Elasticity formulations: theory and practice in System Dynamics

By Max Kleemann, Kaveh Dianati, Matteo Pedercini (Millennium Institute)

<u>*Abstract:*</u>

The adequate representation of non-linear relationships (NLRs) in SD models is often challenging. The implications of using alternative formulations for formalizing NLRs can affect in a significant way both the efficiency of the modeling and calibration process and the accuracy of the results produced. The use of elasticity formulations to represent such relationships is not new to SD as they have been used from the inception of the field, including by Jay Forrester (1956, published in 2003) and are covered briefly in Sterman's popular SD textbook (2000). Elasticity formulations can be very effective and efficient in some circumstances, and their low occurrence in the system dynamics literature may indicate a lack of knowledge of their characteristics and usability that favor the more popular alternative: table functions. Our review indicates that sufficient guidance about how to use elasticity formulations in SD models and about the conditions under which this appears justified/advantageous has thus far been lacking. This paper attempts to address this gap by a theoretical discussion about the properties of elasticity formulations, a brief literature review including criticism of their use, prerequisites for use (including testing), and a comparison of the approach to table functions in terms of advantages and limitations and in terms of practical modeling examples.

1. Introduction

System dynamics modelling, as a way of representing a problematic development as a set of mathematical equations, contains an element of art as well as science: Often there is more than one way to represent the same real-world relationship(s) between two or more variables. More often than not, such relationships are non-linear[1]. While they can in some cases be sufficiently well approximated by linear formulations, such practice can lead to relevant inaccuracies, especially when a system moves beyond its normal operating space. In order to respect non-linearity, in current SD modeling practice, non-linear relationships are typically represented by way of table functions (also known as graphical functions when displayed in X-Y-graph form). These come with advantages and caveats. Advantages of using table functions include that they support the introduction of just about any functional form and do not require particular math skills to be implemented. However, if the functional relationship is known to be monotonic[2] and meets certain other requirements (as discussed below), a constant elasticity formulation can be a simpler and more effective alternative.

The concept of "elasticity" refers to a characteristic of the relationship between two variables, one (Variable X) affecting the other (Variable Y): elasticity is the ratio between the proportional change in Y to the proportional change in X. In other words, elasticity measures how much one variable changes as a consequence of the change in the other. When this ratio tends to be constant over the range of values that X can take, then the elasticity gives us, in one number, an immediate

---

[1] A non-linear relationship is one where the change in output is not proportional to the change in input.

[2] Even for graphical functions, monotony is good modeling practice because, as Sterman (2000, 14.4.3, p.577) points out, non-monotonic table functions imply ambiguous link polarity and thus indicate „multiple causal pathways", that should rather be modeled separately

measure of the strength and the direction of reactions of the relationship. Such a concept could thus be easily used to represent a broad variety of relationships in system dynamics modeling.

However, our literature review indicates that formulations that are constructed around the concept of constant elasticity (constant elasticity formulations or CEF in the remainder of the document) may be scarcely adopted in system dynamics modeling: Searching for 'elasticity' or 'elasticities' in all System Dynamics Review (SDR) volumes rendered 77 results. After excluding the hits that include 'eigenvalue elasticity analysis', which is not our focus here, 48 articles remain. The vast majority of these papers use the concept of elasticity to directly represent existing notions in mainstream economics (e.g., elasticities of demand/supply to price/income) without making any direct references to the notion itself. It should be noted though, that this search was limited to SDR, whereas SD-based articles are also published elsewhere and thus we cannot rule out a higher use of elasticities there.

Perhaps the earliest reference to the concept of elasticity within the SD literature was before the formal conception of the field. In a 1956 manuscript, where Jay Forrester (1956 published in 2003) gives a glimpse into his earliest ideas leading to the genesis of the field, elasticity in systems, along with delays, momentum, stocks, and accelerations, is mentioned as one of 'the fundamental quantities which differential equations have been developed to describe' and as one of 'the quantities which we wish to use for our underlying principles of the economic world.' Here, however, elasticity is referred to as a property of systems rather than a technique used in modelling non-linear relationships.

Some of the early references to elasticities as model variables are within an econometric context (Nathan Forrester 1987, Arif and Saeed 1989, Wirl 1991). Andersen (1990) criticises the use of elasticities in econometric models for being not readily interpretable by the layperson, and for requiring non-trivial mathematical computations and interpretations so they can be turned into estimates that are informative to lay intuition. Within the SD literature, both constant (Kampmann and Sterman 2014, Pierson and Sterman 2013) and dynamic (Brady 2009, Jung and Strohhecker 2009) elasticities have been used. However, a salient theme where elasticities are mentioned is criticising constant elasticities and advocating for dynamic and endogenous elasticities (van Ackere and Smith 1999, Moxnes 1990, Ulli-Beer et al. 2010). Moxnes (1990), for instance, presents a model of fuel substitution for electricity production where he uses endogenously driven elasticities which, he asserts, make his model 'more closely related to reality than the ordinary constant-elasticity models.' Van Ackere and Smith (1999) build an SD model of the UK's National Health Service waiting list, taking an economics perspective. They similarly consider constant elasticities as a limitation and introduce an elasticity of demand with respect to waiting time which varies with waiting time. Furthermore, a distinction is often made between short-run and long-run elasticities (Cavana and Clifford 2006, Inman et al 2020). Inman et al (2020) present a technique for generating supply and demand curves from system dynamics models and show results which indicate that the elasticity of supply with respect to price can be time dependent, such that supply is relatively inelastic in the short term and more elastic in the longer term, reflecting inherent supply-side delays commonly observed.

Some authors have argued in support of the use of elasticities for being well-known and established concepts. Van Ackere and Smith (1999) argue that 'it is more appropriate to use a secure econometric estimate, rather than to attempt to formulate a more speculative behavioural structure.' Inman et al (2020) similarly argue for the benefit of the use of elasticities for being 'more commonly accepted terminology (particularly with respect to the *lingua franca* of supply, demand, and elasticities) employed by many stakeholders who are familiar with basic microeconomic principles.'

Beyond the limited mention of elasticity formulations indicated above, which mostly refer to their use in the context of economic systems, our review indicates that this potentially useful tool is used very little in our field. While we recognize that many historical, cultural, and factors might explain the underutilization of such a practical tool, we believe that a proper description of CE formulations and their properties is necessary in order to facilitate their use. In the remainder of the paper we discuss the underlying mathematical concepts to such formulation; when and how its use is appropriate; as well as what its main advantages, critiques and limitations are. We will also provide a few examples of applications.

2. Constant Elasticity Formulations and their use

From the general definition of elasticities, in this section we derive relevant expressions that can be used to (A) estimate the value of an elasticity in a dataset (or existing table function); and (B) implement a constant elasticity formulation in a model (Equations 12 and 9, respectively).

By definition, an elasticity is the ratio of proportional change in one variable relative to the proportional change in another variable. In simple terms, the elasticity (ε) of Y to change in X can be expressed as:

$$\varepsilon = \frac{\frac{Y_1 - Y_2}{Y_1}}{\frac{X_1 - X_2}{X_1}} \qquad \textit{Equation 1}$$

When applied to discrete points (e.g. on dataset), the measurement of elasticity is carried out on an arc, and thus such estimation would provide the average elasticity over an arc, or the "arc elasticity". Instead, when Equation 1 above is applied to very small marginal increments, we calculate the elasticity for a specific point on a curve, i.e. the "point elasticity".

For infinitely small increments dx and dy we can write the same equation as:

$$\varepsilon = \frac{\frac{dy}{y}}{\frac{dx}{x}} \qquad \textit{Equation 2}$$

Which is a *point elasticity*. This can be rearranged as:

$$\frac{dy}{y} = \varepsilon \cdot \frac{dx}{x} \qquad \textit{Equation 3}$$

Now if we solve this equation for y, using integration and assuming constant elasticity:

$$\int \frac{1}{y} \cdot dy = \varepsilon \cdot \int \frac{1}{x} \cdot dx \qquad \textit{Equation 4}$$

Which results in:

$$ln\,(y) = \varepsilon \cdot ln\,(x) + ln\,(a) = ln\,(a \cdot x^{\varepsilon}) \qquad \textit{Equation 5}$$

Where $ln\,(a)$ is a constant term added as a result of the indefinite integration. This gives our dependent variable y as a function of x:

$$y = a \cdot x^{\varepsilon} \qquad \textit{Equation 6}$$

Now, if we know the value of y at a certain reference point X_r (i.e. Y_r) , we can obtain the constant a, as follows:

$$Y_r = a \cdot X_r^{\varepsilon} \Rightarrow a = \frac{Y_r}{X_r^{\varepsilon}} \qquad \textit{Equation 7}$$

Next, if we replace this known a in Eq. 6, our dependent 'effect' variable is given as follows, as a function of the cause variable x, a constant arc elasticity ε, and the known point (X_r, Y_r):

$$y = \frac{Y_r}{X_r^{\varepsilon}} \cdot x^{\varepsilon} \qquad \textit{Equation 8}$$

Or, in more simplified terms:

$$y = Y_r \cdot \left(\frac{x}{X_r}\right)^{\varepsilon} \qquad \textbf{\textit{Equation 9}}$$

In other words, the ‘effect’ variable y, equals its known value at a reference point Y_r (often the initial point in simulation), times the value of the cause variable relative to its value at the reference point $\frac{x}{X_r}$, to the power of the constant elasticity ε. This CEF equation form can be used in system dynamics models to capture certain non-linear relationships where a power function form appears appropriate, as will be further discussed below.

A power function is also reasonable because it is known to be isoelastic, i.e. its mathematical property is that it has a constant elasticity if the exponent is constant[3].

Equation 9 can be used to represent non-linear relationships in SD models using an elasticity, where *x* is the input variable and *y* is the output variable.

In order to test to what degree a table function or other functional relationship meets the important prerequisite of constant elasticity we can solve equation 9 for ε (equation 12). This is achieved by first rearranging:

$$\frac{y}{Y_r} = \left(\frac{x}{X_r}\right)^{\varepsilon} \qquad \textit{Equation 10}$$

And then taking the natural logarithm of both sides:

$$ln\left(\frac{y}{Y_r}\right) = ln\left(\left(\frac{x}{X_r}\right)^{\varepsilon}\right) = \varepsilon \cdot ln\left(\frac{x}{X_r}\right) \qquad \textit{Equation 11}$$

Finally, solving for ε we obtain:

$$\varepsilon = \frac{ln\left(\frac{y}{Y_r}\right)}{ln\left(\frac{x}{X_r}\right)} \qquad \textbf{\textit{Equation 12}}$$

Equation 12 can be used to estimate an elasticity to be used in place of a different model expression, (e.g. table function or equation), as will be shown in the example further below.

In contrast to equation 2 (*point elasticity*) equation 12 is an *arc elasticity* that is estimated based on two data points (the reference point and the point that is changing). This definition is more limited in application than the point elasticity in that its application in models (CEF as in equation 9) hinges on the assumption of constant elasticity. If the non-linear relationship in reality did not have a constant elasticity, estimating the elasticity of different sets of points would yield different values for the elasticity. This fact allows checking whether the elasticity implied by the table function (or an algebraic formulation) is constant or not by depicting it in a graph over time. This is useful since constancy of an elasticity is a prerequisite for using Equation 9 instead of the table function in the model.

Equation 9 can be extended to include multiple inputs, typically in a multiplicative form[4], in which case their relationship can be represented as in Sterman (2000 p. 526-527): the elasticities ε_i of the affected variable Y to several factors X_i influencing Y are:

$$y = y_r \cdot \prod_{i=1}^{n}\left(\left(\frac{x_i}{x_{ir}}\right)^{\varepsilon_i}\right) \qquad \textit{Equation 13}$$

As the equation indicates, this reference point has to be conceptually the same for all x_i , e.g. referring to the value of x for the same reference year. Typically, the initial year of the simulation is used, although, in special cases another prominent reference point may suggest itself or even be needed. One example of the latter is when the variable that is calculated is not defined or has zero value before a certain point in time (after the simulation start).

As for the wording of the variable names we have found the following convention to be useful in modelling practice: Elasticity variable name: “Elasticity *of* y/y_r *to* x_i/x_r”, e.g.: “Elasticity **of** *relative*

[3] https://en.wikipedia.org/wiki/Isoelastic_function

[4] In some cases, an additive form can also be appropriate when effects of xi on y are expected to be highly independent of each other in reality, (see Sterman 2000, p.528)

road construction cost per km **to** *relative road density*". Note that the term "relative" in the input and output variables *road density* and *road construction cost per km* respectively, indicates that the normalised form is used (with respect to the reference value).

To get an idea as to which functional forms can be represented with this formulation, consider Figure 1.

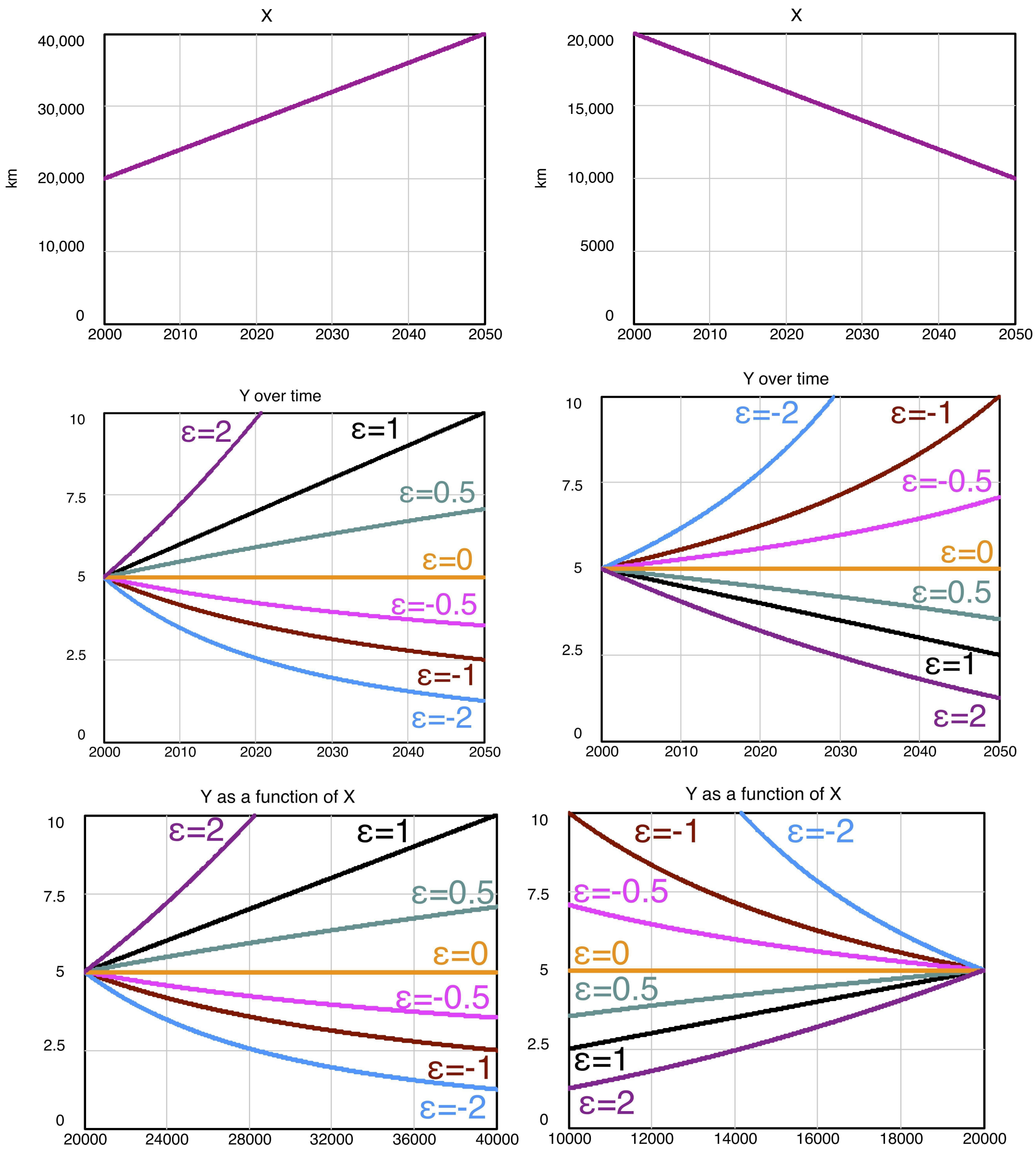

Figure 1: Top: two kinds of inputs a) decreasing by half b) increasing two-fold over the time horizon. Second row: output variable over time for inputs a (left) and b (right), each for different elasticities Third row: Functional form of the output variable for inputs a (left) and b (right), each for different elasticities .

If for example, the input variable doubles over a certain amount of time, then an elasticity >1 implies that the output variable will more than double. It should be kept in mind that the example above uses a linearly in-/decreasing input, so that the response function could look different for more generalized inputs.

The functional forms in Figure 1 can be summed up in the following rules:

- $|\varepsilon|<1$: underproportional reaction of the output variable
- $|\varepsilon|=1$: proportional reaction of the output variable
- $|\varepsilon|>1$: overproportional reaction of the output variable

- $\varepsilon=0$: no reaction
- $\varepsilon>0$: reaction of the output variable in the same direction as the change of the input variable
- $\varepsilon<0$: reaction of the output variable in the opposite direction as the change of the input variable

The simplicity of constant elasticity formulations, and the broad range of response curves they can generate, make them a powerful modeling tool. Nevertheless, they should only be used when the relationship to be represented fulfills certain criteria, as discussed below.

Prerequisites for using elasticity formulations

A. The underlying relationship has to be monotonic, that is, a positive change in a given direction of X is always associated with a change in a certain direction of Y, i.e. the derivative of the function is not changing its sign.
   This implies that a contraindication for elasticity use would e.g. be a variable that reacts strongest for intermediate values of the input, but weakly to both low and high value of the input.
B. The relationship must have an elasticity that is close to constant. If the elasticity was changing over time, that could cause sudden jumps in output
C. The relationship does not involve sudden jumps, e.g. sharp threshold values

The underlying causal relationship should be such that if the input variable does not change neither should the output (ceteris paribus).

Further implementation-aspects of elasticity formulations:

A negative input (relative x) limits elasticities to integers for simulation software that can typical not handle complex numbers. Note that this is only a problem for inputs, where a negative value can occur by definition in reality, e.g. temperature in Celsius/Fahrenheit, bank account balance, electric charge. Often times however, the problem can be avoided through a modified variable definition: temperature can be measured in Kelvin instead, *Trust* can be defined as ranging from 0 to 1 instead of from -1 to 1 (complete distrust to complete trust, respectively for each definition), a bank account balance may be divided into two variables one for debt and one for money actually owned (which may make sense anyway because human decisions may react very differently for changes in debt as compared to changes in money owned…).

Alternatively, a workaround can be used where inputs can go negative by formulating the dependent variable as a stock and letting the independent variable(s) drive the rate of changes in the stock. In this way, the derivative (with respect to time) of the elasticity equation (Equation 9) is used, which allows the elasticity to be multiplied by relative changes in x, rather than used in a power function form, as used for example in Dianati (2022, p184-185).

Functional forms that have a known non-zero roof/floor (saturation effects) for the affected variable cannot readily be represented by elasticities. At least strictly speaking, the monotonically in- or decreasing property of the effect modeled by an elasticity formulation will surpass any roof when time goes to infinity. However, since elasticities with an absolute value between 0 and 1 have diminishing returns as a property, properly chosen values can still approximate such functional forms to some degree within a given range of input values and simulation time.

As indicated above, the decision whether or not an elasticity formulation can be used is not dichotomous, some cases stretch the concept more than others, sometimes the functional form that can be generated by an elasticity formulation is a more or less rough approximation of what the true functional form probably looks like. Sometimes additional equation syntax can enable using an elasticity formulation that would otherwise not be applicable.

Mandatory normalization

One important difference between using elasticities and using table functions is that normalization is *necessary* when using elasticities, while for table functions it is only *recommended* to normalize inputs and conceptualize the table function so that it produces a normalized output. This recommendation is not always followed and using dimensioned in- or outputs does not necessarily lead to inconsistencies if applied with care. In line with that reasoning, modeling softwares do not produce errors if dimensioned arguments are used as input to table functions or dimensioned outputs are produced by them (some softwares e.g. Vensim, do give out *warnings* but not *errors* in such cases though).

When using elasticities normalization is necessary because the formulation can only be expected to produce useful outputs if the inputs are normalized (with a reference value, typically the initial value) and then using the reference/initial value of the output to multiply the normalized output with (see equations 9 and 12).

Cases where the input is undefined for a while after the onset of the simulation

There are cases where the input of the formulation is not defined at the onset of the simulation. Consider for example we desire to model the effect of a pesticide on the mortality of a certain species, but the pesticide has come into use only a certain time after the onset of the simulation. This implies however, that if the initial value of the input is zero, it cannot be used for normalization purposes. Hence, in such cases the value of the input variable at some other reference time must be used instead. In the example above we could use for normalization the amount of the pesticide that was used in a period where a solid estimation was made, and the mortality at that point in time for the output.

Although not always recommended, the first available value for the input variable can be used for normalization. The benefit of such practice is that it can be automated in some modeling softwares. Different applications may differ in their abilities to implement this; in Vensim a *sample-if-true* formulation can be used, where instead of an *initial* variable the *first value*-variable used for initialization jumps from zero to the first appearance of the value and then stays there for the rest of the simulation. In applications where this is not available, if-then-else formulations could be used together with a stock that serves as memory of the value at the previous time step. However, one important limitation of using the first available value for normalization is that some process (e.g. economic production) that might exhibit constant elasticity in the long-run, might still behave differently at the very outset of the process (e.g. because production technology is not mature yet). It is therefore better to choose as anchor for normalization a later value of the input.

Issues of scale

Because constant elasticity formulations do not describe the actual mechanics that lead from a change in input to a change in output, but aggregates it into a single equation, it can be argued that the evaluation of whether or not using an elasticity formulation is appropriate depends to an extent on the choice of level of aggregation and model boundary. However, an elasticity formulation can also be used to represent micro-scale causal relationships in some cases. Mathematically, the requirement is a constant ratio of delta input to delta output, irrespective of aggregation level. For example, if a model is built solely to represent and replicate developments in fertility rate, perhaps using an elasticity for the effect of e.g. income or education might not be 'operational' and transparent enough for this particular purpose, not least because you are losing the dynamics of how the drivers impact fertility rate. If however, the goal is a model of national development on various fronts, at the scale of T21/iSDG[5] for example, then it is not feasible to delve into the operational nitty-gritty of all mechanisms, and therefore for this case the use of

[5]https://www.isdgs.org/isdg

elasticity formulations can be a good choice, given its other advantages (ease of calibration, ease of communication to other disciplines, etc.).

Other implementation aspects

Sterman (2000 Table 14-1 p.553) argues that rigor must be applied to the formulation of table functions in terms of thinking of reference points, the functional form, thresholds etc. It is highly advisable to apply the same rigor to the use of elasticities. Most importantly, they should not simply be auto-calibrated by the modeling software with a search range from $-\infty$ to $+\infty$., but the search range should be limited to plausible ranges. It can to the very least be reasoned first about the sign and whether the absolute value of the elasticity is < or >1 (see rules derived from Figure 1 above).

Moreover, if in doubt on whether an elasticity formulation can be used instead of a table function, it is advisable to first build a table function and then assess whether an elasticity formulation could alternatively be used, e.g. by calculating the elasticity from the table function model using equation 12 as shown in the example further below. In fact, testing the model results obtained with an elasticity formulation versus those obtained using a table function is also advisable.

Also, similarly to the case of implementing table functions, the identification and quantification of elasticities can benefit from expert knowledge, to be elicited according to best practices (Sterman 2000 Table 14.5 p.585ff). In particular, expert knowledge can be used to define reasonable ranges for elasticities, in absolute terms, and also in relative terms (e.g. elasticity "a" is larger than "b"). Note that auto-calibration mechanisms on standard SD-software are typically not able to incorporate such conditional logic, which means that some degree of manual calibration is not only advisable but often necessary to yield meaningful results (This is also a call to software developers to alleviate this shortcoming…).

3. A Classic Example

In order to test the applicability of elasticity formulations in place of table functions, we use a well-known model: the *World-3-03* model that was used for the study *Limits to Growth - the 30 year update* (Meadows et al. 2004). This also has the advantage that the model is freely and easily available as part of the documentation of the Vensim-software. We first review an example of a table function within that model that can very well be replaced by an elasticity formulation; and then cover some examples where this does not seem appropriate without further structural changes.

We will show that an elasticity can appropriately replace the table function *completed multiplier from perceived lifetime table* shown in the Figure 2 below (original formulation of *World-3-03*).

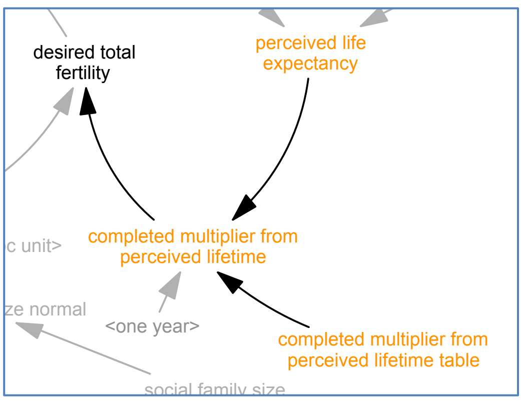

Figure 2: Table function formulation of *completed multiplier from perceived lifetime table* in World-3-03 (color emphasis added)

Figure 3 below shows the quantification and functional form of this table function in World-3-03. It can be seen that the slope is downward and the curvature is convex all along the ranges given by the table function (input *perceived life expectancy*: [0,80] output *completed multiplier from perceived lifetime*: [1,3]. The functional form is close to what can be produced by an elasticity formulation (see Fig. 1). It is thus a good candidate for replacement with an elasticity formulation.

Input	Output
0	3
10	2.1
20	1.6
30	1.4
40	1.3
50	1.2
60	1.1
70	1.05
80	1

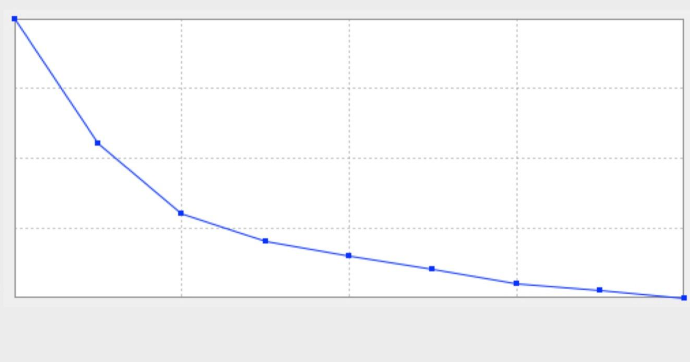

Figure 3: table function *completed multiplier from perceived lifetime table* in World-3-03

Figure 4 below shows a modified structure using an *elasticity of fertility multiplier to life expectancy*. It can be seen that the reference points used are the initial values of the simulation (i.e. year 1900 here).

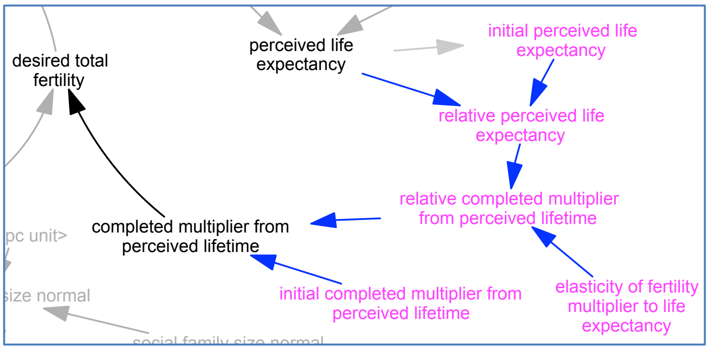

Figure 4: modified structure in World-3-03 using an *elasticity of fertility multiplier to life expectancy* and *initial* and *relative* versions of in- and outputs based on initial values of the in- and outputs

Of course, the original model can also be used to calculate an elasticity using equation 12. In Figure 5 below it can be seen that the elasticity derived from the table-function version of the model (yellow) is nearly constant, which is an important prerequisite of using an elasticity in place of a table function.

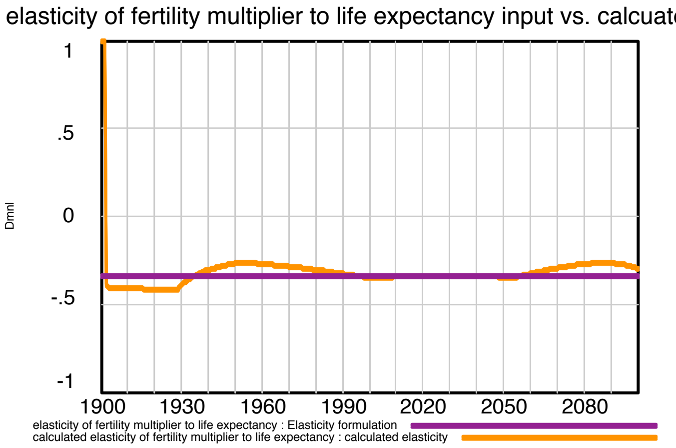

Figure 5: Comparison of elasticity derived from the table-function version of the model (dark yellow) and the constant elasticity used in the modified model with an elasticity formulation (purple). The initial value of 1 can be ignored[6]

The value used for the elasticity in the modified model version (purple line in figure above) could be derived from the calculated one. However, in real world applications, an elasticity value will often be derived without a pre-existing table function simply by calibrating it to yield a good fit with the output, for which data hopefully exists. Here, manual calibration was used to get a good fit to the output from the table function using the *synthesim* feature in Vensim (elasticity of -0.3375). The resulting fit of the *completed multiplier from perceived lifetime* appears more than satisfactory (see Figure 6 below purple being the elasticity formulation and dark yellow being a simulation resulting from the original table function).

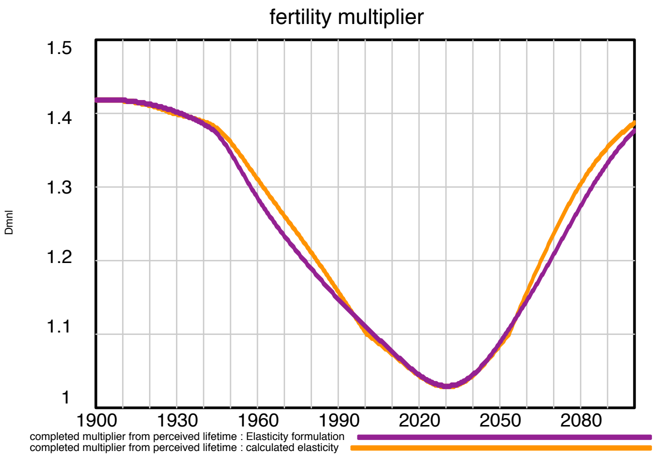

Figure 6: Comparison of the development of the fertility multiplier in the original model with table function (dark yellow) and the modified model with the elasticity formulation (purple)

Figure 7 below shows that the fit is even better for the next variable in the causal chain *total fertility*.

[6] The elasticity has been set to an arbitrary value of 1 when both the relative input and output are =0 , i.e. initially, because otherwise the simulation stops before it starts because of an error due to a division by 0

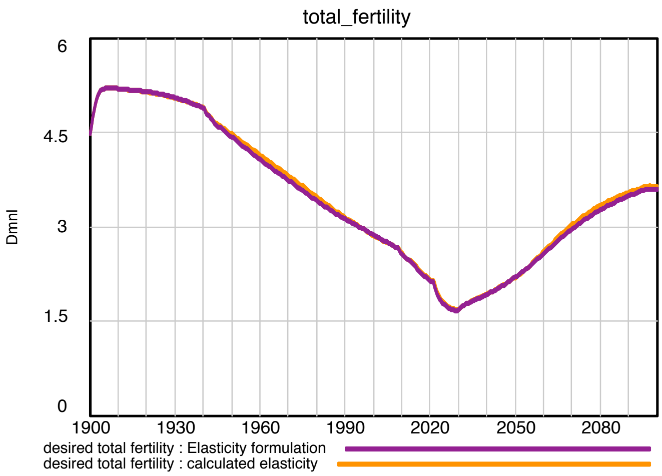

Figure 7: Comparison of the development of the total fertility in the original model with table function (dark yellow) and the modified model with the elasticity formulation (purple)

It was noted before that in some cases other reference points besides the initial values may be used. In this example a meaningful reference point would be the threshold life expectancy (80 years) at or above which *total fertility* is no longer higher than *desired completed family size*, in other words, at or above this age realized family size is no longer adjusted upward due to child mortality (*reference life expectancy without influence on total fertility* in the figure below).

Hence the elasticity formulation could also be based upon this reference point as can be seen in figure 8 below. Note that in this case the reference value for the output *completed multiplier from perceived lifetime at reference life expectancy* is naturally =1.

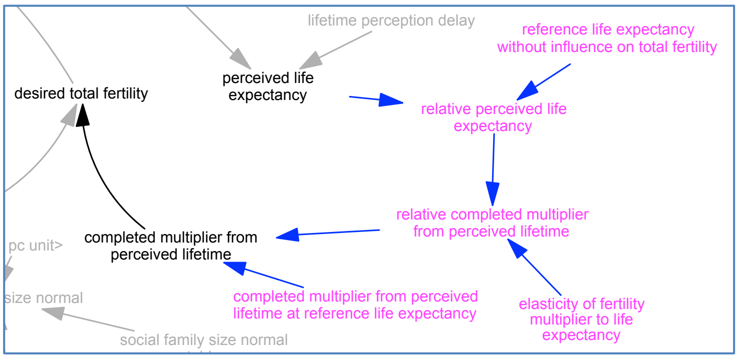

Figure 8: modified structure in World-3-03 using an elasticity of fertility multiplier to life expectancy and initial and relative versions of in- and outputs based on a reference (saturation) life expectancy

A reasonable fit could be obtained using an elasticity of -0.35 as can be seen in figure 9 below.

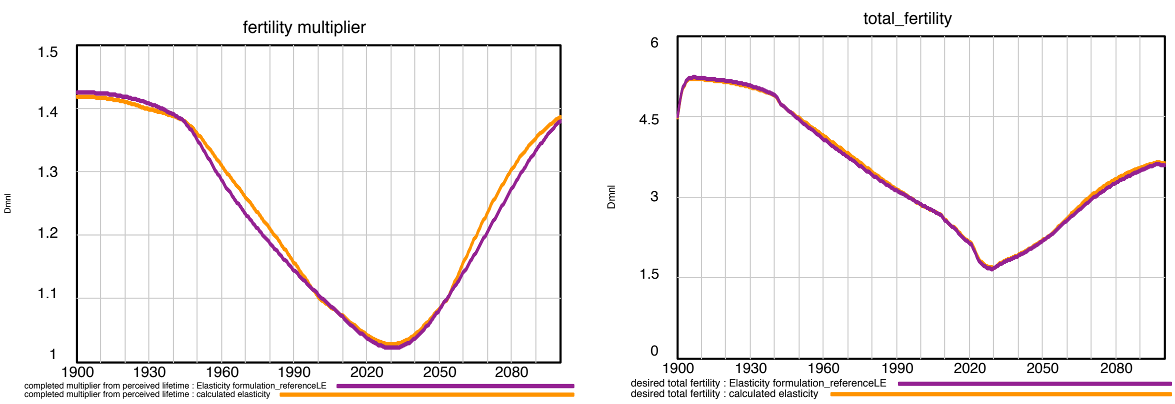

Figure 9: Comparison of the development of the fertility multiplier (left) and of the total fertility (right) in the original model with table function (dark yellow) and the modified model with the elasticity formulation (purple) based on a reference (saturation) life expectancy

However, when using a table-function it is important to also consider how the model will behave beyond the range of its definition. In this case beyond the reference age of 80 years, Vensim[7] will simply keep the last value (=1) constant. For this case that is actually meaningful because the underlying assumption as stated above is that at <u>or beyond</u> this reference age, parents will get more children just because of child mortality (which is implicitly assumed to drop with increasing life expectancy).

In the standard-run of Limits to Growth World-3-03, the life expectancy of 80 years is never surpassed, because it is an overshoot-and-collapse scenario. Hence this run does not allow for comparing the two different formulations with respect to model behavior beyond the reference point. However, the W303S13 Scenario[8] does exhibit life expectancy surpassing 80 years. Using this scenario, figure 10 below examines how the completed multiplier from expected lifetime develops beyond the range of the table function, i.e. beyond the reference point of 80 years for this scenario: The dark yellow curve represents the simulation with the table function model whereas the purple curve shows the simulation with the elasticity formulation. It can be seen that the latter falls below 1, which is not meaningful with respect to the above-mentioned assumption that beyond a certain reference age, child mortality is no longer an issue with respect to family size.

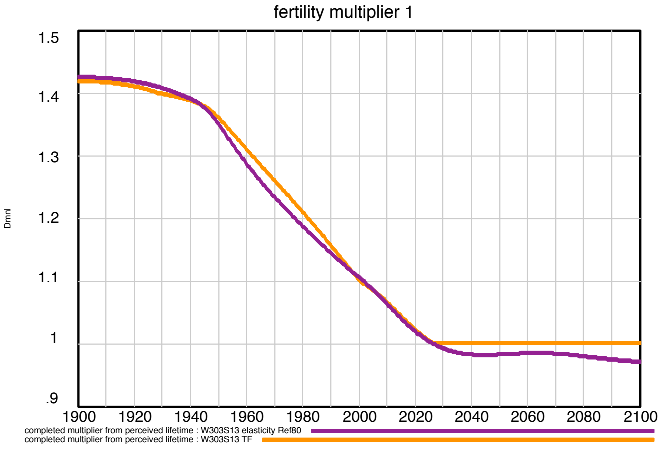

Figure 10: Comparison of the development of the fertility multiplier in the original model with table function (dark yellow) and the modified model with the elasticity formulation (purple) in the W303S13 Scenario

However, one should not take the rest of the model for granted when doing this sort of structural validity testing. We should ask ourselves: beyond 80 years: will there really be no influence at all of child mortality on total fertility, or will that influence more likely just get less and less? When considering family planning only, the assumption that below a certain child mortality, parents would no longer consider having extra children as "backup" (sorry for the formulation) seems justified at first sight. But does that really apply to *ALL* parents? Or could there be a minority who would still consider this (e.g. because they are anxious by nature, live in more dangerous circumstances than other families etc.)? In addition, total fertility could differ from desired family size not only because of family planning but also *as a reaction to* child mortality. Even under medical conditions much better than today, some children may still die, because of accidents (incl. those caused by dangerous hobbies, drugs…) or because of a few remaining diseases that can still not be cured even in a very optimistic future. Hence, if what is in the model *desired completed family size*, was redefined as *<u>reference</u> desired completed family size* and this was defined to be at a life expectancy of 80 years, there are arguments in favor of an elasticity formulation to yield results closer to real-world conditions than the table function could!

---

[7] While some modeling softwares do allow alternative assumptions beyond the range of a table function, constancy is always an option and typically the default assumption

[8] This scenario comes with the model as part of the Vensim documentation, but is not equivalent to any scenario in the book.

Even if one was to disagree with the above-mentioned argumentation of an influence of life expectancy beyond 80years on *desired completed family size*, it also matters to what degree the direct output of the elasticity formulation actually makes a difference on the rest of the model. The impact is actually very low on *desired total fertility* as can be seen in figure 11.

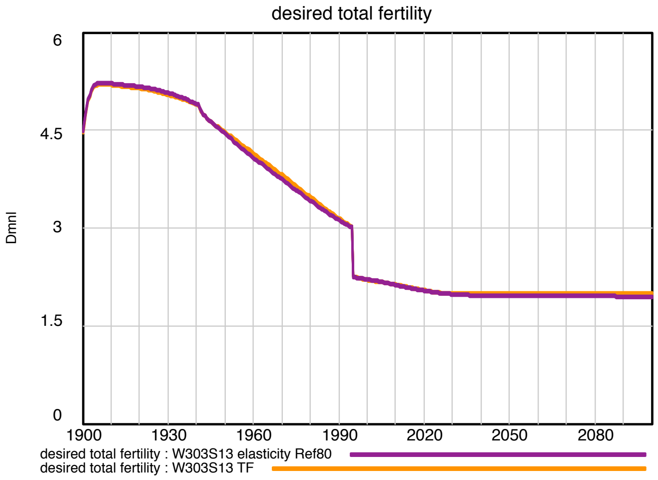

Figure 11: Comparison of the development of total fertility in the original model with table function (dark yellow) and the modified model with the elasticity formulation (purple) in the W303S13 Scenario

In addition, the elasticity formulation could be forced to behave the same way as the table function beyond a *desired life expectancy* of 80 years, by simply using an if-then-else formulation that uses the elasticity formulation below 80 years[9] and otherwise is =1. More sophisticated equation terms could probably be used to smooth the transition between these two ranges of the input variable if desired. This is a simple example of how an elasticity formulation can be used if augmented by other equation syntax.

Next, we look at an example where an elasticity cannot easily replace a table function. Figure 12 shows the original structure calculating an effect of industrial production on life expectancy (*crowding multiplier from industry table*). In Figure 13, the shape of that table function can be seen that as industrial output per capita increases, *crowding* (indicating an adverse effect on life-expectancy) first decreases to even go slightly negative and then slowly increases again.

[9] In the equation it's simpler to define this using the *relative desired life expectancy* <1 vs. else

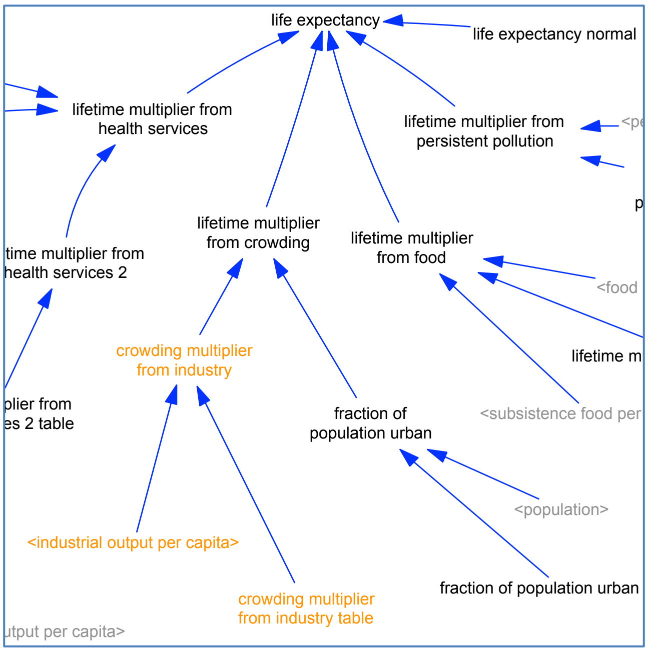

Figure 12: Model structure surrounding table function *crowding multiplier from industry table* in World-3-03 (color emphasis added)

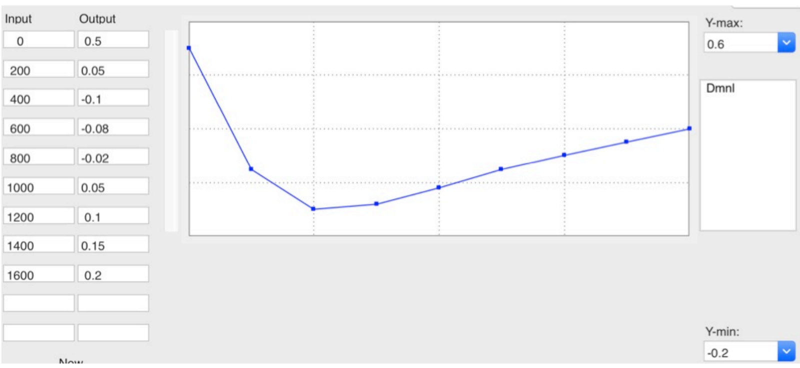

Figure 13: Table function *crowding multiplier from industry table* in World-3-03

This table function is clearly non-monotonic, thus violating one of the prerequisites for use of table functions. In addition, the input may go negative, which does not work with CEFs either (see further above).

One can calculate an elasticity from the simulation using Equation 12 using the structure added in Figure 14.

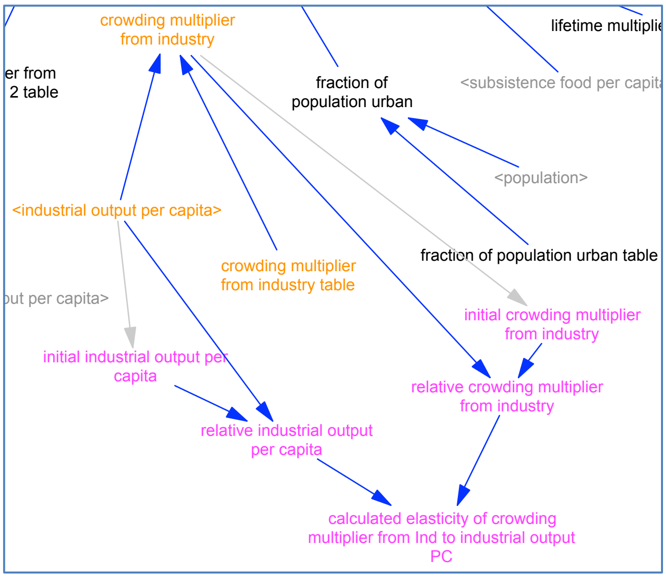

Figure 14: Model structure for calculating an *elasticity of crowding multiplier from Industry to industrial output per capita* (magenta color) added to the original structure

Implication note: Additional equation syntax is necessary to avoid the simulation from stopping prematurely. This may depend on the simulation software used, in Vensim we can use the following formulation:

calculated elasticity of crowding multiplier from Ind to industrial output PC =
IF THEN ELSE (*relative industrial output per capita* = 1 :AND: *relative crowding multiplier from industry* = 1, 1,
IF THEN ELSE(relative crowding multiplier from industry<0 :OR: relative industrial output per capita<0,-999,
LN(*relative crowding multiplier from industry*)/LN(*relative industrial output per capita*)))

Note that the number -999 is used here to indicate negative infinity.

It can be seen in Figure 15 that the elasticity first decreases from just below 0 going to negative infinity for some years (-999 in the model run actually, see note above) and then increases again to just below 0 in 2100.

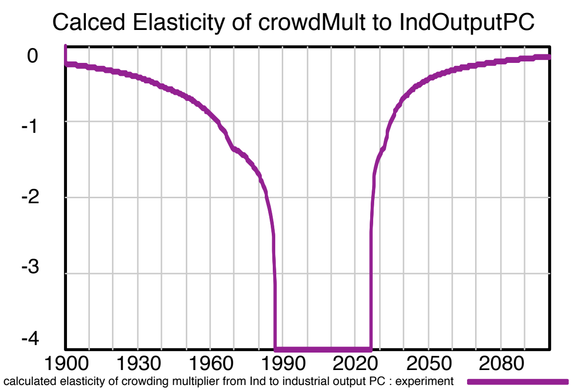

Figure 15: Development of *calculated elasticity of crowding multiplier from Industry to industrial output per capita*

However, as mentioned already in Footnote 2, it is recommended for table functions too, to be monotonic, and if they are not as in this case, there are usually different causal mechanisms expressed in the table function that should better be modeled separately. One could e.g. imagine that the authors of the model thought that industrial output per capita increasing from low levels my imply higher income with associated increases of the standard of living may also increase life expectancy. Since health services and food sufficiency is modeled separately, this multiplier is a residual category representing all other effects on life expectancy e.g. availability of refrigeration of food reducing diseases and hence mortality through avoidance of consumption of spoiled food. ideas behind the table function may be that as industrial output increases, there is at first an effect increase in life expectancy, but then at very high industrial outputs there is an effect decreasing life expectancy again. However, as GDP per capita grows to very high numbers, the authors may have assumed adverse effects besides pollution on life expectancy, .e.g. conflicts. This suggests that these effects could be modeled by at least two separate monotonic table functions or two separate CEF formulations.

4. Discussion

A few arguments have been raised in the past against the use of elasticity formulations in SD models. In a strong critique, Olaya (2015) suggests that this type of formulation stands against operational thinking, which is a cornerstone of SD. In the same vein, Eskinasi (2014) argues that "Elasticity represents the overall statistical correlation strength between economic variables. System dynamics is focused on finding structures generating such correlations. It is therefore arguably more natural to measure the elasticity of system dynamics simulation outcomes ex-post than to use elasticity as a model parameter. The former approach may be helpful in demonstrating congruence between system dynamics models and empirical findings and thus in integrating different strands of research. The latter approach is suspected to lead to unit consistency problems and methodological discussions (Eskinasi 2014, pp125-126)."

While we recognize that elasticity formulations are not always adequate to represent certain types of relationships, we believe that the above critiques are misplaced. Firstly, there is no reason why "elasticity" should necessarily refer only to economic variables: its mathematical definition is universal and not limited to a specific domain of application. Secondly, "elasticity" - intended as a property of a relationship, does not measure correlation: it is simply the ratio of change in a variable to the change in another variable. When implementing a constant elasticity formulation in a causal model (as SD models are) we associate a causal meaning to that relationship, independently on whether the elasticity was estimated using a regression model. In fact, the above arguments may also be used against table functions. They, too, can be used to depict correlations, where the detailed causal mechanisms are not known.

Rather than the generic criticism above, we find that a practical discussion of pros and cons of the use of this type of formulation in system dynamics models is more useful. The discussion that follows is based on the mathematical properties of elasticity formulations, and on practical observations from application in a variety of fields.

Advantages and disadvantages of constant elasticity formulations

While it is clear that, for a monotonic relationship that exhibits constant elasticity, using a CEF or a table function can provide very similar results, in practice there are important advantages and disadvantages to each.

Pro: Less time-consuming

A first, immediate advantage of CEF with respect to table functions is that they can be more rapidly developed, needing only one parameter to be specified, instead of a series of coordinates.

Pro: Ease of calibration

Another advantage of the strength of the relationship being represented by a single value is that it makes calibration easier and possibly more robust, as it supports the use of advanced calibration algorithms. It should be noted that calibration can act as a powerful test of model validity. As Oliva (2003, p557) points out though: "From an operational perspective, however, having a complex error function and multiple parameters to adjust makes the tractability of the errors and the diagnosis of mismatches more difficult." The latter limitation applies to table functions which consist of several values. Since using elasticities calibration, but also sensitivity analyses are much easier, it appears more likely that such rigor is actually applied when elasticities are used as compared to table functions.

Pro: Unbounded

Being defined over the whole potential range of input values (while table functions are defined by a finite set of points) there is no risk of the input falling beyond the range of values pre-set in the model.

Pro: Smoothness

Elasticity formulations clearly have the advantage of built-in steadiness. Table functions typically use single points and linear interpolation in-between. This implies a non-steady function (the points do not have a clear derivative), even if the real functional form is known to be steady. This can lead to artifacts in the simulation outcome. While this can be alleviated for table functions to a certain degree by smoothing the curve using more points, this is not practical, and if it is not supported by data it may be criticized as a pseudo-accuracy. In addition increasing the number of points in a table function exacerbates the issues around calibration and sensitivity analyses outlined above. Hence the number of points in a table function imply a trade-off that can be avoied by using elasticities.

Pro: Easier communication

Using a single value to express the strength of a relationship makes it easy to report in documentation and publications (reporting one value in text compared to reporting a table or even an x-y graph). Also, it facilitates exchange with other disciplines where such a concept is broadly used (e.g. economics).

Con: Harder communication

With an audience who is not aware of the concept of elasticity, using a table function that graphically shows the shape of the underlying curve might be a more effective way to communicate.

Con: Non-negative input

Because of their exponential nature, CEF cannot accept a negative input (with a few exceptions, as discussed above). Input must therefore be normalized in the positive domain.

Con: Never saturating

While a relationship can fulfill the prerequisites for using a CEF (discussed above) for a reasonable range of input values, in reality in many cases bio-physical relationships would exhibit some saturation for extreme input values that CEF do not capture.

Con: Non-zero reference points

Reference points for input and output cannot be zero, otherwise the normalization of input and output is not possible. This might require choosing ad-hoc normalization values that cannot be easily justified.

Con: Initial values as reference points limit choice of timing for the start of the simulation

Some points in time are special - out of the ordinary in some way. When table functions are used that are based on reference points that are only surpassed by the system later in time, the initial values of the system may not be very important for system behaviour at later points in time. If however, elasticity formulations are used with initial values as reference points the initial values have a strong influence on the whole simulation. Hence it is important that the onset of the simulation is at a time when the system is in a *relatively normal* condition. If e.g. a national model of Germany was built, it would not be a good idea to start the simulation right after reunification in 1990, when many things were in a very non-normal state. Similarly, when modeling a production in some way, starting the simulation right after the introduction of a new product may not be advisable either. In both cases starting the simulation somewhat later appears advisable. That said, this limitation also applies to table functions if these are normalized (as it is recommended, as mentioned further above).

5. Conclusions

One of the strengths of the system dynamics method is the ability to account for non-linear relationships between variables. Such relationships are typically represented using table functions, which are highly versatile and support the definition of nearly any functional form. Still, in some cases, table functions are cumbersome to develop and time-consuming to calibrate, leading to edgy curves that cover only a limited range of possible input values. When the relationship to be represented meets certain criteria, the concept of “elasticity” can be used and applied in system dynamics models to more effectively formulate and calibrate non-linear relationships.

Elasticity, the ratio of the proportional change in output over that in input, can provide an immediate measure of the strength of the relationship between two variables. When such a relationship exhibits constant elasticity throughout, a simple power function can be used to represent the relationship, a so-called Constant Elasticity Formulation (CEF). Despite constant elasticity formulations (CEF) being broadly used in economic modeling, our literature review indicates that they are hardly discussed in the system dynamics literature, and that little research has been carried out on the implications of using such formulations in our field.

Our review of the mathematical properties of CEF reveals that they can be effectively used in system dynamics models to represent certain non-linear relationships. By representing the strength of a relationship between two variables with a single number, CEF can be more quickly formulated and calibrated than table functions. In addition, CEF define smooth curves with no sharp angles, and they are defined for an infinite range of input. The practical example that we have discussed, the application of CEF as replacement of a table function in the World-3-03 model, shows that it provides a smoother and more extended curve than the corresponding table function.

While care has to be applied that the relationship to be represented meets the criteria we discussed, elasticity formulations provide a useful alternative to table functions, complementing the toolbox of SD formulations.

6. References

Andersen, D. F. (1990). Analyzing who gains and who loses: The case of school finance reform in new york state. *System Dynamics Review*, *6*(1), 21–43. https://doi.org/10.1002/sdr.4260060103

Arif, M. T., & Saeed, K. (2019). Sustaining economic growth with a nonrenewable natural resource: The case of oildependent Indonesia. *Towards Sustainable Development: Essays on System Analysis of National Policy*, *5*(1), 308–331.

Brady, M. P. (2009). Advertising effectiveness and spillover: Simulating strategic interaction using advertising. *System Dynamics Review*, *25*(4), 281–307. https://doi.org/10.1002/sdr.426

Cavana, R. Y., & Clifford, L. V. (2006). Demonstrating the utility of system dynamics for public policy analysis in New Zealand: The case of excise tax policy on tobacco. *System Dynamics Review*, *22*(4), 321–348. https://doi.org/10.1002/sdr.347

Dianati, K. (2022) London’s Housing Crisis – A System Dynamics Analysis of Long-term Developments: 40 Years into the Past and 40 Years into the Future. PhD Thesis. University College London, UK.

Eskinasi, M. (2014). *Towards Housing System Dynamics*. Radboud Universiteit Nijmegen. Retrieved from http://hdl.handle.net/2066/129859

Forrester, J. W. (2003). Dynamic models of economic systems and industrial organizations. *System Dynamics Review*, *19*(4), 329–345. https://doi.org/10.1002/sdr.284

Forrester, N. B. (1987). The role of econometric techniques in dynamic modeling: Systematic bias in the estimation of stock adjustment models. *System Dynamics Review*, *3*(1), 45–67. https://doi.org/10.1002/sdr.4260030107

Inman, D., Bush, B., Newes, E., Peck, C., & Peterson, S. (2020). A technique for generating supply and demand curves from system dynamics models. *System Dynamics Review*, *36*(3), 373–384. https://doi.org/10.1002/sdr.1663

Jung, T., & Strohhecker, J. (2009). Risk-adjusted pricing strategies for the corporate loans business: Do they really create value? *System Dynamics Review*, *25*(4), 251–279. https://doi.org/10.1002/sdr.429

Kampmann, C. E., & Sterman, J. D. (2014). Do markets mitigate misperceptions of feedback? *System Dynamics Review*, *30*(3), 123–160. https://doi.org/10.1002/sdr.1515

Meadowz, D., Randers, J., & Meadows, D. (2004). Limits to growth the 30-year update. *White River Junction, VT: Chelsea Green Publishing*.

Moxnes, E. (1990). Interfuel substitution in OECD-European electricity production. *System Dynamics Review*, *6*(1), 44–65.

Olaya, C. (2016). Cows , agency , and the signi fi cance of, *31*(4), 183–219. https://doi.org/10.1002/sdr

Oliva, R. (2003). Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, *151*(3), 552–568. https://doi.org/10.1016/S0377-2217(02)00622-7

Pierson, K., & Sterman, J. D. (2013). Cyclical dynamics of airline industry earnings. *System Dynamics Review*, *29*(3), 129–156.

Sterman, J. (2000). Business Dynamics: Systems Thinking and Modelling for a Complex World. McGraw-Hill Higher Education.

Ulli-Beer, S., Gassmann, F., Bosshardt, M., & Wokaun, A. (2010). Generic structure to simulate acceptance dynamics. *System Dynamics Review*, *26*(2), 89–116. https://doi.org/10.1002/sdr.440

Van Ackere, A., & Smith, P. C. (1999). Towards a macro model of National Health Service waiting lists. *System Dynamics Review*, *15*(3), 225–252. https://doi.org/10.1002/(sici)1099-1727(199923)15:3<225::aid-sdr171>3.0.co;2-p

Wirl, F. (1991). Dynamics of commodity taxation: An example of an energy tax. *System Dynamics Review*, *7*(2), 145–158. https://doi.org/10.1002/sdr.4260070204