgescu-Roegen versus Solow/Stiglitz.” *Ecological Economics* 22: 261-266.

Dasgupta, P. S. and G. M. Heal (1979). *Economic Theory and Exhaustible Resources* Oxford: Cambridge University Press.

de Bruyn, S.M. and J. B. Opschoor (1997). “Developments in the throughput-income relationship: theoretical and empirical observations.” *Ecological Economics* 20: 255-268.

Dixit, A., P. Hammond, and M. Hoel (1980). “On Hartwick's rule for regular maximin paths of capital accumulation and resource depletion.” *Review of Economic Studies* 47: 551-556.

Enders, W. (1995). *Applied Econometric Time Series*. New York: John Wiley.

Engle, R. E. and C. W. J. Granger (1987). “Cointegration and error-correction: representation, estimation, and testing.” *Econometrica* 55: 251-276.

Erol, U. and E. S. H. Yu (1987). “On the causal relationship between energy and income for industrialized countries.” *Journal of Energy and Development* 13: 113-122.

Frondel, M. and C. M Schmidt (2002). “The capital-energy controversy: An artifact of cost shares?” *The Energy Journal* 23(3): 53-79.

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

Georgescu-Roegen, N. (1979). “Energy and matter in mankind's technological circuit.” *Journal of Business Administration* 10: 107-127.

Gever J., R. K. Kaufmann, D. Skole, and C. Vörösmarty (1986). *Beyond Oil: The Threat to Food and Fuel in the Coming Decades*. Cambridge, MA: Ballinger.

Glasure, Y. U. (2002). “Energy and national income in Korea: Further evidence on the role of omitted variables.” *Energy Economics* 24: 355-365.

Glasure, Y. U. and A.-R. Lee (1997). “Cointegration, error-correction, and the relationship between GDP and energy: the case of South Korea and Singapore.” *Resource and Energy Economics* 20: 17-25.

Granger, C. W. J. (1969). “Investigating causal relations by econometric models and cross-spectral methods.” *Econometrica* 37: 424-438.

Hall, C. A. S., C. J. Cleveland, and R. K. Kaufmann (1986). *Energy and Resource Quality: The Ecology of the Economic Process*. New York: Wiley Interscience.

Hall, C. A. S., D. Lindenberger, R. Kümmel, T. Kroeger, and W. Eichhorn (2001). “The need to reintegrate the natural sciences and economics.” *BioScience* 51: 663-673.

Hall, C. A. S., P. Tharakan, J. Hallock, C. Cleveland, and M. Jefferson (2003). “Hydrocarbons and the evolution of human culture.” *Nature* 426: 318-322.

Hamilton, J. D. (1983). “Oil and the macroeconomy since World War II.” *Journal of Political Economy* 91: 228-248.

Hamilton, J. D. (2003). “What is an oil shock?” *Journal of Econometrics* 113: 363-398.

Hannon, B. (1973a). “The structure of the ecosystem.” *Journal of Theoretical Biology* 41: 535-546.

Hannon, B. (1973b). “An energy standard of value.” *Annals of the American Academy* 410: 139-153.

Hartwick, J. M. (1977). “Intergenerational equity and the investing of rents from exhaustible resources.” *American Economic Review* 66: 972-974.

Hartwick, J. M. (1995). “Constant consumption paths in open economies with exhaustible resources.” *Review of International Economics* 3: 275-283.

Hicks, J. R. (1932). *The Theory of Wages*. London: Macmillan.

Hondroyiannis, G., S. Lolos, and E. Papapetrou (2002). “Energy consumption and economic growth: assessing the evidence from Greece.” *Energy Economics* 24: 319-336.

Howarth, R. B. (1997). “Energy efficiency and economic growth.” *Contemporary Economic Policy* 25: 1-9.

IBRD (1992). *World Development Report 1992: Development and the Environment*. New York: Oxford University Press.

Islam, S. (1985). “Effects of an essential input on isoquants and substitution elasticities.” *Energy Economics* 7: 194-196.

Jorgenson, D.W. and P. J. Wilcoxen (1993). “Reducing US carbon emissions: an econometric general equilibrium assessment.” *Resource and Energy Economics* 15: 7-25.

Jorgenson, D.W. (1984). “The role of energy in productivity growth.” *Energy Journal* 5(3): 11-26.

Judson, R. A., R. Schmalensee, and T. M. Stoker (1999). “Economic development and the structure of demand for commercial energy.” *The Energy Journal* 20(2): 29-57.

Kander, A. (2002). *Economic Growth, Energy Consumption and CO2 Emissions in Sweden 1800-2000*, Ph.D. Dissertation, Lund University, Lund, Sweden.

Kardashev, N. S. (1964). “Transmission of information by extraterrestrial civilizations.” *Soviet Astronomy* 8: 217.

Kaufmann, R. K. (1987). “Biophysical and Marxist economics: learning from each other.” *Ecological Modelling* 38: 91-105.

Kaufmann, R. K. (1992). “A biophysical analysis of the energy/real GDP ratio: implications for substitution and technical change.” *Ecological Economics* 6: 35-56.

Kaufmann, R. K. (2004). “The mechanisms for autonomous energy efficiency increases: A cointegration analysis of the US energy/GDP ratio.” *Energy Journal* 25(1): 63-86.

Kaufmann, R. K. and I. G. Azary-Lee (1991) “A biophysical analysis of substitution: Does substitution save energy in the U.S. forest products industry?” *Proceedings of Ecological Economics: Implications for Forest Management and Practice*.

Khazzoom, D. J. (1980). “Economic implications of mandated efficiency standards for household appliances.” *Energy Journal*, 1(4): 21-39.

Kraft, J. and A. Kraft (1978). “On the relationship between energy and GNP.” *Journal of Energy and Development* 3, 401-403.

Lovins, A. B. (1988). “Energy saving from more efficient applicances: another view.” *Energy Journal* 10: 157-166.

Masih, A. M. M. and R. Masih (1996). “Energy consumption, real income and temporal causality: results from a multi-country study based on cointegration and error-correction modelling techniques.” *Energy Economics* 18: 165-183.

Masih, A. M. M. and R. Masih (1997). “On the temporal causal relationship between energy consumption, real income, and prices: some new evidence from Asian-energy dependent NICs based on a multivariate cointegration/vector error-correction approach.” *Journal of Policy Modeling* 19: 417-440.

Newell, R. G., A. B. Jaffe, and R. N. Stavins (1999). “The induced innovation hypothesis and energy-saving technological change.” *Quarterly Journal of Economics* 114: 941-975.

O'Connor, M. P. (1993). “Entropic irreversibility and uncontrolled technological change in the economy and environment.” *Journal of Evolutionary Economics* 34: 285-315.

Oh, W. and Lee K. (2004). “Causal relationship between energy consumption and GDP revisited: the case of Korea 1970–1999.” *Energy Economics* 26: 51–59.

Ohanian, L. E. (1988). “The spurious effects of unit roots on vector autoregressions: A Monte Carlo study.” *Journal of Econometrics* 39: 251-266.

Panayotou, T. (1993). *Empirical Tests and Policy Analysis of Environmental Degradation at Different Stages of Economic Development*. Working Paper WP238, Technology and Employment Programme, International Labour Office, Geneva.

Perman, R. and D. I. Stern (2001). “Sustainable development, growth theory, environmental Kuznets curves and discounting.” in J. C. J. M. van den Bergh (ed.) Environmental Economics, in *Encyclopedia of Life Support Systems*, UNESCO.

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

Popp, D. (2002). “Induced innovation and energy prices.” *American Economic Review* 92: 160-180.

Ruth, M. (1993). *Integrating Economics, Ecology, and Thermodynamics*. Dordecht: Kluwer Academic.

Ruth, M. (1995). “Information, order and knowledge in economic and ecological systems: implications for material and energy use.” *Ecological Economics* 13: 99-114.

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

Sims, C.A. (1972). “Money, income and causality.” *American Economic Review* 62: 540-552.

Smulders, S. (1999). “Endogenous growth theory and the environment.” in J. C. J. M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Edward Elgar, Cheltenham, 89-108.

Smulders, S. (in press). “Endogenous technical change, natural resources and growth.” In: R. Ayres, D. Simpson, and M. Toman (eds.), *Scarcity and Growth in the New Millennium*. Washington, DC: Resources for the Future.

Smulders, S. and M. de Nooij (2003). “The impact of energy conservation on technology and economic growth.” *Resource and Energy Economics*, 25: 59–79.

Solow, R. M. (1956). “A contribution to the theory of economic growth.” *Quarterly Journal of Economics*, 70: 65-94.

Solow, R. M. (1974). “Intergenerational equity and exhaustible resources.” *Review of Economic Studies*, Symposium on the Economics of Exhaustible Resources: 29-46.

Solow, R. M. (1978). “Resources and economic growth.” *American Economist* 22: 5-11.

Solow, R. M. (1993). “An almost practical step toward sustainability.” *Resources Policy* 19: 162-172.

Solow, R. M. (1997). “Reply: Georgescu-Roegen versus Solow/Stiglitz.” *Ecological Economics* 22: 267-268.

Soytas, U. and R. Sari (2002). “Energy consumption and GDP: causality relationship in G-& countries and emerging markets.” *Energy Economics* 25, 33-37.

Spreng, D. (1993). “Possibilities for substitution between energy, time and information.” *Energy Policy*, 21: 13-23.

Stern, D. I. (1993). “Energy use and economic growth in the USA, A multivariate approach.” *Energy Economics* 15: 137-150.

Stern, D. I. (1994). *Natural Resources as Factors of Production: Three Empirical Studies*. Ph.D. Dissertation, Department of Geography, Boston University, Boston MA.

Stern, D. I. (1997a). “Limits to substitution and irreversibility in production and consumption: a neoclassical interpretation of ecological economics.” *Ecological Economics*, 21: 197-215.

Stern, D. I. (1997b). “The capital theory approach to sustainability: a critical appraisal.” *Journal of Economic Issues* 31, 145-173.

Stern, D. I. (1999). “Is energy cost an accurate indicator of natural resource quality?” *Ecological Economics* 31: 381-394.

Stern, D. I. (2000). “A multivariate cointegration analysis of the role of energy in the U.S. macroeconomy.” *Energy Economics* 22: 267-283.

Stern, D. I. (2002). “Explaining changes in global sulfur emissions: an econometric decomposition approach.” *Ecological Economics* 42: 201-220.

Stern, D. I. (2004). “Elasticities of substitution and complementarity.” *Rensselaer Working Papers in Economics* 0403.

Stern, D. I. (in press). “The rise and fall of the environmental Kuznets curve.” *World Development*.

Stern, D. I. and M. S. Common (2001). “Is there an environmental Kuznets curve for sulfur?” *Journal of Environmental Economics and Management* 41, 162-178.

Stern, D.I., M. S. Common and E. B. Barbier (1996). “Economic growth and environmental degradation: the environmental Kuznets curve and sustainable development.” *World Development* 24: 1151-1160.

Stiglitz, J. E. (1974). “Growth with exhaustible natural resources: the competitive economy.” *Review of Economic Studies*, Symposium on the Economics of Exhaustible Resources: 139-152.

Tahvonen, O. and S. Salo (2001). “Economic growth and transitions between renewable and nonrenewable energy resources.” *European Economic Review* 45: 1379-1398.

Thompson, H. (1997). “Substitution elasticities with many inputs.” *Applied Mathematics Letters* 10(3): 123-127.

Thompson, P. and T. G. Taylor (1995). “The capital-energy substitutability debate: A new look.” *The Review of Economics and Statistics* 77: 565-569.

Toda, H. Y. and P. C. B. Phillips (1993). “The spurious effect of unit roots on vector autoregressions: an analytical study.” *Journal of Econometrics* 59: 229-255.

Toman, M. A. and B. Jemelkova (2003). “Energy and economic development: an assessment of the state of knowledge.” *Energy Journal* 24(4): 93-112.

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

van den Bergh, J. C.J. M. (1999). “Materials, capital, direct/indirect substitution, and mass balance production functions.” *Land Economics* 75 (4): 547-561.

Yang, H.-Y. (2000). “A note on the causal relationship between energy and GDP in Taiwan.” *Energy Economics* 22: 309-317.

Yu, E. S. H and J-Y Choi (1985). “The causal relationship between energy and GNP: An international comparison.” *Journal of Energy and Development* 10: 249-272.

Yu, E. S. H. and B. Hwang (1984). “The relationship between energy and GNP: Further results.” *Energy Economics* 6: 186-190.

Yu, E. S. H. and J. C. Jin (1992). “Cointegration tests of energy consumption, income, and employment.” *Resources and Energy* 14: 259-266.

Figure 1. Macro-Level Limits to Substitution

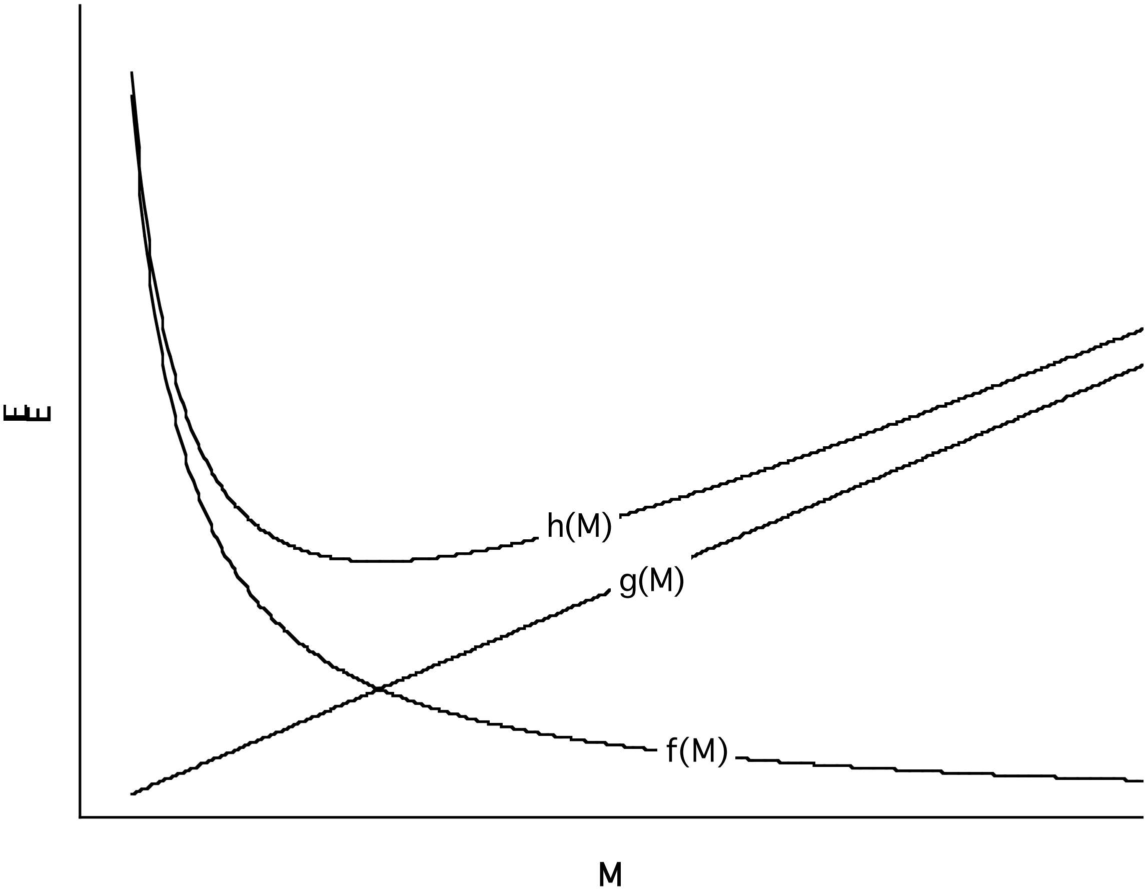
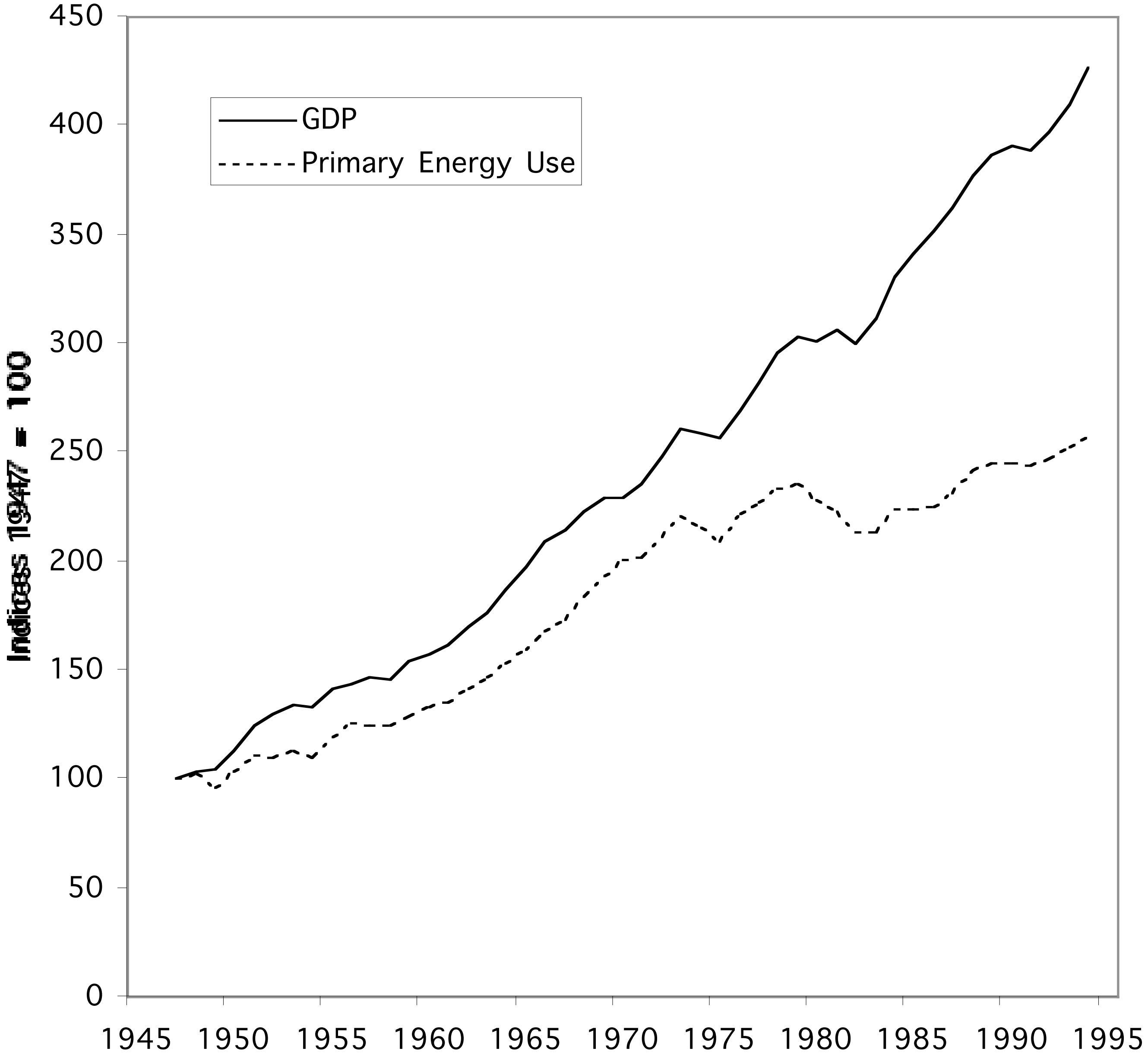


Figure 2. U.S. GDP and Primary Unadjusted Energy Use



Notes: GDP is in constant dollars i.e. adjusted for inflation. Energy use is the sum of primary energy BTUs.

Figure 3. Composition of US Energy Input

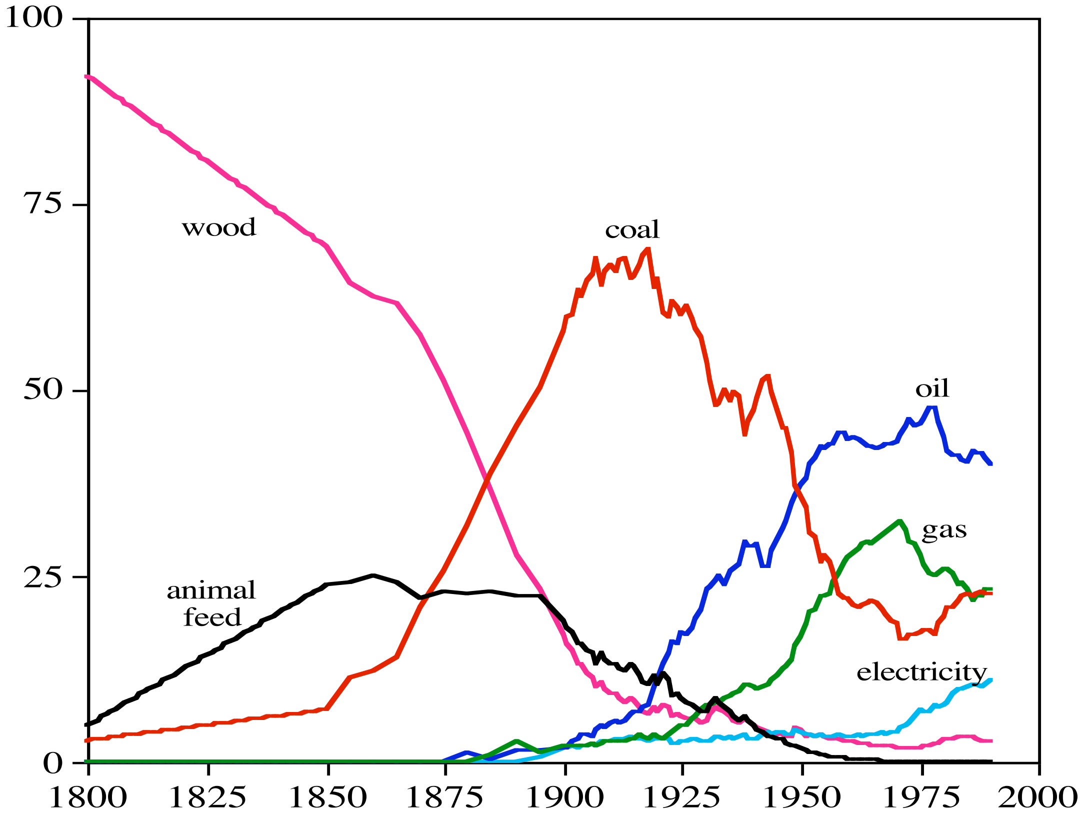
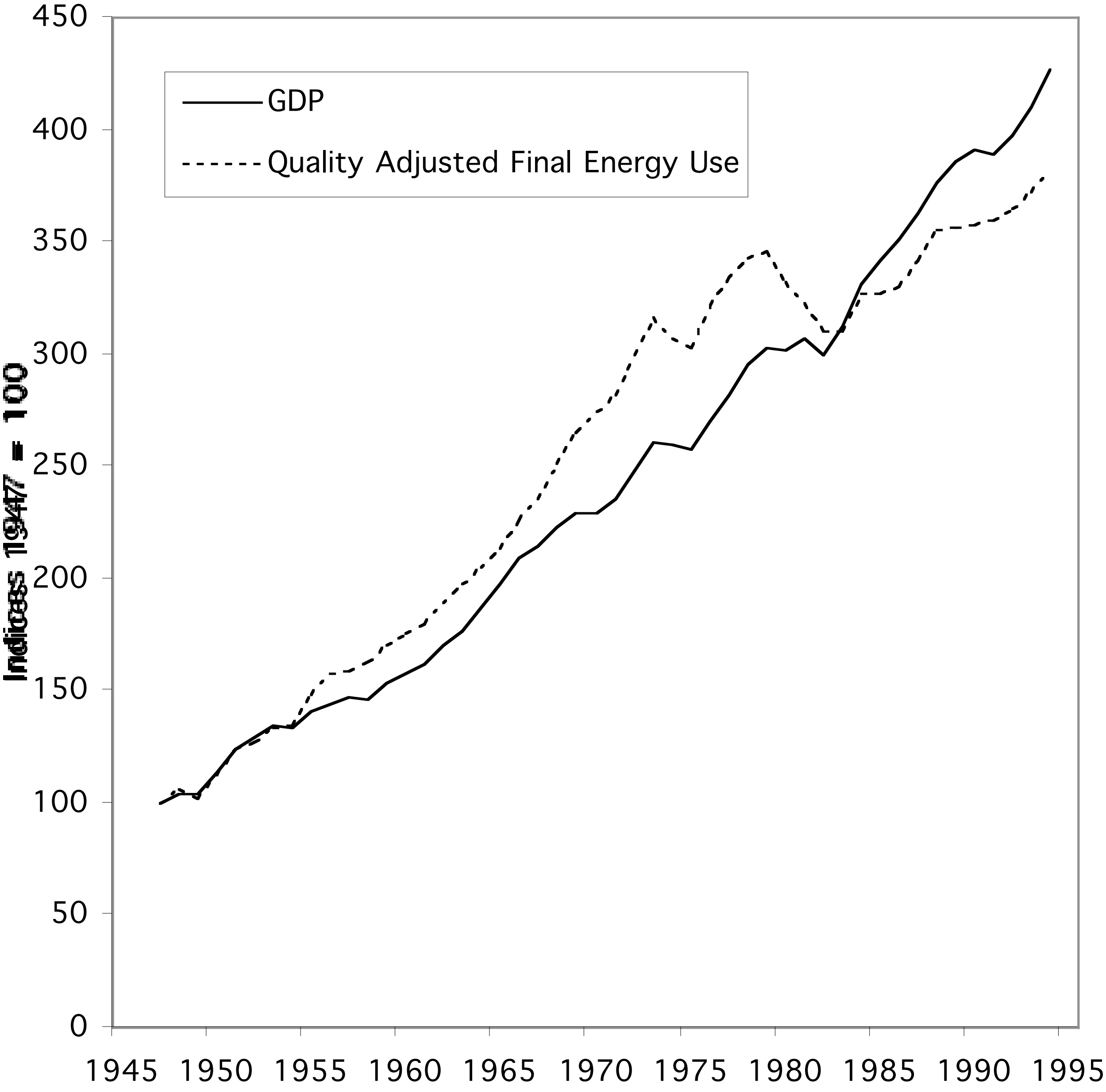


Figure 4. U.S. GDP and Quality Adjusted Final Energy Use



Notes: GDP is in constant dollars i.e. adjusted for inflation. Energy use is a Divisia index of the principal final energy use categories – oil, natural gas, coal, electricity, biofuels etc. The different fuels are weighted according to their average prices.