

New Journal of Physics

The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft  DPG | IOP Institute of Physics

How energy conversion drives economic growth far from the equilibrium of neoclassical economics

Reiner Kümmel¹ and Dietmar Lindenberger²

¹ Institute for Theoretical Physics and Astrophysics, University of Würzburg, D-97074 Würzburg, Germany

² Institute of Energy Economics, University of Cologne, D-50827 Cologne, Germany
E-mail: kuemmel@physik.uni-wuerzburg.de

Received 16 June 2014, revised 23 October 2014

Accepted for publication 4 November 2014

Published 10 December 2014

New Journal of Physics **16** (2014) 125008

doi:[10.1088/1367-2630/16/12/125008](https://doi.org/10.1088/1367-2630/16/12/125008)

Abstract

Energy conversion in the machines and information processors of the capital stock drives the growth of modern economies. This is exemplified for Germany, Japan, and the USA during the second half of the 20th century: econometric analyses reveal that the output elasticity, i.e. the economic weight, of energy is much larger than energy's share in total factor cost, while for labor just the opposite is true. This is at variance with mainstream economic theory according to which an economy should operate in the neoclassical equilibrium, where output elasticities equal factor cost shares. The standard derivation of the neoclassical equilibrium from the maximization of profit or of time-integrated utility disregards technological constraints. We show that the inclusion of these constraints in our nonlinear-optimization calculus results in equilibrium conditions, where generalized shadow prices destroy the equality of output elasticities and cost shares. Consequently, at the prices of capital, labor, and energy we have known so far, industrial economies have evolved far from the neoclassical equilibrium. This is illustrated by the example of the German industrial sector evolving on the mountain of factor costs before and during the first and the second oil price explosion. It indicates the influence of the 'virtually binding' technological constraints on entrepreneurial decisions, and the existence of 'soft constraints' as well. Implications for employment and future economic growth are discussed.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](http://creativecommons.org/licenses/by/3.0/).
Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: energy, economic growth, technological constraints, output elasticities

1. Introduction: Thermodynamics and economics

On 23 October, 2009 a press release³ appeared, titled: ‘The Financial Crisis: How Economists Went Astray. Two Nobel Laureates and over 2000 Signatories Uphold that Economists have mistaken Mathematical Beauty for Economic Truth.’ The signatories signed a web petition in support of an article⁴ by the Nobel Laureate Paul Krugman, saying: ‘Few economists saw our current crisis coming, but this predictive failure was the least of the field’s problems. More important was the profession’s blindness to the very possibility of catastrophic failures in a market economy ... the economics profession went astray because economists, as a group, mistook beauty, clad in impressive-looking mathematics, for truth ...’

This statement reflects a wide-spread feeling among economists that their profession disregards important aspects of real-world economies. It is the purpose of this paper to show that the failure to describe modern economies adequately is *not* due to the introduction of calculus into economic theory by the so-called ‘marginal revolution’ during the second half of the 19th century, when the mathematical formalism of physics decisively influenced economic theory. Rather, the culprit is the disregard of the first two laws of thermodynamics and of technological constraints in the theory of production and growth of industrial economies.

A qualitative summary of the first and the second law of thermodynamics says that nothing happens in the world without energy conversion and entropy production. Entropy production density is positive in all irreversible processes, which, of course, include those of economic production. It consists of particle current densities, driven by specific external forces and by gradients of temperature and chemical potentials, and heat current densities, driven by gradients of temperature. A discussion of the environmental impacts of the corresponding emissions, and a rough mathematical modeling of the limits to growth that may result from them, is given in [1]. It concerns problems of the future like climate change. In the present article we constrain ourselves to elucidating the role of energy in the economy by econometric analyses of the past.

Figure 1 shows the model we use for that purpose. This model had been—and is—the intuitive response to the discussions on the limits to growth, which stimulated research on thermodynamics and economics. It includes energy (more precisely *exergy*)⁵ as a third factor of production on an equal footing with the traditional factors capital and labor. In addition, it introduces the factor ‘creativity,’ which is the specific human contribution to economic growth that cannot be made by any machine capable of learning. Creativity works via ideas, inventions and value decisions and is coupled to the flow of time t . The space available for the evolution of

³ From Professor Geoffrey M Hodgson, The Business School, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK; www.geoffrey-hodgson.info

⁴ New York Times, 2nd September, 2009.

⁵ Energy consists of useful *exergy* and useless *anergy*. Exergy can be converted into any form of useful work, whereas anergy is heat dumped into the environment, for instance. Entropy production enhances anergy at the expense of exergy. Since all primary energy carriers taken into account in our analyses (coal, oil, gas, renewables, and nuclear fuels) are basically 100% exergy, we do not discriminate between energy and exergy, where ‘energy consumption’ means ‘exergy consumption.’

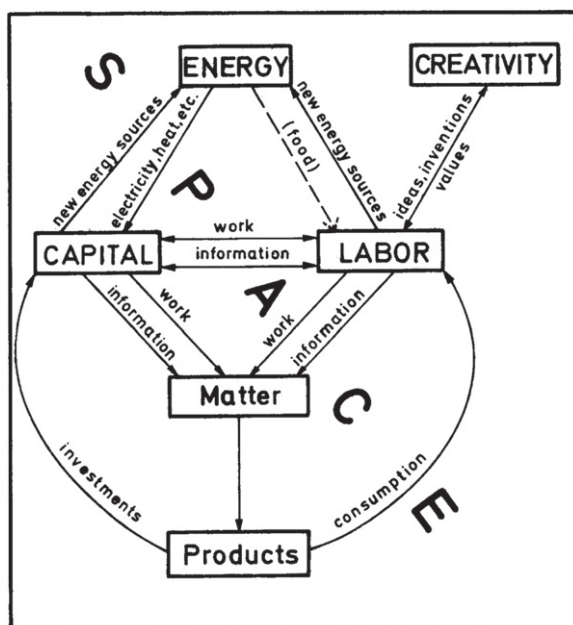


Figure 1. The capital–labor–energy–creativity (KLEC) model of wealth production. Reproduced from [1], chapter 4, p 174, with kind permission of Springer Science + Business Media.

the production system provides natural resources, production sites and absorbs emissions⁶. Work performance and information processing by the production factors (instrumental) capital K , labor L , and energy (conversion) E are the basic processes that produce the material wealth represented by the value added of goods and services. These products represent the output Y of the economy. The capital stock K consists of all energy-converting devices and information processors and the buildings and installations necessary for their protection and operation. Value added per year Y and capital K are measured in constant currency by the national accounts, labor L is given in ‘hours worked per year’ by the national labor statistics, and energy E is measured by the national energy balances in, e.g., ‘petajoules converted per year.’ Obviously, there is no quantitative *ex ante* measure of creativity; nevertheless creativity’s impact can be determined *ex post*, as it is done in section 3.

Economic actors decide how much of the factors K , L , and E are used in order to produce a certain amount of output Y . According to a fundamental behavioral assumption of neoclassical economics the factors are combined in such quantities that the economy operates in an equilibrium that is determined either by the optimization of profit or of time-integrated utility (welfare). We apply these optimization principles to energy-dependent economies in section 2 [1, 3]. In so doing, we take technological constraints on K , L and E into account, which have been ignored so far, and disprove the general validity of the fundamental cost-share theorem of standard economics. This theorem says that the economic weight of a production factor, which is called the *output elasticity* of that factor, should always be equal to the factor’s share in total factor cost. In highly industrialized countries during the second half of the 20th century, the share of energy in total factor cost has been roughly a meager 5%, while the share of labor has

⁶ ‘Space’ expands the traditional factor ‘land’ into the third dimension.

been about 70 and that of capital 25% [4]. Consequently, if energy has been considered at all, only a marginal role has been attributed to it by mainstream economics.

According to the cost-share theorem, reductions of energy inputs by up to 7%, observed during the first energy crisis 1973–1975, could have only caused output reductions of 0.35%, whereas the observed reductions of output in industrial economies were up to an order of magnitude larger. Thus, from this perspective the recessions of the energy crises are hard to understand. In addition, cost-share weighting of production factors has the problem of the Solow residual. The Solow residual accounts for that part of output growth that cannot be explained by the input growth rates weighted by the factor cost shares. It amounts to more than 50% of total growth in many countries. Standard neoclassical economics attributes the discrepancy between empirical and theoretical growth to what is being called ‘technological progress’ or, sometimes, ‘Manna from Heaven.’ The dominating role of technological progress ‘has lead to a criticism of the neoclassical model: it is a theory of growth that leaves the main factor in economic growth unexplained’ [5], as the founder of neoclassical growth theory, Robert A Solow, stated himself.

The observation that since the Industrial Revolution technological progress has manifested itself in increasing numbers of energy-converting devices and information processors, driven by increasing energy inputs and improved by innovations, suggests that ‘technological progress’ results from the cooperation of the production factor energy with capital, labor, and human creativity, as indicated in figure 1. This is the pre-analytic vision that has guided the research presented in this paper.

Econometric analyses of economic growth in Germany, Japan and the USA during the second half of the 20th century are the subject of section 3. They result in output elasticities that are for energy are much larger and for labor much smaller than the cost shares of these factors. That this is compatible with technologically constrained profit maximization is shown in section 4 by the path of the German industrial sector in its cost mountain. ‘Summary and outlook’ points out social and ecological problems that arise from the pivotal role of energy in economic growth.

2. Equilibrium from profit and welfare optimization

Mainstream economics is essentially interested in the behavior of economic actors on markets and models it mathematically in formal analogy with classical mechanics. Söllner [6] presents a modern review of this analogy. He points out why and how the 19th-century neoclassical pioneers like Jevons, Edgeworth, and Walras developed the mathematical formalism of economics in a rather close one-to-one correspondence to classical mechanics, especially Newtonian mechanics of the point-like mass; see also [7]. In this formalism, extremum principles of classical mechanics like the minimization of energy, or Hamilton’s principle of least action, are the godfathers of the principles of profit or welfare optimization used in economics to determine the equilibrium where an economic system is supposed to operate⁷.

In equilibrium, the variables of a system adjust within given constraints in such a way that a system-specific objective becomes an extremum. Let us look into the equilibria that result from profit and welfare optimization.

⁷ Ironically, neoclassical economics uses Hamiltonians as in classical mechanics, but these Hamilton functions have nothing to do with energy, which standard economics disregards as a factor of production.

2.1. Production function and growth equation

A production function $Y(K, L, E; t)$ expresses the value added Y of the goods and services produced in an economic system as a function of the production factors and time. The output Y is the gross domestic product (GDP) of a national economy, or a part of the GDP produced by an economic subsector. Y is a state function⁸ of the inputs K, L, E in the same sense as the potential energy of a particle in a conservative force field, or thermodynamic potentials like Gibb's free energy, are state functions of their spatial or thermodynamic variables.

The growth equation

$$\frac{dY}{Y} = \alpha \frac{dK}{K} + \beta \frac{dL}{L} + \gamma \frac{dE}{E} + \delta \frac{dt}{t - t_0}, \quad \delta \equiv \frac{t - t_0}{Y} \frac{\partial Y}{\partial t}, \quad (1)$$

which is just the total differential of the production function $Y(K, L, E; t)$, divided by Y , is governed by the output elasticities of capital, α , labor, β , and energy, γ , defined as

$$\alpha \equiv \frac{K}{Y} \frac{\partial Y}{\partial K}, \quad \beta \equiv \frac{L}{Y} \frac{\partial Y}{\partial L}, \quad \gamma \equiv \frac{E}{Y} \frac{\partial Y}{\partial E}. \quad (2)$$

α, β , and γ measure, how the growth rate of output, dY/Y , depends on the growth rates $dK/K, dL/L, dE/E$ of the inputs K, L, E . From equations (1) and (2) we see that the *output elasticity* of a production factor measures the productive power of the factor in the sense that (roughly speaking) it gives the percentage of output change when the factor changes by 1%, while the other factors stay constant. Since they give the economic weights (productive powers) of the production factors, the output elasticities are of fundamental importance for the theory of production and growth. δ in equation (1) describes the influence of creativity, with t_0 being the base year of the time-series analysis performed in section 3; we will see that creativity manifests itself in time-changing technology parameters.

At any *fixed* time t an increase of *all* inputs by the same factor κ must increase output by κ , because at the fixed state of technology that exists at the given time t a, say, doubling of the production system doubles output; in other words: two identical factories with identical inputs of capital, labor and energy produce twice as much output as one factory. Thus, the production function must be linearly homogeneous in K, L, E . As a consequence, the Euler relation, in combination with equation (2), yields the so-called 'constant returns to scale' relation:

$$\alpha + \beta + \gamma = 1. \quad (3)$$

2.2. Technological constraints on K, L , and E

Capital, labor, and energy can be treated as independent variables within a certain region of positive K, L, E -space. This region is constrained technologically by the limit to the degree of capacity utilization $\eta(K, L, E)$ and the limit to the degree of automation $\rho(K, L, E)$. As pointed out in [1, 3], we have

⁸ Substantial critique of factor aggregation and of the neoclassical concept of macroeconomic production functions is discussed in appendix 3 to chapter 4 of [1]. Since work performance and information processing are the primary *technical* aggregation principles for output and factors, on which we base production functions (where equivalence factors relate technical quantities to monetary ones) [1, 2, 10], our conceptual foundation differs from the criticized one of neoclassical economics.

$$\eta(K, L, E) = \eta_0^* \left(\frac{L}{K} \right)^\lambda \left(\frac{E}{K} \right)^\nu, \quad \rho(K, L, E) = \eta \frac{K}{K_m(Y)}. \quad (4)$$

The parameters η_0^* , λ , and ν can be determined from empirical data on capacity utilization. An example is given in section 4. $K_m(Y)$ is the capital stock that would be required for maximally automated production of a given output Y ; in this state of the economy, an additional unit of labor would contribute no longer to the growth of output.

Obviously, $\eta(K, L, E)$ cannot exceed unity, because the capital stock cannot work at more than full capacity. (Energy inputs that exceed the power limits the machines are designed for, do not make more productive use of the capital stock. Rather, they may damage the machines; and too much heating or cooling of rooms is also counterproductive. Neither does it make sense to employ more workers than a production system, working at full capacity, requires for its operation and maintenance.) Furthermore, in a given state of technology at time t , the degree of automation of the capital stock has a limit $\rho_T(t)$ at $\eta = 1$. This limit depends on the mass and the volume of the energy-conversion devices and information processors the production system can accommodate when producing the output Y . The outmost limit to automation is 1, when $K = K_m(Y)$ and $\eta = 1$. Thus, the technological constraints on the combinations of capital, labor, and energy are

$$\eta(K, L, E) \leq 1, \quad \rho(K, L, E) \leq \rho_T(t) \leq 1. \quad (5)$$

The behavioral assumptions that firms maximize profit and individuals maximize welfare originate from microeconomics. But they are also applied to macroeconomic systems [4, 8]. Much more important than the difference between these two assumptions is the difference between neglect and non-neglect of the technological constraints (5) when calculating the equilibria in which macroeconomic systems supposedly operate at a given time t . In order to elaborate that, it is sufficient to look into the necessary conditions for profit and welfare maximization.

For notational convenience we identify K, L, E with the components X_1, X_2, X_3 of the vector

$$\mathbf{X} = (X_1, X_2, X_3) \equiv (K, L, E). \quad (6)$$

In this notation, and with the help of the slack variables \mathbf{X}_η and \mathbf{X}_ρ , the constraints (5) can be brought into the form of equations,

$$f_\eta(\mathbf{X}, t) = 0, \quad f_\rho(\mathbf{X}, t) = 0. \quad (7)$$

The slack variables for labor, energy, and capital are L_η, E_η and K_ρ . They define the range in factor space within which the factors can vary independently at time t . Inserting them into equation (4) yields the explicit form of equation (7) as

$$f_\eta(\mathbf{X}, t) \equiv \eta_0^* \left(\frac{L + L_\eta}{K} \right)^\lambda \left(\frac{E + E_\eta}{K} \right)^\nu - 1 = 0, \quad f_\rho(\mathbf{X}, t) \equiv \frac{K + K_\rho}{K_m(Y)} - \rho_T(t) = 0. \quad (8)$$

2.3. Optimization of profit

We assume that the three production factors (X_1, X_2, X_3) have the exogeneously given prices per factor unit (p_1, p_2, p_3) $\equiv \mathbf{p}$, so that total factor cost is $\mathbf{p}(\mathbf{t}) \cdot \mathbf{X}(\mathbf{t}) = \sum_{i=1}^3 p_i(t) X_i(t)$. Economic equilibrium is defined as the state in which profit

$$G(\mathbf{X}, \mathbf{p}, t) \equiv Y(\mathbf{X}, t) - \mathbf{p} \cdot \mathbf{X} \quad (9)$$

is maximum.

The necessary condition for a maximum of profit $G \equiv Y - \mathbf{p} \cdot \mathbf{X}$, subject to the technological constraints (8), is:

$$\vec{\nabla} \left[Y(\mathbf{X}; t) - \sum_{i=1}^3 p_i(t) X_i(t) + \mu_\eta f_\eta(\mathbf{X}, t) + \mu_\rho f_\rho(\mathbf{X}, t) \right] = 0, \quad (10)$$

where $\vec{\nabla} \equiv (\partial/\partial X_1, \partial/\partial X_2, \partial/\partial X_3)$ is the gradient in factor space; μ_η and μ_ρ are Lagrange multipliers. (The sufficient condition for profit maximum involves a sum of second-order derivatives. One assumes that the extremum of profit at finite X_i is the maximum.) Equation (10) yields the three equilibrium conditions

$$\frac{\partial Y}{\partial X_i} - p_i + \mu_\eta \frac{\partial f_\eta(\mathbf{X}, t)}{\partial X_i} + \mu_\rho \frac{\partial f_\rho(\mathbf{X}, t)}{\partial X_i} = 0, \quad i = 1, 2, 3. \quad (11)$$

Multiplication of (11) with X_i/Y , and writing the output elasticities α, β, γ , defined in equation (2), as

$$\epsilon_i \equiv \frac{X_i}{Y} \frac{\partial Y}{\partial X_i}, \quad i = 1, 2, 3, \quad (12)$$

brings the equilibrium conditions into the form

$$\epsilon_i \equiv \frac{X_i}{Y} \frac{\partial Y}{\partial X_i} = \frac{X_i}{Y} \left[p_i - \mu_\eta \frac{\partial f_\eta}{\partial X_i} - \mu_\rho \frac{\partial f_\rho}{\partial X_i} \right], \quad i = 1, 2, 3. \quad (13)$$

One can write these equilibrium conditions as

$$\epsilon_i = \frac{X_i [p_i + s_i]}{\sum_{i=1}^3 X_i [p_i + s_i]}, \quad i = 1, 2, 3, \quad (14)$$

by summing the left and right hand sides of (13) over $i = 1, 2, 3$, observing that $\sum_{i=1}^3 \epsilon_i = 1$ according to (3), and expressing Y by the other terms in the resulting relation. The s_i are

$$s_i \equiv -\mu_\eta \frac{\partial f_\eta}{\partial X_i} - \mu_\rho \frac{\partial f_\rho}{\partial X_i}. \quad (15)$$

They map the constraints into monetary terms in our *nonlinear* optimization. We call them ‘generalized shadow prices’—‘generalized,’ in order to avoid confusion with the concept of ‘shadow prices,’ which refers to terms of the same form one obtains in linear optimization⁹. The Lagrange multipliers μ_η and μ_ρ depend on the factor prices p_i , the production function Y , and the derivatives of $f_\eta(\mathbf{X}, t)$, $f_\rho(\mathbf{X}, t)$. Y cannot be known without knowledge of the output

⁹ The concept of ‘generalized shadow prices’ is introduced here, because, in general, constrained nonlinear optimization may result in an optimum that is situated in the interior of the accessible region, whereas in linear optimization the optimum always lies on a vertex of the boundary, where constraints are binding. If a nonlinear objective function has a number of extrema in its domain, it may also have an extremum in the interior of the restricted domain that remains accessible when constraints are applied. The ‘generalized shadow prices’ are calculated by using the K, L, E magnitudes for the optimum *within* the restricted domain; for details see [1, 3]. Contrary to the ‘shadow price’ from linear optimization, a ‘generalized shadow price’ is not necessarily associated with a binding constraint. Gauvin [9] discusses shadow prices in nonconvex programming, using a nonlinear program with equality and inequality constraints.

elasticities ϵ_i . Equations (14) would *only* be suited for computing output elasticities, if the neoclassical optimum would lie *within* the region of K, L, E space that is accessible in the presence of the technological constraints (5). This is not the case for the prices of capital, labor, and energy we have known so far. Figure 6 shows an example how the constraint from capacity utilization prevents an economic system from ‘rolling downhill’ in its cost mountain—toward the region where profit would be maximum.

But in the *absence* of technological constraints, *all* of positive K, L, E space would be accessible to the economic system, the Lagrange multipliers μ_η, μ_ρ and the generalized shadow prices s_i would be zero, and the equilibrium conditions (14) would turn into the cost-share theorem: on the rhs of equation (14) the numerator would be the cost $p_i X_i$ of the factor X_i , the denominator would be the sum of all factor costs, and the quotient, which is equal to the output elasticity ϵ_i , would represent the cost share of X_i in total factor cost.

This would also justify the neoclassical duality of production factors and factor prices, according to which all relevant economic information on production is in the prices. This duality, which is often used in orthodox growth analyses, is a consequence of the Legendre transformation that results from the requirement that profit $G(\mathbf{X}, \mathbf{p})$ is maximum *without* any constraints on X_1, X_2, X_3 . Then equation (11) would hold with $\mu_\eta = 0 = \mu_\rho$ and yield equilibrium values $X_{1M}(\mathbf{p}), X_{2M}(\mathbf{p}), X_{3M}(\mathbf{p})$. With $\mathbf{X}_M(\mathbf{p})$ the profit function would turn into the price function

$$G(\mathbf{X}_M(\mathbf{p}), \mathbf{p}) = Y(\mathbf{X}_M(\mathbf{p})) - \mathbf{p} \cdot \mathbf{X}_M(\mathbf{p}) \equiv g(\mathbf{p}). \quad (16)$$

The essential information on production would be contained in the price function $g(\mathbf{p})$, which is the Legendre transform of the production function $Y(\mathbf{X})$. (This is in formal analogy to the Hamilton function being the Legendre transform of the Lagrange function in classical mechanics, or to enthalpy and free energy being Legendre transforms of internal energy in thermodynamics.) However, because of the technological constraints and the resulting generalized shadow prices, the cost-share theorem and equation (16) are *not* valid in general. For an understanding of the economy, prices are not enough.

2.4. Optimization of time-integrated utility (welfare)

Maximization of overall welfare is an alternative to the derivation of equilibrium conditions from profit maximization. It tests the sensitivity of the equilibrium relations between output elasticities, on the one hand, and factor quantities and prices, on the other hand, to modified behavioral assumptions. The optimization calculus is performed in analogy to the derivation of the Lagrange equations in classical mechanics; for more details see [1].

In welfare optimization one assumes that society maximizes (undiscounted) time-integrated utility U [8]. The integral W is called (overall) welfare. We take it between the times t_0 and t_1 , during which the relevant variables evolve along a curve symbolized by $[s]$. We limit the calculation to the simplest case that the utility function U only depends on consumption C : $U = U[C]$. A simple function with the property of decreasing marginal utility¹⁰ is, for instance,

$$U(C) = C_0 \ln C/C_0 + U_0. \quad (17)$$

Macroeconomic consumption is the difference between output and capital formation.

¹⁰ The utility of consuming an additional piece of bread is less for a satiated person than for a hungry one.

Thus, the optimization problem is: maximize overall welfare

$$W[s] = \int_{t_0}^{t_1} U[C]dt, \quad (18)$$

subject to the technological constraints (8) and the additional economic constraint that the total cost $\mathbf{p} \cdot \mathbf{X}$ of producing consumption C by means of the factors $(X_1, X_2, X_3) \equiv \mathbf{X}$ must not diverge. Rather, its magnitude $c_f(t)$ must be finite at all times t , where each price per factor unit, p_i , is exogenously given:

$$c_f(t) - \sum_{i=1}^3 p_i(t)X_i(t) = 0. \quad (19)$$

Therefore, besides μ_η and μ_ρ , the variational formalism of intertemporal welfare optimization includes the Lagrange multiplier μ , which takes care of (19). The curve $[s]$, of which $W[s]$ is a functional, depends on the variables that enter consumption C . Output (per unit time) is described by the macroeconomic production function $Y(\mathbf{X}; t)$. Part of Y goes into consumption C and the rest into new capital formation $\dot{X}_1 \equiv dX_1/dt$ plus replacement of depreciated capital. Economic research institutions provide the price of capital utilization p_1 as the sum of net interest, depreciation and state influences. We use this price. Since it already includes depreciation, we can omit explicit reference to the depreciation rate, so that consumption is given by

$$C(\mathbf{X}, \dot{X}_1) = Y(\mathbf{X}; t) - \dot{X}_1. \quad (20)$$

All taken together we have to optimize

$$W[s] = \int_{t_0}^{t_1} dt \left\{ U[C(\mathbf{X}, \dot{X}_1)] + \mu [c_f(t) - \mathbf{p} \cdot \mathbf{X}] + \mu_\eta f_\eta(\vec{\mathbf{X}}, t) + \mu_\rho f_\rho(\vec{\mathbf{X}}, t) \right\}. \quad (21)$$

Thus, $W[s]$ is a functional of the curve $[s] = \{t, \mathbf{X}: \mathbf{X} = \mathbf{X}(t), t_0 \leq t \leq t_1\}$.

Consider another curve $[s, \mathbf{h}] = \{t, \mathbf{X}: \mathbf{X} = \mathbf{X}(t) + \mathbf{h}(t), t_0 \leq t \leq t_1\}$ close to $[s]$, which goes through the same end points so that $\mathbf{h}(t_1) = 0 = \mathbf{h}(t_0)$. Its functional $W[s, \mathbf{h}]$ is obtained from $W[s]$ by changing \mathbf{X} to $\mathbf{X} + \mathbf{h}$ and \dot{X}_1 to $\dot{X}_1 + \dot{h}_1$ everywhere in the integrand of equation (21). Since \mathbf{h} is small, the integrand can be approximated by its Taylor expansion up to first order in \mathbf{h} and \dot{h}_1 .

The necessary condition for the maximum of overall welfare is that the variation $\delta W \equiv W[s, \mathbf{h}] - W[s]$ vanishes. δW contains the time integral of

$$\begin{aligned} \delta U &\equiv U[C(\mathbf{X} + \mathbf{h}, \dot{X}_1 + \dot{h}_1)] - U[C(\mathbf{X}, \dot{X}_1)] \\ &= \frac{dU}{dC} dC = \frac{dU}{dC} \left[\sum_{i=1}^3 \frac{\partial C}{\partial X_i} h_i + \frac{\partial C}{\partial \dot{X}_1} \dot{h}_1 \right]. \end{aligned} \quad (22)$$

Partial integration shifts the time derivative from \dot{h}_1 to the functions multiplying it. Then, the integrand in δW is a sum of terms that multiply h_1, h_2, h_3 . The condition

$$\delta W \equiv W[s, \mathbf{h}] - W[s] = 0 \quad (23)$$

is satisfied, if these terms vanish. Observing (20) one sees that they do, if

$$\frac{dU}{dC} \frac{\partial Y}{\partial X_i} = \mu p_i - \mu_\eta \frac{\partial f_\eta}{\partial X_i} - \mu_\rho \frac{\partial f_\rho}{\partial X_i} - \frac{d}{dt} \left(\frac{dU}{dC} \right) \delta_{i,1}, \quad i = 1, 2, 3; \quad (24)$$

the Kronecker delta $\delta_{i,1}$, is 1 for $i = 1$ and 0 otherwise¹¹.

Equations (24) are the general equilibrium conditions for an economic system subject to cost limits and technological constraints, if the behavioral assumption is that society optimizes time-integrated utility. They correspond to the equilibrium conditions (11) from profit optimization. The difference between the two equilibrium conditions becomes negligible under the conditions indicated below equation (25).

Dividing equations (24) by dU/dC and multiplying them by X_i/Y produces the output elasticities ϵ_i on the lhs, and the rhs involves terms that, after some manipulations, can be brought into the form of (modified) generalized shadow prices

$$s_i \equiv - \frac{\mu_\eta}{\mu} \frac{\partial f_\eta}{\partial X_i} - \frac{\mu_\rho}{\mu} \frac{\partial f_\rho}{\partial X_i} - \delta_{i,1} \frac{1}{\mu} \frac{d}{dt} \left(\frac{dU}{dC} \right), \quad i = 1, 2, 3. \quad (25)$$

With these modified s_i the equilibrium conditions (24) assume the form of equation (14).

The shadow prices (25) differ from those of (15) in two aspects. First, there is the term $\delta_{i,1} \frac{1}{\mu} \frac{d}{dt} \left(\frac{dU}{dC} \right)$. This term originates from equation (20) and is due to taking capital formation into account in intertemporal utility optimization, whereas capital formation is no issue in profit optimization. It vanishes, if one can disregard decreasing marginal utility and approximate the utility function $U[C]$ in (17) by its Taylor expansion up to first order in $C/C_0 - 1$. Second, the ratios μ_η/μ , μ_ρ/μ of Lagrange multipliers take the positions of the Lagrange multipliers μ_η , μ_ρ in (15). If one does profit maximization subject to the additional constraint $\vec{p} \cdot \vec{X} = c_f$, which fixes factor cost, one gets the quotients of the Lagrange multipliers as in (25). Thus, the second difference is rather a formal one.

As in profit maximization, because of the technological constraints the cost-share theorem is not generally valid in intertemporal utility maximization.

2.5. Virtually binding and soft constraints

From here on we consider economies that cannot operate in the neoclassical maximum of profit or welfare, because the barriers from the technological constraints and the associated generalized shadow prizes block the access to this maximum. Consequently, the equilibrium point may be either in the interior or at the boundary of the accessible region in K , L , E space.

Suppose an economy has the maximum of its objective function right at a barrier, where the '=' sign holds in (5). Nevertheless, it might not operate there either, because of several reasons not taken into account so far. For instance, entrepreneurs prefer combinations of K , L , E , in which the quantities of labor L and energy E , which handle and power the capital stock K of a certain degree of automation ρ , produce an output $Y(K, L, E)$ that meets *demand*. Since demand is fluctuating, entrepreneurs have to adjust the degree of capacity utilization η , defined in (4), accordingly. This cost-saving flexibility requires operation in points of factor space that are positioned at some distance from the barrier at $\eta(KL, E) = 1$. In such a situation,

¹¹ If one identifies $U[C(\mathbf{X}, \dot{X}_1)]$ with the Lagrangian $L(\mathbf{X}, \dot{X}_1)$, one sees the formal equivalence of (24) with constrained Lagrange equations of motion in classical mechanics.

the technological constraint on the degree of capacity utilization acts as a ‘virtually binding’ constraint.

Conceivably, there are yet other aspects that keep the economy at some distance from the constraint barrier(s). For instance, entrepreneurs may not wish to automate the capital stock as much as the technological limit $\rho_T(t)$ would allow, because they would have to pay compensations to laid-off workers according to contracts and negotiations with labor unions; legal and social obligations may matter as well. We call these aspects ‘soft’ constraints.

Thus, realistic economic decision makers operate in regions of K, L, E space, where the ‘<’ sign holds in equations (5) and where K, L , and E are independent variables.

3. Energy and economic growth

An alternative method that is needed for computing output elasticities has been developed in studies on energy and economic growth, which are reviewed in [1]. It is sketched subsequently, and its main results are presented.

3.1. Computing output elasticities and production functions

Production functions $Y(K, L, E; t)$, which describe economic evolution, are state functions and integrals of the growth equation (1) along a convenient path in factor space from an initial point (K_0, L_0, E_0) with Y_0 at time t_0 to the point (K, L, E) at the time of interest t . It is convenient to work with quantities that are normalized to their magnitudes in the base year t_0 . Thus, we work with the dimensionless variables

$$k(t) \equiv \frac{K(t)}{K_0}, \quad l(t) \equiv \frac{L(t)}{L_0}, \quad e(t) \equiv \frac{E(t)}{E_0}, \quad y[k, l, e; t] \equiv \frac{Y[kK_0, lL_0, eE_0; t]}{Y_0}. \quad (26)$$

The growth equation (1) and the output elasticities (2) are invariant under the transformations (26).

The production function $y[k, l, e; t]$ as a state function of k, l, e must be twice differentiable with respect to k, l, e . This results in the partial differential equations

$$k \frac{\partial \alpha}{\partial k} + l \frac{\partial \alpha}{\partial l} + e \frac{\partial \alpha}{\partial e} = 0, \quad k \frac{\partial \beta}{\partial k} + l \frac{\partial \beta}{\partial l} + e \frac{\partial \beta}{\partial e} = 0, \quad l \frac{\partial \alpha}{\partial l} = k \frac{\partial \beta}{\partial k} \quad (27)$$

for the output elasticities, where (3) has been taken into account. These equations correspond to the Maxwell relations in thermodynamics. The most general solutions of (27) are

$$\alpha = A\left(\frac{l}{k}, \frac{e}{k}\right), \quad \beta = B\left(\frac{l}{k}, \frac{e}{k}\right) = \int^k \frac{l}{k'} \frac{\partial A}{\partial l} dk' + J\left(\frac{l}{e}\right), \quad \gamma = 1 - \alpha - \beta; \quad (28)$$

here $A(l/k, e/k)$ and $J(l/e)$ are any differentiable functions of their arguments. The output elasticities, and thus the combinations of k, l, e , must satisfy the restrictions

$$\alpha \geq 0, \quad \beta \geq 0, \quad \gamma = 1 - \alpha - \beta \geq 0, \quad (29)$$

which result from the technical-economic requirement that all output elasticities must be non-negative. Otherwise, the increase of an input would result in a decrease of output—a situation the economic actors will avoid.

It is not hard to verify with the help of equations (1) and (28) that the corresponding general form of the twice-differentiable, linearly homogeneous production function is given by

the rhs of

$$y = e\mathcal{F}\left[\frac{l}{k}, \frac{e}{k}; t\right] \equiv e\mathcal{F}[u, v; t], \quad (30)$$

where $u \equiv l/k$, $v \equiv e/k$.

We insert (30) into the growth equation that results from (1) by the transformation (26) to dimensionless variables. This yields

$$\frac{d\mathcal{F}}{\mathcal{F}} = \beta(u, v) \frac{du}{u} - [\alpha(u, v) + \beta(u, v)] \frac{dv}{v} + \delta \frac{dt}{t - t_0}; \quad \delta \equiv \frac{t - t_0}{\mathcal{F}} \frac{\partial \mathcal{F}}{\partial t}. \quad (31)$$

The economic boundary conditions that would allow to determine the output elasticities (28) and the production function (30) *exactly* for a given economic system are not known, and never will be, because—according to the theory of partial differential equations—they would require knowledge of β on a surface and of α on a curve in k, l, e space. In this sense, all output elasticities and production functions are approximations.

The simplest solutions of equation (27) are constant output elasticities α_0 , β_0 , and γ_0 . The corresponding integral of the growth equation is the energy-dependent version of the well-known Cobb–Douglas production function:

$$y_{\text{CDE}} = y_{\text{CDE}}^0 k^{\alpha_0} l^{\beta_0} e^{1-\alpha_0-\beta_0} = y_{\text{CDE}}^0 e \left(\frac{k}{e}\right)^{\alpha_0} \left(\frac{l}{e}\right)^{\beta_0}. \quad (32)$$

The simplest *factor-dependent* solutions are the output elasticities $\alpha_{L1} = a(l + e)/k$, $\beta_{L1} = a(cl/e - l/k)$, and $\gamma_{L1} = 1 - ae/k - acl/e$. With these elasticities integration of the growth equation yields the (first) LinEx production function,

$$y_{L1} = y_{L1}^0 e \exp\left[a\left(2 - \frac{l + e}{k}\right) + ac\left(\frac{l}{e} - 1\right)\right], \quad (33)$$

which depends linearly on energy and exponentially on factor quotients. Here a , c , and y_{L1}^0 are technology parameters: a indicates capital efficiency, and c indicates the energy demand of the fully utilized capital stock. Innovations may change them in time and contribute to δ in equation (1). The output elasticity of capital, α_{L1} , vanishes for vanishing ratios of labor and energy to capital and thus reflects the law of diminishing returns. The output elasticity of labor, β_{L1} vanishes, when the capital stock approaches the magnitude k_m required for maximum automation, and when simultaneously the energy input approaches the quantity $e_m = ck_m$ that is demanded by the fully utilized capital stock k_m .

The technology parameters a , c and y_0 of the LinEx function are determined from the empirical time series of output $y_{\text{empirical}}(t)$. We allow for time dependencies of the technology parameters and model them by logistic functions [10], which are typical for growth in complex systems and innovation diffusion. Alternatively, we have used Taylor expansions of $a(t)$ and $c(t)$ in $(t - t_0)$. The free coefficients of the logistics, or of the Taylor expansions, are determined by minimizing the sum of squared errors

$$\text{SSE} = \sum_i \left[y_{\text{empirical}}(t_i) - y_{L1}(t_i) \right]^2. \quad (34)$$

The sum goes over all years t_i between the initial and the final observation time. It contains the empirical time series of output $y_{\text{empirical}}(t_i)$, and the LinEx function $y_{L1}(t_i)$ with the empirical

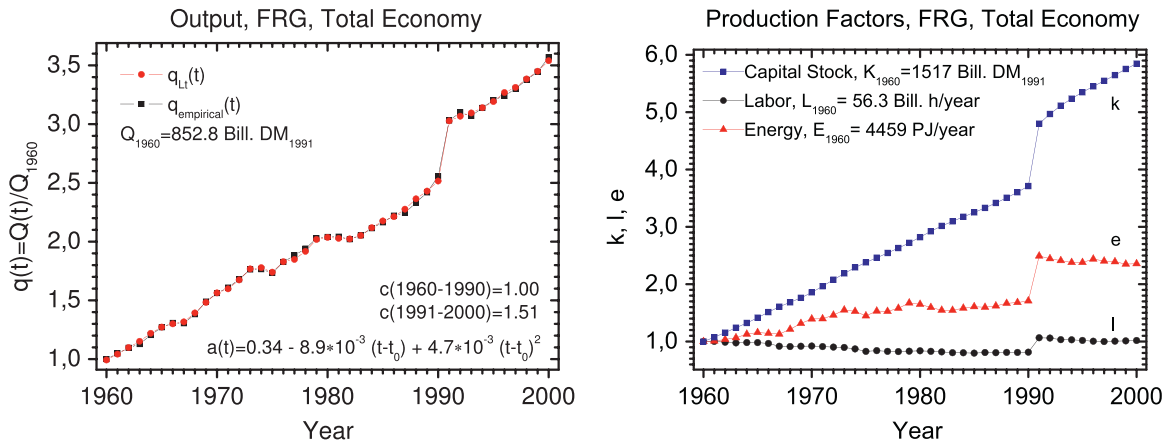


Figure 2. Growth in the total economy of the Federal Republic of Germany (FRG) between the years 1960 and 2000. The five coefficients, shown below the output graph, model the time dependence of the LinEx–function parameters a and c . They reproduce the drastic structural break at German reunification in 1990. Reproduced from [1], chapter 4, pp 206–208, with kind permission of Springer Science + Business Media.

time series of k , l and e as inputs at times t_i . Minimization is subject to the restrictions (29). Methodological details and the empirical time series of y ; k , l , e are given in [1].

3.2. Growth in Germany, Japan, and the USA

The reproduction of economic growth in Germany, Japan, and the USA during the second half of the 20th century by the LinEx function is shown in figures 2–5. These figures are taken from [11]. Their left parts exhibit the empirical growth (squares) and the theoretical growth (circles) of the dimensionless output $y = Y/Y_0 (\equiv q = Q/Q_0)$, and the right parts present the empirical time series of the dimensionless factors capital $k = K/K_0$, labor $l = L/L_0$, and energy $e = E/E_0$. The base year t_0 is 1960 for Germany and the USA, and 1965 for Japan. Note the variations of inputs and outputs in conjunction with the oil price explosions. The price of a barrel of crude oil in inflation-corrected US\$₂₀₁₁ was driven by the OPEC boycott in response to the Yom-Kippur war from 15\$ in 1973 to 53\$ in 1975, and by the war between Iraq and Iran from 48\$ in 1979 to 100\$ in 1981. Then the oil price plummeted to 30\$ in 1986 (with dramatic consequences for the Soviet Union). Between 1997 and 2011 it has risen again, from 18\$ to 110\$.

The time-averaged output elasticities of capital, $\bar{\alpha}$, labor, $\bar{\beta}$, energy, $\bar{\gamma}$, and creativity, $\bar{\delta}$, are presented in table 1 for the economic systems: FRG TE (total economy of the Federal Republic of Germany before and after reunification), FRG I (German industrial sector ‘Warenprouduzierendes Gewerbe’ (GWG)), Japan I (Japanese sector ‘Industries’), and USA TE (total economy of the USA). R^2 is the adjusted coefficient of determination, and d_W is the Durbin–Watson coefficient (whose best value is 2). Both statistical quality measures are quite good.

We note that for all the systems considered the output elasticity of energy is much larger than energy’s cost share of roughly 5%, and the output elasticity of labor is much smaller than labor’s cost share of about 70%. The economic weight $\bar{\delta}$ of creativity, computed from the time changes of $a(t)$ and $c(t)$, is comparable to that of labor, $\bar{\beta}$.

Ayres and Warr have analyzed economic growth in the USA during the 20th century, using ‘useful work’ as the energy variable in the LinEx function. ‘Useful work’ is defined as

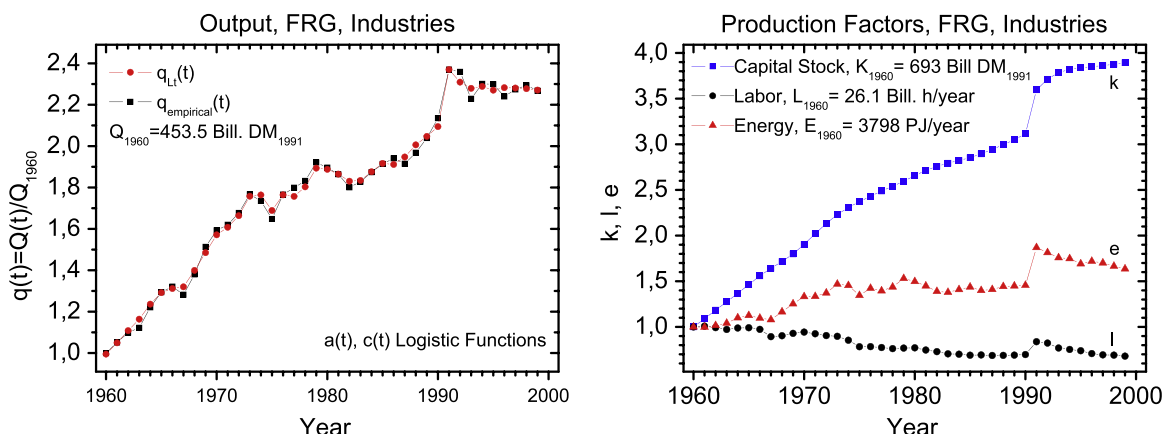


Figure 3. Growth in the German industrial sector ‘Warenproduzierendes Gewerbe’ (GWG) between 1960 and 1999. Reproduced from [1], chapter 4, pp 206–208, with kind permission of Springer Science + Business Media.

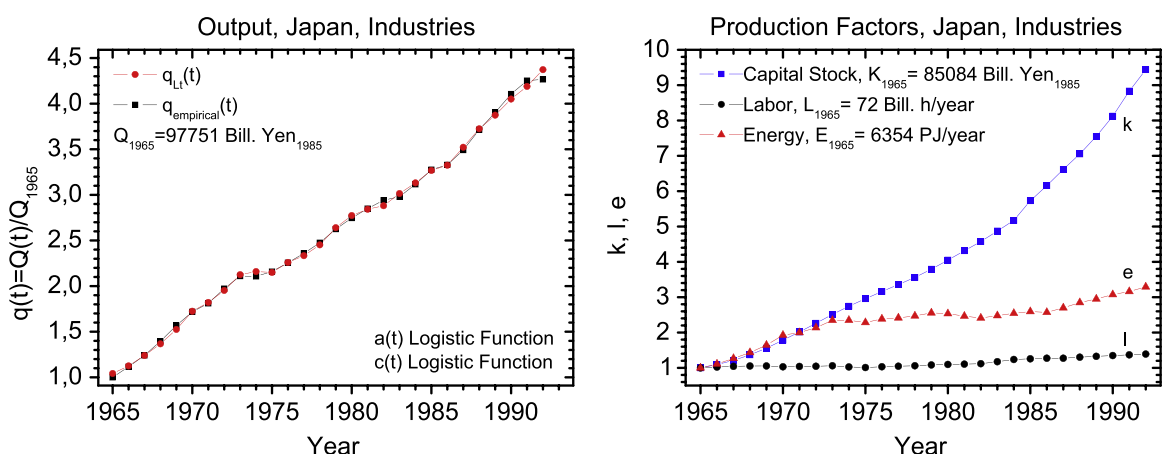


Figure 4. Growth in the Japanese sector ‘Industries,’ which produces about 90% of Japanese GDP, between 1965 and 1992. Reproduced from [1], chapter 4, pp 206–208, with kind permission of Springer Science + Business Media.

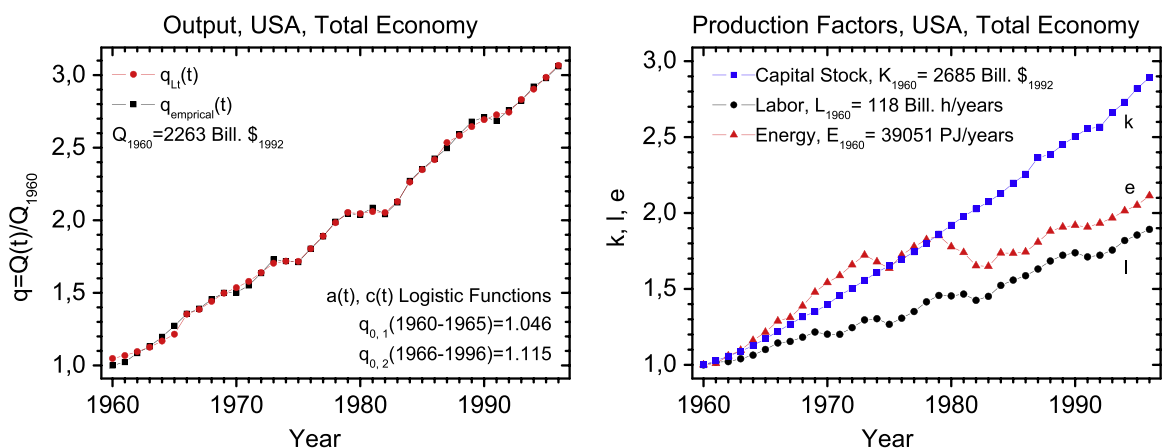


Figure 5. Growth in the total US economy between 1960 and 1996. Reproduced from [1], chapter 4, pp 206–208, with kind permission of Springer Science + Business Media.

Table 1. Time-averaged LinEx output elasticities and statistical quality measures. Reproduced from [1], chapter 4, pp 206–208, with kind permission of Springer Science + Business Media.

System	FRG TE 1960–2000	FRG I 1960–99	Japan I 1965–92	USA TE 1960–96
$\bar{\alpha}$	0.38 ± 0.09	0.37 ± 0.09	0.18 ± 0.07	0.51 ± 0.15
$\bar{\beta}$	0.15 ± 0.05	0.11 ± 0.07	0.09 ± 0.09	0.14 ± 0.14
$\bar{\gamma}$	0.47 ± 0.1	0.52 ± 0.09	0.73 ± 0.16	0.35 ± 0.11
$\bar{\delta}$	0.19 ± 0.2	0.12 ± 0.13	0.14 ± 0.19	0.10 ± 0.17
R^2	>0.999	0.996	0.999	0.999
d_w	1.64	1.9	1.71	1.46

exergy, multiplied by appropriate conversion efficiencies, plus physical work by animals. The ‘useful work’ data [12] already include most of the efficiency improvements that have occurred in the energy converting systems of the USA during the 20th century. In this case two constant technology parameters suffice to reproduce well the gross domestic product of the US economy between 1900 and 1998 [13, 14]. The time averages of the corresponding output elasticities of capital, labor and ‘useful work’ are similar to the ones in table 1.

These findings are supported by cointegration analyses [15]. The energy-dependent Cobb–Douglas function (32), whose constant output elasticities are similar to the time-averaged LinEx output elasticities, also reproduces economic growth more or less satisfactorily, albeit with larger discrepancies between theoretical and empirical growth during and after the 1973–1981 oil-price shocks [1, 14].

The reproduction of economic growth in Germany, Japan, and the USA during up to four decades with small residuals, and output elasticities that are for energy much larger and for labor much smaller than the cost shares of these factors, are the principal results of the KLEC-model applied in this section. The contribution of human creativity to growth appears as comparable to that of labor¹². The fundamental presumption of the model is, that the macroeconomic production function $Y(K, L, E; t)$ (and its dimensionless version $y(k, l, e; t)$ as well) is a state function of the economic system, so that its infinitesimal change is a total differential and the integral of this differential between K_0, L_0, E_0 and K, L, E does not depend on the path in factor space. Such $Y(K, L, E; t)$ must be twice differentiable with respect to K, L, E . From this property follow the growth equation (1) and the partial differential equations (27) for the output elasticities¹³. Essential is that the output elasticities are not set equal to the cost shares, as it is done in mainstream economics, but that instead they are determined by fitting the production functions to the time series of output, where the restrictions (29)—output elasticities must be non-negative—have to be observed.

A number of analyses of past growth with alternative energy-dependent production functions, whose output elasticities satisfy (27), have yielded results similar to the ones presented here [1, 7]. Cobb–Douglas functions are the work-horses of mainstream econometrics. Unfortunately, they ignore the physical restrictions on factor substitution. But

¹² Error propagation produces the large errors in $\bar{\delta}$.

¹³ The energy-dependent versions of the principal production functions of mainstream economics, which had been constructed before equations (1) and (27) were derived, satisfy these equations, too. [7]

since the economic actors are aware of these restrictions, Cobb–Douglas is not too bad for reproductions of past economic growth, despite the insensitivity of the constant output elasticities to input fluctuations. (For scenarios of the future, however, the law of diminishing returns and the approach to the state of maximum automation should be incorporated in the output elasticities, as it is exemplified by the LinEx function (33).) The choice of the base year and the methods to minimize the sum of squared errors SSE (34) have had little influence on the econometric findings so far¹⁴. Energy-dependent service production functions also reproduce the growth of the (West) German service sector quite well, yielding output elasticities of energy and labor that deviate from these factors' cost shares not quite as much as in the total economy [7, 16]. The increasing automation in agriculture and industries shifts more and more of the labor force into the less energy-intensive service sector. Nevertheless, computer- and electricity-based automation is progressing in that sector, too, especially in finance, logistics, and retail business.

Whatever the details of modeling energy's working as a factor of production, one result is common to all models:

If one foregoes cost-share weighting and determines the output elasticities of capital, labor, energy, and creativity econometrically, one gets for energy economic weights that exceed energy's cost share by up to an order of magnitude, and the Solow residual disappears. The production factor energy accounts for most, and creativity for the rest of the growth that neoclassical economics attributes to 'technological progress'.

4. Economic evolution in the cost mountain

We consider the mountain of factor cost that rises above the plane that is spanned by the input ratios 'labor/capital' ($l/k \equiv u$) and 'energy/capital' ($e/k \equiv v$). The cost P_Y of producing a given output $Y(t) = Y[K, L, E; t]$ is obtained by multiplying the time-dependent prices per capital unit, $p_K(t)$, labor unit, $p_L(t)$, and energy unit, $p_E(t)$, by the factor quantities K, L, E required to produce that output: $P_Y = p_K K + p_L L + p_E E = P_K k + P_L l + P_E e$, where $P_K \equiv p_K K_0$, $P_L \equiv p_L L_0$, $P_E \equiv p_E E_0$.

In order to relate the total factor cost at time t to the output at t we rewrite $P_Y = e[P_K k/e + P_L l/e + P_E] = e[P_K/v + P_L u/v + P_E]$, remember that $y(t) \equiv Y(t)/Y_0$ on the lhs of equation (30), rewrite this equation in the form $e = Y(t)/Y_0 \mathcal{F}$, insert this into the equation for P_Y , and obtain the equation of the cost mountain as

¹⁴ Nevertheless, the Levenberg–Marquardt method in combination with the self-consistent iteration scheme (see appendix 6 of chapter 4 of [1]) for SSE minimization, which is used in [11] to produce the results of this paper, has been so far the one that yields the best statistical quality measures. Furthermore, the LinEx function $y_{L1}(k, l, e; t)$ in (33) is the simplest one in the family of LinEx functions, some of whose members are presented in [7]. For the energy-dependent Cobb–Douglas function (32) a shift of the base year results in a transformation of the parameter y_{CDE}^0 , and for the somewhat more complicated LinEx function $y_{L11} = y_{L11}^0 \exp\left[a\left(1 - \frac{l}{k}\right) + \frac{1}{c}\left(1 - \frac{e}{k}\right) + ac\left(\frac{l}{e} - 1\right)\right]$ a base-year shift results in three equations between the parameters y_{L11}^0 , a , and c of the two sets that belong to the two y_{L11} , whose variables are normalized to their values in the two different base years. For the simple y_{L1} , on the other hand, we know (only) from fitting by SSE minimization, that base-year shifts have negligible influence on the results.

$$P_Y = \frac{Y(t)}{Y_0 \mathcal{F}[u, v; t]} \left[P_K \frac{1}{v} + P_L \frac{u}{v} + P_E \right]. \quad (35)$$

The cost mountain rises above the $u - v$ plane, its topography is determined by the prices and ratios of the production factors, and its height scales with $Y(t)/Y_0$.

The profit G_Y obtained from the output Y is $G_Y = Y - P_Y$. Profit maximum means that

$$-G_Y = \frac{Y(t)}{Y_0 \mathcal{F}(u, v; t)} \left[P_K \frac{1}{v} + P_L \frac{u}{v} + P_E \right] - Y(t) \quad (36)$$

is minimum. Since the cost mountain and the mountain of negative profit just differ by the height-shift $Y(t)$, the structure of both mountains is the same.

The negative gradient, $-\mathbf{grad} \equiv -\mathbf{e}_u \partial/\partial u - \mathbf{e}_v \partial/\partial v$, of $-G_Y$ points in the direction where the economy would move, if the economic actors would only have the objective of profit maximization; \mathbf{e}_u and \mathbf{e}_v are the unit vectors parallel to the u -axis and the v -axis that span the $u - v$ plane. Operating with $-\mathbf{grad}$ on $-G_Y$ in equation (36), and observing that equation (31) leads to $\partial \mathcal{F}/\partial u = \beta \mathcal{F}/u$ and $\partial \mathcal{F}/\partial v = -(\alpha + \beta) \mathcal{F}/v$, one obtains the gradient of profit

$$\mathbf{grad} G_Y = -\mathbf{e}_u \frac{\partial P_Y}{\partial u} - \mathbf{e}_v \frac{\partial P_Y}{\partial v} \quad (37)$$

as the negative gradient of the cost mountain with the components

$$-\mathbf{e}_u \frac{\partial P_Y}{\partial u} = -\mathbf{e}_u \frac{Y(t)}{Y_0 \mathcal{F}(u, v; t)} \left[P_L \frac{1}{v} - \frac{\beta}{u} \left(P_K \frac{1}{v} + P_L \frac{u}{v} + P_E \right) \right], \quad (38)$$

$$-\mathbf{e}_v \frac{\partial P_Y}{\partial v} = -\mathbf{e}_v \frac{Y(t)}{Y_0 \mathcal{F}(u, v; t)} \left[\frac{\alpha + \beta}{v} \left(P_K \frac{1}{v} + P_L \frac{u}{v} + P_E \right) - (P_K + P_L u) \frac{1}{v^2} \right]. \quad (39)$$

To visualize the trajectory of a real-life economy on the slope of its cost mountain, the time-changing factor prices, the inputs, and an appropriate production function $e\mathcal{F}(u, v; t)$ are needed in equations (35), (38), and (39). The production function should reproduce economic growth reasonably well with small residuals and acceptable statistical quality measures. As we have seen in section 3, the LinEx function (33) satisfies these criteria. The energy-dependent Cobb–Douglas function (32) with output elasticities that are close to the time-averaged LinEx output elasticities in table 1 may also be considered as an acceptable production function until the second oil-price shock.

In figure 6 the trajectory of the GWG in the cost mountain and the negative cost gradients are projected onto the mountain base, as an example. We use this example, because the sector GWG is the pillar of the German economy—it produced about 50% of (West) German GDP in the 1960s and 1970s—and its aggregated, inflation-corrected energy prices are available from a research project carried out in the 1980s [17]. At the time of this project, factor prices could only be constructed for the times between 1960 and 1981. The directions of the negative cost gradients in figure 6 are relatively little affected by the two oil-price explosions, because the latter varied energy's small cost share only between 4% and 7%. Neither has this direction changed much between 1981 and 1989. The structural break in the German economy at reunification in 1990, which shows in the growth curves of GWG in figure 3, would make it rather difficult to construct a consistent set of updated, inflation-corrected, aggregated energy price data since then. Therefore, we constrain ourselves to the time between 1960 and 1989 in this illustrating example.

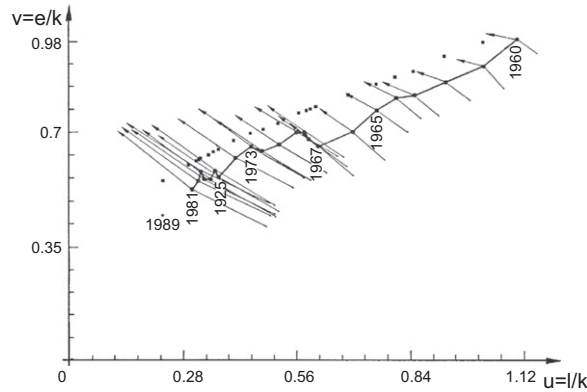


Figure 6. The solid line indicates the path of the German industrial sector ‘Warenproduzierendes Gewerbe’ (GWG) in the cost mountain between the years 1960 and 1981, projected onto the $u - v$ plane; $u \equiv l/k$ and $v \equiv e/k$, where k , l , and e are multiples of capital, labor, and energy in the base year $t_0 = 1960$. The full squares mark the barrier from the limit to capacity utilization. The way of computing them from equation (40) is described below that equation. They complement (the modified) figure 4 of [17].

During the first and the second oil-price shock the aggregate energy price increased in Germany by a factor of 2.5. Similar price shocks hit the other industrial market economies and triggered their first two serious post-war recessions. During these recessions, the variations of economic output closely followed the variations of energy input, as figures 2–5 show. Since then, energy conservation measures have been taken, and energy-intensive, polluting production processes have been shifted abroad. As a result, the technology parameters a and c of the LinEx function acquire a time dependence, which, however, has been negligible for the time span covered by the trajectory in figure 6.

The barrier from the constraint that capacity utilization cannot exceed 1, see equations (5) and (7), is written in terms of the dimensionless inputs as

$$1 = \eta_0 \left(\frac{l}{k} \right)^\lambda \left(\frac{e}{k} \right)^\nu \equiv \eta_0 u^\lambda v^\nu, \quad (40)$$

where the parameters λ , ν , and η_0 have been determined by fitting the rhs of equation (40) to empirical data on capacity utilization in the German economy. These data, published by the ‘Sachverständigenrat für die Gesamtwirtschaft,’ are given in [7]. The parameters $\eta_0 = 1$, $\lambda = -0.152$, $\nu = 0.386$ reproduce the data satisfactorily, although the fit stays somewhat below the maxima in 1965, 1969, and 1970. The barrier from the technological constraint (40) is indicated by the full squares in figure 6. The LinEx function indicates that the barrier from maximum automation is in the region where $v = c \approx 1$, so that $\beta_{L1} = 0$, and where $u \ll 1$.

Direction and length of the arrows above the trajectory indicate directions and strengths of the negative cost gradients, whose components are calculated with the LinEx production function; arrow lengths are for the 1970 output quantity. Below the path, the lines without arrow heads indicate the negative cost gradients obtained with an energy-dependent Cobb–Douglas function whose output elasticities are close to the time-averaged elasticities of the LinEx function. The gradients depend little on the type of production function. What matters is the

magnitude of the output elasticities. The general direction of the trajectory is at a large angle to the negative cost gradients, which point toward the region of small $u \equiv l/k$ and much larger $v \equiv e/k$. Only during times of economic upswings, as from 1967–1970, 1972–1973, and 1975–1979, the path follows the negative cost gradients toward the barrier from the limit to capacity utilization temporarily. This shows that increasing the energy input into the machines that were not fully employed in times of recessions, rapidly increases output and profit¹⁵. The isolated point for 1989 indicates that, until German reunification in 1990, the general direction of the path toward strongly decreasing u and moderately decreasing v continues. It is more or less parallel to the barrier from the limit to capacity utilization, save for the substantial shift of the empirical trajectory toward the barrier in the late 1960s. Obviously, the virtually binding constraint associated with the limit to capacity utilization and soft constraints, which are discussed in section 2, keep the economic system on the uphill side of the barrier. Along the trajectory in the cost mountain, the inputs of expensive labor are reduced and cheaper energy/capital combinations are substituted for them.

The Japanese and US empirical data on capital, labor, and energy in figures 4 and 5 show that until 1973 factor growth in the USA and Japan differs from that of Germany: labor grows in the USA, but more slowly than capital, and remains nearly constant in Japan, whereas capital and energy increase at nearly the same rate in both countries. In Germany, on the other hand, labor always decreases¹⁶, and the input of energy grows much less than the capital stock. But after the first oil price shock, the growth of energy is significantly reduced in all three systems, while the growth of capital continues as before. In the USA and Germany the energy input oscillates in response to the oil price variations, and in Japan it nearly flattens out. Consequently, the US and Japanese path in the $u - v$ plane stays close to a line of constant $e/k \leq 1$ and decreasing l/k until 1973, and thereafter both l/k and e/k decrease. Then, the economic actors in the USA and Japan seem to behave similar to the ones in Germany.

5. Summary and outlook

Energy-dependent production functions, with output elasticities that are for energy much larger and for labor much smaller than the cost shares of these factors, reproduce economic growth in Germany, Japan, and the USA with small residuals and good statistical quality measures—even during the downturns and upswings in the wake of the first and the second oil-price explosion. This does *not* contradict the usual behavioral assumptions that entrepreneurs maximize profit or that society maximizes overall welfare, because real-world economic actors are aware of the technological and the related ‘virtually binding’ constraints on the combinations of capital, labor, and energy. ‘Soft’ constraints from legal and social obligations may also matter. The barriers from the constraints on capacity utilization and automation prevent modern industrial economies from reaching the neoclassical optimum of mainstream economics, where output elasticities would be equal to factor cost shares.

¹⁵ Since the machines of the capital stock are activated by energy, economic causality between output and energy input is bidirectional (at least in the short run, when improvements of energy efficiency have not yet been developed): fluctuations of demand lead to corresponding fluctuations of demand-adjusted production and energy inputs. Fluctuations of energy inputs in the case that an important energy source would dry up and another one would be discovered and opened up, would lead to corresponding fluctuations of output.

¹⁶ This causes the ‘condensation’ of the trajectory points with decreasing u in figure 6.

Since in industrially advanced countries cheap energy is economically much more powerful than expensive labor, there has been the long-time trend toward increasing automation, which replaces expensive labor by cheap energy-capital combinations. In the G7 countries, this trend has drastically reduced the number of people employed in the sectors agriculture and industries during the last four decades, and the contribution of these sectors to GDP as well (see, e.g., tables 4.1 and 4.2 in [1]). Now, in the service sector, electricity-powered computers with appropriate software kill more and more jobs, too. The resulting danger of unemployment in the less qualified part of the labor force is enhanced by the trend toward globalization, where goods and services produced in low-wage countries can be delivered at small cost to high-wage countries thanks to cheap energy and increasingly sophisticated, highly computerized transportation systems.

The nearly unanimous social and political response to this danger is the call for strategies to stimulate economic growth. But these strategies face obstacles from entropy production, which is coupled to energy conversion and its pivotal role in economic growth: (1) there are thermodynamic limits to the improvement of energy efficiency *at unchanged energy services*, because entropy production destroys exergy; (2) emissions associated with entropy production, especially the ones of carbon dioxide from the combustion of fossil fuels, threaten climate stability. The question is, whether society will be willing and able to finance the huge investments that are necessary for the transition to a highly efficient production system powered by non-fossil fuels.

Finally, discussing the possibly decreasing recovery of cheap conventional oil according to the ‘peak-oil’ theory [18], the International Monetary Fund [19] quotes the high output elasticities of energy reported in [1, 10] and [13] and concludes: ‘... if the contribution of oil to output proved to be much larger than its cost share, the effects could be dramatic, suggesting a need for urgent policy action.’ Further research concerning the impact of energy conversion and entropy production on economic evolution is needed to guide that action.

References

- [1] Kümmel R 2011 *The Second Law of Economics: Energy Entropy, and the Origins of Wealth* (New York: Springer) pp 171–273
- [2] Kümmel R 1982 The impact of energy on industrial growth *Energy* **7** 189–203
- [3] Kümmel R, Ayres R U and Lindenberger D 2010 Thermodynamic laws, economic methods and the productive power of energy *J. Non-Equilib. Thermodyn.* **35** 145–79
- [4] Denison E F 1979 Explanation of declining productivity growth *Surv. Curr. Bus* **59** Part II 1–24
- [5] Solow R M 1994 Perspectives on growth theory *J. Econ. Perspect.* **8** 45–54
- [6] Söllner F 1996 *Thermodynamik und Umweltökonomie* (Heidelberg: Physica Verlag)
- [7] Lindenberger D 2000 *Wachstumsdynamik Industrieller Volkswirtschaften: Energieabhängige Produktionsfunktionen und Ein Faktorpreis-Gesteuertes Optimierungsmodell* (Marburg: Metropolis)
- [8] Samuelson P A and Solow R M 1956 A complete capital model involving heterogeneous capital goods *Quart. J. Econ.* **70** 537–62
- [9] Gauvin J 1980 Shadow prices in nonconvex mathematical programming *Math. Program.* **19** 300–12
- [10] Kümmel R, Henn J and Lindenberger D 2002 Capital, labor, energy and creativity: modeling innovation diffusion *Struct. Change Econ. Dyn.* **13** 415–33
- [11] Schmid J 2002 Energie, Kreativität und Wirtschaftswachstum *Diploma Thesis* Fakultät für Physik und Astronomie Universität Würzburg

- [12] Ayres R U, Ayres L W and Warr B 2003 Exergy, power and work in the US economy 1900-1998 *Energy* **28** 219–73
- [13] Ayres R U and Warr B 2005 Accounting for growth: the role of physical work *Struct. Change Econ. Dyn.* **16** 181–209
- [14] Ayres R U and Warr B 2009 *The Economic Growth Engine* (Cheltenham: Edgar Elgar)
- [15] Stresing R, Lindenberger D and Kümmel R 2008 Cointegration of output, capital, labor, and energy *Eur. Phys. J. B* **66** 279–87
- [16] Lindenberger D 2003 Service production functions *J. Econ. (Z. Nationalökon.)* **80** 127–42
- [17] Kümmel R and Strassl W 1985 Changing energy prices, information technology, and industrial growth *Energy and Time in the Economic and Physical Sciences* ed W van Gool and J J C Bruggink (Amsterdam: North-Holland) pp 175–94 The aggregate German energy prices between 1960 and 1981 are listed in this article.
- [18] Murray J and King D 2012 Oil’s tipping point has passed *Nature* **481** 433–35
- [19] International Monetary Fund World Economic Outlook 106–9
see also Kumhof M and Muir D 2012 Oil and the World Economy: some possible futures *International Monetary Fund Working Paper* WP/12/256 doi:[10.5089/9781475586640.001](https://doi.org/10.5089/9781475586640.001)