

Putting decomposition of energy use and pollution on a firm footing - clarifications on the residual, zero and negative values and strategies to assess the performance of decomposition methods

Adrian Muller^{*†‡}

Abstract: I show how the problems with zero and negative values in decomposition can in principle be resolved by avoiding ill-defined mathematical operations used to derive the decomposition formulae (division by zero and taking logarithms of zero and negative values). Referring to integral approximation, which is the basis of any decomposition analysis, I also discuss the residual in decomposition and show that the presence of a non-zero residual is natural and that requiring a zero residual as a strategy to identify optimal decomposition methods is without basis. To nevertheless advise on optimal decomposition methods, I suggest to investigate for which types of functions different decomposition methods are exact or good approximations and how they perform in simulations, where the exact integrals are known. Regarding these criteria the LMDI seems to perform best.

Keywords: decomposition analysis, Divisia Index, logarithmic mean, energy consumption, emissions

JEL: C63, Q41, Q5

^{*}Environmental Economics Unit (EEU), Department of Economics, Göteborg University, e-mail: adrian.muller[-at-]economics.gu.se

[†]Center for Corporate Responsibility and Sustainability CCRS, University of Zürich, Kunstlergasse 15a, 8001 Zürich, Switzerland; phone: 0041-44-634 40 62; e-mail: adrian.mueller[-at-]ccrs.unizh.ch

[‡]Many thanks to Åsa Löfgren and two anonymous referees for very helpful remarks. Financial support from the Environmental Economics Unit at Göteborg University is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

Decomposition¹ of energy use, energy intensity and pollution to identify different drivers for their evolution provides important information for policy makers. Total industrial CO₂ emissions C , for example, can be written as the following sum of (trivial) products:

$$C = \sum_{ij} C_{ij} = \sum_{ij} Q \frac{Q_i}{Q} \frac{E_i}{Q_i} \frac{E_{ij}}{E_i} \frac{C_{ij}}{E_{ij}} = \sum_{ij} Q S_i I_i M_{ij} U_{ij}, \quad (1)$$

where C_{ij} are the emissions from fuel j in sector i , Q_i is the output of sector i , total industrial output is $Q := \sum_i Q_i$, E_{ij} is the consumption of fuel j and E_i total fuel consumption in sector i , $S_i := Q_i/Q$ is the output share and $I_i := E_i/Q_i$ the energy intensity in sector i . $M_{ij} := E_{ij}/E_i$ is the share of fuel j in sector i and the sector- and fuel-wise emission factor is $U_{ij} := C_{ij}/E_{ij}$.

Total emissions are thus expressed as a sum of products of factors referring to size effects Q (total output), structural change S_i (sector output share in total output) and sector-wise technological progress I_i (energy use per output), as well as the sector-wise fuel-structure M_{ij} and sector- and fuel-wise pollution intensity U_{ij} . The main interest lies in the time development of these different factors capturing the different driving forces, and in their respective additive contribution to the evolution of total emissions. Formally, the interest thus lies in the following type of decomposition of changes in C between two time points t and $t + 1$: $C(t + 1) - C(t) = \sum_{ij} (\Delta_Q + \Delta_{S_i} + \Delta_{I_i} + \Delta_{M_{ij}} + \Delta_{U_{ij}})$, where the different deltas “ Δ ” are the part of the change in the total assigned to the variables indicated in subscripts.

The literature on the theoretical aspects of decomposition and on the application to energy use and intensity and pollution still grows (e.g. Cole et al. 2005, Boyd and Roop 2005, Choi and Ang 2003; for a survey, see Ang 1995 and, most recently and encompassing, Liu and Ang 2007). Studies comparing different decomposition methods find that results often depend on the method chosen (e.g. Ang 1995, Greening et al. 1997, Ang 2004, Ang and Liu 2007c). However, very similar results from different methods can also occur. In some cases, this can be traced back to analytical equivalence of the methods (e.g. Choi and Ang 2003, Ang and Liu 2007c).

Furtheron, decomposition methods often lead to a residual (the part of the total energy use, intensity or pollution, which is not captured by the different terms in the decomposition) that remains seemingly unexplained. In

¹In this paper, I primarily use the language and the methods of the so called Index Decomposition Analysis (IDA). I link this to the Structural Decomposition Analysis (SDA) and emphasise formal similarities in the last paragraph before section 2.2.

addition, many methods are unable to properly deal with zero and negative values in the data. New proposals for decomposition methods address these issues and the performance of a method is often measured by the value of its residual, ideally bringing this down to zero, and by the method's convergence properties under replacing zero-values with infinitesimal quantities and taking the limit to zero. Problems with negative values usually can easily be resolved as they can be traced back to the treatment of positive and zero values.

The literature on energy use and intensity and pollution emissions decomposition recently identified the logarithmic mean Divisia index (LMDI) approach as one of the most favorable (Ang 2004; Ang and Liu 2001, 2007a and b, Liu and Ang 2007). This is based mainly on four features of the method, namely its ability to handle zero and negative values, the absence of any residual term and the ease of calculation. In addition, it is invariant under time and factor reversal and fulfills aggregate consistency and proportionality (these terms are explained in section 4 below).

However, several of these motivations actually have no basis as guidelines for the quality of a decomposition method. First, the zero and negative value problems of decomposition analysis stem from ill-defined operations during the calculations and can be avoided. This will be elaborated in section 3 below.

The residual, on the other hand, reflects the fact that any such decomposition is based on integral approximation (Trivedi 1981). This is due to the fact that the functions involved in the decomposition are only known for discrete points of time, e.g. annually. The residual thus cannot be argued to necessarily be zero for an optimal decomposition approach. Forcing the residual to be zero could in principle also involve mutually cancelling terms of opposite sign in different parts of the decomposition (this is the case for the Shapley-value decomposition with more than two factors, for example; cf. Muller 2006). This could make the zero-residual decomposition even less exact than a decomposition with some non-zero residual term based on a good approximation. More on this is presented in section 4 below.

Nevertheless, the LMDI performs better regarding an exact decomposition of different classes of functions than other popular decomposition methods. This can be traced back to the LMDI being exact for a wide range of functions assumed to generate the only discrete data to be decomposed. In addition, an illustrative simulation, where known functions are decomposed using different methods and results are compared to the exact decomposition based on the calculation of the underlying integrals, shows that the LMDI performs well. These claims will be addressed in section 5. It is this property to provide exact decomposition for a wide range of functions that makes

the LMDI a good method for decomposition rather than the four properties referred to above that actually need not hold in the light of decomposition as integral approximation.

From the following analysis, it will also become clear that it is not necessary to treat the different variables involved in decomposition separately (e.g. absolute values, intensities or relative intensities, etc.), as often done in the literature (e.g. Ang 1995). The underlying formulae are the same for all variables and although separate treatment sometimes may be illustrative, it often makes the exposition somewhat lengthy and obfuscates the underlying common formalism. I also want to emphasise that the discussion regarding approximations is different from the discussion of approximate properties in general index theory, as presented in Barnett and Choi (2006), for example. There, the approximation is based on the distance between the discrete time points becoming arbitrarily small, assuming full knowledge of the underlying functions. Such a strategy is not possible here, as the functions are only known for discrete points of time in fixed distance, e.g. annually.

In section 2, I introduce the basic formalism. Section 3 presents the origin of the problems with zero and negative values and how these can be avoided. In section 4, I discuss the residual and why its disappearance cannot serve as a criterion to identify a good decomposition. I also clarify the role of some further theoretical properties claimed to be necessary for any good decomposition. In section 5, I derive for which type of functions the LMDI and other decomposition methods are exact and discuss the role of approximations and simulations for the assessment of the performance of decomposition methods. Section 6 concludes.

2 Basic formalism, integral approximation and weighted means

In this section, I introduce the basic formalism of decomposition analysis and discuss the links to integral approximation.

2.1 Basic formalism

I introduce the formalism of decomposition using the example from Ang and Liu (2007a). Referring equation (1) for CO₂ emissions,

$$C = \sum_{ij} C_{ij} = \sum_{ij} Q \frac{Q_i}{Q} \frac{E_i}{Q_i} \frac{E_{ij}}{E_i} \frac{C_{ij}}{E_{ij}} = \sum_{ij} Q S_i I_i M_{ij} U_{ij}, \quad (2)$$

the general formulae for decomposition of the change in emission level C from time t to $t+1$ into the contributions of different key variables is given by the following equation:

$$\begin{aligned}
\Delta C_{t+1,t} &:= C(t+1) - C(t) = \int_t^{t+1} \frac{dC}{d\tau} d\tau \\
&= \sum_{ij} \int_t^{t+1} \left(S_i I_i M_{ij} U_{ij} \frac{\partial Q}{\partial \tau} + Q I_i M_{ij} U_{ij} \frac{\partial S_i}{\partial \tau} + \right. \\
&\quad \left. + Q S_i M_{ij} U_{ij} \frac{\partial I_i}{\partial \tau} + Q S_i I_i U_{ij} \frac{\partial M_{ij}}{\partial \tau} + Q S_i I_i M_{ij} \frac{\partial U_{ij}}{\partial \tau} \right) d\tau \\
&= \sum_{ij} \int_t^{t+1} Q S_i I_i M_{ij} U_{ij} \left(\frac{1}{Q} \frac{\partial Q}{\partial \tau} + \frac{1}{S_i} \frac{\partial S_i}{\partial \tau} + \frac{1}{I_i} \frac{\partial I_i}{\partial \tau} + \right. \\
&\quad \left. + \frac{1}{M_{ij}} \frac{\partial M_{ij}}{\partial \tau} + \frac{1}{U_{ij}} \frac{\partial U_{ij}}{\partial \tau} \right) d\tau. \tag{3}
\end{aligned}$$

The rationale to assign the contribution of each key variable to a summand in the second and third line is based on the following “infinitesimal form” of the changes: $\frac{dC}{d\tau} = S_i I_i M_{ij} U_{ij} \frac{\partial Q}{\partial \tau} + Q I_i M_{ij} U_{ij} \frac{\partial S_i}{\partial \tau} + Q S_i M_{ij} U_{ij} \frac{\partial I_i}{\partial \tau} + Q S_i I_i U_{ij} \frac{\partial M_{ij}}{\partial \tau} + Q S_i I_i M_{ij} \frac{\partial U_{ij}}{\partial \tau}$. Here, each term captures the contribution of the variable whereof the derivative is taken to the total infinitesimal change. This motivates to assign contributions after the integration involved in (3) in a similar way. This assignment fulfils the intuitive basic requirement as formulated in Vogt (1978, referring to François Divisia), for example, namely that for the contribution of a variable x , changes in the other variables should not make a difference, given x does not change: here, the contribution from Q is always zero in case $\frac{\partial Q}{\partial \tau} = 0$, and the same is valid for the other variables. Finally, equation (3) is often written as follows, employing $\frac{d \ln f}{dx} = \frac{1}{f} \frac{df}{dx}$ to arrive at the “logarithmic form”:

$$\begin{aligned}
\Delta C_{t+1,t} &= \sum_{ij} \int_t^{t+1} Q S_i I_i M_{ij} U_{ij} \left(\frac{\partial \ln Q}{\partial \tau} + \frac{\partial \ln S_i}{\partial \tau} + \frac{\partial \ln I_i}{\partial \tau} + \right. \\
&\quad \left. + \frac{\partial \ln M_{ij}}{\partial \tau} + \frac{\partial \ln U_{ij}}{\partial \tau} \right) d\tau. \tag{4}
\end{aligned}$$

It can readily be seen from formulae (2) to (4) where potential problems with zero or negative values might arise. First, expanding in equation (2) causes a problem if a variable is equal to zero for some sectors, fuels or periods, because in this case some of the factors on the right-hand side of

the equation contain ill-defined division by zero. I emphasize though, that the whole term remains well-defined, as for each denominator equal to zero, there is the corresponding numerator equal zero from the original expansion in one of the other factors. The same problem arises in equations (3) and (4), and the division by zero, resp. the logarithm of zero in the brackets is still well-defined if combined with the corresponding factor equal to zero in front of the brackets. In the logarithmic form, the additional problem of ill-defined logarithms of negative values arises.

In the example, the change in emissions from t to $t + 1$ is decomposed into five parts referring to overall industrial activity or size (the Q -term), industrial structure (the S -term), energy intensity (the I -term), fuel-mix (the M -term) and fuel emission factors (the U -term). Equation (3) is exact, but the data to calculate the exact integrals is never available. Usually, the quantities involved are only known for discrete points t and $t + 1$, e.g. for subsequent years, while the exact shape of the function describing these quantities between these points is unknown. The integrals in (3) have to be approximated. This fact is well-known in the literature (Trivedi 1981; in the energy context, see e.g. Liu et al. 1992). The consequences thereof, however, are largely not acknowledged, as I elaborate below in the discussion of the residual.

The basic task of decomposition is thus to approximate terms of the following structure, where the integrand is only known for the boundary values of the integration interval (referring the example above, we thus have $G = S_i I_i M_{ij} U_{ij}$ and $H = Q$; $G = Q I_i M_{ij} U_{ij}$ and $H = S_i$; etc.).

$$\begin{aligned}
 J := \int_t^{t+1} G(\tau) \frac{\partial H(\tau)}{\partial \tau} d\tau &= \int_t^{t+1} G(\tau) H(\tau) \left(\frac{1}{H(\tau)} \frac{\partial H(\tau)}{\partial \tau} \right) d\tau = \\
 &= \int_t^{t+1} G(\tau) H(\tau) \left(\frac{\partial \ln H(\tau)}{\partial \tau} \right) d\tau. \quad (5)
 \end{aligned}$$

Having established the basic formalism, I make a short detour on the difference between index and structural decomposition analysis (IDA, resp. SDA) usually made in the literature (e.g. in Hoekstra and van der Bergh 2003). The main difference is in the variables a total is decomposed into - which are based on the information from input-output tables in the case of SDA and on more aggregate sector key variables in the case of IDA. Regarding basic formulae, however, there is no difference. The problem is always to decompose a change of a product into a sum of changes of the single factors (see e.g. equations (3) ff in Hoekstra and van der Bergh 2003; (10) ff in Dietzenbacher and Los 1998; (17) ff in Rose and Casler 1996). Thus, the

same methods are in principle applicable. Such a translation of methods is undertaken in Hoekstra and van der Bergh (2003) and this is needed for SDA, as methods there are less elaborate than in IDA. Dietzenbacher and Los (1998) point out the problem of the decomposition methods usually used in SDA, namely their arbitrariness regarding combination of factors evaluated at t and $t + 1$ that cannot be resolved on theoretical grounds and that leads to a considerable range of differing results. The methods they refer to are very similar to some methods used in poverty decomposition (the Datt-Ravaillon approach), as discussed and criticised in Muller (2006). In the light of this criticism and the discussion below, I suggest to refrain from these Datt-Ravaillon type decomposition in SDA and to apply the more refined methods used in IDA.

2.2 Integral approximation and weighted means

There are several methods to approximate the integrals - and derivatives - given knowledge on the functions on the boundaries only. The most simple integral approximation is based on Riemannian sums, i.e. on replacing the integral by the value of the integrand at the right/left-hand-side boundary multiplied with the difference in t between two subsequent data-points (which equals 1 here and is thus dropped in the following):

$$J \approx G(t + 1) \frac{\partial H(\tau)}{\partial \tau} \Big|_{\tau=t+1} \quad \text{and} \quad J \approx G(t) \frac{\partial H(\tau)}{\partial \tau} \Big|_{\tau=t} \quad (6)$$

A combination of the right/left-hand based approach is also common (the so-called trapezoid method):

$$J \approx \frac{1}{2} \left\{ G(t + 1) \frac{\partial H(\tau)}{\partial \tau} \Big|_{\tau=t+1} + G(t) \frac{\partial H(\tau)}{\partial \tau} \Big|_{\tau=t} \right\} \quad (7)$$

In classical approximation with Riemannian sums, the limit $\Delta t \rightarrow 0$ would then be taken and all these approaches are equivalent and exact in the limit. Here, however, where data is available for discrete points only, it is not possible to proceed in this way. The results from the different approximations thus usually differ and other approximation methods can also be employed leading to potentially different results.

Based on first-order Taylor approximation, the most common approach to approximate derivatives replaces them with the difference in the boundary values divided by the difference in t (equalling 1, cf. above). Approximating from the left/right, respectively, gives

$$\frac{\partial H}{\partial \tau} \Big|_{\tau=t} \approx [H(t) - H(t - 1)] \text{ resp. } \approx [H(t + 1) - H(t)]. \quad (8)$$

With the trapezoid method (7), i.e. approximating by a straight line connecting the data, the choice of a right-hand derivative approximation at the lower boundary and a left-hand at the upper may be motivated:

$$J \approx \frac{1}{2} \{G(t+1) + G(t)\} [H(t+1) - H(t)] \quad (9)$$

This is not entirely consistent as an approximation procedure because it treats the term involving derivatives differently from the other term by using a right-hand side approximation for the left boundary and an approximation from the left for the right boundary. It can nevertheless be defended as a viable approach as it gives the same value for the derivative for the interval $(t, t+1)$, i.e. the slope of the straight line joining the boundary points.

For both integral and derivative approximation, restricted data availability makes it impossible to achieve increased accuracy by using higher-order Taylor expansion terms and subsequently decreasing the time-differences to reduce the higher-order errors.

Inspired by the trapezoid method, more general solutions that employ some weighted combination of the values at the upper and lower boundary are suggested (the trapezoid method weights them equal), at the same time replacing the derivative with the difference of the boundary values of the function the derivative is taken of (thus replacing the derivative at both boundaries with the slope of the straight line joining those). These approaches usually are not systematically linked to the basis of decomposition in approximations. If at all, reference to this background is only made in a cursory fashion, not influencing subsequent argumentation. This has led to some problems and confusion. Decomposition has become to be seen as a problem of choosing the right weights for the two known endpoints in calculating the different contributions (cf. e.g. Liu et al. 1992). The following expression (the so-called parametric Divisia Index) has thus been suggested for the integral, replicating an approximation to the general Divisia index (see e.g. Hulten 1973):

$$[G(t) + \alpha(G(t+1) - G(t))] [H(t+1) - H(t)], \quad \alpha \in [0, 1]. \quad (10)$$

Choosing $\alpha = 0/1$ gives the left/right-hand boundary approximation with the mixed approximation for the derivative (replicating the Laspeyres and Paasche Index), $\alpha = \frac{1}{2}$ gives the trapezoid approximation (the Marshall-Edgeworth Index). Any value of α can be seen as a certain choice of weights given to the left/right boundary in the integral approximation. This lacks sound basis from integral and derivative approximation theory, though, especially as it is combined with unweighted approximations for the derivatives,

although those are part of the integrand. Only if more information on the underlying functions is available, some judgement if they are better captured by a step function or the straight line or some other type of approximation might be possible, thus suggesting a most adequate decomposition method or choice of weights, respectively (see Fernandez and Fernandez (2007) for a proposal on how such additional information could be used for parameter specification).

The fundamental role of approximation is sometimes further obfuscated by seeing decomposition as a problem of choosing different integration paths between the known endpoints, without explicitly referring to the fact that there is only one true, although unknown path underlying the data - that could in principle been known given more detailed measurements (Liu et al. 1992, Balk 2005).

Against this background of perceiving the search for an optimal decomposition rather as a problem of weights than of approximations, the development of more flexible, “optimal” weighting schemes based on some basic principles has been a natural path for improvements. The adaptive weighting (Liu et al. 1992) for example builds on the claim that the original and the logarithmic parametric Divisia approach to decomposition are mathematically equivalent. This motivates setting equal the contributions calculated via the original Divisia (eq. (10)) with those calculated via the corresponding logarithmic Divisia approximations (eq. (4)), i.e. $[G(t)H(t) + \alpha(G(t+1)H(t+1) - G(t)H(t))] \ln\left(\frac{H(t+1)}{H(t)}\right)$. This usually uniquely determines α . The mathematical equivalence claimed here however does not hold for these equations that are not exact any longer but approximations, as the derivatives have been replaced by approximations. This method thus lacks sound theoretical underpinning.

Another flexible weighting approach is the refined Divisia method from Ang and Choi (1997), based on a specific choice of weights involving the so-called “logarithmic mean function” $L(x, y) := \frac{x-y}{\ln(x)-\ln(y)}$. There is no motivation whatsoever from integral and derivative approximation to use these weights, though. A big advantage, however, as the authors claim, is that these weights solve the problem of zero-values and residuals, and this is the main motivation to propose this particular weighting scheme. A zero residual is also the main advantage claimed of Boyd and Roop (2005) for their Fisher-Index based approach, which also forces the residual to be zero by definition rather than by reference to exact approximation, though.

A further development of the refined Divisia method is the currently most favourite “logarithmic-mean Divisia Index” LMDI (Ang 2004; Ang and Liu 2001, 2007a and b, Liu and Ang 2007). It is based on the logarithmic form

of the decomposition (i.e. the right-hand expression in eq. (5)) and also employs the “logarithmic mean”:

$$J_{LMDI} := L(G(t+1)H(t+1), G(t)H(t)) \ln\left(\frac{H(t+1)}{H(t)}\right), \quad (11)$$

The problem of zero or negative values as described above arises in this form if $H(t)$ or $H(t+1)$ is zero or negative. The presence of log-mean leads to additional cases when zero or negative values can cause problems, namely for $G(t)$ or $G(t+1)$ being zero or negative. How to deal with these zero values is the topic of the next section.

I link this exposition to the literature by pointing out that equation (11) corresponds to the LMDI decomposition formulae (3) to (7) in Ang and Liu (2007a). To see this, replace G and H with the appropriate variables, sum over all sectors i and fuels j and equate $t+1 = T$, $t = 0$. Furthermore, equation (10) replicates the “additive parametric Divisia method 1” techniques common in the literature, as presented in Ang (1995). Similar reasoning applies to the “parametric Divisia method 2” and the multiplicative approach presented there, being the exponentiated form of the additive approach. The preceding formalism illustrates that decomposition need not be discussed separately for different combinations and types of the variables of interest as it is often done in the literature (e.g. Ang 1995).

3 Zero and negative values

3.1 Zero values

There are basically two solutions suggested in the literature to deal with the zero value problem: the “small-value” strategy and the “analytical limit” (Ang et al. 1998; Wood and Lenzen 2006; Ang and Liu, 2007a). The first strategy recommends to replace zero values by small numbers δ such as 10^{-20} , 10^{-50} , 10^{-100} and to rely on the convergence when δ becomes smaller and smaller. This has the advantage that the general LMDI formulae need not be changed after replacing all zero values in the data with such small numbers. Wood and Lenzen (2006) observe that there are cases that are not robust regarding different choices of δ . In addition, this approach is very unsatisfactory from a mathematical point of view. Wood and Lenzen (2006) are in favor of the “analytical limit”, which is also the mathematically correct approach to such a situation involving divergent terms - taking the limit value in an analytically correct way. Ang et al. (1998) show that these limits exist

for the LMDI and that this thus converges to finite values. The disadvantage is that some formulae where zero values occur have to be changed.

Although the analytical limits exist, the whole problem of zero values is based on ill-defined operations such as divisions by zero that must be avoided from the beginning in any correct formal reasoning. This motivates a third - and the only truly consistent - approach to deal with or, rather, to avoid the problem of zero values: As mentioned above, after equation (4), the problem with zero values ultimately stems from the expansion in equation (3). To avoid this ill-defined operation, the step from the second to the third equation in (3) is not possible in case zero values are involved. Assuming that any approximation of the integrals and derivatives involved will be based on the known values at the end-points t and $t + 1$, a strategy relying on the second rather than on the third equation in (3) or on the logarithmic form in (4) has to be employed in case some of the variable equals zero for the values t , $t + 1$ or both.

If one of the variables Q , S_i , I_i , M_{ij} or U_{ij} is zero for both t and $t + 1$, all terms in an approximation to (3) based on the boundary values at t and $t + 1$ vanish (as also the derivative has to be approximated by such). If the variable is zero only for one of the boundaries, again all terms referring to this boundary vanish in the approximation, while in general all the terms from the other boundary contribute, i.e. there is a contribution from each summand in the integrand. This can also be seen in the more general equation (5), where it is clear that the case $G(t) = 0$, $G(t + 1) \neq 0$ need not lead to $J = 0$.

This differs from the results of the analytical limit approach to zero values in the LMDI. There, only the terms containing the derivative (i.e. the logarithm in the LMDI approximation) of the variables with changes from/to zero to/from positive values and not all the terms contribute to the overall effect (cf. table 4 in Ang and Liu 2007a). This is rather counter-intuitive. Ultimately, the following mechanism lies behind this property: a change from zero to a finite value corresponds to an infinite percentage change and thus dominates any change from a finite to another finite value if measured in percentage changes. The following shows in detail how this result arises because of the properties of the log-mean function that mixes the values from the two boundaries. In case $H(t) = 0$, $H(t + 1) \neq 0$, for example, we have (for $H(t + 1) \neq 0$, $G(t + 1) \neq 0$, $G(t) \neq 0$)

$$\begin{aligned}
J_{LMDI} &= \frac{G(t+1)H(t+1) - G(t)H(t)}{\ln[G(t+1)H(t+1)] - \ln[G(t)H(t)]} \ln\left(\frac{H(t+1)}{H(t)}\right) = \\
&= \frac{(G(t+1)H(t+1) - G(t)H(t))\left(\frac{\ln H(t+1)}{\ln H(t)} - 1\right)}{\frac{\ln G(t+1)}{\ln H(t)} + \frac{\ln H(t+1)}{\ln H(t)} - \frac{\ln G(t)}{\ln H(t)} - 1} = \\
&\longrightarrow G(t+1)H(t+1) \quad \text{for } H(t) \rightarrow 0. \tag{12}
\end{aligned}$$

For one of the other variables being zero, i.e. $G(t) = 0, G(t+1) \neq 0, H(t) \neq 0, H(t+1) \neq 0$, for example,² we have

$$\begin{aligned}
J_{LMDI} &= \frac{(G(t+1)H(t+1) - G(t)H(t))\left(\frac{\ln H(t+1)}{\ln G(t)} - \frac{\ln H(t)}{\ln G(t)}\right)}{\frac{\ln G(t+1)}{\ln G(t)} + \frac{\ln H(t+1)}{\ln G(t)} - 1 - \frac{\ln H(t)}{\ln G(t)}} = \\
&\longrightarrow 0 \quad \text{for } G(t) \rightarrow 0. \tag{13}
\end{aligned}$$

Summarizing, to avoid the zero value problem in general, refrain from using the manipulations that lead to the second or third term in (5) in case zero values are present and base the decomposition on the first term in this equation. This, however, does not in general provide a solution on how to proceed from this first term.

In case of the LMDI, I thus recommend to treat zero values as follows: a) if no zero values are involved - use the LMDI formulae; b) in case a variable shows zero values for both t and $t+1$ - drop the corresponding term from the decomposition. This gives the same results as the “analytical limit” (cf. table 4 in Ang and Liu 2007a); c) if a zero value is involved for t or $t+1$ only, refrain from the manipulations that lead to the logarithmic form in equation (5) and to the LMDI formula (11) and stick to the second equation in (3), resp. the first in (5). Starting from this equation, any approximation of the integrals given by some weighted combinations of the values at the boundaries involves only one term because the other equals zero. But all variables contribute, not only the term with the variable of interest in the derivative as in the LMDI. This deviates from the LMDI “analytical limit” prescription, because the operations that led to the LMDI formulae are ill defined in this case: starting from the LMDI formulae (11), the analytical limit is correct, but starting from the general decomposition formulae (3), the LMDI formulae cannot be defined. Thus, directly taking the zero values into account in the general decomposition leads to different results. In particular,

²This captures the other summands in (3) as the variable for which “H” stood above is now part of “G”, i.e. the factor multiplying the derivative factor.

the finding that each term involving zero values contributes equally to the final result (Ang and Liu 2007a) and that the contributions of all the other terms are zero cannot be justified.

The question on how to proceed from the second equation in (3) remains. One approach would be to base strategies for zero values on the requirement that they are exact for the types of functions the LMDI is exact for (see section 5 below). This however fails for the exactness conditions for the LMDI established below in equation (20). There, $H = 0$ implies $G = 0$ and vice versa, which is a special case of the zero value problem only.

For the LMDI, I thus propose to search for solutions to the zero value problem using the other strategies for the assessment of the performance of decomposition methods presented in section 5, namely simulations and identification of types of functions for which the LMDI is a good approximation. I am aware that I am short in offering solutions here, but I think it is important to present in detail the problems behind zero values in decomposition and currently proposed solutions and to show potential paths to new solutions although such are not yet found. Depending on the results of this exercise, it can well be that some other decomposition approach emerges as optimal with zero values. The treatment of zero values should then be based on this and not on the LMDI, which is not directly applicable for zero values, as elaborated above.

3.2 Negative values

The problem of negative values arises from the general logarithmic form of decomposition in case $H < 0$ (cf. eq. (5)). It is thus clear that this problem does not arise in the form without logarithms and that it can thus always be avoided from the beginning for any decomposition method not involving logarithms.

Specifically in the LMDI, the problem with negative values arises from the definition of the log-mean, involving logarithms of the total C_{ij} in (2) to (4) (i.e. “ GH ” in the general formulae) that is negative if an impair number of factors is negative, and from the logarithms of $H(t + 1)$ and $H(t)$ in case one of these is negative.

Ang and Liu (2007b) suggest to treat the problem of negative values in the LMDI approach by replacing them by the corresponding positive values in the logarithms if a change from a negative to a negative value from t to $t + 1$ is involved³. If changes from/to a negative value to/from zero or a

³They do not make this substitution in the linear numerator of the logarithmic mean L ; they do not explicitly state this, but replacing all negative values with their positive

positive value are involved, they suggest to trace it back to the treatment of zero-values as described above and in Ang and Liu (2007a) and at the end of the following paragraph, respectively. The remedy for negative values in the LMDI is consistent in case of a change from negative to negative values, as it is, as just mentioned, only an expansion by (-1) . The treatment in the other cases is inconsistent, though, as it involves the treatment of zero values as described and criticized above.

Referring to the general case, i.e. the logarithmic form of decomposition equation (5) in case $H < 0$, it can be seen that the principle of the remedy suggested by Ang and Liu applies in general as well: If $H < 0$ for the whole range of integration, the problem can be resolved for the logarithmic form by replacing the negative value in the logarithm by the same but positive value: $\frac{1}{H} \frac{\partial H}{\partial \tau} = \frac{1}{-H} \frac{\partial(-H)}{\partial \tau} = \frac{\partial \ln(-H)}{\partial \tau}$, where $-H$ is positive and thus all the expressions are well-defined. Doing this, it is important to pay attention not to replace the H multiplying the derivative-term in eq. (5) with $-H$. Again, the remedy is thus actually not a general replacement of H by $-H$, but an expansion of H by $(-1)(-1)$ that cancels directly for the H multiplying the derivative term and leads to replacement of H with $-H$ (and the other factors ‘ -1 ’ cancel) in the logarithm. If $H < 0$ only for parts of the range of integration, we may split this range into the parts where $H \geq 0$ and where $H \leq 0$. The former can then be treated as the general zero value problem, and the latter can be transformed to such by applying the manipulation to replace H with $-H$ as described before.

4 The residual in decomposition

As already mentioned above, the main step in decomposition is to approximate integrals and derivatives such as in equation (3) or, more general, (5), where the integrand is only known at the endpoints t and $t + 1$. I also mentioned that there are several approaches to integral approximation in the decomposition context. These approaches differ in the weights they give to the right or left boundary or how they combine these boundary values and in how they approximate the derivative in the integrand. One conclusion is that the choice of a certain method cannot be motivated unless more information on the underlying functions is directly available, or without information on

values and adding an overall factor -1 just amounts to this (cf. their formulae (4) to (6)). Thus, the strategy is actually not a substitution of the negative values by their positive but rather an expansion of all negative values by $(-1)(-1)$, which is, in contrast to the substitution with negative values, a well-founded approach, as it is based on mathematical equivalence.

which weights or combinations lead to an exact decomposition or a good approximation for a wide range of functions, where the true unknown function can reasonably be assumed to fall under.

Many authors search for an optimal decomposition method: Besides the LMDI and earlier attempts by Ang et al. (e.g. the refined Divisia Index described in Ang and Choi 1997), see, for example, the mean rate-of-change index of Chung and Rhee (2001) or the method proposed by Sun (1998). The various suggestions however never strive for optimality based on additional knowledge on the underlying functions but rather on some theoretical properties seen as desirable, foremost on a zero residual (equivalent to the so-called factor-reversal property from general index number theory), on being consistent in aggregation⁴, and on being easy to understand and to calculate. The last two reasons need not much discussion - but the first two do.

The residual is the difference between the change in the total emission as they are measured, i.e. $\Delta C_{t+1,t}$ and the value the various integrals on the right-hand side of equation (3) sum up to after approximation. Due to the integral approximations being involved, it is only natural to expect some residual different from zero. The residual only reflects the lack of knowledge on the underlying functions. A zero residual thus rather reflects the fact that any error in approximation has implicitly been distributed among the various terms⁵. This bears the danger that the single terms in the decomposition might even be less exact for a method with zero residual than for another method with some residual, as mutually cancelling contributions of any magnitude can have been attributed to the different terms.

Consequently, the often produced argument that the presence of a residual leads to problems in interpretation of the result is not convincing. The residual only reflects the approximate character of the whole decomposition exercise. In this light, rather a zero residual is difficult to interpret, as there is no clue on how much of the approximation error has been attributed to which term. Clearly, a big residual results in a poor explanatory power of the decomposition as a big part correspondingly remains unexplained. But

⁴This is the property that a certain decomposition approach applied on some intermediate disaggregate level and then using the same method to calculate the effects for the total based on this intermediate results leads to the same result as the direct decomposition of the total effect.

⁵Sun (1998) proceeds like this, explicitly assigning the residual to the different terms in the decomposition, based on equal treatment of all factors. This attribution of the residual to the different contributions has no further theoretical foundation. Ang et al. (2003) observe that the decomposition proposed in Sun (1998) is equivalent to the Shapley-value decomposition. The lack of foundation in assigning the residual thus applies to the latter as well. More on the Shapley-value, how it relates to other decomposition methods and what is problematic with it can be found in Muller (2006) and references therein.

this should be corrected for by striving for better approximations of the underlying integrals - which automatically leads to smaller residual terms -, rather than by the criterion of striving for a zero residual, irrespective of how this is achieved and of how good the strategy is as an approximation strategy.

Some illustration for this might be given by referring to regression analysis, where a correct estimation of the various coefficients is much more important than a high coefficient of determination R^2 in total (i.e. than to have a high proportion of the variance explained). Striving for high R^2 can corrupt results as too many or too few variables may be included, and as collinearity problems can obscure the identification of the effects of single variables. This latter point is analogous to what can happen if the residual is forced to be zero and thus might be assigned to the different contributions without possibility to judge how this might happened.

A similar critique as against the zero residual can be put forward for the criterion of consistency in aggregation (cf. footnote 4), as e.g. framed in Ang and Liu (2001), where a first aggregation is made for subgroups $k = 1, \dots, k_0$ with n_k sectors each. Thus, the following equation should hold:

$$\sum_{k=1}^{k_0} \tilde{w}_k \sum_i^{n_k} \sum_j \tilde{w}_{ij}^k \ln \left(\frac{S_{ij}(t+1)}{S_{ij}(t)} \right) = \sum_{ij} w_{ij} \ln \left(\frac{S_{ij}(t+1)}{S_{ij}(t)} \right). \quad (14)$$

This equation is valid if the sufficient (but not necessary condition) $\tilde{w}_k \tilde{w}_{ij}^k = w_{ij}$ is met, for example. This is the case for the LMDI as can be shown by inserting the corresponding log-mean expression for the weights w . Seeing the weighted sums as integral approximations, this condition need not be met, as the different error terms from the different levels usually will disrupt its exact validity also in case the approximations are good. A mismatch in aggregation only reflects the lack of knowledge on different levels of aggregation. As with the residual, however, a decomposition resulting in a too large deviation from the criterion being met might be judged to perform badly.

Time-reversal (the criterion that the parts in the decomposition switch sign if t is replaced by $-t$) and proportionality (homogeneity of degree one) are two further theoretical criteria judged important by Ang (2004). Again, referring to the approximate quality of all formulae involved, strict validity of these criteria need not be granted.

Besides the inconsistencies related to the treatment of zero and negative values in the LMDI, the reservations towards a zero residual and consistency in aggregation are additional reasons that challenge the LMDI as the default best method for decomposition. These properties are among the main theoretical arguments for the LMDI approach given in the literature (Ang and

Liu 2001, Ang 2004). As they fail due to the very quality of decomposition as an approximation to unknown integrals, reference to them is actually no valid motivation to use the LMDI for the integral approximation.

5 Determining the performance of decomposition methods

In the previous sections, I motivated that general properties such as a zero residual are not a valid criterion to identify an optimal decomposition method. To nevertheless arrive at a judgement on which method is best without referring to additional information on the underlying function - as such is not likely to be available⁶ - I suggest to investigate the performance of the different decomposition methods regarding different broad types of functions G and H . This discussion can be based on three strategies. First, taking the formulae of a decomposition method, it can be assessed for which type of functions the method is exact, i.e. when it is equal to the true value of the integral that it tries to approximate. Similarly, it can be identified, for which type of functions a given method is a good approximation, leading to only minor errors. Finally, simulations can be employed to further illustrate how different decomposition methods perform, comparing their results to the exact solutions known due to the known functions underlying the data used in the simulations. Ideally, these strategies would allow to identify the classes of functions for which a certain decomposition method is optimal, viz exact or leading to only minor errors compared to the exact solution. I illustrate these strategies primarily by application to the LMDI and focusing on exactness conditions.

Establishing exactness for the LMDI is identical to finding conditions for

$$\int_t^{t+1} G \frac{\partial H}{\partial \tau} d\tau = \frac{G(t+1)H(t+1) - G(t)H(t)}{\ln\left(\frac{G(t+1)H(t+1)}{G(t)H(t)}\right)} \ln\left(\frac{H(t+1)}{H(t)}\right). \quad (15)$$

The left-hand-side is an additive combination of terms only dependent on $t+1$ or t , $\int_t^{t+1} G \frac{\partial H}{\partial \tau} d\tau = F(t+1) - F(t)$, for the antiderivative F , while on

⁶Fernandez and Fernandez (2007) present a detailed proposal on how to incorporate additional information for some variables in decomposition analysis. They base their exposition on an extension of Sun's (1998) Shapley value decomposition. This leads to improvements given more information is available, but does not solve the problems of the Shapley value for the variables where such is not available (cf. footnote 5). It is, however, a strategy that could in principle be adapted to other decomposition methods as well to incorporate additional information on some variables.

the right-hand-side, dependence on $t + 1$ and t is mixed. The equation can thus only hold in case the whole expression equals zero or the terms causing the mixing on the right-hand-side cancel. This implies the necessary but not sufficient conditions

$$H(t + 1) = H(t) \quad \text{or} \quad G(t + 1)H(t + 1) = G(t)H(t) \quad \text{or:} \quad (16)$$

$$\frac{\ln(H(t + 1)) - \ln(H(t))}{\ln(H(t + 1)G(t + 1)) - \ln(H(t)G(t))} = c = \text{constant}, \quad (17)$$

where we assume $c \neq 1$. The case $c = 1$ is treated below. Equation (17) leads to $\ln(H(t + 1)) - \ln(H(t)) = c[\ln(H(t + 1)) + \ln(G(t + 1)) - \ln(H(t)) - \ln(G(t))]$. Taking derivatives with respect to the lower boundary t gives (an equivalent derivation can be done taking derivatives with respect to $t + 1$)

$$\frac{\partial}{\partial t} \ln(H(t)) = \frac{c}{1 - c} \frac{\partial}{\partial t} \ln(G(t)). \quad (18)$$

Integrating and exponentiating yields

$$H(t) = e^{-K} G(t)^{\frac{c}{1-c}}, \quad (19)$$

where K is the constant from the integration. Equation (19) is thus a general necessary condition that the LMDI decomposition approach is exactly true in case (16) does not hold.

The special case $c = 1$ is equivalent to $G(t + 1) = G(t)$ and the LMDI then is correct in case $\int_t^{t+1} G \frac{\partial H}{\partial \tau} d\tau = G(t + 1)(H(t + 1) - H(t))$. Using partial integration, this is equivalent to $\int_t^{t+1} H \frac{\partial G}{\partial \tau} d\tau = 0$, which is valid only for special cases of G and H , e.g. for the case when G is a constant. It goes without saying that $H(t + 1) = H(t)$ or $G(t + 1)H(t + 1) = G(t)H(t)$ also are valid in special cases only. Equation (19), however, is of considerable generality. Written differently, and after establishing sufficiency for this condition by inserting it into the original equation (15), it follows that the LMDI is exact for all

$$H(t) = \kappa G(t)^\gamma, \quad \text{where } \kappa, \gamma \in \mathbb{R}. \quad (20)$$

Other conditions for G and H could be derived for the case when (16) or (17) with $c = 1$ hold ($c = 1$ corresponds to the boundary cases $\gamma = \pm\infty$ in (20)): the LMDI is for example exact for all G if H is constant and for all H if G is constant - these are trivial cases, though. I restrict the discussion to the case (20), which is not trivial and shows that the LMDI is exact for a wide range of possible underlying functions. Choose, for example, $G(t) = a(t + b)^d$, $a, b, d \in \mathbb{R}$. Among many other paths, this allows a wide

range of both concave or convex but always monotone paths joining the endpoints at t and $t + 1$ that thus cover a class of potential true underlying functions that are much more general than the step-function or the straight line at the base of the Paasche-, Laspeyres or Divisia-Index. Choosing higher polynomials or periodic functions for G allows for a wide range of functions with local extrema as well - as they may occur in reality due to seasonal patterns. Clearly, H is related to G by only two additional parameters, but the possibility to take arbitrary powers of G makes this very flexible. A drawback of equation (20) is only, that, due to its functional form, $G = 0$ implies $H = 0$ and vice versa. It thus does not inform on how to treat zero values (cf. section 3.1 above).

The exactness for a wide range of functional forms is the main reason why I suggest the LMDI as a currently most reliable method. Furthermore, proceeding similarly leads to the result that both the original and logarithmic parametric Divisia are exact for a much more restricted set of functional forms only. This is an additional reason to suggest the LMDI as the method currently to be preferred. The Divisia-Index with $\alpha = \frac{1}{2}$, for example, is exact in case $\int_t^{t+1} G \frac{\partial H}{\partial \tau} d\tau = \frac{1}{2} \{G(t+1) + G(t)\} [H(t+1) - H(t)]$. Arguing as above that no terms with both arguments $t+1$ and t must occur on the right hand side, necessary conditions are either $G(t+1) = G(t)$ or $G(t)H(t+1) = G(t+1)H(t)$. This leads to G being a constant, which is also sufficient, or to $H = \kappa G$ for $\kappa \in \mathbb{R}$, which is more general, but by far not as general a condition as the one established for the LMDI. Similarly, the conditions for other decomposition methods can be investigated. The logarithmic Divisia-Index with $\alpha = \frac{1}{2}$, for example, is exact only for the trivial cases $G = 0$ or $H = \text{constant}$ and the Laspeyres- and Paasche-Indices are exact for $G = \text{constant}$ only.

I emphasize that the derivation above establishes exactness of the LMDI for a wide range of functions without knowing more details on concrete parameter values. This contrasts with Fernandez and Fernandez (2007), whose method is also exact for a wide range of functions (parametrized by one parameter), but the knowledge of the true value of the parameter is necessary and exactness can thus only be established given additional information (cf. footnote 6).

Besides establishing conditions for exactness of a certain decomposition method, additional information on their performance can be gained via investigation of conditions under which decomposition methods are good approximations. First, the exactness conditions for decomposition methods such as established above can be used to assess for which type of functional forms the respective method may not be exact but a good approximation. Second, in

case of small relative changes of the variables over one period, one could assess how well the different methods perform if applied to polynomial or other expansions of the underlying functions. Given the similarities in results for some methods in application to real data and in simulations (see below), this may not add much additional insight though. In addition, situations with small changes in all the variables involved are not of prime interest.

Finally, additional information can be gained via simulations, where the exact result is known due to knowledge of the underlying simulated functions. This allows a detailed comparison between the exact decomposition and any decomposition method for the specific case the simulation is based on.

Simulations might best be undertaken on a systematic basis, e.g. for types of functions that have some potential to capture essential features of the true underlying functions. Examples could be functional forms gained by moving averages or moving aggregations, as the true functions G and H have these properties: being reported as annual values, for each time τ , they actually refer to the average value of the variable of interest back over the time-length of a period (i.e. over $[\tau - 1, \tau]$) or the sum over this time-interval.

To address simulations systematically would be subject to future research. Here, I provide the results from a simulation for illustrative reasons only (see Appendix A for details). This simulation shows that the LMDI performs quite well and considerably better than other methods such as the original or logarithmic parametric Divisia indices with $\alpha = \frac{1}{2}$, especially for large time-intervals.

Besides these strategies to assess the performance of single decomposition methods, it is interesting to compare results of several methods when applied to the same data. Quite a number of such studies are available in the literature (see Liu and Ang 2007 for a recent and encompassing review). Generally, they find that indices based on the Laspeyres-Index differ considerably from various types of Divisia-Indices, which are more similar and sometimes even equal among themselves (Ang 1994, Ang et al. 1998, Choi and Ang 2003, Greening et al. 1997, Ang 2004, Ang and Liu 2007c). Only Greening et al. (1997) report that one of the Divisia-Indices chosen differs sometimes significantly from other Divisia-Index based methods. Ang (1995) looks at a Paasche-type index as well and finds that this is different from a Laspeyres- and Divisia-based index. I have calculated some further comparisons and find that for the data employed in Hammar and Löfgren (2001) and the data from in Chung and Rhee (2001) the several terms of the decomposition using the logarithmic Divisia-Index with $\alpha = \frac{1}{2}$, the Divisia-related decomposition method used in Hammar and Löfgren (2001) and the LMDI were almost identical. This is in line with the findings from the literature, that the various Divisia-Indices usually give similar results.

The fact that the LMDI performed better than other methods in the illustrative simulation for larger time intervals (cf. Appendix A) suggests that the similar performance of the logarithmic Divisia-Index with $\alpha = \frac{1}{2}$ with real data might depend on the availability of relatively short time-intervals. The differing results from the Laspeyres- and Paasche-Indices with respect to Divisia-based indices may be linked to the fact that the former two are only exact for the trivial case that $G = \text{constant}$. It is then intuitive that they perform worse than the Divisia-Indices. These points should however be systematically addressed in future research on optimal decomposition methods.

6 Conclusion

In this paper, I provided a firm footing for decomposition based on approximation of integrals and derivatives. This adds to the ongoing discussion of optimal decomposition methods by clarifying the underlying formalism. In consequence, the absence of a residual term and the ability to treat zero and negative values, which are the main arguments produced in favour of optimal decomposition methods, are seen in a new light. A non-zero residual term is a consequence of the (unavoidable) errors due to the underlying approximations and the problems of zero and negative values can be avoided by refraining from undertaking ill-defined operations.

This criticism also applies to the LMDI method currently favored in the literature. Both analytical treatment and a simulation comparing the LMDI to exactly known integrals and their decomposition show how its property of zero residual cannot be defended on sound formal grounds. In addition, the treatment of zero and negative values by the LMDI leads to wrong results.

Nevertheless, the analytical treatment and an illustrative simulation show that the LMDI often performs well and that it is exact for a wide range of functions. This makes the LMDI a favourable decomposition method. This is however no reason to apply the LMDI uncritically - still often lacking a sound basis, its adequacy has to be assured in each situation anew. But its exactness for a wide range of functions increases chances for the LMDI to be adequate. Furthermore, it is exact for a wider range of functions than other decomposition methods and it performs better in the simulation I undertook. Whether it performs better in simulations in general is, however, an open question to be addressed in future research. On the basis of its exactness properties I expect it to perform reasonably well in many simulations, though.

A drawback on the LMDI is the erroneous treatment of zero and negative values that can in principle be corrected, but for which I cannot yet provide a reliable solution. Paths on how to pursue possible solutions can be indicated,

namely to look for zero value treatment that performs similar as the LMDI in simulations and under approximation. The strategy to identify optimal zero value treatment based on exactness conditions however fails for the LMDI.

More generally, to establish a sounder basis for the choice of optimal decomposition methods, I propose to refrain from the criteria emphasised in the literature such as a zero residual and suggest to pursue three strategies, namely assessment of conditions under which the methods proposed are exact, under which they provide good approximations and how they perform under simulations. This should be systematically addressed in future research. New decomposition methods could even be designed on the base of exactness or approximation criteria for a particularly wide range of functions or for certain types of functions that are particularly important for decomposition, such as moving averages or moving aggregates of functions with seasonal periodicity, for example. This could also provide guidance for treatment of potential problems with zero values for certain decomposition methods, if not avoided from the beginning by avoiding ill-defined operations. While exactness and approximation properties thus may be first in assessing the performance of different methods, simulations provide valuable additional insight, especially if based on types of functions that cannot be treated analytically.

In this light, it would be ideal to find decomposition methods with some linearity properties, such that the decomposition of linear combinations of functions could be traced back to the decomposition of the single terms. In this case, methods optimal for exact decomposition of certain sets of basis functions that span the complete space of functions one is interested in could be sought. Examples of such basis functions would be trigonometric functions, Legendre-polynomials, and others. I suspect, however, that such linearity constraints would be too strong in restraining potential forms of decomposition methods, such that this path may not lead to better methods - but also this should be assessed systematically.

A further strategy to get more information on the various decomposition methods is to systematically collect data sets and to compare the results of different decomposition methods on those - not only by comparing published studies, which is very informative (Liu and Ang 2007), but also by redoing calculations employing the same decomposition methods to all the different data, thus increasing comparability. It may turn out that differences are not that big for some classes of methods (such as those based on the Divisia-Index). This may help to assess how much effort to find better methods is adequate from a practitioners point of view. The LMDI may be good enough for most cases, for example. Furthermore, such systematic comparison can reveal additional information for which type of data which methods might

be similar or differ. This could also inform the search for types of functions where certain methods are particularly adequate.

Appendix

A An illustrative simulation

In this appendix, I present an illustrative, albeit somewhat arbitrary example of a simulation to compare the performance of different decomposition methods. For time t , I assume the following:

$$E = \frac{10}{t+1}, \quad Y = \frac{t+2}{20} \quad \Rightarrow \quad I := \frac{E}{Y} = \frac{200}{(t+1)(t+2)}. \quad (21)$$

$$\frac{\partial}{\partial t} E = -\frac{10}{(t+1)^2}, \quad \frac{\partial}{\partial t} Y = \frac{1}{20}, \quad \frac{\partial}{\partial t} I = -\frac{200(2t+3)}{(t+1)^2(t+2)^2}. \quad (22)$$

where E is the total energy consumption, Y the total output and I is the energy intensity. Employing equations (5), this gives

$$\begin{aligned} & \frac{-10\Delta T}{(T+\Delta T+1)(T+1)} - 10 \ln \left(\frac{(T+\Delta T+1)(T+2)}{(T+\Delta T+2)(T+1)} \right) \\ \text{and} \quad & 10 \ln \left(\frac{(T+\Delta T+1)(T+2)}{(T+\Delta T+2)(T+1)} \right) \end{aligned} \quad (23)$$

for the exact intensity- (G is Y and H is I) and size (total output)-term (G is I and H is Y) in the decomposition, where ΔT is the difference between subsequent periods, above usually chosen to equal 1. Using (11) it leads to

$$\begin{aligned} & \frac{-10\Delta T \ln \left(\frac{(T+1)(T+2)}{(T+\Delta T+1)(T+\Delta T+2)} \right)}{(T+\Delta T+1)(T+1) \ln \left(\frac{T+1}{T+\Delta T+1} \right)} \\ \text{and} \quad & \frac{-10\Delta T \ln \left(\frac{T+\Delta T+2}{T+2} \right)}{(T+\Delta T+1)(T+1) \ln \left(\frac{T+1}{T+\Delta T+1} \right)} \end{aligned} \quad (24)$$

for the LMDI-intensity- and -size-contribution. It can readily be seen that this does not coincide in general, i.e. that the LMDI although exhibiting a zero residual does not lead to the true decomposition (cf. section 5 for the conditions for the LMDI to equal the exact decomposition - they are not met here). If this is calculated for some range of t , say, from 1 to 30, with

$\Delta T = 1$, the LMDI is almost equal to the exact decomposition and deviations lie below 1 percent. The same is true for the Divisia decompositions with $\alpha = \frac{1}{2}$, for example. If, however, the change between the single periods is bigger, e.g. reflected by choosing $\Delta T = 29$, differences become significant and the LMDI deviates from the exact values by 5 to 10 percent. In this case, the LMDI performs much better than the Divisia-Indices, though.

References

- Ang, B.W., 1994. Decomposition of industrial energy consumption - The energy intensity approach. *Energy Economics* 16(3): 163-174.
- Ang, B.W., 1995. Decomposition methodology in industrial energy demand analysis, *Energy* 20(11), 1081-1095.
- Ang, B.W., 2004. Decomposition analysis for policymaking in energy: which is the preferred method? *Energy Policy* 32, 1131-1139.
- Ang, B.W., Choi, K., 1997. Decomposition of aggregate energy and gas emission intensities for industry: a refined Divisia Index method. *The Energy Journal* 18(3), 59-73.
- Ang, B.W., Liu, N., 2001. A new energy decomposition method: perfect in decomposition and consistent in aggregation. *Energy* 26, 537-548.
- Ang, B.W., Liu, N., 2007a. Handling zero values in the logarithmic mean Divisia index decomposition approach. *Energy Policy* 35(1): 238-246.
- Ang, B.W., Liu, N., 2007b. Negative value problems of the logarithmic mean Divisia index decomposition approach. *Energy Policy* 35(1): 739-742.
- Ang, B.W., Liu, N., 2007c. Energy decomposition analysis: IEA model versus other methods. *Energy Policy* 35(5): 1426-1432.
- Ang, B.W., Zhang, F.Q., Choi, K.H., 1998. Factorizing changes in energy and environmental indicators through decomposition. *Energy* 23, 489-495.
- Ang, B.W., Liu, F.L., Chew, E.P., 2003. Perfect Decomposition Techniques in Energy and Environmental Analysis. *Energy Policy* 31, 1561-1566.
- Balk, B.M., 2005, Divisia price and quantity indices: 80 years after. *Statistica Neerlandica* 59(2)L: 119-158.
- Barnett, W.A., Choi, K.H., 2006. Operational identification of the complete class of superlative index numbers: an application of Galois theory. *Journal of Mathematical Economics*, in press.
- Boyd, G.A., Roop, J.M., 2005. A note on fisher ideal index decomposition for structural change in energy intensity. *The Energy Journal* 25(1); 87-101.
- Choi, K.H., Ang, B.W., 2003. Decomposition of aggregate energy intensity changes in two measures: ratio and difference. *Energy Economics* 25(6): 615-624.

- Chung, H., Rhee, H., 2001. A residual-free decomposition of the sources of carbon dioxide emissions: a case of the Korean Industries. *Energy* 26, 15-30.
- Cole, M.A., Elliot, R.J.R. and Shimamoto, K., 2005. A Note on Trends in European Industrial Pollution Intensities: A Divisia Index Approach. *The Energy Journal* 26(3): 61-73.
- Dietzenbacher, E., Los, B., 1998. Structural decomposition techniques: Sense and sensitivity. *Economic Systems Research* 10(4): 307-323.
- Fernandez, E., Fernandez, P., 2007. An extension to Sun's decomposition methodology: The Path Based approach. *Energy Economics* (in press), doi: 10.1016/j.eneco.2007.01.004.
- Greening, L.A., Davis, W.B., Schipper, L. and Khrushch, M., 1997. Comparison of six decomposition methods: application to aggregate energy intensity for manufacturing in 10 OECD countries. *Energy Economics* 19: 375-390.
- Hammar, H., Löfgren, Å., 2001. The determinants of sulfur emissions from oil consumption in Swedish manufacturing industry, 1976-1995. *The Energy Journal* 22(2), 107-126.
- Hoekstra, R., van der Bergh, J.J.C.J.M., 2003. Comparing structural and index decomposition analysis. *Energy Economics* 25: 39-64.
- Hulten, C.R., 1973. Divisia Index Numbers. *Econometrica* 41(6): 1017-1025.
- Liu, N., Ang, B.W., 2007. Factors shaping aggregate energy intensity trend for industry: Energy intensity versus product mix. *Energy Economics* 29: 609-635.
- Liu, X.Q., Ang, B.W., Ong, H.L., 1992. The application of the Divisia index to the decomposition of changes in industrial energy consumption. *The Energy Journal* 13(4): 161-177.
- Muller, A., 2006. Clarifying poverty decomposition. *Scandinavian Working Papers in Economics* No. 217 (submitted).
- Rose, A., Casler, S., 1996. Input-output structural decomposition analysis: A critical appraisal. *Economic Systems Research* 8(1): 33-62.
- Sun, J.W., 1998. Changes in energy consumption and energy intensity: a complete decomposition model. *Energy Economics* 20, 85-100.
- Trivedi, P.K., 1981. Some discrete approximations to Divisia integral indices. *International Economic Review* 22(1), 71-77.
- Vogt, A., 1978. Divisia Indices on Different Paths. In: Eichhorn, Henn, Opitz and Shepard (eds), *Theory and Application of Economic Indices*, Physica.
- Wood, R., Lenzen, M., 2006. Zero-value problems for the logarithmic mean Divisia index decomposition method. *Energy policy* 34(12): 1326-1331.