Issues with Constant Elasticity of Substitution functions for Energy Modeling in General Equilibrium Integrated Assessment Models

Abdulla Kaya[1], Denes Csala[2], Sgouris Sgouridis[2*]

[1] Masdar Institute, Abu Dhabi, United Arab Emirates

[2] Lancaster University, United Kingdom

Abstract

The use of Constant Elasticity of Substitution (CES) functions in General Equilibrium Integrated Assessment Models (GE-IAMs) to forecast the long-term energy system evolution under emissions constraints introduces unjustified biases. Applying CES, an economic modeling method, for the substitution of technical factor inputs where an entire factor (fossil fuels) should be substituted fails to match historically observed transition patterns and is very sensitive to the structure of CES implementation (nesting) and parameter choice. The resulting methodology-related artefacts include: (i) the extension of the status quo technology shares for future energy supply relying on fossil fuels with carbon capture, biomass and nuclear, (ii) monotonically increasing marginal abatement costs of carbon, and (iii) substitution of energy with non-physical inputs (e.g. knowledge and capital) without conclusive evidence that this is possible to the extent modeled. We demonstrate these issues using simple examples and analyze how they are relevant in the case of three major CES-based GE-IAMs. To address this, we propose alternative formulations either by opting for perfect substitution for alternative energy options or by introducing dynamically variable elasticity of substitution as a potential intermediate solution. Nevertheless, complementing the economic analysis with physical modeling accounting for storage and resource availability at a high resolution spatially and temporally would be preferable.

Keywords: societal energy efficiency, sustainable energy transition; energy economics; renewable energy.

* Corresponding Author: Associate Professor, Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates, Tel: +971 2 810 9156, Fax: +971 2 810 9901, email: ssgouridis@masdar.ac.ae

1 Introduction

A sustainable energy transition that eliminates atmospheric carbon emissions within a few decades is a critical component of climate change mitigation strategies. Identifying the optimal mix of energy resources to reach the carbon mitigation targets is one of the aims of Integrated Assessment Models (IAMs), models that combine climate and economic considerations. This paper discusses how IAMs that use constant elasticity of substitution (CES) functions – also referred to as general equilibrium IAMs (GE-IAMs) – and their variants for modeling the energy system may generate results that are inconsistent with empirically observed dynamics of previous energy transitions. GE-IAMs thus tend to preserve the share of conventional resources into the future, consequently underestimating the share of non-fossil energy resources that currently have a small share, and tend to forecast an exponentially increasing carbon price. We show that these outcomes are driven by the use of CES functions rather than from an assessment of the relative characteristics of the resources and their technical substitution potential.

IAMs study the co-evolution of the climate and economic systems several decades into the future. By comparing the costs of climate change against the costs of preventative actions, they intend to inform climate and energy policy. In order to fulfill their stated objective, IAMs should (Pindyck 2013):

(i) Represent the economic system over long periods.

(ii) Relate economic activity into corresponding greenhouse-gas (GHG) emissions.

(iii) Estimate the impact of GHG emissions on the climate usually summarized in the degrees of change of average temperature from the pre-industrial baseline.

(iv) Calculate how climate change impacts economic activity. These impacts can be summarized as the social cost of carbon (SCC) – i.e. the net present value of the impacts of GHG allocated per unit of emissions.

(v) Estimate the costs incurred by reducing or substituting GHG emitting activities in order to comply with a desired level of GHG atmospheric concentration. The technical costs of emissions mitigation are represented by the marginal abatement cost (MAC) of GHG for meeting a given level of atmospheric

GHG concentration. In an efficient, universal trading system, the carbon price (e.g. price of emission permits) should be equal to the MAC (Kuik et al. 2009).

The estimation of the SCC is effectively exogenous to the IAM and dependent on three fundamental assumptions that can lead to widely variable results: the choice of discount rate, the formulation of the climate damage function, and the assessment of the probability of catastrophic events (Pindyck 2013). The upper bounds of climate sensitivity are difficult to deduce from historical data. In particular, a fat-tail climate response, where the likelihood of high temperature changes remains substantial even for moderate GHG concentrations, cannot be excluded (Allen and Frame 2007). This leads Weitzman (2009) to consider as misleading climate damage functions that disregard the fat-tail distribution properties of catastrophic climate change.

While the issues related to IAMs' SCC estimations have been critically assessed, relatively less attention has been paid to their energy modeling aspects. Barreto and Kemp (2008) confirm that energy models fail to take into account the empirically-solid observations of S-curve innovation and technology diffusion patterns but they did not identify specifically the inconsistencies arising from this omission. Pearce and Weyant (2008) point out the great sensitivity of IAMs' cost estimates on "structural characteristics and assumptions embedded in the model" even when the key assumptions are the same. Truong (2009) notes a potential flaw that can lead to overestimation of the MAC indicating that "*the CES production function […] can have unclear biophysical implications […] imply[ing] a pattern of the output elasticity […] which can be inaccurate and unrealistic from a biophysical or technological viewpoint*". He also points to the need to separate technical from economic efficiency in monetary models as their confusion may lead to physically impossible results. Frei et al. (2003) notes that "*structural change, i.e. the 'natural' emerging and phasing out of technologies, is incompatible with the neoclassical concept of smooth (i.e. differentiable) substitution*" that is assumed in a CES. Recently, Pietzcker et al. (2017) attempted to address identified deficiencies in the representation of variable renewable sources in IAMs and noted that CES "*create[s] a preference for base-year calibration shares […that] can lead to physically implausible aggregation.*" Finally, Rosen and Guenther (2015) highlight that the complexity and uncertainty surrounding economic and technological evolution makes it difficult to defend simplistic conclusions of economic costs drawn by IAMs as their assumptions do not capture the synergistic benefits of technology development, climate mitigation, and increased reliance on renewable sources.

Building on these critiques, we investigate the structural issues in GE-IAMs, arising from a CES that can bias (i) the estimation and shape of the MAC, and (ii) the technological composition of future energy scenarios. The second case includes the exaggeration of the energy efficiency potential and the longevity of the fossil fuel status quo with only minor variation in the technology shares of future energy supplies. We show that while the MAC is highly sensitive to the parameter values (elasticity of substitution, initial shares, and assumptions on economic costs), the technology composition is also strongly dependent on the model structure per se. A common factor to both issues is that they are endogenous outcomes of the modeling formulation, preventing the reproduction of the empirical behavior of energy technology transitions.

Section 2 presents the MAC and technology share biases by analyzing two CES function examples, one retrospective and one prospective. Section 3 discusses how CESs are implemented in several state-of the art GE-IAMs through a nesting methodology and structurally explains the results of these IAMs regarding the MAC and the technology shares pointing out the issues. Section 4 suggests alternative formulations based on empirical research on energy transitions, including a modified application of CES.

2 The CES function and its introduced artefacts when applied in long-term technical transitions

CES functions were introduced as a general production function form for capital and labor substitution (Arrow et al. 1961) that could provide flexibility increasing the options between a Leontief function (where the elasticity of substitution is zero i.e. the two factors cannot be substitutes) and a Cobb-Douglas function (where the elasticity of substitution is unitary). In this setting, neither labor nor capital would be reasonably expected to dominate and the range of possible substitution between them would be limited. Bounded this way, the use of CES was justifiable and representative. Nevertheless, convenience and ubiquitous use in general and partial equilibrium models of what became a conventional approach have led to extending the use of CES functions to model processes with technical factor inputs. As we discuss below, this extension is not supported empirically because technical processes can fully substitute each other and in fact the substitution tends to become easier as the penetration of an alternative technology increases, thus changing the elasticity of the substitution. This section demonstrates in detail how CES operates and explains why some policy-critical results are artefacts of the equation without a corresponding physical justification and are highly sensitive to the following parameters: initial shares, choice of the elasticity of substitution, technology cost improvement.

2.1 The CES function and physical input shares

In the form shown in Eq. 1, (α) is the share of factor of production (F) and ρ defines the elasticity of substitution (σ) as $\sigma = 1/(1+\rho)$. While the two-input form can be expanded into multi-parameter general forms, there are limitations in solving these, especially if the elasticities are not equal, so in practice computable general equilibrium (CGE) models use nested CES functions to model multiple factors of production with different elasticities of substitution. This is discussed in Section 3.1.

$$Y = [\alpha \cdot F^{-\rho} + (1-\alpha) \cdot R^{-\rho}]^{-1/\rho} \quad (1)$$

A CES function describes the amount of change required in one input in response to a certain amount of change in other input(s) to maintain a given utility level. CES functions are limited by the original structure of the inputs of the production function. If an input is not initially utilized and therefore not designed-in, then it will not be used by the model in its future projections. Importantly, in a CES function the marginal rate of technical substitution (MRTS) of one factor with respect to another (i.e. $\Delta R/\Delta F$) depends on the factor proportional value (R/F) as shown in Eq.2 (derivation in SM).

$$\mathrm{MRTS}_{R,F} = -\frac{\Delta R}{\Delta F} = -\frac{\alpha}{1-\alpha} \cdot \left(\frac{R}{F}\right)^{1+\rho} \quad (2)$$

A fundamental underlying assumption is that the two inputs are not perfect substitutes. Initially in order to substitute one unit of the larger factor requires a comparatively small amount of the substitute input, for a large initial share (α) and a positive ρ this can be significantly less than one. By its mathematical definition though, CES subsequently requires exponentially increasing amounts from the alternate factor presenting an exponentially increasing obstacle to the substitution process. In other words, *CES functions tend, by design, towards factor share-preservation*. As the original share (α_i) of a factor *i* increases, other factors need to be utilized exponentially more to replace a marginal reduction of *i* while maintaining the output (Y) constant as the substitution progresses.

In order to demonstrate the problematic nature of this formulation for large-scale energy transitions, we apply it retrospectively to the coal/oil factor substitution in the US and global economy that occurred between 1880 and 1980. We take the position of a modeler in 1900 assuming perfect foreknowledge of the shape and timing of the relative share (logistics curve) of coal as well as the growth rate in total energy demand. Fig. 1 shows a CES formulation that replicates a coal-oil substitution nest as used in the IAMs of Section 3.

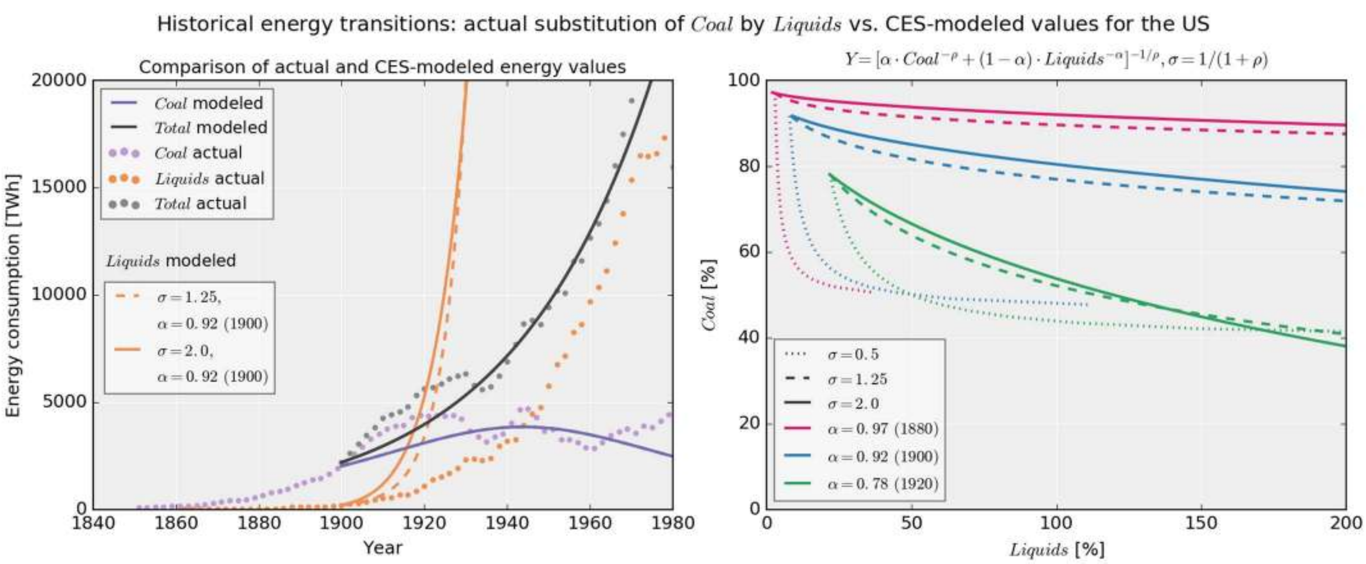

Figure 1 Retrospective CES model of substitution for coal by liquids for the US starting in 1900 showing the physical trajectory versus the modeled one (Left) and in terms of relative shares of the total output (Right), using different values for coal's initial share of the energy supply (α) and the elasticity of substitution (σ).

The absolute factor input values are an important and tangible indicator that can be used to examine how aligned model forecasts are with historical experience. In this example, the CES formulation forecast contradicts any reasonably expected technical substitution behavior as it only matches the historical trajectory for less than a decade, thereafter vastly overestimating the growth in liquid energy supply necessary to compensate for the decrease in coal energy. In the modeled case, the oil curve is estimated with coal at an initial share α=92% (solid orange line in Fig. 1L) and σ=2. The ratio of the liquids and coal modeled values to the modeled total energy output (the black solid line in

Fig. 1L) for different initial share assumptions is shown in Fig. 1R. For initial coal share (α) to drop from 92% to 80% of the output, the CES model requires liquids to grow from 8% to 100% of the output (Fig. 1R, blue solid line). As the substitution progresses, the required energetic output from liquids exceeds the total demand since they are considered an *imperfect substitute to coal*. This assumption eventually leads to radical differences between the modeled and observed values (compare the orange solid line with the orange dots in Fig. 1L). This occurs progressively as for every unit of coal that is substituted, it requires exponentially more liquids due to a decreasing marginal rate of technical substitution. Such a result is contrived, estimated with no reference to the technical capabilities and specifications of the two resources it inevitably fails empirical validation.

In practice, the unrealistic physical substitution requirements the CES generates in high-share transitions are circumvented in three ways: (i) assuming substitution by a non-energy factor of production which is theoretically unbounded – e.g. capital, labor, or knowledge, (ii) applying a back-stop technology that is activated once substitution costs cross a threshold, and (iii) using nested CES that can increase the number of factor inputs. Effectively the nested models "optimize" between the shares without letting the physical "substitutes" become too expensive as the substitution costs also increase. We discuss this behavior of nested CES models in Section 3.1, but it is useful to also illustrate how the cost dynamics of the technologies are handled and how the MAC curve is generated in the case of a single CES function.

2.2 The CES function and mitigation costs: inferring the MAC

We calculate Eq. 1 values for the substitution of a portfolio of fossil sources (F) by a portfolio of renewable energy (R) sources generating output (Y) aiming to reduce system emissions. We initialize and solve the CES function assuming the market cost of the fossil resources (c_F) as constant unit of comparison. The initial cost of RE (p_R) is calculated by the initial conditions through Eq. 3 and subsequently may decrease at rate (ζ). The amount of R needed to substitute F at any point corresponds to a price of fossil resources (p_F) necessary for forcing the transition (Eq. 3). A result of the Lagrangian function solution of the CES with the assumption of a perfectly competitive two-goods market for energy and unrestricted income is derived in the Supplementary Information (SI) in line with Gohin and Hertel

(2003). The MAC is the difference between the calculated fossil fuel price (p_F) that is required to effect the substitution and their market cost (c_F) shown in Eq. 4.

$$\frac{p_F}{p_R} = \frac{\alpha}{1-\alpha} \cdot \left(\frac{F}{R}\right)^{-\rho-1} \quad (3)$$

$$MAC = p_F - c_F = p_R(1-\zeta) \cdot \frac{\alpha}{1-\alpha} \cdot \left(\frac{F}{R}\right)^{-\rho-1} - c_F \quad (4)$$

Fig. 2 graphs the values of the input resources and resulting MAC for different initial shares (α) and elasticity of substitution (σ) values in line with the range of values used in IAMs assuming a constant growth rate for the output (for a constant output see Fig. S1). As we have seen in Section 2.1, (i) the substitution curve F-R shows an exponentially increasing relationship, and (ii) the substitution rate reaches different levels depending on the initial share (1-α) but remains far from a complete substitution. The MAC pattern is a typical exponential curve (see Fig. 2-right) that is *very* sensitive to the initial share, the choice of the elasticity of substitution parameter, to the initial share (α) and to the cost reduction rate (ζ). The sensitivity of MAC is mapped in Fig. S2 where a seemingly small variation in the choice of σ can lead to differences of up to four orders of magnitude in the final MAC value and up to seven if we include the variation of α.

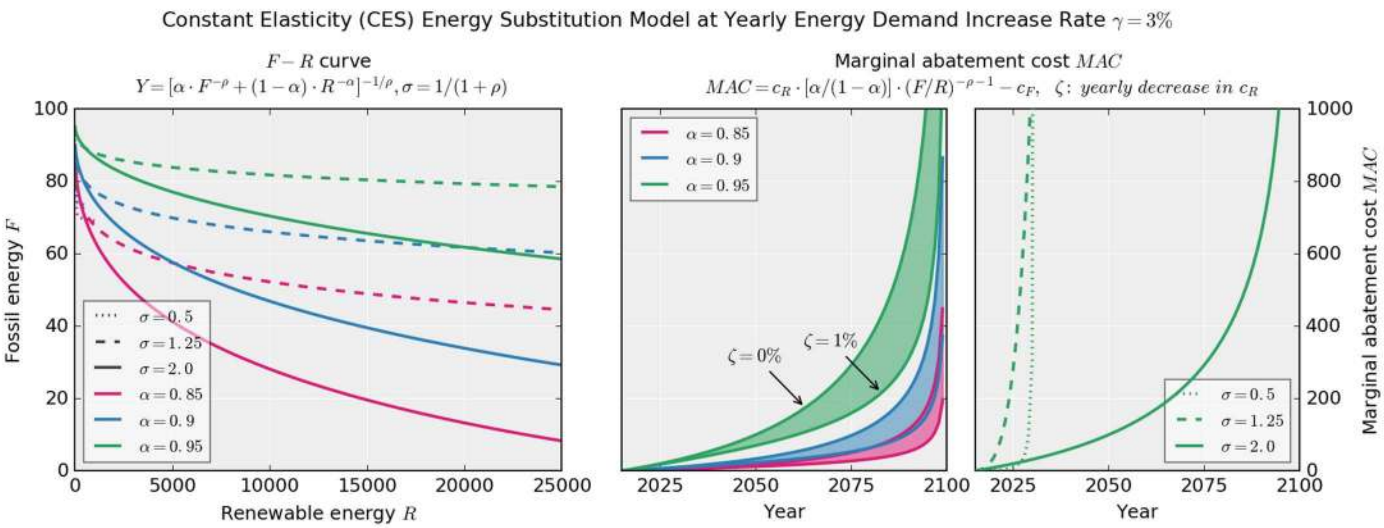

Figure 2 Substitution quantities of fossil (F) and renewable (R) energy in energy units (left) and corresponding marginal abatement cost levels in monetary units needed to "force" the modeled using a CES function (right) for energy output increasing at rate (γ)=3% with variable initial shares (α), elasticities of substitution (σ), and rates of RE cost decrease (ζ).

We note that using prices instead of physical inputs in Eq. 3 veils the behavior of the physical quantities. The MAC trajectory shows that it becomes exponentially harder to switch from fossils to renewables without even reaching complete substitution, with MAC levels that are one thousand times the cost of the fossil energy resource. Nevertheless, the resultant MAC is also extremely sensitive to the rate of RE cost reduction (ζ), which can reduce its value significantly. Yet a high assumed rate of RE cost reduction (ζ) changes neither the assumed technical share nor the exponential MAC behavior.

These stylized CES examples demonstrate the fundamental characteristics and sensitivity of CES and how the resulting share preservation and exponentially increasing carbon price dynamics are artefacts not based on the characteristics and dynamics of technological substitutions. We examine such instances in current GE-IAMs in the next section.

3 CES –based artefacts and their implications

Table S1 presents an indicative collection of GE-IAMs and the method for energy economy modeling. While there is a large proliferation of models there are fundamental similarities in how the models approach technology substitution as can be seen by the similarities in the handling of wind and solar, the most scalable renewable energy (RE) options. While the broad similarities of the CES-based results lead us to believe that the underlying mechanisms described in Section 2 apply, we do not extend our analysis to all permutations of CES-based IAMs. We focus on four current, state-of-the-art CES-based models: World Induced Technical Change Hybrid Model WITCH (Bosetti et al. 2006), Emissions Predictions and Policy Analysis (EPPA) (Paltsev et al. 2005), the Global Trade and Environment Model GTEM-C (Cai et al. 2015) and the Regional Model of Investments and Development REMIND model (XXX) in order to illustrate how the structural choices of the models replicate the artefacts discussed in Section 2 in their results.

3.1 Propagating the Status-quo: Structure and implications of CES nests

The nested CES structure of the WITCH model is reproduced in Fig. S3, indicating both the type and the elasticity of substitution of each nest. In Fig. S8, we have also included the original structure of REMIND from (XXX). For the case of WITCH, Energy (EN) is a CES nest with input factors electricity (EL) and non-electricity (NEL) which are imperfect substitutes (σ=0.5). The NEL nest by construction can only be supplied by biomass/biofuel or fossil fuels (with/without CCS). As a result, this formulation forces (i) the use of significant NEL indefinitely and (ii) reliance on biomass as the only way to economically provide the NEL fraction in the event of a fossil fuel phase-out, effectively making options like widespread electrification of transport or a hydrogen-based economy *de facto* impossible. This is directly visible in the RE shares for the RCP2.6 results shown in Fig. S4; even for the power sector, RE does not reach 20% (Bosetti et al. 2015). Moving to the EL nest, while hydro can perfectly substitute all sources of electricity, nuclear, RE (as ELW&S) and fossil fuels (ELFF) are imperfect substitutes (σ = 2). In other words, again simply by the structure of the model, for the share of RE to grow it would soon require very large amounts of resources and even if it costs very little, it could never fully substitute neither FF nor nuclear. As a final observation, energy services (ES) has as a factor input a representation of knowledge/efficiency (HE) along with EN, and in turn is a factor input along with economic factors (K, L) for the total output. This structure and the presence of non-physical substitutes (HE, K, L) imply that if the EN costs become too high, the model will choose to introduce additional knowledge or additional capital/labor (neither of which have a physical limitation in the model universe) allowing the energy intensity of the economy to improve without examining if such a feat is technically feasible.

REMIND has a similar nested CES structure to that of WITCH and it also raises energy to the level of substitutability (albeit relatively hard) with capital and labor. It segregates energy into practically unsubstitutable (σ = 0.3) stationary and energy for transport categories. It does allow for electrification of transport, but only for the case of light-duty vehicles. Stationary energy is split by electric and non-electric energy use, making electric power relatively inelastic (σ = 1.3) compared to non-electric stationary energy. Final energy demand is then calculated through a linear resource distribution model using region-specific fossil fuel extraction curves and renewable energy resource potentials, through techno-economic cost assumptions. This raises further problems from the CES nested structure as it has exogenous GDP growth and interest rates, as well as exogenous technology advancement rates – all three of which have been notoriously hard to estimate accurately over decadal timescales (XXX).

However, a very recent model inter-comparison study (XXX) that included detailed analyses using modifications of both WITCH (XXX) and REMIND (XXX) to include curtailment and storage effects of variable renewables, presented renewables shares of 37–75% of electricity supply by 2050, and 53–89% by 2100 (Fig. S5). This has been enabled by XXX.

EPPA is another complex, multi-regional CGE IAM model that incorporates global economy interactions, trade and GHGs abatement dynamics. It is constructed using the CES function and variants for all type of production output, including energy, and final consumption in the characteristic nested structure but also includes perfect substitution for the (Figure S6upper). While some technologies are represented as perfect substitutes, EPPA treats RE "*as imperfect substitutes represent[ing] the unique aspects of these renewable technologies. While they can be well-suited to some remote locations, they also suffer from intermittency that can add to their cost if they were to provide a large share of electricity production*" (Paltsev et al. 2005). The same biases are present in the household use of energy in which fossil electricity is an imperfect substitute for fossil fuels, which are the only possible input for transport (Figure S6lower). It is therefore unsurprising that its results also essentially perpetuate the current situation, simply choosing the "efficiency" pathway in the same mechanism as described for WITCH and shown in Fig. S7.

The elasticity of substitution chosen in EPPA permits gradual but limited adoption of renewable energy only as prices of other technologies rise. Recognizing this EPPA limitation, Cheng (2005) developed an approach for integrating intermittent sources in CGE modeling with counterintuitive but edifying results. Assuming that all intermittent sources require fossil back-up, i.e. ignoring non-fossil storage options, ends up *eliminating* all RE sources in a high carbon price low emissions scenario.

GTEM-C is a hybrid CGE model designed to combine the top-down (CGE) model with the bottom-up engineering details of energy production and consumption based on an earlier version of the Global Trade and Environment Model (GTEM). They use a technology bundle approach in addition to the classical CES and a back-stop technology. For electricity generation, the bundle includes nuclear, hydro, wind, solar, biomass, waste, and other renewables in addition to fossil fuels. GTEM-C uses different variants of the CES function in several calculation stages, including switching between production factors for the total economic output but also between different technologies for energy generation. As reproduced in Figure S8, GTEM-C utilizes a combination of Leontief functions, constant return of elasticities homothetic (CRESH) functions (Hanoch 1971), and typical CES to create a nested structure for industrial output. The technology bundle approach allows the power sector to largely transition to "clean" sources although without specifying in the results the relative shares of the technologies. Nevertheless the upper nests' structural requirements prevent any effective transition in the rest of the energy sectors, effectively decarbonizing less than 20% of the energy system.

Other CES-based models also tend to propagate the status-quo in energy shares, finding that CCS and biomass with CCS (i.e. negative emissions) is a significant part of any viable technology mix for achieving a low concentration pathway, c.f. Fig. S10 RCP2.6 in van Vuuren et al. (2011). In a comparative study of three IAMs (GCAM, MERGE and EPPA) the carbon price was exogenously determined either as a quadratic increase in carbon tax or a carbon shock (Wilkerson et al. 2015). Since GCAM is based on logit functions for estimating market share, is not within the scope of this paper. MERGE and EPPA under the tested assumptions forecast that fossil fuels remain the primary energy input in the future, albeit with CCS. Notably, this study showed *extreme* rate changes of nuclear and biomass without recognizing the practical constraints of the task and timeframes described. GCAM estimates a quadrupling of nuclear energy within 5 years, while MERGE which is CES-based deals with the shock by increasing biomass with CCS ten-fold within a decade (see Fig. S11 from Wilkerson et al. (2015)).

Kriegler et al. (2015) compared a larger number of models with similar relevant results. Using the metric of transformation index (with 0 no change and 2 total change) the final energy mix scores values from 0.3 to 1.1 and primary energy mix from 0.9 to 1.6. In other words even when renewables include biomass, they still do not effect a complete transition by 2100 even under high carbon prices. That biomass is conventionally considered as an easier substitute for fossil fuels than RE is reinforced in a recent model survey (Luderer et al. 2013).

When it comes to MAC implications, in section 3 of the SI, we postulate an explanation for the exponentially rising marginal abatement costs, rising from the structural presentation of the models given in Section 2.2 and presented Fig. 2.

3.2 Policy implications of the RE modeling under CES

As we saw, the use of CES for the energy system can lead to a systematic expectation of exponentially increasing costs imposed by a transition to a zero-carbon energy system that translates into an exponentially increasing carbon price and the reliance on perceived as known substitutes: biomass, fossil fuels with CCS, and nuclear without a physical consideration of their ability to scale. These results may elicit three responses in policy-makers: *price diversion, sticker shock, and technological myopia*.

Price diversion refers to focusing climate policy on setting a price for GHG. It supports an expectation that known technologies can scale fast in response to carbon price introduction and changes as in (Wilkerson et al. 2015). At the same time, an exponentially rising MAC in conjunction with optimistic assumptions of a growing economy and the comparably low cost estimates from climate change allows limited action to become an optimal policy as in (Nordhaus 1992). Using the MAC to guide policy decisions in a very dynamic technological transition environment where new technology costs drop faster than any prior technology forecast and commodity prices fluctuate seems counterproductive.

Sticker shock is a reaction to exponentially increasing carbon prices and the implied "costs" of mitigations. Policy makers are likely to perceive setting a carbon price measure in the thousands of dollars as politically futile when it is barely possible to pass legislation that raises it to the low tens. This attitude would be different if the need for a carbon price was shown to be temporary and with an eventually declining slope. In fact alternative economic models do find much lower carbon prices with non-exponential shape as sufficient to induce successful transitions, e.g. Mercure et al. (2014). Additionally, a high carbon price creates the unfounded perception that building alternatives is an endeavor that creates costs to the economy that are additional to business-as-usual. This is not necessarily true in absolute terms as we discuss in Section 4 because it fails to capture the implications of the complexity of the economic system (Rosen and Guenther 2015) and the dual dividend of efficiency and innovation that has been observed to arise when pollution externalities are radically constrained (Porter and Van der Linde 1995).

Finally, *technology myopia* is manifest in the belief that the technologies favored by the model can in fact expand at the forecasted rate while other alternatives that are perceived as imperfect substitutes become too expensive and unable to cover all energy needs. It also implies that non-physical inputs of knowledge and capital can somehow substitute for energy thus reducing the economic energy intensity. Reinforced by the structure of the CES nests, the costs of incorporating new technologies soon become prohibitive reverting to perceived as perfect substitute technologies: fossil CCS, nuclear, or biomass. In the models we have studied, we do not find explicit supporting evidence of these substitutes' technical ability to scale or their resource availability (especially for biomass) nor of the substitution potential by non-technical factors (e.g. knowledge and capital).

4 Alternatives for technical substitution: towards a physical basis

One would assume that a review of empirical findings should be a critical first step when modeling transitions. Yet, (Rosen and Guenther 2015) found "*no literature comparing investment decisions for energy-consuming equipment implicit in IAMs with real-world trends in the past*", a test the CES approach fails in Section 2.1. A recent ex-post analysis of cost-optimization models for the UK found significant deviations of the actual system from the cost-optimal or near-cost-optimal estimates (Trutnevyte 2016). Looking forward, the reliance on traditional technologies plus CCS which is the consensus from CES-based GE-IAMs in Section 3 is contrasts with both current deployment trends (Breyer et al. 2017) and the literature that uses bottom-up, physically constrained analysis to model a 100% carbon free economy and that considers solar and wind RE as the scalable workhorse technologies with some reliance on biomass depending on the location but with no nuclear or CCS e.g. (Lund and Mathiesen 2009) for Denmark, (Ben Elliston et al. 2013) for Australia, and (Singer 2010; Jacobson and Delucchi 2011; Sgouridis et al. 2016) for global transitions. These studies recognize the physical and political scale limitations of nuclear, biomass, and CCS, and the strong learning curve in scalable RE technologies, solar (Görig and Breyer 2016), wind and storage, which allow them to present a complete energy system alternative. The RE deployment deficit in IAMs is recognized by Pietzcker et al. (2017). They investigate how changes to IAM model specifications allows for greater penetration of RE primarily through the use of residual load curves in combination with multinomial logit models (Ueckerdt et al. 2015). This offers a welcome change but is still limited by looking only at the electricity system, assuming that the load curve is unresponsive and power-to-liquids (if included) only used for storage rather than providing economic uses outside the electricity system.

Generally, physical models conform better with historical observations that, as the technology substitution progresses, the transition accelerates before slowing down near saturation. This S-curve shape of technological transitions, well established across economic sectors (Schumpeter 1943; Christensen 1997) but also in energy system transitions as shown in Figures S12, S13 (Ayres et al. 2003), (Fouquet 2010), (Grubler et al. 1999), implies that the elasticity of substitution is not constant. In the widest documented sampling of technology deployment trajectories, the pattern of exponential deployment rate and exponential cost reduction in technology diffusion was consistently observed (Nagy et al. 2013). Substitution dynamics were explored early on by economists using the logistic function (Mansfield 1961) and historical energy transitions on a technological/social basis (Marchetti and Nakićenović 1979) but also on a physical depletion basis with the well-known Hubbert curve (Maggio and Cacciola 2009). Differential equation-based models like the Bass Diffusion model (Norton and Bass 1987), were also utilized.

Since an S-curve transition cannot be modeled with a CES function because it fundamentally presents a relationship with a *dynamically* evolving elasticity that should eventually end with a complete substitution. Taking this as a starting point, we suggest the use of either perfect substitution with modeled storage or using a dynamic elasticity of substitution (DES) with σ increasing over time. This approach also allows for a complete substitution and ameliorates the issue of very high increases in the modeled quantities of substitutes even from very small initial shares thus offering a viable compromise that can be retrofit to existing formulations (see Fig. S14 for the input shares). It also has the additional advantages of providing an inflection point in the curve of the MAC and drastically reduced estimates of MAC values as shown in Fig. 4.

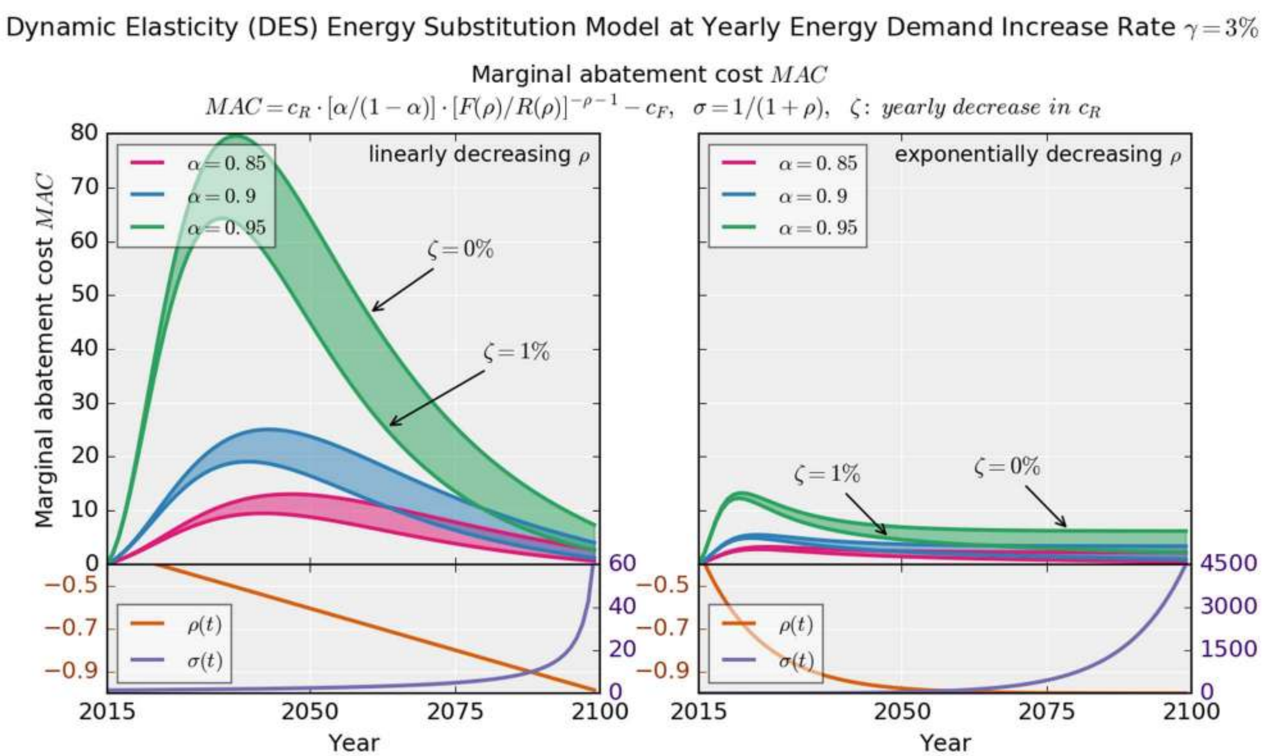

Figure 3 Using a Dynamic Elasticity Energy Model with linear and exponential decrease of the parameter (ρ) for increasing output

The MAC peak observed by using a dynamic elasticity of substitution may be counterintuitive to neoclassical thinking, but in practice, stems from the dynamics of technical diffusion. Economies of scale, scope, learning, and the initial adopters' pull of followers define technology diffusion. In the worst case, the carbon price would be capped by the cost of the back-stop technology (e.g. CSP with thermal storage, PV with chemical or mechanical storage) with no reason to increase exponentially beyond. Empirically, we note that RE generation capacity is expanding at a significant rate in markets that do not have a carbon price (e.g. in China and India) and are fully cost competitive at utility scale with other power system options. This does not mean that RE will expand automatically – it still requires support, primarily in the form of leveling the playing field but also for establishing a similar momentum in large-scale energy storage and power-to-liquids (P2L).

In key sectors like transportation, changes like a modal shift from air travel to high-speed rail provide important avenues to utilize the renewable energy supply (Sgouridis et al. 2010) competing directly and even entirely dominating in certain markets. For automobiles and residential energy storage, battery technologies are well on their way to become directly competitive (Nykvist and Nilsson 2015). On the supply side, the levelized cost of energy (LCOE) trajectories for wind and especially for solar photovoltaic have been consistently declining, reaching grid-parity compared to both residential and wholesale electricity prices in several regions. While it is correctly pointed out that LCOE of RE does not account for the full cost of renewables integration (Joskow 2011), it is possible to operate such systems with no fossil input as the combined LCOE of RE plus short and long-term storage, including power to liquids that can fuel any non-electrifiable application, fully represent the system costs and there are no further "integration" costs. Studies that take this into consideration confirm the existence of manageable upper bounds at global (Pleßmann et al. 2014), or country level analyses e.g. Germany (Ueckerdt et al. 2015) and Japan (Komiyama and Fujii 2015). Notably, while Pleßmann et al. disadvantaged renewables by not considering long-distance transmission options and also disregarded the use of hydro resources, the costs of 100% RE-based electricity do not exceed double the current retail cost of electricity for the majority of the world population.

5 Conclusions

While complex IAMs have been developed that represent multiple economic sectors and geographical regions, those that rely on CES functions (GE-IAMs, nested or not) for modeling the energy system introduce mathematical artefacts that are inconsistent with long-term energy transitions. These models produce exponential increases in MAC as the transition continues and rely on what the modeler perceives as perfect substitutes to the current fossil fuel paradigm (fossil with CCS, nuclear, and biomass) ignoring historical observations and current technological trends. As such, these mathematical artefacts should be carefully scrutinized before basing policy decisions on. We suggest that using different formulations to recognize perfect substitution for renewable energy with storage or at least introduce a dynamic elasticity function offer two plausible options. Dynamic elasticity allows for a peaking MAC at reasonable values preventing the exponential MAC growth resulting from CES. Eschewing CES and adopting concepts like the residual load-curve and multinomial logit is useful but still the electricity system focus should be expanded to include all economic energy uses. This is in line with the empirical evolution of past transitions and is in accordance with the downward sloping costs of the renewable energy supply and storage technologies. In the longer term, energy-based models should be reconciled with economics-based GE-IAM results.

Acknowledgments

We thank Joshua Msika and Maury Markowitz for their comments and editorial suggestions. The careful reading of the two reviewers and the editor for their critical input. Finally we acknowledge the financial support by Masdar Institute in conducting this study.

References

Allen MR, Frame DJ (2007) Call off the quest. Science 318:582–583.

Arrow KJ, Chenery HB, Minhas BS, Solow RM (1961) Capital-labor substitution and economic efficiency. The Review of Economics and Statistics 43:225–250.

Ayres RU, Ayres LW, Warr B (2003) Exergy, power and work in the US economy, 1900–1998. Energy 28:219–273. doi: 10.1016/S0360-5442(02)00089-0

Barreto L, Kemp R (2008) Inclusion of technology diffusion in energy-systems models: some gaps and needs. Journal of Cleaner Production 16:S95–S101. doi: 10.1016/j.jclepro.2007.10.008

Ben Elliston, MacGill I, Diesendorf M (2013) Least cost 100% renewable electricity scenarios in the Australian National Electricity Market. Energy Policy 59:270–282. doi: 10.1016/j.enpol.2013.03.038

Bosetti V, Carraro C, Galeotti M, Massetti E (2006) WITCH: A World Induced Technical Change Hybrid Model. The Energy Journal 27:13–37.

Bosetti V, Marangoni G, Borgonovo E, et al (2015) Sensitivity to energy technology costs: A multi-model comparison analysis. Energy Policy 80:244–263. doi: 10.1016/j.enpol.2014.12.012

Breyer C, Bogdanov D, Gulagi A, et al (2017) On the role of solar photovoltaics in global energy transition scenarios. Prog Photovolt: Res Appl 6:545–20. doi: 10.1002/pip.2885

Cai Y, Newth D, Finnigan J, Gunasekera D (2015) A hybrid energy-economy model for global integrated assessment of climate change, carbon mitigation and energy transformation. Applied Energy 148:381–395. doi: 10.1016/j.apenergy.2015.03.106

Cheng AYC (2005) Economic Modeling of Intermittency in Wind Power Generation. Massachusetts Institute of Technology

Christensen CM (1997) The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail. Harvard Business School Press, Boston, MA

Fouquet R (2010) The slow search for solutions: Lessons from historical energy transitions by sector and service. Energy Policy 38:6586–6596. doi: 10.1016/j.enpol.2010.06.029

Frei CW, Haldi PA, G Sarlos (2003) Dynamic formulation of a top-down and bottom-up merging energy policy model. Energy Policy 31:1017–1031. doi: 10.1016/s0301-4215(02)00170-2

Gohin A, Hertel TW (2003) A Note on the CES Functional Form and Its Use in the GTAP Model. Center for Global Trade Analysis, Purdue University

Görig M, Breyer C (2016) Energy learning curves of PV systems. Environ Prog Sustainable Energy 35:914–923. doi: 10.1002/ep.12340

Grubler A, Nakicenovic N, Victor DG (1999) Dynamics of energy technologies and global change. Energy Policy 27:247–280.

Hanoch G (1971) CRESH production functions. Econometrica 39:695–712.

Jacobson MZ, Delucchi MA (2011) Providing all global energy with wind, water, and solar power, Part I: Technologies, energy resources, quantities and areas of infrastructure, and materials. Energy Policy 39:1154–1169. doi: 10.1016/j.enpol.2010.11.040

Joskow PL (2011) Comparing the costs of intermittent and dispatchable electricity generating technologies. American Economic Review 238–241.

Komiyama R, Fujii Y (2015) Long-term scenario analysis of nuclear energy and variable renewables in Japan's power generation mix considering flexible power resources. Energy Policy 83:169–184. doi: 10.1016/j.enpol.2015.04.005

Kriegler E, Petermann N, Krey V, et al (2015) Diagnostic indicators for integrated assessment models of climate policy. Technological Forecasting & Social Change 90:45–61. doi: 10.1016/j.techfore.2013.09.020

Kuik O, Brander L, Tol RSJ (2009) Marginal abatement costs of greenhouse gas emissions: A meta-analysis. Energy Policy 37:1395–1403. doi: 10.1016/j.enpol.2008.11.040

Luderer G, Bosetti V, Jakob M, et al (2011) The economics of decarbonizing the energy system—results and insights from the RECIPE model intercomparison. Climatic Change 114:9–37. doi: 10.1007/s10584-011-0105-x

Luderer G, Krey V, Calvin K, et al (2013) The role of renewable energy in climate stabilization: results from the EMF27 scenarios. Climatic Change 123:427–441. doi: 10.1007/s10584-013-0924-z

Lund H, Mathiesen BV (2009) Energy system analysis of 100% renewable energy systems—The case of Denmark in years 2030 and 2050. Energy 34:524–531. doi: 10.1016/j.energy.2008.04.003

Maggio G, Cacciola G (2009) A variant of the Hubbert curve for world oil production forecasts. Energy Policy 37:4761–4770. doi: 10.1016/j.enpol.2009.06.053

Mansfield E (1961) Technical change and the rate of imitation. Econometrica 29:741–766.

Marchetti C, Nakićenović N (1979) The Dynamics of Energy Systems and the Logistic Substitution Model. International Institute for Applied Systems Analysis, Laxenburg, Austria

Mercure JF, Pollitt H, Chewpreecha U, et al (2014) The dynamics of technology diffusion and the impacts of climate policy instruments in the decarbonisation of the global electricity sector. Energy Policy 73:686–700. doi: 10.1016/j.enpol.2014.06.029

Nagy B, Farmer JD, Bui QM, Trancik JE (2013) Statistical Basis for Predicting Technological Progress. PLoS ONE 8:e52669. doi: 10.1371/journal.pone.0052669.g001

Nordhaus WD (1992) An optimal transition path for controlling greenhouse gases. Science 258:1315–1319.

Norton JA, Bass FM (1987) A diffusion theory model of adoption and substitution for successive generations of high-technology products. Management science 33:1069.

Nykvist B, Nilsson M (2015) Rapidly falling costs of battery packs for electric vehicles. Nature Climate change 5:329–332. doi: 10.1038/nclimate2564

Paltsev S, reilly J, Jacoby H, et al (2005) The MIT Emissions Prediction and Policy Analysis (EPPA) Model: Version 4. MIT Joint Program on the Science and Policy of Global Change

Pearce J, Weyant JP (2008) Insights Not Numbers: The Appropriate Use of Economic Models. Pew Center for Global Climate Change

Pietzcker RC, Ueckerdt F, Carrara S, et al (2017) System integration of wind and solar power in integrated assessment models: A cross-model evaluation of new approaches. Energy Economics 1–17. doi: 10.1016/j.eneco.2016.11.018

Pindyck RS (2013) Climate Change Policy: What Do the Models Tell Us? Journal of Economic Literature 51:860–872. doi: 10.1257/jel.51.3.860

Pleßmann G, Erdmann M, Hlusiak M, Breyer C (2014) Global energy storage demand for a 100% renewable electricity supply. Energy Procedia 46:22–31. doi: 10.1016/j.egypro.2014.01.154

Porter ME, Van der Linde C (1995) Toward a new conception of the environment-competitiveness relationship. Journal of Economic Perspectives 9:97–118. doi: 10.1257/jep.9.4.97

Rosen RA, Guenther E (2015) The economics of mitigating climate change: What can we know? Technological Forecasting & Social Change 91:93–106. doi: 10.1016/j.techfore.2014.01.013

Schumpeter JA (1943) Capitalism, Socialism & Democracy, 5 edn. Routledge

Sgouridis S, Bonnefoy PA, Hansman RJ (2010) Air transportation in a carbon constrained world: Long-term dynamics of policies and strategies for mitigating the carbon footprint of commercial aviation. TRANSPORTATION RESEARCH PART A 1–15. doi: 10.1016/j.tra.2010.03.019

Sgouridis S, Csala D, Bardi U (2016) The Sower's Way: Quantifying the Narrowing Net-Energy Pathways to a Global Energy Transition. (Accepted Manuscript). Environmental Research Letters 1–14.

Singer S (2010) The energy report: 100% renewable energy by 2050.

Truong PT (2009) Constant elasticity of substitution (CES) production function can greatly overestimate the economic costs of climate policies. Institute of Transport and Logistics Studies, University of Sydney

Trutnevyte E (2016) Does cost optimization approximate the real-world energy transition? Energy 106:182–193. doi: 10.1016/j.energy.2016.03.038

Ueckerdt F, Brecha R, Luderer G, et al (2015) Representing power sector variability and the integration of variable renewables in long-term energy-economy models using residual load duration curves. Energy 90:1799–1814. doi: 10.1016/j.energy.2015.07.006

van Vuuren DP, Stehfest E, Elzen den MGJ, et al (2011) RCP2.6: exploring the possibility to keep global mean temperature increase below 2°C. Climatic Change 109:95–116. doi: 10.1007/s10584-011-0152-3

Weitzman ML (2009) On Modeling and Interpreting the Economics of Catastrophic Climate Change. Review of Economics and Statistics 91:1–19. doi: 10.1257/jel.45.3.703

Wilkerson JT, Leibowicz BD, Turner DD, Weyant JP (2015) Comparison of integrated assessment models: Carbon price impacts on U.S. energy. Energy Policy 76:18–31. doi: 10.1016/j.enpol.2014.10.011