The Weight of Energy in Economic Growth

Robert U. Ayres[a,b], Jeroen C.J.M. van den Bergh[c,d,e], Reiner Kümmel[f], Dietmar Lindenberger[g], Benjamin Warr[a]

[a] European School of Business Administration, INSEAD, 77305 Fontainebleau, France.

[b] International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria.

[c] ICREA, 08010 Barcelona, Spain.

[d] Institute for Environmental Science and Technology, and Department of Economics and Economic History, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.

[e] Faculty of Economics and Business Administration, and Institute for Environmental Studies, Vrije Universiteit, Amsterdam, The Netherlands.

[f] Institute for Theoretical Physics and Astrophysics, University of Würzburg, D-97074 Würzburg, Germany.

[g] Institute of Energy Economics, Cologne University, D-50923 Cologne, Germany.

June 2009

Abstract
According to neoclassical economic theory, growth can – and will – continue at past rates regardless of the availability or cost of energy. This bizarre conclusion follows from the widespread assumption in economic models that capital and labor are the only important factors of production. That assumption can be traced to a textbook "theorem" which says that the output elasticity of energy in the economy must be proportional to the cost share of energy in the GDP. Since primary energy accounts for a very small fraction of the GDP – around 5 percent – it seems to follow that it cannot be an important factor of production. We argue that both ends of that proposition are wrong: the original "cost-share theorem" was derived for an oversimplified economic model that is not applicable to the real world. It follows that (1) energy is actually a much more important factor of production than its small cost share would indicate, and (2) that perpetual future growth cannot safely be assumed. A future scenario of shrinking reserves of fossil fuels and an increasingly stringent climate policy, with associated rising energy prices, has very negative implications for economic growth worldwide. We argue that "recovery" in the sense of returning to pre-crisis growth rates is unlikely.

Keywords:
Cost share, Energy, Economic growth, Limits to growth, Production.

1. Introduction

Virtually all of economic growth theory since (1,2) assumes that GDP growth per capita is driven partly by capital deepening, but mostly by an underlying, though unexplained long-term growth dynamic commonly interpreted as "technological progress". Modern theorists mostly attribute growth to an accumulation of "knowledge capital" (3,4). In either case, the driver of growth is uni-directional: up only. The standard theory of growth – as distinct from business cycle theory – does not reflect energy availability or prices and cannot explain declines (except as a consequence of reduced labor hours, which really are a consequence not a cause of the decline).

Nonetheless, economic declines (recessions) are a fact of economic life. While some past recessions may be attributable to non-economic events such as wars, others, including the present example, have evidently been caused by the collapse of "bubbles", which are unquestionably economic phenomena (5, 6). Given that the "Dot-Com" bubble of 1998-2000 and the real estate bubble of 2003-2007 were not predicted, or even recognized, by most economists until after the collapse, how much credence can we give to another very standard forecast of standard uni-directional energy-free growth theory: namely that there are not limits to growth and our grandchildren will be a lot richer than we are? We argue, on the contrary, that it is not safe to assume that long-term growth will continue along the historical "track" and that future growth, especially in the industrial world, at best will be much slower than in the past.

2. Energy: the neglected factor of production

There is good reason to doubt that past GDP growth per capita is entirely explained by capital accumulation or non-specific knowledge accumulation. In the first place, those factors of production rarely, if ever, decline, as noted above. More important, the standard classification of the national accounts divides all payments into just two categories: payments to labor (wages, salaries) and payments to capital (interest, dividends, rents and royalties). Nevertheless, while the national accounts do not reflect payments to or from a category called "energy", it is obvious that neither labor nor capital can function without inputs of energy, either as food or animal feed, or as fuel for engines, or electric power for light, communications and appliances of all sorts. In other words, a flow of energy, in some form, is just as essential for economic output (production) as capital or labor. It should, logically, be regarded as a factor of production, along with labor and capital.

Yet energy has been largely ignored by economic theory, since the first attempts at quantification by the French Physiocrats in the eighteenth century (7). The physiocrats regarded agriculture as the basis of every economy, and agriculture was – for them – a function of land and labor (including labor by working animals). The energy inputs to the land by sun and rain were never considered separately. Their value was hidden in the value of arable land. Similarly, the value of the solar energy (actually exergy[1]) embodied in fossil fuels is implicitly assumed to be reflected in the price, and thus in the profits to mine owners and oil companies. In the past, these energy-related rents have been both hidden and very small, considered as a fraction of GDP, because the energy resources in question were abundant.

That situation has continued, even though coal and later oil became increasingly important and dominant inputs to economic activity in the past two centuries. The first attempts to treat energy as an explicit factor of production was a response to economic concerns raised by the Arab oil embargo and the accompanying “energy crisis” of 1973-74 and again by the Iranian Revolution in 1979-1980. Both associated price spikes were followed by deep recessions. Several economists introduced the KLEM production function, where K refers to capital, L to labor, E to energy and M to materials (8, 9, 10, 11). But the measurement difficulties were not resolved at the time, or since. Moreover, several critics, notably Denison (12), the “dean” of growth accountants, argued that energy prices could not have a significant impact on GDP because the cost share of energy (and materials) in the national accounts was so small – only 4 or 5 percent for most OECD countries.

3. The cost-share confusion

The Denison criticism was widely accepted, but it was based on a drastically oversimplified model of the economy: namely an economy consisting of a single sector producing a single product serving both as a consumption good and as a capital good. Mankiw (13, p.30) imagines an economy consisting of a large number of small bakeries producing bread from rented capital and rented labor. In such an economy, it is an easy textbook exercise to show that each input factor (capital and labor) will be used in proportion to its marginal productivity. In this simple model economy it follows that the output elasticity (close to a measure of marginal productivity) of each factor will be exactly equal to its share of all costs (payments to capital and labor), i.e. to its cost share in the national accounts. We call this the “cost-share theorem”.

The strange logic of the above argument might have been challenged back in the early 1980s, except for the fact that the cost shares of capital and labor for the US economy have been relatively constant at 0.3 and 0.7 respectively throughout the twentieth century. Thus the equality between cost shares and output elasticities also justifies the standard Cobb-Douglas (C-D) production function. This empirical observation has seemed like confirmation both of the over-simplified model-economy used to derive the cost-share theorem, and the use of the C-D function in economic forecasting models.

However the single-sector, single-good model is clearly unrealistic. When the economy is considered as a multi-sector, multi-good system, it becomes obvious that the impact of a cut in one essential input can have a much bigger effect on the whole economy than its cost share. Raw material inputs go only to the primary extractive sectors, even though value is added by a sequence of downstream sectors. If an *essential* input to the primary sector is absent, and if there is no substitute, the whole system must fail. For instance, consider the example of Schelling (14). He supposes that US agricultural output drops by a factor of two, due to some climate disaster. The quantitative loss to the economy *from that sector alone* would be fairly small. Since agriculture accounts for only about 4 percent of US GDP, he notes that a 50 percent loss would only be 2 percent of US GDP (disregarding price increases). That loss, as Schelling notes, might be made up thanks to a single year’s economic growth – if the growing economy had only a single sector producing a single product.

But in the real multi-sector economy, an impact to the agricultural sector would also be felt by food and beverage processors, truckers, wholesalers, retailers, hotels and restaurants, not to mention consumers. The overall loss to the economy would be many times greater than the 2 percent attributable to loss of agricultural output alone. Presumably a good applied general equilibrium model with an input-output structure, would capture these indirect impacts. But certainly the overall impact of a 50 percent drop in agricultural output would be a lot greater than 2 percent of the GDP.

Similarly, a few years ago petroleum inputs to the economy also accounted for around 4 percent of the US GDP. Schelling's argument, as applied to agriculture would apply equally to oil. Suppose oil inputs were cut by a factor of two, as in Schelling's agricultural example. The overall impact in that hypothetical case would also be 2 percent of GDP. But in the real, multi-sector economy, a 50 percent cut in petroleum supplies would cut oil refinery output and petrochemical output. Car, truck and air transportation would be cut by virtually the same amount (because there is no immediate, practical substitute for liquid hydrocarbon fuels) and all other sectors depending on non-electrified transportation services would be affected to a similar degree. The overall impact on the economy, as captured by a good general equilibrium model would be far greater than 2 percent. The multiplier effect in this case would probably be closer to a factor of ten, if not more. Putting it another way, *the marginal productivity (output elasticity) of petroleum must be far greater than its tiny cost share*.

It turns out that the traditional cost share theorem as taught in textbooks is not even true for a single sector economy if there are constraints on input factor combinations. Traditionally, labor and capital are assumed to be perfectly substitutable, but, in reality, the range of substitutability is fairly narrow. There is an optimal ratio: too much labor, or too little, will result in under-utilization of labor or of machines, respectively.

This is equally true in the three factor case. We know that both capital and labor would be totally unproductive without a flow of energy (exergy) to nourish the workers and drive the machines and computers. For instance, in the bakery case, we postulate a need for fuel gas for the bakery ovens. And suppose we insist that there is a certain fixed requirement of gas fuel by each oven. If the flow of fuel is too great it will be wasted (or even harmful – the baker will be poisoned or the bread will be burnt). More generally, suppose the optimal requirements of capital, labor and energy can be expressed as functional relationships, such as ratios.[2] Again, this implies non-substitutability, except perhaps over a narrow range.

To make a long story short, it turns out that the simple "cost-share theorem" is not valid. The output elasticity of each of the three factors depends not only on the cost share of the factor but also on the "shadow prices" due to technological constraints. The bottom line is that the output elasticity need not be equal to the cost share. It can be much larger (or smaller). The revised version of the cost-share theorem is derived in the Appendix.

4. Energy and economic growth

As we have mentioned earlier, shortages and price spikes must have a negative impact on economic growth. A suitably modified theory of growth should be able to explain the many observations of growth slowdowns following price spikes (15, 16).

Parenthetically, it seems very likely to us that the very high oil prices during the spring and summer of 2008 hastened the end of the real-estate price boom by squeezing household expenditures at a time when many people already had big credit card debts and no savings, which in turn, caused an uptick in foreclosures. That may have triggered the financial meltdown, which followed from the realization that mortgage-based securities could not be priced realistically, which meant that many financial institutions and banks were over-leveraged.

More important, in the long run, the forthcoming advent of “peak oil”, whether it has already happened or whether it occurs ten or twenty years in the future, must have a significant negative impact on future global economic growth. The reason is that energy in general, and oil in particular, are essential to virtually all economic activity, with marginal productivity (output elasticity) far greater than its still small – though increasing – cost share. As the prices of oil and oil substitutes rise, the demand for energy intensive products will fall, as happened in the autumn of 2008. That brings the price of oil temporarily back down, which encourages renewed consumption but discourages investment in energy conservation measures that depend on higher prices. This, in turn, delays needed economic adjustment while accelerating the onset of the next crisis.

In a realistic multi-sector, multi-product economy we find that the output elasticity of an *essential* (non-substitutable) input, like petroleum, or more generally, energy, tends to be much larger than its cost-share, whereas the output elasticity of labor tends to be much smaller than its cost share. This discrepancy can be interpreted in a more down-to-earth way. It can be argued that raw (unskilled) labor is over-priced in modern economies whereas flows of energy, especially petroleum, have been relatively under-priced up to now. This suggests a connection with the policy proposal to shift taxes from labor to energy, as a means of environmental regulation with an additional benefit, namely reducing labor market distortions and thus lowering average unemployment rates (17). Indeed, past studies may have underestimated the potential of such a policy due to underrating the role of energy.

The non-equality of output elasticities and cost shares has important consequences for the standard theory of economic growth. The first implication is that the standard Cobb-Douglas production function must be discarded because it assumes that output elasticities are equal to cost shares and that the latter are constant. Dropping this assumption implies that the output elasticities of factor inputs must be functions of all the input variables, namely capital, labor and energy or energy services. (18) have shown that the simplest functional form for a production function that allows for non-constant output elasticities, takes into account the energy flows in a physically plausible way, and permits an explicit parametric formulation of the constraints, is the so-called LINEX production function.

When growth theory is suitably modified to reflect the true importance of energy as an input, it turns out that the primary driver of growth, apart from capital deepening, is the increasing supply of “useful work” (mechanical work, chemical work, electrical work, etc.) in the economy (19-22). This has been a consequence of two past trends: (1) the discovery of oil (and gas) reserves, and (2) the increasing efficiency of conversion of primary energy (fossil fuels) into various forms of useful work, such as electric power and motive power.

The advent of “peak oil” means that, as the supply of oil and gas cannot be expected to continue to increase in the future, driving energy prices down – as it did for most of the last two centuries – future economic growth will depend more than in the past on technological progress, especially in the area of increasing energy (exergy) efficiency in the economy. Yet, the rate of exergy efficiency increase (in the US, at least) has been slowing down since the 1970s. The bottom line, here, is that either US economic growth will slow down permanently (with global consequences) or effective measures to increase the rate of increase of exergy efficiency must be undertaken to compensate for the coming decline in natural resource availability. An additional problem is that energy conservation and improved efficiency will invite rebound effects from both consumers and producers, which will partly undo the original reductions in energy use (23).

5. Conclusions

The non-equivalence of cost shares and output elasticities has enormous consequences for the analysis of economic growth in the presence of energy and environmental constraints. For instance, the underpricing of energy (exergy) resources accounts for our “addiction” to oil (as President Bush put it) and our over-dependence on the internal combustion engine. This, in turn, accounts for most of the atmospheric pollution, especially of greenhouse gases (GHGs), that has accompanied our global industrialization process. It further accounts for the logical inference that energy taxes or pollution taxes (or both) may be the way forward in terms of confronting the challenge of climate change by reducing GHG emissions.

The assumption of painless and perpetual GDP growth implies (to some) that “our grand-children will be a lot richer than we are”. This proposition must be challenged and discarded for at least two reasons. In the first place, it depends on the assumption that growth will continue as in the past, even if energy becomes scarce and ever more expensive. The rather standard assumption that economic growth is independent of energy availability must be discarded. It is not tenable.

The second reason for challenging this proposition is that it depends on the false assumption that GDP is truly a measure of human welfare (24-26). Yet it is clear already that climate change will impose huge environmental damages, due to more intense storms, floods, rising sea levels, the onset of new diseases (because micro-organisms evolve faster than large animals develop immunity) and mass migrations from threatened areas. Moreover, the economic activities undertaken to prevent repair or compensate for damages – fighting new diseases, building fences to keep out refugees or dikes to protect coastal cities, for example – will increase the GDP, without increasing anyone’s welfare. Indeed, there is now a very real potential for resource wars, as nations try to secure long-term energy supplies for themselves (27). Many believe that the Iraq war was such an attempt.

In summary, we believe that the future will not be a straightforward continuation of past trends. Deeper consideration is needed.

Appendix: Economic equilibrium under technological constraints

Suppose an economic system produces output Y with three factors of production X_1,X_2,X_3, whose combinations are subject to technological constraints, labeled by the index a and expressed by the equations $f_a(X_1,X_2,X_3,t) = 0$ with the help of slack variables. Then profit maximization under constant returns to scale results in three equilibrium conditions

$$\varepsilon_i = \frac{X_i}{Y}\frac{\partial Y}{\partial X_i} = \frac{X_i(p_i + s_i)}{\sum_{i=1}^{3} X_i(p_i + s_i)} \qquad i = 1,2,3. \tag{1}$$

These conditions relate the output elasticities ε_i of factors X_i to market prices p_i per factor unit and the factor shadow prices

$$s_i = -\sum_a \frac{\mu_a}{\mu}\frac{\partial f_a}{\partial X_i} \qquad i = 1,2,3.$$

where μ_a and μ are the Lagrange multipliers of the technological and fixed-cost constraint equations in the optimization calculus. Thus, the output elasticities in eq. (1) are equal to "shadowed" cost shares. (Intertemporal utility optimization yields that for decreasing marginal utility of consumption, $dU/dC<0$, the shadow price of capital contains an additional term proportional to the time derivative of dU/dC. This term vanishes in linear approximation.)

If there were no technological constraints, all Lagrange multipliers μ_a would be zero, the shadow prices s_i would vanish, and one would have the usual factor cost shares on the r.h.s of eq. (1). In the presence of technological constraints and non-zero shadow prices, on the other hand, the ratios μ_a/μ of Lagrange multipliers are finite and functions of the output elasticities ε_i. The dependence of the shadow prices on the output elasticities prevents the calculation of the latter from the equilibrium conditions.

If the production factors are capital, labor, and energy(exergy) there are two technological constraints on the combinations of these factors:

1) One cannot feed more energy into the machines of the capital stock than they are designed for. Or, in other words, the degree of capacity utilization of the capital stock cannot exceed one.
2) The state of technology determines the possible degree of automation which cannot be exceeded by combinations of capital and energy that substitute for routine labor in the process of automation.

The detailed equations for constraints and shadow prices are presented at http://www.ewi.uni-koeln.de/fileadmin/user/WPs/ewiwp0802.pdf, where it is also shown how output elasticities and the LINEX function are calculated independently from the equilibrium conditions.

Notes

[1] Exergy is defined as available energy, meaning energy that can do physical work. More precisely, it is the maximum work that can be done by a system reversibly approaching thermodynamic equilibrium. It is what most people mean when they speak of energy. Whereas energy is conserved and cannot be used up, exergy is destroyed by transformation processes (as entropy increases).

[2] Evidently the constraints are somewhat “fuzzy”, which is probably why their importance has been overlooked. Capital-labor relationships and capital exergy relationships are more like distribution functions, with a range of possible GDP output values around a central peak. Substitution of inputs is possible within a narrow range. It is evident that the economy at any point in time does require all three factors to be present in reasonable proportions.

References

1. Solow RM (1956) A contribution to the theory of economic growth. *Quart J Econ* 70: 65-94.
2. Solow RM (1957) Technical change and the aggregate production function. *Rev Econ Statist* 39: 312-320.
3. Romer PM (1990) Endogenous technological change. *J Polit Economy* 98(5): S71-S102.
4. Romer PM (1994) The origins of endogenous growth. *J Econ Perspect* 8(1): 3-22.
5. Kindleberger CP (1989) *Manias, Panics and Crashes: A History of Financial Crises,* 2nd ed. (Basic Books, New York).
6. Streeter WJ (1984) *The Silver Mania* (D. Reidel/Kluwer, Dordrecht).
7. Hall CAS, Lindenberger D, Kümmel R, Kroeger T, Eichhorn W (2001). The need to reintegrate the natural sciences with economics. *BioSci* 51: 663-73.
8. Hudson EA, Jorgenson DW (1974) U.S. energy policy and economic growth 1975-2000. *Bell J Econ* 5: 461-514.
9. Jorgenson DW (1978). The role of energy in the US economy. *Nat Tax J* 31: 209-220.
10. Jorgenson DW (1984) The role of energy in productivity growth. *Amer Econ Rev* 74 (2): 26-30.
11. Berndt ER, Wood DO (1979) Engineering and econometric interpretations of energy-capital complementarity. *Amer Econ Rev* 69: 342-354.
12. Denison EF (1979). Explanations of declining productivity growth. *Survey of Current Business* 59(Part II): 1-24.
13. Mankiw NG (1997) *Macroeconomics*. (Worth Publishing, London).
14. Schelling TC (1997) The cost of combatting global warming. *Foreign Aff* 76(6): 8-14.
15. Hamilton JD (1983). Oil and the macroeconomy since World War II. *J Polit Economy* 91: 228-248.
16. Hamilton JD (2003) What is an oil shock? *J Econometrics* 113: 363-398.
17. Patuelli R, Nijkamp P, Pels E (2005) Environmental tax reform and the double dividend: A meta-analytical performance assessment. *Ecolog Econ* 55(4): 564-583.
18. Kümmel R, Henn J, Lindenberger D (2002). Capital, labor, energy and creativity: modeling innovation diffusion. *Struct Change Econ Dynam* 13: 415-433.
19. Ayres RU, Ayres LW, Warr B (2003) Exergy, power and work in the US economy, 1900-1998. *Energy* 28(3): 219-273.

20. Ayres RU, van den Bergh JCJM (2005) A theory of economic growth with material/energy resources and dematerialization: Interaction of three growth mechanisms. *Ecolog Econ* 55 (1): 96-118.
21. Ayres RU, Warr B (2005) Accounting for growth: The role of physical work. *Struct Change Econ Dynam* 16(2): 181-209.
22. Ayres RU, Warr B (2009) *The Economic Growth Engine: How Energy and Work Drive Material Prosperity* (Edward Elgar, Cheltenham).
23. Sorrell S, Dimitropoulos J (2008) The rebound effect: Microeconomic definitions, limitations and extensions. *Ecolog Econ* 65(3): 636-649.
24. Nordhaus W, Tobin J (1972) in *Economic Growth: Fiftieth Anniversary Colloquium* (National Bureau of Economic Research, New York), pp. 1-80.
25. Daly HE, Cobb J (1989) *For the Common Good* (Beacon Press, Boston).
26. van den Bergh JCJM (2009) The GDP paradox. *J Econ Psych* 30: 112-135.
27. Klare MT (2001) *Resource Wars: The New Landscape of Global Conflict*. (Henry Holt & Co., New York).