Energy-Extended CES Aggregate Production: Current Aspects of Their Specification and Econometric Estimation

Paul E. Brockway 1,*, Matthew K. Heun 2, João Santos 3 and John R. Barrett 1

1 Sustainability Research Institute, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK; j.r.barrett@leeds.ac.uk
2 Engineering Department, Calvin College, Grand Rapids, MI 49546, USA; mkh2@calvin.edu
3 MARETEC—Marine, Environment, and Technology Center, Environment and Energy Scientific Area, Department of Mechanical Engineering, Instituto Superior Técnico (IST), University of Lisbon, Avenida Rovisco Pais, 1 1049-001 Lisbon, Portugal; joao.dos.santos@tecnico.ulisboa.pt
* Correspondence: p.e.brockway@leeds.ac.uk; Tel.: +44-113-343-2846

Academic Editor: Erik Gawel
Received: 18 November 2016; Accepted: 4 February 2017; Published: 10 February 2017

Abstract: Capital–labour–energy Constant Elasticity of Substitution (CES) production functions and their estimated parameters now form a key part of energy–economy models which inform energy and emissions policy. However, the collation and guidance as to the specification and estimation choices involved with such energy-extended CES functions is disparate. This risks poorly specified and estimated CES functions, with knock-on implications for downstream energy–economic models and climate policy. In response, as a first step, this paper assembles in one place the major considerations involved in the empirical estimation of these CES functions. Discussions of the choices and their implications lead to recommendations for CES empiricists. The extensive bibliography allows those interested to dig deeper into any aspect of the CES parameter estimation process.

Keywords: CES production function; econometrics; estimation; elasticity of substitution; energy policy; energy

1. Introduction

1.1. The Growing Use of CES Aggregate Production Functions

Production functions seek to explain economic output arising from input factors of production, and are central to growth accounting (i.e., the study of the components of economic growth), empirical investigations versus economic theory, and macroeconomic modelling. For our purposes, we define aggregate production functions as those applied at sector [1,2] or economy-wide [3–5] levels.

The two most common aggregate production functions are the Cobb–Douglas (C-D) and Constant Elasticity of Substitution (CES) functions [6,7], as shown by the Google Scholar results illustrated in Figure 1. (Google Scholar was preferred to searching in Scopus or Web of Science, as it enabled access to wider “real-world” production function literature such as central bank reports). Their central position in macroeconomic models mean “these functions play an important role in the [government’s] economic forecasts and policy” ([8], p. 1).
The C-D function in its famous 1928 formulation [9] is given in Equation (1), which according to conventional economic theory ascribes economic output ($Y_t$) to two primary factors—capital ($K_t$) and labour ($L_t$):

$$Y_t = \theta e^{\lambda t} K_t^\alpha L_t^\beta$$

where $\alpha$ and $\beta$ are the elasticities of output ($Y_t$) with respect to capital and labour, respectively (noting also typically $\alpha + \beta = 1$ to meet constant returns-to-scale assumption), $\theta$ captures a scale parameter, $e^{\lambda t}$ is the exogenous Solow residual, i.e., the part of economic output not explained by the endogenous factors of production, and $t$ is time relative to an initial year. $\lambda$ is a parameter capturing (exogenous) productivity growth, as defined by Equation (2), where $\dot{Y}$, $\dot{K}$, and $\dot{L}$ are time derivatives of $Y$, $K$, and $L$, respectively.

$$\lambda = \frac{\dot{Y}}{\dot{Y}} - \frac{\dot{K}}{K} - \frac{\dot{L}}{L}$$

An important parameter in economics is the elasticity of substitution ($\sigma$), a measure of the ease by which one production factor (e.g., labour) may be substituted by another (e.g., capital). For aggregate production functions, it is most commonly measured by the Hicks Elasticity of Substitution (HES), as given in Equation (3), where $\partial Y / \partial X_i$ is the marginal productivity of input $X_i$, and $\partial Y / \partial X_j$ is the marginal productivity of input $X_j$. The HES is thus a measure of the curvature of the production function isoquant, or as Stern writes, the "difficulty of substitution" ([10], p. 80).

$$HES_{ij} = -\frac{\partial \ln \left( \frac{X_i}{X_j} \right)}{\partial \ln \left( \frac{\partial Y}{\partial X_i} / \frac{\partial Y}{\partial X_j} \right)}$$

In a C-D function, the elasticity of substitution has a fixed unity value. This significant constraint is overcome by the CES function, introduced in 1956 by Solow [11], and subsequently generalized in the "ACMS" paper by Arrow, Chenery, Minhas and Solow [12] in 1961. The CES function in Equation (4) has $\delta$ as a share parameter, $\rho$ as a substitution parameter (leading to the HES, $\sigma = 1/(1 - \rho)$), $\nu$ as a returns-to-scale parameter, $\theta$ as a scale parameter, and $\lambda$ is (as before) the exogenous productivity growth parameter. The CES is therefore more flexible than the C-D function, with several special cases depending on the value of $\sigma$ as noted by Arrow et al. [12]: Leontief ($\sigma = 0$), C-D ($\sigma = 1$) and Linear ($\sigma = \infty$) functions.

$$Y_t = \theta e^{\delta t} [\delta K_t^\nu + (1 - \delta) L_t^\nu]^{\nu - \rho}$$

In an empirical study, historical time-series data (of the factors of production and economic output) is added to the functional form (e.g., Equation (4)) to form an analytical model, whose econometric estimation obtains values for the unknown CES function parameters. Solow’s 1957 US
study [3] using the C-D function was the first time-series empirical study of its kind and “a landmark in the development of growth accounting” ([13], p. 1) that was followed by others including Arrow et al. [12] and Denison [14]. Whilst many studies follow this neo-classical C-D approach [4,15,16], many researchers—famously including Solow [3]—found that increases in capital and labour factors of production commonly explained only a minority of output growth, with the remainder ascribed to exogenous growth parameter λ. Abramovitz ([17], p. 11) described the unknown component of economic output (i.e., the Solow residual) as “a measure of our ignorance”, whilst Hulten ([18], p. 9) advocates it “covers many components, some wanted (such as the effects of technical and organizational innovation), others unwanted (such as measurement error, omitted variables, aggregation bias, and model misspecification)”. As a result, a focus on growth accounting (including the Solow residual) has remained a priority for researchers including Jorgenson [19], Denison [20] and Hulten [18,21].

1.2. Adding Energy as a Factor of Production

Neo-classical capital–labour aggregate production functions ignore the possible role of energy as a factor of production, since it is viewed as an intermediate product (of capital and labour), rather than a primary input. The 1970s oil crises focussed attention on the role of energy in economic growth, and thus provided an opportunity for researchers to add energy (E) as an input [22–24], typically amending the C-D function in Equation (1) to that shown in Equation (5):

\[ Y_t = \theta e^{\lambda t} K_t^\alpha L_t^\beta E_t^\gamma \]  

(5)

where γ is the elasticity of output with respect to energy, and \( \alpha + \beta + \gamma = 1 \) to meet constant returns to scale assumption.

More recently, adding energy as a factor of production in aggregate production functions has regained popularity [25]. One possible reason is practicality, in that “increasing attention on the energy and environmental issues has evoked a revival of the relevant macroeconomic modelling” ([26], p. 793)—in other words, the effects of energy in an energy–economic model cannot be studied unless it is included as an endogenous factor of production. Another possible reason is the growing evidence base that energy is tightly linked to economic growth [27–29].

Energy (E) can be placed inside a nested CES function by augmenting Equation (4) as shown in Equation (6), with capital and labour in an inner (K-L) nest, and energy in an outer (KL_E) nest.

\[ Y_t = \theta e^{\lambda t} [(\delta_1 K_t^p + (1 - \delta_1)L_t^p)^{\rho_1}]^{\rho_1} + (1 - \delta)E_t^\rho \]  

(6)

where ρ and \( \rho_1 \) are substitution parameters which lead to the inner nest \( \sigma_1 \) within the inner (K-L) nest and an outer nest \( \sigma \) between the inner (K-L) composite and energy (E). Our inner nest (\( \delta_1, \rho_1 \)) and inner-to-outer nest (\( \delta, \rho \)) share and substitution parameter notation follows Henningsen and Henningsen [30]. For completeness and relevant to the “nesting” discussion in Section 3.3, with three factors of production (K, L, and E), the CES function has two other possible nests in addition to the KL(E) structure—EK(L) in Equation (7) and LE(K) in Equation (8):

\[ Y_t = \theta e^{\lambda t} [(\delta_1 E_t^p + (1 - \delta_1)K_t^p)^{\rho}]^{\rho_1} + (1 - \delta)L_t^\rho \]  

(7)

\[ Y_t = \theta e^{\lambda t} [(\delta_1 L_t^p + (1 - \delta_1)E_t^p)^{\rho}]^{\rho_1} + (1 - \delta)K_t^\rho \]  

(8)

1.3. Aim and Scope of Paper

Three propositions provide the rationale for our paper. First, capital–labour–energy CES aggregate production functions are important to macroeconomic models which inform climate and economic policy. Second, this places a due weight of responsibility on the CES empiricist to make the most appropriate choices regarding the many aspects of their econometric specification and estimation. It also places a responsibility on the “downstream” users of empirical CES study results, to be aware of such aspects and their implications. Third, though single aspect literature of CES...
production function theory and empirical usage [5,26,30–34] exists, without a succinct collation of the most important issues and options, analytical blindspots and poorly specified functions are more likely, which may have significant impacts on the estimated parameters, and ultimately energy policy.

In response, this paper comprises three novel components. First, we evaluate the proposition that CES functions have become the most important energy-extended aggregate production function in empirical use. Second, we assemble in one place the major aspects of the econometric specification and estimation for capital–labour–energy CES functions. Third, the merits of choices relating to these aspects are discussed, and recommendations made where evidence or consensus exists. Whilst the primary audience are those CES analysts involved in empirical studies, the succinct and accessible paper is open to all along the energy modelling-to-policy chain, including macroeconomic modellers and energy policymakers. The extensive bibliography is deliberate: permitting those interested to dig deeper into issues than space allows in this single journal paper.

The paper starts with a broader review the applications of C-D and CES aggregate production function studies in Section 2. This provides the context for the narrowing of focus in Section 3 to consider the specification of the empirical CES model: comprising the design of the function form and the input time-series datasets. Next, parameter estimation techniques are examined in Section 4, before recommendations and conclusions are given in Section 5.

Before we begin, a note on our study boundary. First, our focus is predominantly at the economy-wide (i.e., national) scale, though many aspects considered are also relevant to sectoral-level functions. Second, spatial constraints mean we cannot empirically test the collated aspects. Instead, this is undertaken by Heun et al. [35], which is an empirical complement to this landscape paper, where four key modelling choices are examined to establish the differences in resulting CES parameter values, and the potential effects on downstream energy policy. Third, by considering only C-D and CES aggregate production functions, we exclude further discussion on less popular aggregate production functions (e.g., translog [36], variable elasticity of substitution (VES) [37], linear exponential (LINEX) [38], linear [39] and Leontief [40] functions) and cost functions—which are a price-based alternative to production functions [36,41–43]. We also limit widespread further discussion on the important class of capital–labour–energy–material (KLEM) CES-based production functions. Whilst becoming increasingly popular [44–46], we retain our focus on KLE-based CES functions for practical reasons: spatial constraints and the need for brevity, combined with the reality that our paper is very relevant for those working with KLEM-based CES functions, since they will share many of the specification and estimation aspects raised regarding KLE functions. Lastly, we also exclude computational general equilibrium (CGE) based studies [47,48], since CGE models do not estimate CES aggregate production function parameters.

2. Applications of C-D and CES Aggregate Production Functions

We now briefly review common applications of C-D and CES aggregate production functions. The inclusion of the C-D function (in addition to the CES function) is intended to show the changing focus of application as the C-D function is increasingly replaced by the CES function. We start with a sample survey, and then move to a wider literature search.

2.1. Sample Survey

We studied a small sample of the Google-Scholar results in Figure 1, seeking to identify similarities and differences in applications. Whilst Google-Scholar returned results for all production function types (i.e., firm level to sectoral to economy-wide scales), it nevertheless provides a guide as to the context and application of production function studies. We reviewed 46 studies [1,4,6,15,16,22,23,25,46,49–85], with 29 C-D and 17 CES studies, in proportion with their number in the total returned results. To make the best of the tiny, biased sample (0.1% of 40,000 Google-Scholar references obtained for Figure 1), we selected studies based on three criteria. First, studies were selected based on Google Scholar’s “highest returned relevance”—which is a metric based on the publication’s full text, the source publication and author, and the number of scholarly citations.
Secondly, this is then filtered to only include empirical studies at an aggregate (sector or economy-wide) scale. Thirdly, we selected studies in proportion with the number of CES and C-D studies in each decade in Figure 1—which was intended to get some sense of the change in direction/focus of the studies, and to give insights into the transition (in popularity) from C-D to CES functions.

Figure 2 shows a histogram of the different purposes driving the C-D and CES sample studies. For C-D studies, the most common purpose was analysing historical changes in exogenous component of economic output (Solow residual), and studying new factors of production in addition to capital and labour. As the CES studies allow non-unity elasticities of substitution, and are weighted (in number) towards more recent studies, this helps explain their focus on elasticities of substitution and computational methods (e.g., use of new parameter estimation algorithm).

![Figure 2. Primary study rationale in the sample.](image)

The output measure was almost exclusively GDP, with the key differentiator being whether it was GDP in constant prices (30 No.) or GDP per worker (14 No.).

Figure 3 shows the wide variation for choice of factors of production. For the conventional factors of production of capital and labour, capital stock and number of workers were the most common metrics. Energy was the most popular additional factor of production, appearing first in the post oil-crises 1970s [22,23], and reappearing in our sample in the 1990s [1,25,57,64,79,83]. It is interesting to see in our sample—but unclear why—the C-D studies favoured price-adjusted energy (e.g., values in £), versus the CES studies which used thermal energy content of primary or final energy (e.g., values in Joules).

![Figure 3. Factors of production in the sample.](image)
2.2. Wider Literature Search

Including energy as a factor of production starts from the idea that variables in addition to labour and capital—such as energy [22,23,70], materials [46,86] or money balance [52,55]—help explain economic output. As far back as 1974, Binswanger and Ledergerber [87] suggested that “the decisive mistake of traditional economics is the neglect of energy as a factor of production”. However, including energy as a factor of production remains controversial. One argument is that energy is not an independent, primary input, but instead as an intermediate quantity made by labour and capital is thereby redundant (see Dales’ Biophysical GEMBA model as an example reflecting this argument [88,89]). To counter, the same argument could be applied to capital (i.e., you cannot make capital without labour), and authors including Stern [90] advocate energy as an independent factor of production. Some authors go further: Kümmel [91] suggests energy is the only factor of production, with capital and labour therefore intermediate products (of energy). Denison [20] suggests a second argument: that energy’s low “cost-share” (typically below 10% of GDP [92,93]) means it can only make a correspondingly small contribution to economic growth. However, authors including Stresing et al. [94] have sought to debunk this argument, whilst Aucott and Hall [95] show how—despite its low “cost-share”—small variations in energy prices have significant impacts on economic output.

Aggregate production functions themselves are not without criticism. Indeed Mishra ([96], p. 20) suggests they are “the most turbulent area of research in the economics of production”. Criticism occurs on three main fronts. First is the accounting identity critique [97,98], which infers the C-D function can be derived from an income accounting identity: output equals wages plus profits. This is held to explain the excellent historical fits, with observed correlation coefficients ($R^2$) commonly above 0.99 [15,16]. Later, Felipe and McCombie [99] extended the accounting identity argument to include the CES function. Second, are concerns about measuring capital: Robinson [100] and Fisher [101] were among a group involved in the 1950s–1970s “Cambridge-controversy”, who suggested aggregate capital could not be measured, thereby invalidating the use of aggregate production functions. Third are empirical concerns, since factors of production typically explain only a minority of economic growth, leading Solow ([3], p. 312) to remark “it takes something more than the usual ‘willing suspension of disbelief’ to talk seriously of the aggregate production function”.

Despite on-going critiques [102–104], the practical reality is that: (1) “economists have continued using the aggregate production function in both theoretical and applied works” ([98], p. 262); and (2) that energy is increasingly used as a factor of production by a wide set of studies beyond academia, including government agencies [8,68,82,105] and central banks [67,106–109]. Several reasons may explain this. First, is the “pull” from energy-related questions including macroeconomic energy rebound [110,111], the contribution of energy to reducing exogenous growth [112], and climate and economic implications of energy transitions [5,113]. Second, since the elasticity of substitution ($\sigma$) is an important parameter in economics [114,115], significant effort in capital–labour–energy CES empirical analysis is directed to estimate values of $\sigma$ [1,33,46,83]. Third, the comparison between CES and C-D functions is an important study focus—whether for cross country comparisons [6], specific countries [81], sectors [105] or business cycles [84]. Fourth, general equilibrium models are an important application of the empirical CES study results, as highlighted by van der Werf [25], and are widely used to assess the impact of policy [116,117]. CGE models are the most popular, and are commonly CES-based [30,48,117,118] since this allows non-unity elasticity of substitution values, but may also include C-D modules [47,119,120]. Dynamic Stochastic General Equilibrium (DSGE) models are less common, but also use CES production functions [84,121].

Overall, capital–labour–energy CES aggregate production functions have emerged into widespread usage, serving as a good compromise between complexity (of the analysis) and flexibility (i.e., wider range of available parameters). For example, Stern and Kander ([5], p. 58) noted their choice of CES over translog production function was because “we decided that it was better to model some of the main features more reliably or believably [in a CES function] than to attempt to model many features of the data less reliably [in a translog function]”.

3.1. Economic Output (Y)

Three broad classes of economic output (the dependent variable, Y) exist: Gross Domestic Product (GDP), Gross Value Added (GVA)—also called Net Output, which is equal to GDP minus subsidies and taxes, and Gross Output (GO)—equal to GDP plus intermediate inputs. (Note also the valuable in-depth work on this by Hulten [122] and Cobbold [123]). In our case, it seems initially straightforward (at an economy-wide level) to select economic output as GDP, since this is the most common metric nationally reported, and has been widely adopted in mainstream growth accounting [69,124], which adopts labour and capital as the two production inputs. However, when energy is added to the production function, things change, as this “intermediate” input is used by industry to produce final “products” for end consumers. Ideally, the output measure should be selected “so that the total value of output is equal to the total value of inputs” ([5], p. 59). Those working with capital–labour–energy–materials (KLEM) productivity databases, such as O’Mahony and Timmer [125], use gross output as the economic output measure. In the capital–labour–energy case, lacking materials, a measure somewhere between GVA and gross output seems logical, and this is indeed the path taken by Kander and Stern [126] and Van der Werf [25], who adopt a modified “gross output” measure as GVA plus the value (cost) of energy. It is revealing though that this is not a common approach, as empirical K-L-E CES studies adopt a wide range of output metrics, as shown in Table 1:

<table>
<thead>
<tr>
<th>Author/Study</th>
<th>Year</th>
<th>Energy Measure</th>
<th>Output Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kemfert [79]</td>
<td>1998</td>
<td>Final energy</td>
<td>GVA</td>
</tr>
<tr>
<td>Kemfert and Welsch [83]</td>
<td>2000</td>
<td>Final energy</td>
<td>GVA</td>
</tr>
<tr>
<td>Van der Werf [25]</td>
<td>2008</td>
<td>Final energy</td>
<td>GVA + energy cost</td>
</tr>
<tr>
<td>Koesler and Schymura [46]</td>
<td>2012</td>
<td>Final energy</td>
<td>Gross Output</td>
</tr>
<tr>
<td>Turner et al. [127]</td>
<td>2012</td>
<td>Final energy</td>
<td>Gross Output</td>
</tr>
<tr>
<td>Dissou et al. [1]</td>
<td>2012</td>
<td>Final energy</td>
<td>Industry GVA</td>
</tr>
<tr>
<td>Sun [128]</td>
<td>2012</td>
<td>Final energy</td>
<td>GVA + energy cost</td>
</tr>
<tr>
<td>Kander and Stern [5]</td>
<td>2014</td>
<td>Primary energy</td>
<td>GVA + energy cost</td>
</tr>
<tr>
<td>Shen and Whalley [33]</td>
<td>2014</td>
<td>Primary energy</td>
<td>GDP</td>
</tr>
<tr>
<td>Zha and Zhou [26]</td>
<td>2014</td>
<td>Final energy</td>
<td>Not stated</td>
</tr>
</tbody>
</table>

A second important issue is whether to specify output in (more common) constant prices [15,69,70,74] or Purchasing Power Parity (PPP) prices [25,82]. Since PPP places a higher weight on GDP in non-OECD countries—one $US Dollar in China buys more goods than in the US—PPP may be useful in cross-country studies [129,130] by providing a more level playing field for comparisons.

3.2. Factors of Production (K, L, and E)

3.2.1. Unadjusted (Basic) Factors

Studies commonly adopt capital stock (K), labour (L) and primary energy (E), which we can consider as unadjusted (or basic) factors of production, i.e., they are measured without taking into account qualitative differences. Capital stock (the estimated market value, in currency units, of assets involved in production) is most commonly derived via the Perpetual Inventory Method (PIM), where an assumed initial capital stock valuation changes each year via additions (new stock) minus subtractions. Gross capital stock (GCS) defines subtractions as retirements of existing assets; whilst Net Capital Stock (NCS) is equal to GCS less depreciation of existing assets. With NCS and GCS data published by statistical agencies [131,132], CES studies have adopted both NCS [74,133] and GCS [22,25] datasets. For labour, three options for unadjusted values of workforce labour exist, listed here in descending accuracy as a measure of labour input: work-hours [22,73], numbers of workers [33,105], or population (for economy-wide studies only) [85]. Unadjusted energy—typically given in
energy units as terajoules (TJ) or million tonnes of oil equivalent (mtoe)—can be based on primary energy values or final (purchased) energy. Economy-wide studies most commonly use primary energy [33], whilst sector-level studies only use final energy [1,25,46] since primary energy values are not reported at that level.

Overall, these unadjusted variables remain very popular for empirical production function analysis, due to the availability of national and international time series across countries and sectors [134–137].

3.2.2. Quality-Adjusted Factors

Quality-adjusted values for capital (K*), labour (L*) and energy (E*) seek to better represent the productive effect of the basic factors of production (K, L, and E) on economic output (Y). Since quality-adjusted factors of production typically grow faster than unadjusted values [35,133], in such cases their use at an economy-wide scale assigns more of the increase in economic output to the growth in factors of production, and less to exogenous technical change (i.e., Solow residual).

Quality adjustment of capital can be achieved by estimating “capital services”, defined as “a flow of productive services from the cumulative stock of past investments” ([138], p. 7). Consider a machine in a factory: its capital service can be measured by multiplying the price of the goods by the amount of goods produced by the machine in each year. As national-level time-series of capital services emerge [139], their use and application in empirical CES studies is increasing [69,133]. A less common alternative is capital utilisation, which estimates how productively capital equipment is used following economic cycles (i.e., less in recessions, more at other times), as shown in Paquet and Robidoux’s Canadian study [140].

Quality adjustment of labour multiplies (unadjusted) work-hours by a quality index—commonly of worker schooling or skills. As international datasets of such quality metrics—such as Barro and Lee [141]—have become more available, quality-adjusted labour appears more widely used in CES studies [58,65,72,142].

Quality-adjusting energy seeks to capture “the relative economic usefulness of different fuels and electricity” ([143], p. 302). This can be done on either a physical or economic basis. An example physical approach can consider the amount of exergy (available energy) of the energy carrier nearer the end of the energy conversion stage as useful work [144], when it is lost in exchange for energy services. Regarding economic approaches, Cleveland et al. [143] suggest higher fuel prices are indicators of higher quality, whilst Stern ([145], p. 1474) introduces a substitution method whereby quality can be measured by “how much of one fuel is required to replace another”. Weighting can range from simple aggregation to Divisia indices. Including quality-adjusted energy in empirical aggregate production function studies are rare: Ayres and Warr used the physical approach by including useful work data in economy-wide C-D and LINEX functions [144,146], whilst Stern and Kander [5] provide an economic-based CES example, using a Divisia weighted price based method for energy quality.

Despite the apparent merits of quality adjustment [63], caution is needed. For capital services, Inklaar [147] raises concerns about the accuracy of the methodology, such that the Penn World Tables (PWT) retains capital stock for its capital data [124]. For energy, economic approaches can be problematic—for example energy price data varies with sector and end use and may be distorted by taxes and subsidy effects, whilst simple price weighting is biased as it assumes no restrictions on substitutability between energy inputs [24]. As for physical approaches to energy quality, few national datasets exist of thermodynamic “useful energy”, leaving researchers to time-consumingly construct their own datasets [148,149]. The result is that most CES empirical studies continue to use unadjusted energy, i.e., primary or final energy datasets [25,83].

Interestingly, empirical studies involving only capital and labour expend significant effort to quality adjust at least one variable [57,85,108,150], but those introducing energy as a third variable typically use unadjusted values for capital and labour [1,25,79,151]. This seems surprising, but perhaps reflects the significant effort required to develop or obtain time-series of quality-adjusted variables.
3.3. Nesting and Elasticity of Substitution

Nesting, and elasticity of substitution, are interlinked aspects of CES function specification: the choice in one affects the other—so they are presented and discussed together in this section.

3.3.1. Nesting

Once the CES function has more than two factors of production, the issue of whether—and how—to nest them, arises. To see why, let us view the non-nested CES function introduced by McFadden [152], and used by Edenhofer et al. [153]. It is given in Equation (9), using the notation of nested Equation (6):

\[ Y_t = \theta e^{\lambda\left(K_t^\rho + L_t^\rho + E_t^\rho\right)} \]

This formulation assumes all factors of production are equal substitutes (\(\sigma = 1/(1 - \rho)\), which is highly restrictive and as Broadstock et al. ([154], p. 55) note, this “appears unlikely in practice and also excludes the possibility of complementarity”. As a result, various authors [25,26,44] report this structure is rarely used, with instead CES studies preferring to “nest” the factors of production, which is more flexible by allowing different elasticities of substitution to exist between production factors. A nested three factor format typically has two factors of production placed within an “inner” nest and one in an “outer” nest. Such a nest is shown in Figure 4, which portrays the KL(E) nesting structure of Equation (6), given earlier, where capital–labour is in the inner nest, and energy sits in the outer nest.

Van der Werf [25] reviewed numerous capital–labour–energy (KLE) and capital–labour–energy–materials (KLEM) production functions used in climate-based models. He found whilst most studies analysed a single, KL(E) nest—a view supported by Zha and Zhou [26]—there was also considerable variation in nesting structure. This presents two routes forward for analysts. The first is to constrain the empirical analysis to a single nesting structure, based on a-priori theoretical or other considerations. For example, Saunders [31] suggests the KL(E) nesting is the nesting structure that permits the fullest range of energy rebound (Re): from hyperconservation (Re < 0) to backfire (Re > 1). The second, less common approach, is to separately estimate and report all three types of nesting [1,25,33,79]—though care is needed in interpretation, since certain solution aspects (such as elasticity of substitution) will not be comparable between different nestings.

3.3.2. Elasticity of Substitution, \(\sigma\)

Interwoven with the issue of nesting is the elasticity of substitution, \(\sigma\), which tells us the ease by which one factor of production (e.g., labour) is substitutable by another (e.g., capital). For aggregate production functions and downstream macroeconomic models, we most commonly assign \(\sigma = \frac{1}{1+\rho}\) as the Hicks Elasticity of Substitution (HES). Taking the CES function in Equation (6), this leads to the special cases where capital and labour have zero substitutability (i.e., are complements)
in a Leontief function \((\sigma = 0)\); some substitutability in a C-D function \((\sigma = 1)\), and are perfect substitutes in a linear function \((\sigma = \infty)\). Chirinko and Mallick [114,155,156] highlight the importance that conventional economics places on the elasticity of substitution between capital and labour—which appears borne out by Thomas Piketty’s recent work [157] and the subsequent flurry of academic debate [158–160]. (However, it also reveals how orthodox economists continue with capital–labour aggregate production functions that ignore energy as a factor of production.).

With multiple factors of production several key issues appear regarding the elasticity of substitution. The first relates to the confusion and mis-use stemming from multiple definitions of elasticity of substitution in common use—Stern [10] for example reviews ten different elasticities, including the Allen Elasticity of Substitution (AES), Cross-price elasticity (CPE), and Morishima Elasticity of Substitution (MES). Whilst different elasticities may be appropriate for different purposes—for example Klump and de La Grandville [161] recommended the use of MESs when studying economic growth, whereas Sancho employs HESs for CGE model calibration [118]—the multiple definitions are confusing. The choice of elasticity matters, since whilst some elasticities (e.g., AES = HES) are equal for two-input functions [162], they are not in our three factor \((K, L, \text{and } E)\) CES function case. This creates the situation where downstream mis-use of elasticities occurs. For example, Sorrell is critical of the use of non-HES elasticities for CGE modelling, since “estimates of the [more commonly estimated] AES, CPE or MES between two inputs provide little guidance in choosing the appropriate values of the HES between those inputs that are required for the nested CES functions used in CGE models.” ([163], p. 2863). Meanwhile, Van der Werf [25] (p. 2965), argues that even if HES values are selected from the literature, they are likely to be incorrect since “in most applied dynamic climate policy models, neither the production structure nor the accompanying elasticities of substitution have an empirical basis.”

The second issue is the impact of nesting structure on the elasticity of substitution. Sato’s [164] two-level nest CES function in Equation (6) permits separate values for inner-nest (K-L) elasticity \(\sigma_{KL} = \sigma_{KL}^\ell\) and outer-nest (KL-E) elasticity \(\sigma = \sigma_{KLE}\),—which he tells us can be used to justify nesting choice:

\[
\text{Introspection tells us that the [inner-nest] elasticities of substitution should be substantially higher than the [outer-nest] elasticity. After all, we justify the aggregation by the fact that aggregated factors are similar in techno-economic characteristics. One of such similarities is obviously the ease of substitution.} \ [164] \ (p. 203)
\]

Sorrell [163] (p. 2863) picks up the implication of this important point, suggesting “estimates of substitution elasticities are likely to be biased if separability is assumed where not supported by the data”. This means that the choice of nesting structure matters (e.g., KL(E) versus EK(L)), and amounts to imposing separability on the factors of production—since they are forced into nesting structures that may not match the data. Van der Werf [25] continues, illustrating how the estimated elasticity between two factors of production (e.g., K-L) vary significantly depending on the nesting structure.

Third, is that, as Sato [164] showed, in a two-level CES function only one of the two partial elasticities is actually constant over time, except in restrictive cases. An example is the constrained CES function based on Hogan and Manne [165], where the capital–labour inner-nest is assumed as a Cobb–Douglas function \((\sigma = 1)\), as given in Equation (10). Saunders [31] adopts this approach, as do some CGE models [166,167]. However, by setting (pre-analysis) elasticity of substitution values for the inner nest, this constrains the available values for all parameters to be estimated, including the outer nest elasticity of substitution.

\[
Y_t = 0 e^{xt} \left[ \delta(K_t^\ell L_t^{1-\alpha}E_t^{\rho})^p + (1 - \delta)E_t^\rho \right]^\frac{1}{p}
\]

(10)

All of this matters, since estimated parameters—such the elasticity of substitution in empirical CES studies [25,26,46,163]—can have a large influence on macroeconomic model results. For example, with a KL(E) nest, Jacoby et al. [168] found changes to the elasticity of substitution was the main driver of differences in their CGE model results, whilst in relation to energy rebound, Saunders,
who first suggested the sensitivity (and thus importance) of elasticity of substitution to energy rebound [169] subsequently found empirical support for this assertion [110].

3.4. Other CES Function Parameters

3.4.1. Productivity/Technical Change Coefficients

The exogenous part of economic output (as captured by the term $e^{\lambda t}$ in Equation (6) can also be stated as Hicks-neutral technical progress [170], with $\lambda$ a measure of its rate of change. This means productivity changes are neutral—rather than biased—across factors of production. Whilst many studies employ this assumption [79, 83], it is restrictive since it assumes the productivity of labour, energy and capital all increase at the same rate, which may simply not be true.

To overcome this restraint, separate productivity coefficients ($\tau_K$, $\tau_L$, and $\tau_E$) can be introduced (omitting the time dependant suffix from $\tau$ for ease of viewing) as first shown in a 3-factor CES specification by Saunders [169], and estimated for each factor of production. In our case, this modifies Equation (6) such that $\tau_K$, $\tau_L$ and $\tau_E$ replace the (now redundant) term $e^{\lambda t}$, as shown in Equation (11). The productivity coefficients represent technological changes of each production factor while leaving the productivity of the others unchanged. Sorrell [163] describes this as giving the separate coefficients’ ability to assign bias in technical change to specific production factors. Note if $\tau_K = \tau_L = \tau_E$, Equation (11) returns to the Hicks-neutral Equation (6).

\[
Y_t = \theta \left[ \delta_1 \left( \delta \tau_K K_t^{\rho_1} + (1 - \delta)\tau_L L_t^{\rho_1} \right)^{\rho/\rho_1} + (1 - \delta_1)\tau_E E_t^{\rho_1} \right]^{\sigma} \tag{11}
\]

In a capital–labour–energy CES production function context, van der Werf [25] and Dissou et al. [1] provide examples of this method, estimating directly the technical change parameters assigned to the factors of production. Papagerogiou et al. [171] extend this approach, by splitting fossil fuel and renewables adopting separate technical productivity coefficients.

However, a central caveat is that the factor-augmenting technical change parameters are likely to overlap with the use of quality-adjusted inputs given earlier, for example “effective labour” (as depicted by $\tau_L$) is closely related to quality-adjusted labour (human capital index × labour).

3.4.2. Returns to Scale, ($\nu$)

Empirical CES studies almost exclusively assume unity returns-to-scale ($\nu = 1$), which is an important economic assumption, and matches the popular CES specification set by Arrow et al. [12]. This restrictive assumption is tested by Layson [172], who finds the generalized CES function (where $\nu$ is unconstrained) allows an “explosive” case (where $\nu$ and $\sigma$ are >1). Szeto [82]—who estimated $\nu = 1.09$, and Duffy and Papageorgiou [6]—who estimated $\nu = 0.97–1.00$, provide rare generalized CES function examples. Curiously, both these economy-wide studies were then returned to $\nu = 1$, since as Szeto [82] (p. 7) noted “theory suggests that there are constant returns to scale in production, we will impose this restriction in the remainder of our empirical analysis”. One caveat attached to the unconstrained approach is that whilst the results will indicate how well the model supports the unity returns-to-scale assumption, the model will have fewer degrees of freedom, meaning the parameter estimates will be less precise.

3.4.3. Output Share Parameters, $\delta, \delta_1$

In the capital–labour C-D function given in Equation (1), it is a mathematical result that the partial output elasticity for capital ($\alpha$) and labour ($\beta$) is equal to the respective cost-shares of aggregate output (typically around 0.3 for capital, 0.7 for labour)—under the neoclassical assumptions that firms are profit maximizing and markets are perfectly competitive. However, in the nested capital–labour–energy CES function, the output share parameters ($\delta, \delta_1$) in Equations (6)–(8) are not equal to (and therefore cannot be set as) the time-varying output elasticities ($\alpha_K, \alpha_L, \alpha_E$), as shown in Equation (12)—adapted from Heun et al. [35]. (The exception is the limiting C-D case where $\rho = 0$).
3.5. Normalisation

A historical complaint about aggregate production functions is that they combine different units, e.g., capital (\$), labour (h), and energy (TJ), generating “production function parameters [that] have no economic interpretation” [108] (p. 7). One approach to overcome this critique is to normalise the factors of production prior to estimating the unknown parameters, since this “removes the problem that arises from the fact that labour and capital are measured in different units” ([104], p. 30). The method indexes time-series data to the base year, so \( y = Y_t/Y_0; \ k = K_t/K_0; \ l = L_t/L_0; \ e = E_t/E_0 \); with the resultant normalised (lower case) version of Equation (6) shown as Equation (13):

\[
\hat{y}_t = \theta e^{2\delta_1} \left[ (\delta_k^{p_1} + (1 - \delta_l^{p_1})L_t^{p_1} + (1 - \delta_e^{p_1})e_t^{p_1} \right]^{\frac{1}{p_1}}
\]  

(13)

For the estimation of aggregate production functions, the introduction and use of normalised variables (\( y, k, l, \) and \( e \)) means different estimated values can be obtained for the leading coefficient (theta) and share (delta) parameters. The latter could affect economic interpretation of the results. That said, the effects of normalisation may differ for empirical studies which base their research on estimating first-order conditions rather than the overall production function.

Advocates of normalization include La Grandville, Klump, and co-authors [32,108,173], who suggest advantages including a more comparable basis to study elasticities of substitution between different studies [33,115]. That said, caveats do exist: for example Temple [34] comments on the misuse of normalization for certain applications. Returning to capital–labour–energy CES empirical studies, normalization has yet to make significant inroads—Shen and Whalley [33] and Heun et al. [35] provide rare examples—but this may change as the use and publication of indexed aggregate datasets such as those of Jorgenson [174] should aid their dissemination and usage in growth accounting.

4. Empirical CES Model—Parameter Estimation

4.1. Estimation Methods

The C-D function in Equation (5) is typically estimated as a linear equation by ordinary least squares (OLS), after first taking logarithms. This simple, linear solution method is one reason for its enduring popularity. However, the CES function in Equation (6) cannot be transformed in the same simple manner to a linear equation without approximation, and so numerous other techniques have been developed, as evidenced by the CES sample studies shown in Figure 5.

Figure 5. CES estimation method in sample papers.
The most popular technique in the sample—used by over half the sample [1, 6, 46, 77–79, 82, 83]—is direct non-linear estimation of the aggregate function. Though complex, its popularity appears to be increasing, which may be due to the increased availability of econometric guidance [175], off-the-shelf programmes [30], and advances in computing power. A common method is a grid based search over relevant parameter ranges, such as that set out by Henningsen and Henningsen [30,176], and used by Heun et al. [35]. Though non-linear techniques appear attractive to solve the inherently non-linear CES function, and have been found to perform better than linear alternatives [46]—care needs to be taken: for example Papageorgiou et al. [171] (p. 26) note that “results of non-linear estimation procedures maybe sensitive to the choice of starting values of the estimation parameters”.

A second method indirectly estimates the parameters, since the solution to the non-linear function is not directly estimated. Instead, three linear simultaneous equations—one for each factor of production—are derived, based on applying the important first-order economic assumption of equality between factor prices and marginal products to the CES function—also known as Shephard’s Lemma. This method is a common approach where the sole parameter of interest is the elasticity of substitution (\( \sigma \)), as Van der Werf [25] and Dissou et al. [1] show. However, systems estimation carries with it additional risks, for example misspecification in one equation can have deleterious consequences for estimates in other equations in the system.

Third, is a hybrid indirect-direct method, based on Nerlove’s 1967 two-step process [177]. Bonga Bonga [81] provides a rare, recent example, which in the first step estimates the elasticity of substitution (\( \sigma \)) and distribution parameter (\( \delta \)), based on the estimated ratio of marginal productivities under perfect competition, and then in the second step inserts \( \sigma \) and \( \delta \) back into the CES equation, reducing it to a linear equation which is then directly estimated.

A fourth method used is direct linear approximation, based on Kmenta’s 1967 simplification of the non-linear CES equation [178]. However, since the Kmenta approximation cannot be used to lineairise CES production functions with more than two factors of production [176], it is found only in our samples for two factor (capital–labour) studies [6, 74, 85], and is therefore not a valid estimation technique for the energy-extended (capital–labour–energy) CES function.

Outside of the sample CES studies, Klump et al. [108] raise an additional, “system approach” method, which involves the non-linear solution of a “system” comprising the aggregate production function and linear first-order conditions. Two other estimation points are notable. The first relates to the “endogeneity problem” raised by Mundlak [179]—where an explanatory variable is correlated with the error term in a regression. Malleck [156] is amongst those who used Instrument Variables (IV) as a way around this issue. Second is that the production function is a long-run equilibrium concept, meaning as Chirinko [114] (p. 671) notes, there is a “fundamental tension between the short-run data that are available and the long-run parameter that is required”. Cointegration [180] and filtering techniques [181] offer potential routes forward.

4.2. Statistical Reporting

Statistical techniques and reporting provides important context to the empirical results. Three key aspects are considered here. The first are the common statistical tests on the fitted function and its econometrically estimated coefficients. From the sample, the majority report goodness-of-fit via the coefficient of determination (R²) [74, 75, 78, 83] and the Durbin–Watson (D-W) value—testing for autocorrelation of the residuals of the regression [25, 73, 77, 83]. (However, we also note that not all econometric techniques can generate the same statistics. For example, R² is not applicable with Seemingly Unrelated Regressions). The overall F-test—giving the statistical significance of the overall relationship—was less commonly reported [75, 77, 81]. Within our sample, only Easterly and Fischer [78] and Duffy and Papageorgiou [6] reported tests for heteroskedasticity in the error term (i.e., the fitted residual).

The second important aspect is the reporting of p-values on the statistical significance of individual coefficients, which is also common [6, 25, 77, 82], but should be used (and viewed) with caution. This is because p-values are a measure of the evidence against a null hypothesis, with small p-values indicate overwhelming evidence against the null. Statistical fitting software typically
assumes the null hypotheses that fitted parameters are zero, which may not be meaningful for some
parameters of the CES production function, for example the share parameter, δ. In short, analysts
should be very careful that reported p-values accord with the purposes of a study.

A third aspect relates to the reporting of standard errors, which adds important information
about the precision with which parameters are estimated. For example, in a study examining the
substitutability of energy for the capital/labour composite in a (KL)E nesting structure (Equation (4)),
the value of sigma is central. If sigma is reported as 0.5 with standard error of 0.3, it will be hard to
claim whether KL and E are substitutes or complements. If, instead, sigma is found to be 0.95 with
standard error 0.02, it could reasonably be claimed that KL and E are substitutable. Bootstrap
resampling can also be used to gain valuable insights on the precision of the estimated parameter
values. This may be particularly relevant for study of parameters estimated close to economically-
meaningful boundaries (e.g., delta = 0.96), which are unlikely to have symmetric distributions. Whilst
none of the sample studies used this technique, it is entering the wider growth accounting literature
[109,171,182], and can be applied to empirical CES analyses, as shown by Heun et al. [35].

Overall, whilst statistical reporting can strengthen the empirical results and provide better
context for comparison of results between studies, it seems an aspect of the estimation process that is
under-reported at present.

5. Recommendations

From the collated aspects in Section 3, we provide a summary of CES specification
recommendations in Table 2.

<table>
<thead>
<tr>
<th>Section</th>
<th>Aspect</th>
<th>Options</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 3.1</td>
<td>Output measure, Y</td>
<td>• Gross Domestic Product&lt;br&gt;• Gross Value Added&lt;br&gt;• Gross Output</td>
<td>“Modified” gross output metric, measured as GVA + value of intermediate inputs (i.e., GVA + cost of energy).</td>
</tr>
<tr>
<td>Section 3.2.2</td>
<td>Quality-adjusted inputs</td>
<td>• Do not quality-adjust&lt;br&gt;• Quality adjust</td>
<td>Yes, quality-adjust where possible.</td>
</tr>
<tr>
<td>Section 3.3</td>
<td>Nesting</td>
<td>• Estimate one nesting structure&lt;br&gt;• Estimate all three nests (KL-E, EK-L, KE-L)</td>
<td>Estimate and report parameters for all three nesting options.</td>
</tr>
<tr>
<td>Section 3.3</td>
<td>Substitution parameters, η</td>
<td>• Constrained fitting&lt;br&gt;• Unconstrained fitting</td>
<td>Estimate unconstrained parameters first. Then re-estimate (for comparison) with constrained substitution parameters.</td>
</tr>
<tr>
<td>Section 3.4.1</td>
<td>Technical change parameters τ L, τ K, τ E</td>
<td>• Introduce τ L, τ K, τ E&lt;br&gt;• Exclude τ L, τ K, τ E</td>
<td>Introduce if values are known/available, but be wary of conflict with quality-adjusted input data.</td>
</tr>
<tr>
<td>Section 3.4.2</td>
<td>Returns-to-scale ν</td>
<td>• Specify a-priori constant returns-to-scale ν = 1&lt;br&gt;• Unconstrained fitting</td>
<td>Estimate unconstrained parameters first. Then re-estimate (for comparison) with constant returns-to-scale (ν = 1) parameter.</td>
</tr>
<tr>
<td>Section 3.4.3</td>
<td>Share parameters, δ</td>
<td>• Specify a-priori (perhaps by cost-share)&lt;br&gt;• Constrained fitting (i.e., between 0 and 1),&lt;br&gt;• Unconstrained fitting (could be &lt;0 or &gt;1),</td>
<td>Estimate function parameters with constrained (between 0 and 1) share parameters. Exact values to be determined by the estimation process.</td>
</tr>
<tr>
<td>Section 3.5</td>
<td>Normalisation of Y, K, L, E</td>
<td>Normalise or not</td>
<td>Always normalize for three factor CES functions.</td>
</tr>
</tbody>
</table>

Two key aspects of CES specification are worthy of further discussion here. First is the linked
issue of output measure and normalization. We saw earlier the variation in Table 1 relating to choice
of output measure in empirical energy-extended CES studies. This was a surprising finding, and one
we would expect perhaps to change in future, if a consensus emerges—e.g., towards a “modified”
gross output (GVA + energy cost). Allied to this is the effect (and thus importance) of normalization. If absolute measures of economic output (Y) are used when fitting the CES aggregate production function, a change from GDP to GVA or gross output will affect parameter estimates if GVA or gross output are different from GDP, as they are likely to be. If indexed (normalized) measures of economic output (Y) are used when fitting the CES aggregate production function, a change from GDP to GVA or gross output will affect parameter estimates only if the indexed value of GVA or gross output is different from the indexed value of GDP. i.e., estimates of fitted parameters will not change if GVA or gross output is different from GDP by a constant multiplicative factor only. Second—and related to the first—is the issue of quality-adjusted inputs. Whilst desirable, two caveats can be attached: (1) that quality-adjusting inputs can bring a direct conflict/duplication with technical change parameters, τ; and (2) there are different methods of quality-adjusting each input (capital, labour and energy). In sum, this means great care is required to construct the inputs properly, and to align them to the chosen output measure.

Moving to CES estimation, two recommendations from Section 4 are made. First is the advocation—with caveats—for the common use of off-the-shelf non-linear coding to solve the aggregate function such as Henningsen and Henningsen [30]. The caveat is that caution is required: there are issues such as near boundary solutions and multiple solutions which require attention—see Heun et al. [35]. Second is the desire for the deepening of statistical reporting via: (1) reporting all nesting structure results; and (2) greater reporting of the estimated precision of parameter values, such as through bootstrap resampling.

Finally, looking forward, to ultimately provide better inputs to energy policy, effort in two key areas is required. First, more detailed research by the energy economics community is required specifically to investigate the impacts of modelling choices (e.g., normalization, or quality adjusting inputs) on estimated parameter results. This will help identify which of the CES modelling choices have the most significant impacts on results. The empirical “sister” to this current paper by Heun et al. [35] is an intended next step in this direction. By improving understanding of modelling parameters, this will also better inform users of the analytical results—e.g., CGE modellers and policy-makers—as to the robustness and sensitivity of assumed parameters. Second, empirical CES studies currently exhibit great variation in the type of results reported. Developing a consistent approach to reporting would add interpretative value to the study itself, and enable improved inter-study comparisons. For example, studies could consistently report results showing the effect of: (1) new datasets (e.g., new quality-adjusted verses incumbent non-adjusted variable); (2) different nesting structures; and (3) the application of broader statistical techniques (e.g., standard errors and bootstrapping).

Acknowledgments: We gratefully acknowledge the support of Engineering and Physical Sciences Research Council (EPSRC) and Arup for contributing to the PhD CASE (Collaborative Award in Science and Engineering) scholarship of Paul Brockway. Paul Brockway’s time was also funded in the later stages of this work as part of the research programme of the UK Energy Research Centre (UKERC), supported by the UK Research Councils under EPSRC award EP/L024756/1. João Santos was financially supported by FCT/MCTES (PIDDAC) through project UID/EEA/50009/2013. The support of the Economic and Social Research Council (ESRC) is also gratefully acknowledged. The work contributes to the programme of the ESRC Centre for Climate Change Economics and Policy. We also gratefully acknowledge comments and inputs from Timothy Foxon, Randy Prium, Marco Sakai, Steve Sorrell, and Julia Steinberger during the development of the paper. Finally, we would like to thank the three anonymous reviewers whose comments have strengthened the paper.

Author Contributions: Matthew K. Heun, Paul E. Brockway conceived this landscape paper at the 2014 International Exergy Economics Workshop in Leeds, England, 19–20 May. Paul E. Brockway led the narrative framing for the paper and coordinated its writing. All co-authors provided substantial inputs to writing and significant comments on numerous drafts.

Conflicts of Interest: The authors declare no conflict of interest.

References


