

value to the other category of income of population subgroups on which has revived our interest in this distribution of Income Network, contract caveat applies.

which has revived our interest in this distribution of Income Network, contract caveat applies.

on structures. *Int. J. Game Theory* **3**,

*l'inégalité des revenus à l'aide de la théorie des Sfax* (1992)

deced contributions and hierarchies of

ition of income by factor components.

ian, P. (eds.) *The New Palgrave: Game*

value. In: Roth, A. (ed.) *The Shapley Value*. Cambridge University Press (1988)

*Econometrica* **57**, 589–614 (1989)

6–442 (1976)

2–111 (1976)

urce: a new approach and applications

R., Moeschlin, O. (eds.) *Mathematical*

*e Notes in Economics and Mathematics*

Lloyd S. Shapley. Cambridge University

, Tucker, A.W. (eds.) *Contributions to Mathematics Studies*. Princeton University

inequality measures. *Econometrica* **48**,

ponents. *Econometrica* **50**, 193–211

n subgroups. *Econometrica* **51**, 1369–

its of income. *Econ. Lett.* **2**, 197–244

games with coalition structure. *Games*

## Decomposition procedures for distributional analysis: a unified framework based on the Shapley value

Anthony F. Shorrocks

Received: 19 May 2011 / Accepted: 26 May 2011 / Published online: 7 January 2012  
© Springer Science+Business Media, LLC 2012

**Keywords** Inequality · Poverty · Decomposition · Shapley value

### 1 Introduction

Decomposition techniques are used in many fields of economics to help disentangle and quantify the impact of various causal factors. Their use is particularly widespread in studies of poverty and inequality. In poverty analysis, most practitioners now employ decomposable poverty measures—especially the Foster et al. [10] family of indices—which enable the overall level of poverty to be allocated among subgroups of the population, such as those defined by geographical region, household composition, labour market characteristics or education level. Recent examples include Grootaert [11], Szekely [26], Thorbecke and Jung [28]. Other dynamic decomposition procedures are used to examine how economic growth contributes to a reduction in poverty over time, and to assess the extent to which the impact of growth is reinforced, or attenuated, by changes in income inequality: see for example, Ravallion and Huppi [20], Datt and Ravallion [6] and Tsui [29]. In the context of income inequality, decomposition techniques enable researchers to distinguish the “between-group” effect due to differences in average incomes across subgroups (males and females, say), from the “within-group” effect due to inequality within the population subgroups. Decomposition techniques have also been developed in order to measure the importance of components of income such as earnings or transfer payments.

I dedicate this paper to the memory of my mother, Vera Florence Shorrocks.

A. F. Shorrocks (✉)  
Global Economic Perspectives Ltd., London, UK  
e-mail: Shorrocks@wider.unu.edu

Despite their widespread use, these procedures have a number of shortcomings which have become increasingly evident as more sophisticated models and econometrics are brought to bear on distributional questions. Four broad categories of problems can be distinguished. First, the contribution assigned to a specific factor is not always interpretable in an intuitively meaningful way. As Chantreuil and Trannoy [4] and Morduch and Sinclair [15] point out, this is particularly true of the decomposition by income components proposed by Shorrocks [24]. In other cases, the interpretation commonly given to a component may not be strictly accurate. Foster and Shneyerov [8], for example, question the conventional interpretation of the between-group term in the decomposition of inequality by subgroups.

The second problem with conventional procedures is that they often place constraints on the kinds of poverty and inequality indices which can be used. Only certain forms of indices yield a set of contributions that sum up to the amount of poverty or inequality that requires explanation. Similar methods applied to other indices require the introduction of a vaguely defined residual or "interaction" term in order to maintain the decomposition identity. The best known example is the subgroup decomposition of the Gini coefficient, which has exercised the minds of many authors including Pyatt [19] and Lambert and Aronson [14].

A less familiar, but potentially much more serious, problem concerns the limitations placed on the types of contributory factors which can be considered. Subgroup decompositions can handle situations in which the population is partitioned on the basis of a single attribute, but have difficulty identifying the relevant contributions in multi-variate decompositions. Nor is there any established method of dealing with mixtures of factors, such as a simultaneous decomposition by subgroups (into, say, males and females) and income components (say, earnings and unearned income). As more sophisticated models are used to analyse distributional issues, these limitations have become increasingly evident. The studies by Cowell and Jenkins [5], Jenkins [12], Bourguignon et al. [3], and Bouillon et al. [1] illustrate the range of problems faced by those trying to apply current techniques to complex distributional questions.

The final criticism of current decomposition methods is that the individual applications are viewed as different problems requiring different solutions. No attempt has been made to integrate the various techniques within a common overall framework. This is the main reason why it is impossible at present to combine decompositions by subgroups and income components. Yet the individual applications share certain features and objectives which enable a common structure to be formulated. Let  $I$  represent an aggregate statistical indicator, such as the overall level of poverty or inequality, and let  $X_k$ ,  $k = 1, 2, \dots, m$ , denote a set of contributory factors which together account for the value of  $I$ . Then we can write

$$I = f(X_1, X_2, \dots, X_m), \quad (1.1)$$

where  $f(\cdot)$  is a suitable aggregator function representing the underlying model. The objective in all types of decomposition exercises is to assign contributions  $C_k$  to each of the factors  $X_k$ , ideally in a manner that allows the value of  $I$  to be expressed as the sum of the factor contributions.

The aim of this paper is to offer a solution to this general decomposition problem and to compare the results with the specific procedures currently applied to a number

es have a number of shortcom-  
more sophisticated models and  
questions. Four broad categories  
ution assigned to a specific factor  
ningful way. As Chantreuil and  
ut, this is particularly true of the  
y Shorrocks [24]. In other cases,  
nt may not be strictly accurate.  
ie conventional interpretation of  
equality by subgroups.

ires is that they often place con-  
ndices which can be used. Only  
is that sum up to the amount of  
imilar methods applied to other  
ed residual or "interaction" term.  
The best known example is the  
which has exercised the minds of  
I Aronson [14].

us, problem concerns the limita-  
tions which can be considered. Subgroup  
population is partitioned on the  
ififying the relevant contributions  
established method of dealing with  
position by subgroups (into, say,  
earnings and unearned income).  
e distributional issues, these lim-  
itudes by Cowell and Jenkins [5],  
et al. [1] illustrate the range of  
techniques to complex distributional

ods is that the individual applica-  
fferent solutions. No attempt has  
in a common overall framework.  
sent to combine decompositions  
individual applications share certain  
structure to be formulated. Let  $I$   
s the overall level of poverty or  
et of contributory factors which  
ite

$$(1.1)$$

enting the underlying model. The  
o assign contributions  $C_k$  to each  
e value of  $I$  to be expressed as the

s general decomposition problem  
res currently applied to a number

of distributional questions. In broad terms, the proposed solution considers the marginal effect on  $I$  of eliminating each of the contributory factors in sequence, and then assigns to each factor the average of its marginal contributions in all possible elimination sequences. This procedure yields an exact additive decomposition of  $I$  into  $m$  contributions.

Posing the decomposition issue in the general way indicated by Eq. 1.1 highlights formal similarities with problems encountered in other areas of economics and econometrics. Of particular relevance to this paper is the classic question of cooperative game theory, which asks how a certain amount of output (or costs) should be allocated among a set of contributors (or beneficiaries). The Shapley value [22] provides a popular answer to this question. The proposed solution to the general decomposition problem turns out to be formally equivalent to the Shapley value, and is therefore referred to as the *Shapley decomposition*. Rongve [21] and Chantreuil and Trannoy [4] have both applied the Shapley value to the decomposition of inequality by income components, but fail to realise that a similar procedure can be used in all forms of distributional analysis, regardless of the complexity of the model, or the number and types of factors considered. Indeed, the procedure can be employed in all areas of applied economics whenever one wishes to assess the relative importance of the explanatory variables.

The paper begins with a description of the general decomposition problem and the proposed solution based on the Shapley value. Section 3 shows how the procedure may be applied to three issues concerned with poverty: the effects of growth and redistribution on changes in poverty; the conventional application of decomposable poverty indices; and the impact of population shifts and changes in within-group poverty on the level of poverty over time.

Section 4 looks in more detail at the features of the Shapley decomposition in the context of a hierarchical model in which groups of factors may be treated as single units. This leads to a discussion of the two-stage Shapley procedure associated with the Owen value [18]. A number of results in this section establish the conditions under which the Shapley and Owen decompositions coincide, and indicate several ways of simplifying the calculation of the factor contributions. These results are then used to generate the Shapley solution to the multi-variate decomposition of poverty by subgroups, a problem which has not been solved before.

In Sections 5 and 6, attention turns to inequality analysis, beginning with decomposition by subgroups using the Entropy and Gini measures of inequality. This is followed by a discussion of the application of the Shapley rule to decomposition by source of income.

The main purpose of these applications is to see how the Shapley procedure compares with existing techniques in the context of a variety of standard decomposition problems. The overall results are encouraging. In all cases, the Shapley decomposition either replicates current practice or (arguably) provides a more satisfactory method of assigning contributions to the explanatory factors. However, the greatest attraction of the procedure proposed in this paper is that it overcomes all four of the categories of problems associated with present techniques. As a consequence, it offers a unified framework capable of handling any type of decomposition exercise. After summarising the principal findings of this paper the Appendix section briefly discusses the wide range of potential applications to issues which have not previously been considered candidates for decomposition analysis.

## 2 A general framework for decomposition analysis

Consider a statistical indicator  $I$  whose value is completely determined by a set of  $m$  contributory factors,  $X_k$ , indexed by  $k \in K = \{1, 2, \dots, m\}$  so that we may write

$$I = f(X_1, X_2, \dots, X_m), \quad (2.1)$$

where  $f(\cdot)$  describes the underlying model. In the applications examined later, the indicator  $I$  will represent the overall level of poverty or inequality in the population, or the change in poverty over time. The factor  $X_k$  may refer to a conventional scalar or vector variable, but other interpretations are possible and often desirable; for the moment it is best regarded as a loose descriptive label capturing influences like “uncertain returns to investments”, “differences in household composition” or “supply-side effects”.

In what follows, we imagine scenarios in which some or all of the factors are eliminated, and use  $F(S)$  to signify the value that  $I$  takes when the factors  $X_k$ ,  $k \notin S$ , have been dropped. As each of the factors is either present or absent, it is convenient to characterise the *model structure*  $\langle K, F \rangle$  in terms of the set of factors (or, more accurately, “factor indices”),  $K$ , and the function  $F : \{S | S \subseteq K\} \rightarrow \mathbb{R}$ . Since the set of factors completely accounts for  $I$ , it will also be convenient to assume throughout that  $F(\emptyset) = 0$ : in other words, that  $I$  is zero when all the factors are removed.<sup>1</sup>

A *decomposition* of  $\langle K, F \rangle$  is a set of real values  $C_k$ ,  $k \in K$ , indicating the contribution of each of the factors. A *decomposition rule*  $C$  is a function which yields a set of factor contributions

$$C_k = C_k(K, F), \quad k \in K, \quad (2.2)$$

for any possible model  $\langle K, F \rangle$ .

In seeking to construct a decomposition rule, several desiderata come to mind. First, that it should be *symmetric* (or *anonymous*) in the sense that the contribution assigned to any given factor should not depend on the way in which the factors are labelled or listed. Secondly, that the decomposition should be *exact* (and additive), so that

$$\sum_{k \in K} C_k(K, F) = F(K), \quad (2.3)$$

for all  $\langle K, F \rangle$ . When condition (2.3) is satisfied, it is meaningful to speak of the proportion of observed inequality or poverty attributable to factor  $k$ .

It is also desirable that the contributions of the factors can be interpreted in an intuitively appealing way. In this respect, the most natural candidate is the rule which equates the contribution of each factor to its (first round) marginal impact

$$M_k(K, F) = F(K) - F(K \setminus \{k\}), \quad k \in K. \quad (2.4)$$

This decomposition rule is symmetric, but will not normally yield an exact decomposition. A second possibility is to consider the marginal impact of each of the factors when they are eliminated in sequence. Let  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_m)$  indicate

---

<sup>1</sup>If  $F(\cdot)$  is derived from an econometric model, this constraint will usually mean that one of the factors represents the unexplained residuals. When the constraint is not satisfied automatically,  $I$  can always be renormalised so that it measures the “surplus” due to the identified factors.

npletely determined by a set of  $m$ , ...,  $m$ ) so that we may write

$$\dots, m), \quad (2.1)$$

applications examined later, the ty or inequality in the population,  $X_k$  may refer to a conventional are possible and often desirable; riptive label capturing influences ces in household composition" or

h some or all of the factors are at  $I$  takes when the factors  $X_k$ , s is either present or absent, it is ,  $F$ ) in terms of the set of factors he function  $F : \{S | S \subseteq K\} \rightarrow \mathbb{R}$ . will also be convenient to assume ' is zero when all the factors are

values  $C_k$ ,  $k \in K$ , indicating the n rule  $C$  is a function which yields

$$\vdash K, \quad (2.2)$$

several desiderata come to mind. in the sense that the contribution i the way in which the factors are n should be *exact* (and additive),

$$\vdash K, \quad (2.3)$$

it is meaningful to speak of the utable to factor  $k$ .

factors can be interpreted in an natural candidate is the rule which round) marginal impact

$$\vdash \}, \quad k \in K. \quad (2.4)$$

not normally yield an exact de- the marginal impact of each of . Let  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_m)$  indicate

t will usually mean that one of the factors not satisfied automatically,  $I$  can always identified factors.

the order in which the factors are removed, and let  $S(\sigma_r, \sigma) = \{\sigma_i | i > r\}$  be the set of factors that remain after factor  $\sigma_r$  has been eliminated. Then the marginal impacts are given by

$$C_k^\sigma = F(S(k, \sigma) \cup \{k\}) - F(S(k, \sigma)) = \Delta_k F(S(k, \sigma)), \quad k \in K, \quad (2.5)$$

where

$$\Delta_k F(S) \equiv F(S \cup \{k\}) - F(S), \quad S \subseteq K \setminus \{k\}, \quad (2.6)$$

is the marginal effect of adding factor  $k$  to the set  $S$ . Using the fact that  $S(\sigma_r, \sigma) = S(\sigma_{r+1}, \sigma) \cup \{\sigma_{r+1}\}$  for  $r = 1, 2, \dots, m-1$ , we deduce

$$\begin{aligned} \sum_{k \in K} C_k^\sigma &= \sum_{r=1}^m C_{\sigma_r}^\sigma = \sum_{r=1}^m [F(S(\sigma_r, \sigma) \cup \{\sigma_r\}) - F(S(\sigma_r, \sigma))] \\ &= F(S(\sigma_1, \sigma) \cup \{\sigma_1\}) - F(S(\sigma_m, \sigma)) = F(K) - F(\emptyset) = F(K). \end{aligned} \quad (2.7)$$

The decomposition (2.5) is therefore exact. However, the value of the contribution assigned to any given factor depends on the order in which the factors appear in the elimination sequence  $\sigma$ , so the factors are not treated symmetrically. This "path dependence" problem may be remedied by considering the  $m!$  possible elimination sequences, denoted here by the set  $\Sigma$ , and by computing the expected value of  $C_k^\sigma$  when the sequences in  $\Sigma$  are chosen at random. This yields the decomposition rule  $C^S$  given by

$$\begin{aligned} C_k^S(K, F) &= \frac{1}{m!} \sum_{\sigma \in \Sigma} C_k^\sigma = \frac{1}{m!} \sum_{\sigma \in \Sigma} \Delta_k F(S(k, \sigma)) \\ &= \sum_{s=0}^{m-1} \sum_{\substack{S \subseteq K \setminus \{k\} \\ |S|=s}} \frac{1}{m!} \sum_{\substack{\sigma \in \Sigma \\ S(k, \sigma)=S}} \Delta_k F(S) \\ &= \sum_{s=0}^{m-1} \sum_{\substack{S \subseteq K \setminus \{k\} \\ |S|=s}} \frac{(m-1-s)!s!}{m!} \Delta_k F(S). \end{aligned} \quad (2.8)$$

Using  $\pi(s, m-1) = (m-1-s)!s!/m!$  to indicate the relevant probability,<sup>2</sup> Eq. 2.8 is expressed more succinctly as

$$C_k^S(K, F) = \sum_{S \subseteq K \setminus \{k\}} \pi(|S|, |K \setminus \{k\}|) \Delta_k F(S) = \mathcal{E}_{S \subseteq K \setminus \{k\}} \Delta_k F(S), \quad k \in K, \quad (2.9)$$

where  $\mathcal{E}_{S \subseteq L}$  is the expectation taken with respect to the subsets of  $L$ .

From Eq. 2.7 it is clear that  $C^S$  is an exact decomposition rule, and also one which treats the factors symmetrically. Furthermore, the contributions can be interpreted as the expected marginal impact of each factor when the expectation is taken over all the possible elimination paths.

Expression 2.8 will be familiar to many readers, since it corresponds to the Shapley value for the cooperative game in which "output" or "surplus"  $F(K)$  is shared amongst the set of "inputs" or "agents"  $K$  (see, for example, Moulin [16], Chapter 5). The application to distributional analysis is quite different from the context in which

<sup>2</sup> $\pi(|S|, |M|)$  is the probability of randomly selecting the subset  $S$  from  $M$ , given that all subset sizes from 0 to  $|M|$  are equally likely.

the Shapley value was conceived, and the results therefore need to be reinterpreted. Nevertheless, it seems convenient and appropriate to refer to Eq. 2.8 as the *Shapley decomposition rule*.<sup>3</sup>

### 3 Applications of the Shapley Decomposition to Poverty Analysis

To illustrate how the Shapley decomposition operates in practice, this section looks at three simple applications to poverty analysis.

#### 3.1 The impact of growth and redistribution on the change in poverty

An important issue in development economics concerns the extent to which economic growth helps to alleviate poverty. With a fixed real poverty standard, growth is normally expected to raise the incomes of some of the poor, thereby reducing the value of any conventional poverty index. However, this tendency can be moderated, or even reversed, if economic growth is accompanied by redistribution in the direction of increased inequality.

Datt and Ravallion [6] suggest a method for separating out the effects of growth and redistribution on the change in poverty between two points of time. Given a fixed poverty line, the poverty level at time  $t$  ( $t = 1, 2$ ) may be expressed as a function  $P(\mu_t, L_t)$  of mean income  $\mu_t$  and the Lorenz curve  $L_t$ . Denoting the growth factor by  $G = \mu_2/\mu_1 - 1$ , and the redistribution factor by  $R = L_2 - L_1$ ,<sup>4</sup> the problem becomes one of identifying the contributions of growth  $G$  and redistribution  $R$  in the decomposition of

$$\begin{aligned}\Delta P &= P(\mu_2, L_2) - P(\mu_1, L_1) \\ &= P(\mu_1(1+G), L_1 + R) - P(\mu_1, L_1) = F(G, R).\end{aligned}\quad (3.1)$$

Figure 1 illustrates the basic structure of the Shapley decomposition for this example, which is particularly simple given that there are just two factors,  $G$  and  $R$ , and hence only two possible elimination sequences. Eliminating  $G$  before  $R$  produces the path portrayed on the left, with the marginal contribution  $F(G, R) - F(R)$  for the growth factor, and the contribution  $F(R)$  for the redistribution effect. Repeating the exercise for the right-hand path, and then averaging the results, yields the Shapley contributions

$$\begin{aligned}C_G^S &= \frac{1}{2}[F(G, R) - F(R) + F(G)] \\ C_R^S &= \frac{1}{2}[F(G, R) - F(G) + F(R)]\end{aligned}\quad (3.2)$$

<sup>3</sup>Several characterisations of the Shapley value are available, and may be reinterpreted in the present framework. For example, Eq. 2.8 is the only symmetric and exact decomposition rule which, for each  $k$ , yields contributions  $C_k(K, F)$  that depend only on the set of marginal effects  $\{\Delta_k F(S) | S \subseteq K \setminus \{k\}\}$  relating to factor  $k$  [30].

<sup>4</sup>This is a slight abuse of notation, as the growth and redistribution *factors* ought to be distinguished from the *variables* representing growth and redistribution. However, in this example, the growth and redistribution factors are eliminated by setting  $G$  and  $R$  equal to zero, so no serious confusion arises. Factors and variables are distinguished more carefully in later sections.

before need to be reinterpreted. I refer to Eq. 2.8 as the *Shapley*

## Poverty Analysis

In practice, this section looks

change in poverty

cerns the extent to which economic real poverty standard, growth of the poor, thereby reducing the his tendency can be moderated, aned by redistribution in the

rating out the effects of growth en two points of time. Given a ) may be expressed as a function curve  $L_t$ . Denoting the growth factor by  $R = L_2 - L_1$ ,<sup>4</sup> the s of growth  $G$  and redistribution

$$, L_1) = F(G, R). \quad (3.1)$$

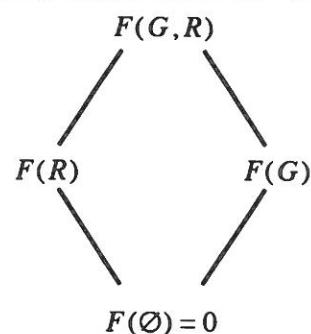
Shapley decomposition for this e are just two factors,  $G$  and  $R$ . Eliminating  $G$  before  $R$  produces tribution  $F(G, R) - F(R)$  for the istribution effect. Repeating the g the results, yields the Shapley

$$\begin{aligned} & \cdot F(G) \\ & \cdot F(R) \end{aligned} \quad (3.2)$$

nd may be reinterpreted in the present act decomposition rule which, for each marginal effects  $\{\Delta_k F(S) | S \subseteq K \setminus \{k\}\}$

ation factors ought to be distinguished never, in this example, the growth and to zero, so no serious confusion arises. sections.

**Fig. 1** The Shapley decomposition for the growth and redistribution components of the change in poverty



When growth is absent,  $G$  takes the value 0 and the change in poverty becomes

$$F(R) = P(\mu_1, L_2) - P(\mu_1, L_1) = D_L P(\mu_1, L), \quad (3.3)$$

where  $D_L P(\mu, L) = P(\mu, L_2) - P(\mu, L_1)$  indicates the rise in poverty due to a shift in the Lorenz curve from  $L_1$  to  $L_2$ , holding mean income constant at  $\mu$ . Conversely, eliminating the redistribution factor by setting  $R = 0$  yields

$$F(G) = P(\mu_2, L_1) - P(\mu_1, L_1) = D_\mu P(\mu, L_1), \quad (3.4)$$

where  $D_\mu P(\mu, L) = P(\mu_2, L) - P(\mu_1, L)$ , is the rise in poverty due to a change in mean income from  $\mu_1$  to  $\mu_2$ , with the fixed Lorenz curve  $L$ . The Shapley contributions for the growth and redistribution effects are therefore given by

$$\begin{aligned} C_G^S &= \frac{1}{2}[F(G, R) - F(R) + F(G)] \\ &= \frac{1}{2}[P(\mu_2, L_2) - P(\mu_1, L_2) + P(\mu_2, L_1) - P(\mu_1, L_1)] \\ &= \frac{1}{2}[D_L P(\mu_1, L_1) + D_L P(\mu_2, L_2)] \\ C_R^S &= \frac{1}{2}[D_L P(\mu_1, L) + D_L P(\mu_2, L)]. \end{aligned} \quad (3.5)$$

These contributions sum up, as expected, to the overall change in poverty, and have intuitively appealing interpretations. The growth component,  $C_G^S$ , indicates the rise in poverty due to a shift in mean income from  $\mu_1$  to  $\mu_2$ , averaged with respect to the Lorenz curves prevailing in the base and final years, while the redistribution effect,  $C_R^S$ , represents the average impact of the change in the distribution of relative incomes, with the average taken with respect to the mean income levels in the two periods.

Despite the attractions of the Shapley decomposition values given in Eq. 3.5, these are not the contributions proposed by Datt and Ravallion [6]. Instead, they associate the growth and redistribution effects with the marginal change in poverty starting

from the base year situation. This yields the contributions  $C_G = D_\mu P(\mu, L_1)$  and  $C_R = D_L P(\mu_1, L)$ .<sup>5</sup> These do not sum to the observed change in poverty, so Datt and Ravallion are obliged to introduce a residual term  $E$  into their decomposition equation

$$\Delta P = C_G + C_R + E. \quad (3.6)$$

They acknowledge the criticisms which can be levelled against the residual component, and note that it can be made to vanish by averaging over the base and final years, as is done in Eq. 3.5. However, this solution is rejected as being arbitrary ([6], footnote 3). Far from being arbitrary, the above analysis suggests that this is exactly the outcome which results from applying a systematic decomposition procedure to the growth-redistribution issue. Furthermore, the general framework outlined in Section 2 offers the chance of extending the analysis to cover not only changes in the poverty line, but also more disaggregated influences such as changes in mean incomes and income inequality within the modern and traditional sectors.

### 3.2 Decomposable poverty indices

Another standard application of decomposition techniques involves the use of decomposable poverty indices. When assigning contributions to subgroups of the population, such indices enable the overall degree of poverty,  $P$ , to be written

$$P = \sum_{k=1}^m v_k P_k, \quad (3.7)$$

where  $v_k$  and  $P_k$  respectively indicate the population share and poverty level associated with subgroup  $k \in K = \{1, 2, \dots, m\}$ . Indices with this property—especially the family of measures proposed by Foster et al. [10]—are nowadays used routinely to study the way in which differences according to region, household size, age, and education attainment contribute to the overall level of poverty.

In many respects, this is the simplest and most clear-cut application of decomposition techniques. Suppose, for instance, that the population is partitioned into  $m$  regions. Then factor  $k$  can be interpreted as “poverty within region  $k$ ”, and the question of interest is the contribution which this factor makes to poverty in the whole country. Adopting the notation of Section 2, the model structure  $\langle K, F \rangle$  is defined by

$$F(S) = \sum_{k \in S} v_k P_k, \quad (3.8)$$

so

$$\Delta_k F(S) = v_k P_k, \quad \text{for all } S \subseteq K \setminus \{k\}. \quad (3.9)$$

Since eliminating poverty in region  $k$  reduces aggregate poverty by the amount  $v_k P_k$  regardless of the order in which the regions are considered, it follows from Eqs. 2.4

---

<sup>5</sup>These terms correspond to the first round marginal effects; in other words,  $C_G = M_G(\{G, R\}, F)$  and  $C_R = M_R(\{G, R\}, F)$  in the notation of Eq. 2.4.

or equivalently, in the notation of Eq. 3.7,

$$C_k^S(K \cup L, F) = \frac{1}{2} v_k P_k \quad (3.17)$$

Thus, in this two attribute example, each region is assigned exactly half the contribution that would be obtained in a decomposition by region alone. A similar result applies to the age group factors. More generally, in a simultaneous decomposition by  $n$  attributes, each factor is allocated one  $n$ th of the contribution obtained in the single attribute decomposition.<sup>6</sup>

This result will be comforting to those who use decomposable poverty indices, for it shows that nothing is lost by looking at each attribute in isolation; the outcomes of multi-attribute decompositions can be calculated immediately from the series of single attribute results, and the *relative* importance of different subgroups remains the same, regardless of the number of attributes considered.

### 3.3 Changes in poverty over time

Decomposable poverty indices can also be used to identify the subgroup contributions to poverty changes over time. If  $v_{kt}$  and  $P_{kt}$  represent the population share and poverty level of subgroup  $k \in K$  at time  $t$  ( $t = 1, 2$ ), Eq. 3.7 yields

$$\Delta P = \sum_{k \in K} [v_{k2} P_{k2} - v_{k1} P_{k1}] . \quad (3.18)$$

The aim here is to account for the overall change in poverty,  $\Delta P$ , in terms of changes in poverty within subgroups,  $\Delta P_k = P_{k2} - P_{k1}$ ,  $k \in K$ , and the population shifts between subgroups,  $\Delta v_k = v_{k2} - v_{k1}$ ,  $k \in K$ .

The subgroup poverty values can be changed independently, so the poverty change factors may be indexed by the set  $K_p = \{p_k, k \in K\}$ . However the population shifts necessarily sum to zero. To avoid complications at this stage, the population shift factors will be treated as a single composite factor, denoted by  $K_s$ . The model structure  $\langle K_p \cup \{K_s\}, F \rangle$  is then given by

$$F(S \cup T) = \sum_{k \in K} [v_k(T) P_k(S) - v_{k1} P_{k1}], \quad S \subseteq K_p, \quad T \subseteq \{K_s\}, \quad (3.19)$$

where

$$P_k(S) = \begin{cases} P_{k2} & \text{if } p_k \in S \\ P_{k1} & \text{if } p_k \notin S \end{cases}; \quad v_k(T) = \begin{cases} v_{k2} & \text{if } T = K_s \\ v_{k1} & \text{if } T = \emptyset \end{cases}. \quad (3.20)$$

For each  $p_k \in K_p$  we have

$$\begin{aligned} \Delta_{p_k} F(S \cup T) &= v_k(T) P_k(S \cup \{p_k\}) - v_k(T) P_k(S) \\ &= v_k(T) \Delta P_k, \quad S \subseteq K_p \setminus \{p_k\}, \quad T \subseteq \{K_s\}, \end{aligned} \quad (3.21)$$

---

<sup>6</sup>See Proposition 4 in the next section.

$$P_k \quad (3.17)$$

assigned exactly half the contribution by region alone. A similar result for a simultaneous decomposition of the contribution obtained in the

decomposable poverty indices, for example in isolation; the outcomes will immediately from the series of different subgroups remains considered.

identify the subgroup contributions represent the population share and Eq. 3.7 yields

$$P_{k1}]. \quad (3.18)$$

ge in poverty,  $\Delta P$ , in terms of  $P_{k1}$ ,  $k \in K$ , and the population

independently, so the poverty  $k \in K\}$ . However the population remains at this stage, the population factor, denoted by  $K_s$ . The model

$$\subseteq K_p, \quad T \subseteq \{K_s\}, \quad (3.19)$$

$$\begin{cases} v_{k2} & \text{if } T = K_s \\ v_{k1} & \text{if } T = \emptyset \end{cases}. \quad (3.20)$$

$$v_k(T)P_k(S) \\ \backslash \{p_k\}, \quad T \subseteq \{K_s\}, \quad (3.21)$$

and setting  $M = \{K_s\} \cup K_p \setminus \{p_k\}$ , it follows that

$$\begin{aligned} \Delta_{p_k} F(S) + \Delta_{p_k} F(M \setminus S) &= v_k(S \cap \{K_s\}) \Delta P_k + v_k((M \setminus S) \cap \{K_s\}) \Delta P_k \\ &= (v_{k1} + v_{k2}) \Delta P_k, \end{aligned} \quad (3.22)$$

for all  $S \subseteq M$ . So, using Eq. 3.14, the Shapley contribution associated with the change in poverty within subgroup  $k$  is given by

$$\begin{aligned} C_{p_k}^S &= \mathcal{E}_{S \subseteq M} \Delta_{p_k} F(S) = \frac{1}{2} \mathcal{E}_{S \subseteq M} [\Delta_{p_k} F(S) + \Delta_{p_k} F(M \setminus S)] \\ &= \frac{1}{2} \mathcal{E}_{S \subseteq M} (v_{k1} + v_{k2}) \Delta P_k = \frac{v_{k1} + v_{k2}}{2} \Delta P_k. \end{aligned} \quad (3.23)$$

Conversely

$$\Delta_{K_s} F(S) = \sum_{k \in K} [v_k(K_s) - v_k(\emptyset)] P_k(S) = \sum_{k \in K} P_k(S) \Delta v_k, \quad S \subseteq K_p. \quad (3.24)$$

So

$$\begin{aligned} C_{K_s}^S &= \mathcal{E}_{S \subseteq K_p} \Delta_{K_s} F(S) = \frac{1}{2} \mathcal{E}_{S \subseteq K_p} [\Delta_{K_s} F(S) + \Delta_{K_s} F(K_p \setminus S)] \\ &= \frac{1}{2} \mathcal{E}_{S \subseteq K_p} \sum_{k \in K} [P_k(S) + P_k(K_p \setminus S)] \Delta v_k \\ &= \frac{1}{2} \mathcal{E}_{S \subseteq K_p} \sum_{k \in K} [P_{k1} + P_{k2}] \Delta v_k = \sum_{k \in K} \frac{P_{k1} + P_{k2}}{2} \Delta v_k \end{aligned} \quad (3.25)$$

This is a very natural allocation of contributions given that we seek a decomposition which treats the factors in a symmetric way, and given that Eq. 3.18 may be rewritten

$$\Delta P = \sum_{k \in K} \frac{v_{k1} + v_{k2}}{2} \Delta P_k + \sum_{k \in K} \frac{P_{k1} + P_{k2}}{2} \Delta v_k. \quad (3.26)$$

#### 4 Hierarchical structures

Despite its attractive properties, the Shapley decomposition has one major drawback for distributional analysis: the contribution assigned to any given factor is usually sensitive to the way in which the other factors are treated. In many applications, certain groups of factors naturally cluster together. This leads to a hierarchical structure comprising a set of primary factors, each of which is subdivided into a (possibly single element) group of secondary factors. For example, when income inequality is decomposed by source of income (see Section 6 below), one may first wish to regard income as the sum of labour income, investment income and transfers. Then investment income, say, might be split into interest, dividends, capital gains and rent. The Shapley decomposition does not guarantee that the contribution assigned to earnings will be the same if investment income is treated as a single entity or viewed in terms of its separate components. Nor does it ensure that the inequality contributions assigned to the components of investment income sum to the contribution of investment income treated as a single unit.

To study this issue in more detail, consider a *hierarchical model*  $\langle K, A, F \rangle$  consisting of a set of  $m$  secondary factors indexed by  $K$ , and a partition of  $K$  into

a set of primary factors  $A = \{L_j, j \in J\}$ . The fine (i.e. secondary factor) structure of the model is denoted by  $\langle K, F \rangle$ . Replacing each set of secondary factors with its corresponding primary factor produces the aggregated model  $\langle A, F^A \rangle$  defined by

$$F^A(T) = F(K_T), \quad T \subseteq A, \quad (4.1)$$

where

$$K_T = \cup_{L \in T} L, \quad T \subseteq A, \quad (4.2)$$

denotes the set of secondary factors covered by the subset  $T$  of primary factors. More generally, substituting a subset  $B \subseteq A$  of primary factors for their corresponding groups of secondary factors produces the partially aggregated model  $\langle B \cup K \setminus K_B, F^B \rangle$  defined by

$$F^B(S \cup T) = F(S \cup K_T), \quad S \subseteq K \setminus K_B, \quad T \subseteq B. \quad (4.3)$$

A decomposition rule for hierarchical models is a function  $C^*$  which assigns the contribution  $C_k^*(K, A, F)$  to each secondary factor  $k \in K$ , and the contribution  $C_L^*(K, A, F)$  to each primary factor  $L \in A$ . It will be said to be *aggregation consistent* for the model  $\langle K, A, F \rangle$  if

$$C_L^*(K, A, F) = \sum_{k \in L} C_k^*(K, A, F), \quad \text{each } L \in A, \quad (4.4)$$

or in other words, if the contribution of each primary factor is the sum of the contributions of its constituents.

Applying the Shapley decomposition both to the fine model structure  $\langle K, F \rangle$  and to the aggregated model  $\langle A, F^A \rangle$  produces the hierarchical decomposition rule

$$C_k^S(K, A, F) = C_k^S(K, F), \quad k \in K; \quad C_L^S(K, A, F) = C_L^S(A, F^A), \quad L \in A. \quad (4.5)$$

As already mentioned, this procedure does not ensure aggregation consistency. However, the problem can be overcome by adopting a sequential Shapley approach along the lines proposed by Owen [18]. First, contributions are allocated as above to each of the primary factors using the Shapley decomposition of the aggregated model  $\langle A, F^A \rangle$ . This yields

$$\begin{aligned} C_L^O(K, A, F) &= C_L^S(A, F^A) = \mathcal{E}_{T \subseteq A \setminus \{L\}} [F^A(T \cup \{L\}) - F^A(T)] \\ &= \mathcal{E}_{T \subseteq A \setminus \{L\}} [F(K_T \cup L) - F(K_T)] = \bar{F}_L(L), \quad \text{each } L \in A, \end{aligned} \quad (4.6)$$

where

$$\bar{F}_L(S) = \mathcal{E}_{T \subseteq A \setminus \{L\}} [F(K_T \cup S) - F(K_T)], \quad S \subseteq L, \quad L \in A. \quad (4.7)$$

The contribution of each primary factor  $L$  is then allocated amongst its constituents, by applying the Shapley decomposition to  $\langle L, \bar{F}_L \rangle$ :

$$\begin{aligned} C_k^O(K, A, F) &= C_k^S(L, \bar{F}_L) = \mathcal{E}_{S \subseteq L \setminus \{k\}} \Delta_k \bar{F}_L(S) \\ &= \mathcal{E}_{T \subseteq A \setminus \{L\}} \mathcal{E}_{S \subseteq L \setminus \{k\}} \Delta_k F(K_T \cup S), \quad k \in L, \quad L \in A. \end{aligned} \quad (4.8)$$

(i.e. secondary factor) structure set of secondary factors with its model  $\langle A, F^A \rangle$  defined by

$$\vdash A, \quad (4.1)$$

$$\vdash A, \quad (4.2)$$

the subset  $T$  of primary factors. primary factors for their corre- the partially aggregated model

$$\vdash K_B, \quad T \subseteq B. \quad (4.3)$$

is a function  $C^*$  which assigns factor  $k \in K$ , and the contribution said to be *aggregation consistent*

$$\text{each } L \in A, \quad (4.4)$$

primary factor is the sum of the fine model structure  $\langle K, F \rangle$  and hierarchical decomposition rule

$$\vdash = C_L^S(A, F^A), \quad L \in A. \quad (4.5)$$

ensure aggregation consistency. ing a sequential Shapley approach. contributions are allocated as above to position of the aggregated model

$$\vdash \cup \{L\}) - F^A(T)] \\ = \bar{F}_L(L), \quad \text{each } L \in A, \quad (4.6)$$

$$\vdash ], \quad S \subseteq L, \quad L \in A. \quad (4.7)$$

then allocated amongst its con-  $\vdash \langle L, \bar{F}_L \rangle$ :

$$\vdash \bar{F}_L(S) \\ \vdash \cup S), \quad k \in L, \quad L \in A. \quad (4.8)$$

As the Shapley decomposition is exact, it follows that

$$\begin{aligned} \sum_{k \in L} C_k^O(K, A, F) &= \sum_{k \in L} C_k^S(L, \bar{F}_L) = \bar{F}_L(L) \\ &= C_L^O(K, A, F), \quad \text{each } L \in A. \end{aligned} \quad (4.9)$$

So this two-stage procedure is always aggregation consistent.<sup>7</sup>

Although the hierarchical form of the Shapley rule is not usually aggregation consistent, there is one important exception. Let us say that the function  $F : \{S | S \subseteq K\} \rightarrow \mathbb{R}$  is *separable over  $L \subseteq K$*  if

$$\Delta_k F(S \cup T) = \Delta_k F(S), \quad \text{all } k \in L, \quad T \subseteq L \setminus \{k\}, \quad S \subseteq K \setminus L; \quad (4.10)$$

in other words, the marginal contribution of each factor  $k \in L$  does not depend on the other factors in  $L$ . Note that if  $F$  is separable over  $L$ , then  $F$  is also separable over any subset of  $L$ . Note also that if  $T$  is written as  $T = \{k_1, k_2, \dots, k_r\}$ , then it follows from (4.10) that

$$\begin{aligned} F(S \cup T) - F(S) &= \sum_{r=1}^r \Delta_{k_r} F(S \cup \{k_1, \dots, k_{r-1}\}) \\ &= \sum_{k \in T} \Delta_k F(S), \quad \text{all } T \subseteq L, \quad S \subseteq K \setminus L, \end{aligned} \quad (4.11)$$

so the marginal effect of introducing any group  $T$  of factors from the subset  $L$  is the sum of the marginal effects of introducing each factor separately.

We now obtain:

**Proposition 1** Consider the model  $\langle K, F \rangle$ , and suppose  $F$  is separable over  $L \subseteq K$ . Then

$$C_k^S(K, F) = C_k^S(\{L\} \cup K \setminus L, F^{\{L\}}), \quad k \in K \setminus L \quad (4.12a)$$

$$\sum_{k \in L} C_k^S(K, F) = C_{\{L\}}^S(\{L\} \cup K \setminus L, F^{\{L\}}) \quad (4.12b)$$

$$C_k^S(K, F) = C_k^S(K \setminus L \cup \{k\}, F), \quad k \in L \quad (4.12c)$$

where  $F^{\{L\}}$  is defined in Eq. 4.3.

*Proof* See Appendix. □

Proposition 1 establishes three things. Equation 4.12a shows that treating a separable subset  $L$  as a single entity in the Shapley decomposition does not affect the contributions assigned to the factors outside  $L$ . As a consequence, the sum of the contributions of the factors in  $L$  must equal the contribution of the grouped factor in the more aggregated model, as indicated in Eq. 4.12b. Finally, Eq. 4.12c shows that the contribution of any factor  $k$  in a separable set  $L$  can disregard the complementary set of factors  $L \setminus \{k\}$ .

<sup>7</sup>The two-stage decomposition can be extended to a multi-stage procedure if the secondary factors are divided into tertiary factors, and so on.

Framed in terms of hierarchical structures, Proposition 1 implies that any separable set of secondary factors can be replaced by its corresponding primary factor without altering the contributions of the other factors. This process may be repeated for further separable groups of factors, thereby establishing:

**Proposition 2** Consider the hierarchical model  $\langle K, A, F \rangle$ , and suppose that  $F$  is separable over each  $L \in A$ . Then, for all  $L \in A$  and all  $k \in L$ , we have

$$C_k^S(K, F) = C_k^O(K, A, F) = \mathcal{E}_{T \subseteq A \setminus \{L\}} \Delta_k F(K_T) = \Delta_k \bar{F}_L(\emptyset). \quad (4.13)$$

So the Shapley and Owen decompositions coincide, and the Shapley decomposition is aggregation consistent.

*Proof* See Appendix. □

The results of Proposition 2 enable several short-cuts to be implemented in the calculation of the Shapley contributions. While the requirement that  $F$  is separable over each  $L \in A$  may seem very restrictive, it should be noted that  $F$  is (trivially) separable over any single element subset of  $K$ . So it is always possible to apply Proposition 2, by treating any non-separable subsets of factors as a set of single factors in the partition  $A$  of  $K$ . If  $F$  is not separable over the primary factor  $L$ , then Proposition 2 leads us to expect that the Shapley contributions of the secondary factors  $k \in L$  will not be obtained by the Owen two-stage method. In such situations, it may well be the case that the Owen procedure is favoured, in order to ensure an aggregation consistent decomposition. However, there are likely to be several alternative ways in which secondary factors can be grouped together into primary factors, leaving room for judgements about the most appropriate arrangement.

In the general structural model denoted by  $\langle K, F \rangle$ , factors can interact in complex ways, and there is nothing to prevent some of the factors being redundant in the sense that a proper subset of  $K$  completely accounts for the initial level,  $F(K)$ , of the statistic under examination. This will be captured by saying that the set of factors  $L \subseteq K$  is sufficient if  $F(S) = 0$  for all  $S \subseteq K \setminus L$ .

The sufficiency property leads to a powerful result when combined with the results of Proposition 2. For if each of the primary factors  $L \in A$  are sufficient in the hierarchical model  $\langle K, A, F \rangle$  then, for all  $L \in A$ , all  $T \subseteq A \setminus \{L\}$ , and all  $S \subseteq L$ , we have

$$F(K_T \cup S) = 0 \quad (4.14)$$

unless  $T = A \setminus \{L\}$  and  $S \neq \emptyset$ . So

$$\bar{F}_L(S) = \frac{1}{|A|} F(K_{A \setminus \{L\}} \cup S) = \frac{1}{|A|} F(K \setminus L \cup S), \quad S \subseteq L, \quad (4.15)$$

and, if  $F$  is separable over  $L$ , then

$$\Delta_k \bar{F}_L(S) = \frac{1}{|A|} \Delta_k F(K \setminus \{k\}) = \frac{1}{|A|} M_k(K, F), \quad k \in L, \quad S \subseteq L \setminus \{k\}, \quad (4.16)$$

sition 1 implies that any sepa-  
corresponding primary factor  
This process may be repeated  
ishing:

$A, F\rangle$ , and suppose that  $F$  is  
 $\forall k \in L$ , we have

$$\varphi(K_T) = \Delta_k \bar{F}_L(\emptyset). \quad (4.13)$$

the Shapley decomposition is

□

cuts to be implemented in the  
requirement that  $F$  is separable  
be noted that  $F$  is (trivially)  
it is always possible to apply  
s of factors as a set of single  
le over the primary factor  $L$ ,  
contributions of the secondary  
age method. In such situations,  
favoured, in order to ensure  
there are likely to be several  
grouped together into primary  
appropriate arrangement.  
factors can interact in complex  
actors being redundant in the  
for the initial level,  $F(K)$ , of  
by saying that the set of factors

when combined with the results  
 $\forall L \in A$  are sufficient in the  
ll  $T \subseteq A \setminus \{L\}$ , and all  $S \subseteq L$ ,

$$(4.14)$$

$$\forall L \cup S, \quad S \subseteq L, \quad (4.15)$$

$$k \in L, \quad S \subseteq L \setminus \{k\}, \quad (4.16)$$

in the notation of Eq. 2.4. Combined with Proposition 2, this yields:

**Proposition 3** For the hierarchical model  $\langle K, A, F \rangle$ , suppose that  $F$  is separable over each  $L \in A$ , and that each  $L \in A$  is sufficient. Then

$$C_k^S(K, F) = C_k^S(L, \bar{F}_L) = \frac{1}{|A|} M_k(K, F), \quad k \in L, \quad L \in A. \quad (4.17)$$

Thus, in the context described, the Shapley contributions are determined solely by the number of primary factors,  $|A|$ , and the first round marginal effects,  $M_k(K, F)$ .

The implications of Proposition 3 are well illustrated by returning to the example of decomposable poverty indices discussed in Section 3.2, which can now be extended easily to the general multivariate case by defining the primary factors in terms of the attributes (region, household size, etc.), and the secondary factors in terms of the attribute subgroups. To be specific, suppose  $K$  is partitioned into  $A = \{L_j, j \in J\}$ , where  $L_j$  refers to attribute  $j$ , and  $k \in L_j$  is the factor representing poverty within category  $k$  of attribute  $j$ . Then the hierarchical model  $\langle K, A, F \rangle$  is characterised by

$$F(S) = \sum_{k \in L_j \cap S} v_k^j P_k^j(S), \quad \text{each } j \in J, \quad S \subseteq K, \quad (4.18)$$

where  $v_k^j$  and  $P_k^j(S)$  respectively indicate the population share and poverty level associated with category  $k$  of attribute  $j$  after the factors in the set  $K \setminus S$  have been removed, and where

$$P_k^j(S \cup T) = P_k^j(S \cup \{k\}), \quad \text{all } j \in J, \quad S \subseteq K \setminus L_j, \quad T \subseteq L_j, \quad (4.19)$$

since the poverty level associated with category  $k \in L_j$  is not affected by eliminating poverty in the categories  $L_j \setminus \{k\}$ . Condition (4.19) implies that the function  $F$  is separable over each  $L_j \in A$ . Furthermore,  $F(S) = 0$  for  $S \subseteq K \setminus L_j$ , so each of the primary factors  $L_j, j \in J$ , is sufficient. It therefore follows from Proposition 3 and Eq. 4.16 that

$$C_k^S(K, F) = \frac{1}{|J|} M_k(K, F) = \frac{1}{|J|} v_k^j P_k^j(K), \quad \text{all } k \in L_j, \quad j \in J. \quad (4.20)$$

This establishes:

**Proposition 4** When a decomposable poverty index is employed in a multivariate poverty decomposition with  $n$  attributes, the Shapley contribution associated with category  $k$  of attribute  $j$  is given by

$$C_k^S = \frac{1}{n} v_k^j P_k^j, \quad (4.21)$$

where  $v_k^j$  is the population share associated with category  $k$  of attribute  $j$  and  $P_k^j$  is the poverty level observed for this category.

The intuition behind this result is clear. Each of the  $n$  attributes accounts for the overall poverty level  $P$ , and must therefore be assigned the contribution  $P/n$ , given that the factors are treated symmetrically. Furthermore, the secondary factors associated with any attribute operate independently (in the sense captured by the

separability property), so the contribution of each attribute is allocated amongst its constituent factors in proportion to their marginal effect.

## 5 Inequality decomposition by subgroups

The results obtained in the preceding section assist in the analysis of some of the other standard applications of decomposition methods. We first consider the question of decomposing inequality by subgroups, a topic pioneered by Theil [27] and later developed by Bourguignon [2], Shorrocks [23, 25], and Foster and Shneyerov [8, 9], amongst others.

The problem may be posed in terms of a set of individuals  $N = \{1, 2, \dots, n\}$  with income vector  $\mathbf{y}$  and mean income  $\mu$ , which is partitioned into a set of subgroups  $N_k$  ( $k = 1, 2, \dots, m$ ) with vectors  $\mathbf{y}^k$  and means  $\mu_k$ . Without loss of generality, it may be assumed that the subgroups are numbered so that  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_m$ , and that each of the subgroup income vectors is arranged in increasing order. For each subgroup  $k$ , denote the (ordered) vector of relative incomes by  $\mathbf{w}^k = \mathbf{y}^k / \mu_k$ , the relative mean income by  $b_k = \mu_k / \mu$ , and the share of the population by  $v_k$ . Then for any inequality index  $I(\cdot)$  which is symmetric and scale invariant (i.e. homogeneous of degree zero), the overall level of inequality can be expressed as

$$\begin{aligned} I(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m) &= I(\mathbf{w}^1 b_1, \mathbf{w}^2 b_2, \dots, \mathbf{w}^m b_m) \\ &= \bar{I}(\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^m, \mathbf{b}), \end{aligned} \quad (5.1)$$

where  $\mathbf{b} = (b_1, \dots, b_m)$ . In this framework, decomposition of inequality by subgroups is typically viewed as the exercise which assigns contributions to inequality within each subgroup (as captured in the vectors  $\mathbf{w}^k$ ), and to the “between-group” effect (as captured by  $\mathbf{b}$ ). We will think of these as the within-group *factors*, indexed by  $K = \{1, 2, \dots, m\}$ , and the between-group *factor* indexed by the (single element) set  $L$ .

### 5.1 Entropy indices

Subgroup inequality decomposition is most often undertaken using an inequality measure drawn from the entropy family

$$E_c(\mathbf{y}) = E_c(y_1, \dots, y_n) = \frac{1}{n} \sum_{i \in N} \phi_c(y_i / \mu), \quad (5.2)$$

where  $\phi_c(t) = (t^c - 1)/[c(c - 1)]$ ,  $c \neq 0, 1$ ;  $\phi_1(t) = t \ln t$ ; and  $\phi_0(t) = -\ln t$ . These indices yield the decomposition equation

$$\begin{aligned} E_c(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m) &= \bar{E}_c(\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^m, \mathbf{b}) \\ &= \sum_{k=1}^m v_k b_k^c E_c(\mathbf{w}^k) + \sum_{k=1}^m v_k \phi_c(b_k). \end{aligned} \quad (5.3)$$

It is standard practice to allocate the contribution

$$W_k = v_k b_k^c E_c(\mathbf{w}^k), \quad k = 1, 2, \dots, m, \quad (5.4)$$

attribute is allocated amongst its effect.

sist in the analysis of some of methods. We first consider the topic pioneered by Theil [27] and [25], and Foster and Shneyerov

individuals  $N = \{1, 2, \dots, n\}$  with ioned into a set of subgroups  $N_k$  tout loss of generality, it may be  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_m$ , and that each easing order. For each subgroup  $y \mathbf{w}^k = \mathbf{y}^k / \mu_k$ , the relative mean on by  $v_k$ . Then for any inequality e. homogeneous of degree zero),

$$\dots, \mathbf{w}^m b_m) \\ , \mathbf{w}^m, \mathbf{b}), \quad (5.1)$$

sition of inequality by subgroups ontributions to inequality within d to the “between-group” effect within-group factors, indexed by indexed by the (single element)

undertaken using an inequality

$$\sum_i \phi_c(y_i / \mu), \quad (5.2)$$

$\ln t$ ; and  $\phi_0(t) = -\ln t$ . These in-

$$\mathbf{w}^m, \mathbf{b}) \\ ) + \sum_{k=1}^m v_k \phi_c(b_k). \quad (5.3)$$

$$, 2, \dots, m), \quad (5.4)$$

to the within-group factors, on the grounds that  $W_k$  is the amount by which overall inequality falls when incomes within subgroup  $k$  are redistributed equally. The remaining “between group component”

$$B = \sum_{k=1}^m v_k \phi_c(b_k) \quad (5.5)$$

is the level of inequality which results when the incomes of all individuals are replaced by the mean of the subgroup to which they belong, and is usually regarded as the contribution of the between-group factor.

Although this procedure yields an exact decomposition of inequality, the interpretation of the between-group component,  $B$ , is questionable. As Foster and Shneyerov [8] point out, eliminating the between-group factor not only removes the component  $B$  but also changes the weights attached to the subgroup inequality values in the within-group component. If the between-group factor is eliminated by setting each  $b_k = 1$ ,<sup>8</sup> then the (first round) marginal effect is

$$B' = \sum_{k=1}^m v_k (b_k^c - 1) E_c(\mathbf{w}^k) + \sum_{k=1}^m v_k \phi_c(b_k). \quad (5.6)$$

Removing the within-group factors in subsequent rounds produces the contributions

$$W'_k = v_k E_c(\mathbf{w}^k), \quad k = 1, 2, \dots, m. \quad (5.7)$$

The expressions for  $B$  and  $B'$  coincide only when  $c = 0$ , corresponding to the mean logarithmic deviation index,  $E_0$ . In all other cases the standard practice of assigning the contributions according to Eqs. 5.4 and 5.5 rests on the implicit assumption that the between-group factor is eliminated last.

The Shapley decomposition treats the factors symmetrically, and consequently yields an intermediate solution. Defining  $K$  and  $L$  as above, and setting  $A = \{K, L\}$ , yields the hierarchical model  $\langle K \cup L, A, F \rangle$  where

$$F(S \cup T) = \sum_{k \in S} W_k(T) + B(T), \quad S \subseteq K, T \subseteq L, \quad (5.8)$$

with  $W_k(L) = W_k$ ;  $W_k(\emptyset) = W'_k$ ;  $B(L) = B$ ;  $B(\emptyset) = B'$ . Since

$$\Delta_K F(S \cup T) = W_k(T), \quad k \in K, S \subseteq K \setminus \{k\}, T \subseteq L, \quad (5.9)$$

the function  $F$  is separable over  $K$ , and also (trivially) over  $L$ . So by Proposition 2 the Shapley contributions may be obtained via the Owen two-stage procedure. This gives

$$C_K^S(K \cup L, A, F) = \frac{1}{2}[F(K \cup L) - F(L) + F(K)] = \frac{1}{2} \sum_{k \in K} [W_k + W'_k] \\ C_L^S(K \cup L, A, F) = \frac{1}{2}[B + B'] \quad (5.10)$$

<sup>8</sup>While it is usual to eliminate the between group factor by setting each  $b_k = 1$ , there is no compelling reason for doing so. We follow standard practice here, but note that setting each  $b_k = \beta > 0$  results in only minor modifications to the analysis.

as the contributions of the primary factors, and

$$C_k^S(K \cup L, A, F) = \frac{1}{2} [\Delta_k F(\emptyset) + \Delta_k F(L)] = \frac{1}{2} [W_k + W'_k], \quad k \in K, \quad (5.11)$$

as the contributions of the individual within-group factors.

As already mentioned, standard practice assigns the within- and between-group inequality contributions given by Eqs. 5.4 and 5.5. The Shapley decomposition generates the same assignment when the mean logarithmic deviation index,  $E_0$ , is chosen as the measure of inequality, but the results will not be the same when other indices are employed.<sup>9</sup> While the Shapley decomposition departs from common practice, there is a compelling logic behind the assignment rule, and to that extent it offers a potential improvement over current methods.

## 5.2 The Gini coefficient

Numerous attempts have been made to decompose the Gini coefficient along similar lines to Eq. 5.3. Using the notation described earlier in this section, the most common method may be formulated by supposing that person  $i$  occurs in the  $i$ th position when the distribution is written  $\mathbf{y} = (\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m)$ , and in position  $r_i$  when all incomes are arranged in increasing order.<sup>10</sup> The value of the Gini coefficient is then given by

$$G(\mathbf{y}) = \frac{2}{n^2 \mu} \sum_{i \in N} r_i (y_i - \mu) \quad (5.12)$$

and yields the decomposition equation

$$\begin{aligned} G &= G(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m) = \frac{2}{n^2 \mu} \sum_{k=1}^m \sum_{i \in N_k} r_i (y_i - \mu) \\ &= \frac{2}{n^2 \mu} \sum_{k=1}^m \left\{ \sum_{i \in N_k} i (y_i - \mu_k) + \sum_{i \in N_k} i (\mu_k - \mu) + \sum_{i \in N_k} (r_i - i) y_i \right\} \\ &= W + B + R, \end{aligned} \quad (5.13)$$

where

$$W = \frac{2}{n^2 \mu} \sum_{k=1}^m \sum_{i \in N_k} i (y_i - \mu_k) = \sum_{k=1}^m v_k^2 b_k G(\mathbf{y}^k) = \sum_{k=1}^m v_k^2 b_k G(\mathbf{w}^k) \quad (5.14)$$

is a weighted sum of the within-group inequality values, and

$$B = \frac{2}{n^2 \mu} \sum_{k=1}^m \sum_{i \in N_k} i (\mu_k - \mu) = \sum_{k=1}^m b_k v_k \left[ \sum_{j=1}^k v_j - \sum_{j=k}^m v_j \right] \quad (5.15)$$

<sup>9</sup>The methods may be reconciled by explicitly recognising the different treatment of the within- and between-group factors. This may be done by redefining the problem so that the between-group term is separated out, and the question becomes one of allocating contributions to the within-group factors in the decomposition of  $I - B$ .

<sup>10</sup>For convenience it is assumed that all incomes are distinct, and hence that  $r_i$  is uniquely defined.

$$W_k + W'_k], \quad k \in K, \quad (5.11)$$

ctors.

he within- and between-group . The Shapley decomposition rithmic deviation index,  $E_0$ , is ill not be the same when other position departs from common nment rule, and to that extent ds.

the Gini coefficient along similar in this section, the most common l occurs in the  $i$ th position when position  $r_i$  when all incomes are coefficient is then given by

$$(5.12)$$

$\mu$ )

$$k - \mu) + \sum_{i \in N_k} (r_i - i) y_i \} \\ (5.13)$$

$$= \sum_{k=1}^m v_k^2 b_k G(\mathbf{w}^k) \quad (5.14)$$

es, and

$$\left[ \sum_{j=1}^k v_j - \sum_{j=k}^m v_j \right] \quad (5.15)$$

different treatment of the within- and blem so that the between-group term tributions to the within-group factors

id hence that  $r_i$  is uniquely defined.

is the “between-group component”, indicating the value of the Gini coefficient when all incomes are replaced by the mean income of the subgroup to which they belong. The final term,  $R$ , in Eq. 5.13 is a residual or “interaction” effect which vanishes when the subgroup income ranges do not overlap (so that  $r_i = i$ , for all  $i$ ), and is otherwise strictly positive.

The Gini decomposition (5.13) is less satisfactory than the corresponding Entropy formulation (5.13), because the interaction term introduces a third, vaguely specified, element into the equation. It is difficult to predict how the interaction effect will respond to changes in subgroup characteristics, such as a narrowing of income differentials between subgroups. As a consequence, the overall Gini value may react perversely to such changes: for example, a reduction in inequality in every subgroup may cause overall inequality to rise, even when the subgroup means and sizes are held constant. The Shapley decomposition cannot overcome this “subgroup inconsistency” problem, since this is a fundamental property of the Gini index. However, it does remove the need for a separate interaction term, by absorbing this component into the contributions of the within- and between-group factors.

While the results may be obtained straightforwardly via a suitable computer algorithm, they do not produce simple analytical formulae. To gain some idea of the outcome, consider the 2-factor decomposition based on the within-group primary factor,  $K$ , and the between group primary factor,  $L$ . The elimination sequence  $(K, L)$  yields the marginal contributions

$$C_K = W + R; \quad C_L = B, \quad (5.16)$$

so in this case the whole interaction effect is allocated to the within-group factor. However, the situation becomes more complex when the between group factor is removed in the first round, since setting each  $b_k = 1$  not only eliminates the between-group component  $B$  in Eq. 5.13, but also changes  $W$  and  $R$  to

$$W' = \sum_{k=1}^m v_k^2 G(\mathbf{w}^k)$$

and

$$R' = \frac{2}{n^2 \mu} \sum_{k=1}^m \sum_{i \in N_k} (r'_i - i) w_i^k, \quad (5.17)$$

respectively, where  $r'_i$  is the position of person  $i$  when the vector  $(\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^m)$  is rearranged in increasing order. The elimination path  $(L, K)$  therefore produces the marginal contributions

$$C'_L = G - W' - R'; \quad C'_K = W' + R', \quad (5.18)$$

and the Shapley decomposition is given by

$$C_K^S = \frac{1}{2} [C_K + C'_K] = \frac{1}{2} [W + W' + R + R'] \\ C_L^S = \frac{1}{2} [C_L + C'_L] = B + \frac{1}{2} [W - W' + R - R']. \quad (5.19)$$

This assignment retains the basic features of the Shapley decomposition: the contributions sum to the overall Gini value, and they correspond to the marginal

effect of removing each factor, averaged over all the elimination sequences. However there seems little prospect of gaining insights from further inspection of the formulae. One glimmer of hope is provided by the fact that the contributions in Eq. 5.19 may be rewritten as

$$\begin{aligned} C_K^S &= C_K + \frac{1}{2} [G - C_K - C'_L] \\ C_L^S &= C'_L + \frac{1}{2} [G - C_K - C'_L] \end{aligned} \quad (5.20)$$

So the contributions effectively begin with the first round marginal effects  $C_K$  and  $C'_L$ , and then allocate half the “surplus” to each of the factors. This is a general consequence of applying the Shapley decomposition to two factors. However the property does not generalise easily when the within-group effects are treated separately; and since the model is not separable with respect to the set of within-group factors (unlike the Entropy case in Section 5.1), the individual within-group effects are not expected to sum to the combined within-group effect derived above.

## 6 Inequality decomposition by source of income

The last of the conventional decomposition problems concerns the situation in which income is divided into components such as earnings, investment income, taxes and transfers, and we seek to identify the contribution of these income sources to overall income inequality. If  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  denotes the distribution of income for a population of size  $n$ , and  $\mathbf{y}^k$  is the distribution of income from source  $k \in K = \{1, 2, \dots, m\}$ , the original model may be written

$$I = I(\mathbf{y}) = I\left(\sum_{k \in K} \mathbf{y}^k\right), \quad (6.1)$$

where  $I(\cdot)$  is an inequality index. This leads naturally to the model structure  $\langle K, F \rangle$ , where the factors represent “incomes from source  $k$ ” and are indexed by  $K$ , and the function  $F(\cdot)$  is defined by

$$F(S) = I\left(\sum_{k \in S} \mathbf{y}^k\right), \quad S \subseteq K, \quad (6.2)$$

with the understanding that  $F(\emptyset) = 0$ . A slightly different formulation results if the factors are interpreted as “differences in incomes from source  $k$ ”. Denoting mean income by  $\mu$ , the mean income from source  $k$  by  $\mu^k$ , and the  $n$ -tuple of 1’s by  $\mathbf{e}$ , the model structure now becomes  $\langle K, \tilde{F} \rangle$ , where

$$\tilde{F}(S) = I\left(\sum_{k \in S} \mathbf{y}^k + \sum_{k \notin S} \mu^k \mathbf{e}\right), \quad S \subseteq K. \quad (6.3)$$

The distinction between Eqs. 6.2 and 6.3 is subtle, but important. It is best appreciated by considering a component of income which is equally distributed—for instance, a poll tax or subsidy. Since there are no differences across individuals, the marginal impact of removing these differences is always zero. As a consequence, the Shapley decomposition based on Eq. 6.3 suggests that any equally distributed component of income makes no contribution to overall inequality. In contrast, in the model based on Eq. 6.2, eliminating a poll subsidy will typically increase income inequality, so here the Shapley decomposition yields a negative contribution,

elimination sequences. However further inspection of the formulae, contributions in Eq. 5.19 may

$$C_L] \quad (5.20)$$

st round marginal effects  $C_K$  each of the factors. This is a position to two factors. However ifthin-group effects are treated h respect to the set of within-1), the individual within-group in-group effect derived above.

concerns the situation in which investment income, taxes and these income sources to overall e distribution of income for a income from source  $k \in K =$

$$), \quad (6.1)$$

to the model structure  $(K, F)$ , and are indexed by  $K$ , and the

$$\subseteq K, \quad (6.2)$$

erent formulation results if the om source  $k$ ". Denoting mean and the  $n$ -tuple of 1's by  $\mathbf{e}$ , the

$$), \quad S \subseteq K. \quad (6.3)$$

otile, but important. It is best which is equally distributed— differences across individuals, always zero. As a consequence, ts that any equally distributed overall inequality. In contrast, subsidy will typically increase i yields a negative contribution,

suggesting an equalising effect (see Proposition 5 below). This probably accords better with intuition, although, as already indicated, the distinction really turns on whether one is interested in the contribution of a particular source of income, or in the contribution of variations in incomes from that source.

Before the results of the Shapley decomposition are discussed, it is worth reviewing the methods currently used to decompose income inequality by source. If the variance is employed as the inequality index, Eq. 6.1 becomes

$$\sigma^2(\mathbf{y}) = \sigma^2 \left( \sum_{k \in K} \mathbf{y}^k \right) = \sum_{k \in K} \text{cov}(\mathbf{y}^k, \mathbf{y}). \quad (6.4)$$

This suggests the factor contributions:

$$C_k = \text{cov}(\mathbf{x}^k, \mathbf{x}), \quad k \in K, \quad (6.5)$$

an assignment rule which Shorrocks [24] calls "the natural decomposition of the variance". Similarly, using Eq. 5.12, the Gini index may be written

$$G(\mathbf{y}) = \frac{2}{n^2 \mu} \sum_{i=1}^n r_i(y_i - \mu) = \frac{2}{n\mu} \text{cov}(\mathbf{r}, \mathbf{y}) = \sum_{k \in K} \frac{2}{n\mu} \text{cov}(\mathbf{r}, \mathbf{y}^k), \quad (6.6)$$

suggesting the "natural decomposition of the Gini" given by

$$C_k = \frac{2}{n\mu} \text{cov}(\mathbf{r}, \mathbf{y}^k), \quad k \in K. \quad (6.7)$$

Shorrocks [24] shows that many other decomposition rules can be constructed, but narrows down the options using a set of axioms. In combination these yield a unique decomposition rule in which the *relative* contribution of each income component is given by the natural decomposition of the variance, regardless of the choice of inequality index.

For our purposes, the feature of most interest is the fact that both Eqs. 6.5 and 6.7—and more generally any allocation based on "natural decompositions"—assign a zero contribution to any component of income which is equally distributed. In general terms, this means that current methods conform more with the model based on Eq. 6.3, which looks for the contribution of income differences, rather than the model based on Eq. 6.2, which seeks the contributions of income levels.

The Shapley decomposition is able to handle both interpretations, and therefore provides a richer range of possibilities. The results for the variance are the most easy to derive. In this particular case, formulations (6.2) and (6.3) coincide and yield

$$F(S) = \tilde{F}(S) = \sigma^2 \left( \sum_{k \in S} \mathbf{y}^k \right), \quad S \subseteq K, \quad (6.8)$$

so

$$\begin{aligned} \Delta_k F(S) &= \sigma^2 \left( \mathbf{y}^k + \sum_{t \in S} \mathbf{y}^t \right) - \sigma^2 \left( \sum_{t \in S} \mathbf{y}^t \right) \\ &= 2\text{cov}(\mathbf{y}^k, \sum_{t \in S} \mathbf{y}^t) + \sigma^2(\mathbf{y}^k), \quad S \subseteq K \setminus \{k\}. \end{aligned} \quad (6.9)$$

Using Eq. 3.4, it then follows that

$$\begin{aligned} C_k^S(K, F) &= C_k^S(K, \tilde{F}) = \frac{1}{2} \mathcal{E}_{S \subseteq K \setminus \{k\}} [\Delta_k F(S) + \Delta_k F(K \setminus S \setminus \{k\})] \\ &= \mathcal{E}_{S \subseteq K \setminus \{k\}} \left[ \text{cov} \left( \mathbf{y}^k, \sum_{t \in K \setminus \{k\}} \mathbf{y}^t \right) + \sigma^2(\mathbf{y}^k) \right] \\ &= \text{cov}(\mathbf{y}^k, \mathbf{y} - \mathbf{y}^k) + \sigma^2(\mathbf{y}^k) = \text{cov}(\mathbf{y}^k, \mathbf{y}), \quad k \in K. \end{aligned} \quad (6.10)$$

Thus, when the variance is used to measure inequality, the Shapley decomposition of inequality by source generates the usual “natural” decomposition values given in Eq. 6.5, regardless of whether the contributions are interpreted along the lines of Eq. 6.2 or Eq. 6.3.

Similar conclusions do not hold for any other index, although the result is partially true when the square of the coefficient of variation is selected as the inequality index. In this case, interpretation (6.3) yields

$$\tilde{F}(S) = \sigma^2 \left( \sum_{k \in S} \mathbf{y}^k \right) / \mu^2, \quad S \subseteq K, \quad (6.11)$$

and repeating the above steps establishes that

$$C_k^S(K, \tilde{F}) = \text{cov}(\mathbf{y}^k, \mathbf{y}) / \mu^2, \quad k \in K. \quad (6.12)$$

So the relative contribution of each factor again conforms with the natural decomposition of the variance when the factors are viewed as income differences from the various sources.

Under the alternative scenario (6.2) based on income levels, the Shapley decomposition does not appear to produce informative formulae for indices other than the variance. It is possible, however, to draw one useful conclusion regarding the contribution of a source of income which is distributed equally across the population. Assume that  $\mu^k > 0$  for all  $k \in K$ , and that the index  $I(\cdot)$  is scale invariant and strictly Schur-convex. Then, since equal income increments are equalising, we have

$$I \left( \sum_{t \in S} \mathbf{y}^t + \alpha \mathbf{e} \right) < I \left( \sum_{t \in S} \mathbf{y}^t \right), \quad \text{for all } \alpha > 0, \text{ and all } S \subseteq K. \quad (6.13)$$

So if  $\mathbf{y}^k = \mu^k \mathbf{e}$  we can deduce that

$$\Delta_k F(S) < 0, \quad \text{for all } S \subseteq K \setminus \{k\}, \quad (6.14)$$

from which it follows that  $C_k^S(K, F) < 0$ . Thus

**Proposition 5** Consider the decomposition of income inequality where the factors represent incomes from various sources. Suppose that the mean income from each source is positive, and that the inequality index is scale invariant and strictly Schur-convex. Then the Shapley decomposition will assign a negative inequality contribution to any component of income which is distributed equally across the population.

## 7 Concluding remarks

The main objective of this paper was to describe a general method of assessing the contributions of a set of factors which together account for the observed value of

$$\begin{aligned} & ) + \Delta_k F(K \setminus S \setminus \{k\}) \\ & + \sigma^2(\mathbf{y}^k) \Big] \\ & v(\mathbf{y}^k, \mathbf{y}), \quad k \in K. \end{aligned} \quad (6.10)$$

ty, the Shapley decomposition decomposition values given in interpreted along the lines of

, although the result is partially elected as the inequality index.

$$S \subseteq K, \quad (6.11)$$

$$k \in K. \quad (6.12)$$

forms with the natural decom- is income differences from the

me levels, the Shapley decom- rmuulae for indices other than useful conclusion regarding the l equally across the population. (.) is scale invariant and strictly re equalising, we have

$$0, \text{ and all } S \subseteq K. \quad (6.13)$$

$$\setminus \{k\}, \quad (6.14)$$

e inequality where the factors it the mean income from each le invariant and strictly Schur- negative inequality contribution ly across the population.

general method of assessing the ount for the observed value of

some aggregate statistic. The proposed solution involves calculating the marginal impact of each of the factors as they are eliminated in succession, and then averaging these marginal effects over all the possible elimination sequences. The resulting formula is formally identical to the Shapley value in cooperative game theory, and has therefore been referred to as the Shapley decomposition.

The Shapley procedure has several basic features which make it an attractive candidate for a general decomposition rule. It treats the factors in a symmetric manner; the contributions sum to the amount which needs to be "explained"; and the contributions can be interpreted as the expected marginal effects. This paper has demonstrated that it also generates sensible results when applied to the standard decomposition problems encountered in distributional analysis. In three classic situations, the Shapley rule exactly replicates current practice: the application of decomposable poverty indices to population subgroups; inequality decomposition by subgroups using the mean logarithmic deviation index; and inequality decomposition by source of income using the variance as the measure of inequality. In other applications, such as the growth-redistribution issue discussed in Section 3.1, and the Gini decomposition considered in Section 5.2, it improves upon existing methods by avoiding the need to introduce a residual component into the decomposition equation. The paper has also shown how the Shapley procedure can provide solutions to problems which have previously been difficult to address, such as multi-variate poverty decomposition discussed in Section 4.

Most of these applications are concerned with specific situations where previous work has suggested simple expressions for the factor contributions, and where the Shapley decomposition also yields explicit formulae, enabling the results to be compared. But the great advantage of the procedure proposed in this paper is that it can be applied to a wide range of problems which cannot be solved with existing techniques. Applications using other aggregate indicators or more complex models are unlikely to yield simple analytic expressions for the Shapley contributions, and will therefore require an algorithm to calculate the values (and, ideally, also their standard errors). In many situations, there will be sets of factors which group naturally together, suggesting a hierarchical model of the type described in Section 4, and the replacement of the Shapley rule by the two-stage Owen procedure. While it is difficult to predict the properties of the factor contributions in these general circumstances, the results of Section 4 and elsewhere will help researchers understand why certain features are observed in practice. For example, Propositions 1 and 2 will help explain why groups of factors can be treated as a single entity without affecting their total contribution, and Proposition 5 tells us to expect that a negative inequality contribution will be attached to any income component which is distributed roughly evenly across the population.

Many other topics are obvious candidates for application of the Shapley decomposition procedure. These include the division of income mobility into structural and exchange components; a breakdown of the distributional impact taxes and benefits; the decomposition of wage inequality along the lines proposed by Juhn et al. [13], and the measurement of discrimination due to Oaxaca [17]. In the longer run, however, the applications with the greatest potential are the standard econometric formulations of applied economics, which all conform to the general specification (1.1) indicated at the outset. Fields [7] recognises the link between conventional OLS regressions and the problem of decomposing income inequality by source. The results of this paper suggest that the link can be extended to any econometric

specification used in applied economics, in order to supplement the standard measures of statistical significance with an assessment of the relative importance of the explanatory variables.

## Appendix

To demonstrate Propositions 1 and 2, we first define

$$\alpha(r, n) = \frac{n!}{r!(n-r)!}, \quad n \geq r, \quad (8.1)$$

and recall that

$$\pi(r, n) = \frac{r!(n-r)!}{(n+1)!} = \frac{1}{(n+1)\alpha(r, n)}, \quad n \geq r. \quad (8.2)$$

Identifying the coefficient of  $x^r$  in the expansion of  $(1-x)^{-s-1}(1-x)^{-(n-s)-1}$  reveals that

$$\begin{aligned} \sum_{t=0}^r \alpha(t, r) \pi(s+t, n+r) &= \sum_{t=0}^p \frac{r!(s+t)!(n+r-s-t)!}{t!(r-t)!(n+r+1)!} \\ &= \frac{s!(n-s)!}{(n+1)!} = \pi(s, n) \\ &= \sum_{t=1}^{r+1} \alpha(t-1, r) \pi(s+t-1, n+r) \end{aligned} \quad (8.3)$$

for all  $n \geq s \geq 0$  and all  $r \geq 0$ . The proofs of Propositions 1 and 2 are now established via the following two Lemmas

**Lemma 1** Consider the model  $\langle K, F \rangle$ , and suppose  $F$  is separable over  $L \subseteq K$ . Then, for all  $S \subseteq K \setminus L$  and all  $n \geq s \geq 0$ , we have

$$\sum_{T \subseteq L} \pi(s + |T|, n + |L|) F(S \cup T) = \sum_{T \subseteq \{L\}} \pi(s + |T|, n + |L|) F(S \cup K_T) \quad (8.4)$$

*Proof* Let  $p = |L|$  and  $t = |T|$ . Given that  $F$  is separable over  $L$ , condition (4.11) yields

$$\begin{aligned} \sum_{\substack{T \subseteq L \\ |T|=t}} F(S \cup T) &= \alpha(t, p) F(S) + \sum_{\substack{T \subseteq L \\ |T|=t}} \sum_{k \in T} \Delta_k F(S) \\ &= \alpha(t, p) F(S) + \alpha(t, p) \frac{t}{p} \sum_{k \in L} \Delta_k F(S) \\ &= \alpha(t, p) \left[ \frac{p-t}{p} F(S) + \frac{t}{p} F(S \cup L) \right] \end{aligned} \quad (8.5)$$

plement the standard measure of the relative importance of the

$$r, \quad n \geq r. \quad (8.1)$$

$$-x)^{-s-1}(1-x)^{-(n-s)-1} \text{ reveals} \quad (8.2)$$

$(n+r-s-t)!$

$t!(n+r+1)!$

$$(s, n)$$

$$r)\pi(s+t-1, n+r) \quad (8.3)$$

ns 1 and 2 are now established

's separable over  $L \subseteq K$ . Then,

$$, n + |L|)F(S \cup K_T) \quad (8.4)$$

table over  $L$ , condition (4.11)

$$\begin{aligned} & \sum_{k \in T} \Delta_k F(S) \\ & \frac{t}{p} \sum_{k \in L} \Delta_k F(S) \\ & - \left[ \frac{t}{p} F(S \cup L) \right] \end{aligned} \quad (8.5)$$

Hence, using Eq. 8.3,

$$\begin{aligned} & \sum_{T \subseteq L} \pi(s + |T|, n + |L|)F(S \cup T) \\ & = \sum_{t=0}^p \pi(s + t, n + p) \sum_{\substack{T \subseteq L \\ |T|=t}} F(S \cup T) \\ & = \sum_{t=0}^{p-1} \alpha(t, p-1) \pi(s + t, n + p) F(S) \\ & \quad + \sum_{t=1}^p \alpha(t-1, p-1) \pi(s + t, n + p) F(S \cup L) \\ & = \pi(s, n+1) F(S) + \pi(s+1, n+1) F(S \cup L) \\ & = \sum_{T \subseteq \{L\}} \pi(s + |T|, n + |L|)F(S \cup K_T) \end{aligned}$$

This completes the proof of Lemma 1.  $\square$

**Lemma 2** Given the hierarchical model  $\langle K, A, F \rangle$ , consider any  $B \subseteq A$  such that  $F$  is separable over each  $L \in B$ . Then, for all  $S \subseteq K \setminus K_B$  and all  $n \geq s \geq 0$ , we have

$$\begin{aligned} & \sum_{T \subseteq K_B} \pi(s + |T|, n + |K_B|)F(S \cup T) \\ & = \sum_{T \subseteq B} \pi(s + |T|, n + |B|)F(S \cup K_T) \end{aligned} \quad (8.6)$$

*Proof* Let  $B = \{L_1, L_2, \dots, L_r\}$ . Then repeated application of Lemma 1 yields

$$\begin{aligned} & \sum_{T \subseteq K_B} \pi(s + |T|, n + |K_B|)F(S \cup T) \\ & = \sum_{T_1 \subseteq L_1} \sum_{T_2 \subseteq L_2} \dots \sum_{T_r \subseteq L_r} \pi \left( s + \sum_{j=1}^r |T_j|, n + \sum_{j=1}^r |L_j| \right) F \left( \bigcup_{j=1}^r T_j \cup S \right) \\ & = \sum_{T_1 \subseteq \{L_1\}} \sum_{T_2 \subseteq L_2} \dots \sum_{T_r \subseteq L_r} \pi \left( s + \sum_{j=1}^r |T_j|, n + 1 + \sum_{j=2}^r |L_j| \right) F \left( \bigcup_{j=2}^r T_j \cup K_{T_1} \cup S \right) \\ & = \sum_{T_1 \subseteq \{L_1\}} \sum_{T_2 \subseteq \{L_2\}} \dots \sum_{T_r \subseteq \{L_r\}} \pi \left( s + \sum_{j=1}^r |T_j|, n + r \right) F \left( \bigcup_{j=1}^r K_{T_j} \cup S \right) \\ & = \sum_{T \subseteq B} \pi(s + |T|, n + |B|)F(S \cup K_T) \end{aligned} \quad (8.7)$$

This completes the proof of Lemma 2.  $\square$

We now proceed to demonstrate Propositions 1 and 2.

*Proof of Proposition 1* Let  $N = K \setminus L$ ,  $m = |K|$ , and  $n = |N|$ . Then, using Lemma 1, we have

$$\begin{aligned} C_k^S(K, F) &= \sum_{S \subseteq N \setminus \{k\}} \sum_{T \subseteq L} \pi(|S \cup T|, m - 1) \Delta_k F(S \cup T) \\ &= \sum_{S \subseteq N \setminus \{k\}} \sum_{T \subseteq \{L\}} \pi(|S \cup T|, n) \Delta_k F(S \cup K_T) \\ &= \mathcal{E}_{S \subseteq \{L\} \cup N \setminus \{k\}} \Delta_k F^{\{L\}}(S) = C_k^S(\{L\} \cup N, F^{\{L\}}) \end{aligned} \quad (8.8)$$

for all  $k \in N$ , as required in Eq. 4.12. In addition, since the Shapley decomposition is exact, it follows that

$$\begin{aligned} C_{\{L\}}^S(\{L\} \cup N, F^{\{L\}}) &= F^{\{L\}}(\{L\} \cup N) - \sum_{k \in N} C_k^S(\{L\} \cup N, F^{\{L\}}) \\ &= F(K) - \sum_{k \in N} C_k^S(K, F) = \sum_{k \in L} C_k^S(K, F) \end{aligned} \quad (8.9)$$

as required for Eq. 4.12b.

Finally, for all  $k \in L$  and all  $s$  such that  $n \geq t \geq 0$ , separability over  $L$  implies

$$\begin{aligned} &\sum_{S \subseteq L \setminus \{k\}} \pi(|S| + t, m - 1) \Delta_k F(S \cup T) \\ &= \sum_{s=0}^{m-n-1} \sum_{\substack{S \subseteq L \setminus \{k\} \\ |S|=s}} \pi(s + t, m - 1) \Delta_k F(T) \\ &= \sum_{s=0}^{m-n-1} \alpha(s, m - n - 1) \pi(s + t, m - 1) \Delta_k F(T) \\ &= \pi(t, n) \Delta_k F(T), \end{aligned} \quad (8.10)$$

using Eq. 8.3. Hence

$$\begin{aligned} C_k^S(K, F) &= \sum_{T \subseteq N} \sum_{S \subseteq L \setminus \{k\}} \pi(|S \cup T|, m - 1) \Delta_k F(S \cup T) \\ &= \sum_{T \subseteq N} \pi(|T|, n) \Delta_k F(T) = \mathcal{E}_{T \subseteq N} \Delta_k F(T) \\ &= C_k^S(N \cup \{k\}, F), \end{aligned} \quad (8.11)$$

as required for Eq. 4.12c, and the proof of Proposition 1 is complete.  $\square$

12.  
 $=|N|$ . Then, using Lemma 1,

$$\Delta_k F(S \cup T) = C_k^S(K, F) = C_k^S(\{L\} \cup N, F^{(L)}) \quad (8.8)$$

the Shapley decomposition is

$$C_k^S(\{L\} \cup N, F^{(L)}) = \sum_{k \in L} C_k^S(K, F) \quad (8.9)$$

separability over  $L$  implies

$$\Delta_k F(T) = (n-1)\Delta_k F(T) \quad (8.10)$$

$$\Delta_k F(S \cup T) = \sum_{T \subseteq N} \Delta_k F(T) \quad (8.11)$$

is complete.  $\square$

*Proof of Proposition 2* Consider any  $L \in A$  and any  $k \in L$ , and let  $m = |K|$  and  $n = |K \setminus L|$ . Then, Eq. 8.9 and Lemma 2 yield the Shapley contributions

$$\begin{aligned} C_k^S(K, F) &= \sum_{T \subseteq K \setminus \{L\}} \sum_{S \subseteq L \setminus \{k\}} \pi(|S \cup T|, m-1) \Delta_k F(S \cup T) \\ &= \sum_{T \subseteq K \setminus \{L\}} \pi(|T|, n) \Delta_k F(T) = \sum_{T \subseteq A \setminus \{L\}} \pi(|T|, |A|-1) \Delta_k F(K_T) \\ &= \mathcal{E}_{T \subseteq A \setminus \{L\}} \Delta_k F(K_T) = \Delta_k \bar{F}_L(\emptyset). \end{aligned} \quad (8.12)$$

Since  $F$  is separable over  $L$ , it follows from Eq. 4.8 that

$$\begin{aligned} C_k^O(K, A, F) &= \mathcal{E}_{T \subseteq A \setminus \{L\}} \mathcal{E}_{S \subseteq L \setminus \{k\}} \Delta_k F(K_T \cup S) \\ &= \mathcal{E}_{T \subseteq A \setminus \{L\}} \mathcal{E}_{S \subseteq L \setminus \{k\}} \Delta_k F(K_T) = \mathcal{E}_{T \subseteq A \setminus \{L\}} \Delta_k F(K_T). \end{aligned} \quad (8.13)$$

So the proof of Proposition 2 is complete.  $\square$

## References

1. Bouillon, C., Legovini, A., Lustig, N.: Rising inequality in Mexico: returns to household characteristics and the 'Chiapas effect'. Paper presented at the LACEA Conference, Buenos Aires (1998)
2. Bourguignon, F.: Decomposable income inequality measures. *Econometrica* **47**, 901–920 (1979)
3. Bourguignon, F., Fournier, M., Gurgand, M.: Distribution, development and education: Taiwan, 1979–1994. Paper presented at the LACEA Conference, Buenos Aires (1998)
4. Chantreuil, F., Trannoy, A.: Inequality decomposition values. Mimeo, Université de Cergy-Pointoise (1997)
5. Cowell, F.A., Jenkins, S.P.: How much inequality can we explain—a methodology and an application to the United-States. *Econ. J.* **105**, 421–430 (1995)
6. Datt, G., Ravallion, M.: Growth and redistribution components of changes in poverty measures—a decomposition with applications to Brazil and India in the 1980s. *J. Dev. Econ.* **38**, 275–296 (1992)
7. Fields, G.S.: Accounting for differences in income inequality. Mimeo, Cornell University (1995)
8. Foster, J.E., Shneyerov, A.A.: Path independent inequality measures. Discussion Paper No. 97-W04, Department of Economics, Vanderbilt University (1996)
9. Foster, J.E., Shneyerov, A.A.: A general class of additively decomposable inequality indices. Discussion Paper No. 97-W10, Department of Economics, Vanderbilt University (1997)
10. Foster, J.E., Greer, J., Thorbecke, E.: A class of decomposable poverty indices. *Econometrica* **52**, 761–765 (1984)
11. Grootaert, C.: Structural change and poverty in Africa: a decomposition analysis for Côte d'Ivoire. *J. Dev. Econ.* **47**, 375–402 (1995)
12. Jenkins, S.P.: Accounting for inequality trends: decomposition analyses for the UK, 1971–86. *Economica* **62**, 29–64 (1995)
13. Juhn, C., Murphy, K.M., Pierce, B.: Wage inequality and the rise in returns to skill. *J. Polit. Econ.* **101**, 410–442 (1993)
14. Lambert, P.J., Aronson, J.R.: Inequality decomposition analysis and the Gini coefficient revisited. *Econ. J.* **103**, 1221–1227 (1993)
15. Morduch, J., Sinclair, T.: Rethinking inequality decomposition, with evidence from rural China. Mimeo, Stanford University (1998)
16. Moulin, H.: Axioms of Cooperative Decision Making. Cambridge University Press (1988)
17. Oaxaca, R.: Male-female wage differentials in urban labour markets. *Int. Econ. Rev.* **14**, 693–709 (1973)
18. Owen, G.: Values of games with priori unions. In: Heim, R., Moeschlin, O. (eds.) Essays in Mathematical Economics and Game Theory. Springer, New York (1977)

19. Pyatt, G.: On the interpretation and disaggregation of Gini coefficients. *Econ. J.* **86**, 243–255 (1976)
20. Ravallion, M., Huppi, M.: Measuring changes in poverty: a methodological case study of Indonesia during an adjustment period. *World Bank Econ. Rev.* **5**, 57–84 (1991)
21. Rongve, I.: A Shapley decomposition of inequality indices by income source. Discussion Paper #59, Department of Economics, University of Regina (1995)
22. Shapley, L.: A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) *Contributions to the Theory of Games*, vol. 2. Princeton University Press (1953)
23. Shorrocks, A.F.: The class of additively decomposable inequality measures. *Econometrica* **48**, 613–625 (1980)
24. Shorrocks, A.F.: Inequality decomposition by factor components. *Econometrica* **50**, 193–211 (1982)
25. Shorrocks, A.F.: Inequality decomposition by population subgroups. *Econometrica* **52**, 1369–1385 (1984)
26. Szekely, M.: Poverty in Mexico during adjustment. *Rev. Income Wealth* **1995**(3), 331–348 (1995)
27. Theil, H.: *Statistical Decomposition Analysis*. North Holland, Amsterdam (1972)
28. Thorbecke, E., Jung, H.S.: A multiplier decomposition method to analyze poverty alleviation. *J. Dev. Econ.* **48**, 279–300 (1996)
29. Tsui, K.-Y.: Growth-equity decomposition of a change in poverty: an axiomatic approach. *Econ. Lett.* **50**, 417–424 (1996)
30. Young, H.P.: Monotonic solutions of cooperative games. *Int. J. Game Theory* **14**, 65–72 (1985)