Journal of Economic Geography **5** (2005) pp. 59–81
Advance Access originally published online on 7 January 2005 doi:10.1093/jnlecg/lbh054

Spatial decomposition of inequality

Anthony Shorrocks and Guanghua Wan***

Abstract

This paper reviews the theory and application of the decomposition methods commonly used to measure the impact of spatial location on income inequality. It establishes some new theoretical results with potentially wide applicability, and examines empirical evidence drawn from a large number of countries.

Keywords: income inequality, decomposition, spatial, regional
JEL classification: C43, D31, D63, R12
Date submitted: January 2004 **Date accepted:** May 2004

1. Introduction

Spatial disparities in living standards have been the subject of a great deal of attention in recent years. At the global level there has been concern at the prospect of rising inequality in the world distribution of income and the extent to which this is fuelled by factors linked to globalization: see, for example, Milanovic (2002); Sala-i-Martin (2002); Bourguignon and Morrison (2002); Fischer (2003); Kremer and Maskin (2003). Similar concerns surface within individual countries, especially those countries where income inequality has been rising over time and where average incomes vary considerably across regions or provinces. In China, for example, unease with the growing disparity between the living standards in the coastal areas and the inland regions has prompted the Chinese government to launch a campaign to develop the western regions (Kanbur and Zhang, 2005). The problem becomes a more intense political issue when spatial inequality is perceived to be related to discrimination against particular groups of citizens such as rural farmers (compared to urban residents), ethnic minorities concentrated in remote areas, migrants in certain districts, or religious groups in particular regions (e.g., Muslims in Xinjiang Region in China).

Residential location is not, of course, the only factor which accounts for differences in living standards: there are typically wide disparities in incomes within, as well as between, regions. To assess the significance of location on income inequality therefore requires a method of identifying and measuring the contribution of the spatial factors that affect living standards. Empirical regression studies routinely include regional dummy variables in order to control for location. This helps to determine whether the influence of spatial location is statistically significant. However, little or no attention is given to measuring the quantitative impact of spatial factors on the level or trend in aggregate income inequality—at least in terms of the inequality measures commonly employed elsewhere.

* World Institute for Development Economics Research, United Nations University (UNU-WIDER), Katajanokanlaituri 6 B, 00160 Helsinki, Finland.
email <Shorrocks@wider.unu.edu>
** World Institute for Development Economics Research, United Nations University (UNU-WIDER), Katajanokanlaituri 6 B, 00160 Helsinki, Finland.
email <wan@wider.unu.edu>

The main alternative approach applies the techniques developed for subgroup decomposition of inequality to the spatial context. Typically this involves partitioning a household data sample into a set of geographical regions or districts and then calculating two components of aggregate inequality: a weighted average regional inequality value known as the 'within-group' component; and a 'between-group' term which captures the inequality due to variations in average incomes across regions. Certain inequality indices have the convenient property that overall inequality can be expressed as the sum of these two terms, in a similar way to the age-old analysis of variance procedure. Other inequality indices—in particular, the Gini coefficient—require a third term that reflects 'interactions' or 'overlaps' between the regional income distributions.

The aim of this paper is to review and extend the current state of knowledge regarding inequality decomposition in a spatial context.[1] As will become apparent, particular attention is focussed on the properties of the between-group component of inequality, B, expressed as a proportion of total inequality, I. The between-group term represents the level of inequality that would be observed if the income of each person is replaced by the mean income of his or her respective region. It therefore provides the most immediate answer to the counterfactual questions: 'how much inequality would be observed if there was no inequality within regions?' or 'how much inequality would occur if regional income differences are the only source of inequality?'.

If policies are undertaken to reduce regional differences in living standards, then B, I, and B/I are all expected to fall. Conversely, if spatial variations become more evident, as has happened in Russia and China, for example, during the last decade, then between-group inequality is likely to increase both in absolute terms and as a share of total inequality. Monitoring the level and trend of B and B/I is therefore central to our appreciation of the distributional impact of spatial factors.

The paper begins in Section 2 with a review of the basic methodology of inequality decomposition. This material is included for completeness, but most of the details can be skipped by those readers who are more interested in the practical applications. Section 3 discusses a number of basic properties of the between-group inequality term concerning, for example, the way in which B depends on the number of identified regions, and the way in which B is related to the distribution of incomes across regions. These theoretical insights are not only interesting in themselves, they also help in the interpretation and assessment of empirical evidence on spatial decomposition.

Section 4 assembles data on spatial inequality decompositions for a number of countries and analyses the results. The empirical evidence indicates that the magnitude of the between-group term tends to increase with the number of identified regions, as predicted. But, the results also suggest that B is very sensitive to the way that the spatial categories are defined; a two-way rural-urban split, for example, appears to be much more significant than a two-way north-south or east-west division. Within the limited range of inequality indices captured in the empirical studies, the share of between-group inequality is relatively insensitive to the choice of inequality index, with the exception of the Gini coefficient which tends to produce larger values.

Section 5 concludes with some suggestions for future research topics.

1 The broad thrust of our results will also apply to inequality decomposition by subgroups partitioned according to gender, age, education level, or other criteria.

2. Theoretical foundations of spatial decomposition of inequality

The analysis of spatial inequality typically begins with a measure of living standards for a population of individuals or households. We follow common practice in referring to the measure of living standards as 'income', although it should be stressed that the income concept must be interpreted broadly to encompass not only home production and non-pecuniary income, but also all the advantages and disadvantages associated systematically with geographical location, including climate, regional price levels, local public good provision and environmental quality. In essence, the analysis assumes that individuals with the same income at different locations are equally well-off.

The formalities below are framed in terms of a (homogeneous, equally weighted) population of individuals represented by $N = \{1, 2, \ldots, n\}$, with incomes given by the vector $\mathbf{y} = (y_1, \ldots, y_n)$ and mean income denoted by μ. Income inequality is captured by an *inequality index* $I(\mathbf{y})$ which is assumed throughout to satisfy the following five basic properties:[2]

(A1) symmetry (or anonymity);
(A2) the Pigou-Dalton *principle of transfers* (or strict Schur convexity): a mean-preserving progressive transfer reduces inequality;[3]
(A3) scale invariance (or homogeneity of degree zero);
(A4) replication invariance; and
(A5) zero normalization: the minimum value of I is zero (achieved when all incomes are identical).

The decomposition of inequality according to a partition of the aggregate population into geographical regions (or, more generally, into any set of mutually exclusive and exhaustive subgroups) is most often undertaken with one of the entropy indices popularized by Theil (1967, 1972) and later explored in more detail by Bourguignon (1979), Shorrocks (1980, 1984, 1988), Cowell and Kuga (1981), and Foster and Shneyerov (2000), amongst others. The single parameter entropy family may be written:

$$E_c(\mathbf{y}) = \frac{1}{c(c-1)n} \sum_{i \in N} \left\{ \left(\frac{y_i}{\mu} \right)^c - 1 \right\}, \quad c \neq 0, 1, \tag{1a}$$

$$E_1(\mathbf{y}) = \frac{1}{n} \sum_{i \in N} \frac{y_i}{\mu} \ln \frac{y_i}{\mu}, \tag{1b}$$

$$E_0(\mathbf{y}) = \frac{1}{n} \sum_{i \in N} \ln \frac{\mu}{y_i}, \tag{1c}$$

or in the more condensed form:

$$E_c(\mathbf{y}) = \frac{1}{n} \sum_{i \in N} \varphi_c(y_i/\mu), \tag{2}$$

2 These are all standard properties of measures of relative inequality. For more details see Silber (1999).
3 Note that (A2) extends to situations in which frequencies $f_1, f_2, \ldots, f_n$ are attached to the income levels $y_1, y_2, \ldots, y_n$. Successive application of the principle of transfers implies that inequality will be reduced by any equalization of two income levels which preserves the overall (weighted) mean income.

where $\varphi_c(t) = (t^c - 1)/[c(c-1)]$, $c \neq 0$, 1; $\varphi_1(t) = t \ln t$; and $\varphi_0(t) = -\ln t$. Special cases include the Theil coefficient (corresponding to $c = 1$), the mean logarithmic deviation ($c = 0$), and one half of the square of the coefficient of variation ($c = 2$).

The decomposition properties of this class of measures are best illustrated by considering the index E_0 and by supposing that the set of individuals, N, is partitioned into m proper subgroups N_k ($k = 1, 2, \dots, m$), with respective income vectors $\mathbf{y}^k$, mean incomes μ_k, population sizes n_k, and population shares $\nu_k = n_k/n$. It will also be convenient to let $\bar{\mathbf{y}}^k$ denote the distribution obtained by replacing each income in the vector $\mathbf{y}^k$ with the subgroup mean μ_k. Then

$$\begin{aligned} E_0(\mathbf{y}) = E_0(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m) &= \frac{1}{n}\sum_{k=1}^{m}\sum_{i\in N_k} \ln\frac{\mu}{y_i} \\ &= \sum_{k=1}^{m}\frac{n_k}{n}\frac{1}{n_k}\sum_{i\in N_k} \ln\frac{\mu_k}{y_i} + \frac{1}{n}\sum_{k=1}^{m}\sum_{i\in N_k} \ln\frac{\mu}{\mu_k} \\ &= \sum_{k=1}^{m}\nu_k E_0(\mathbf{y}^k) + \sum_{k=1}^{m}\nu_k \ln\frac{\mu}{\mu_k} = \mathrm{W} + \mathrm{B}, \end{aligned} \tag{3}$$

where

$$\mathrm{W} = \sum_{k=1}^{m}\nu_k E_0(\mathbf{y}^k) \tag{4a}$$

is a weighted average of subgroup inequality values, traditionally referred to as the 'within-group' component of inequality; and

$$\mathrm{B} = \sum_{k=1}^{m}\nu_k \ln\frac{\mu}{\mu_k} = E_0(\bar{\mathbf{y}}^1, \bar{\mathbf{y}}^2, \dots, \bar{\mathbf{y}}^m) \tag{4b}$$

is the 'between-group' contribution to inequality, representing the level of inequality obtained by replacing the income of each person with the mean income of their respective subgroup. Thus—for the index E_0 at least—the overall level of inequality for a country can be expressed in an intuitively appealing fashion as an exact sum of the average inequality within regions and the inequality due purely to differences in average incomes between regions.

To appreciate the special attraction of the decomposition indicated by (3), it may be noted that for any inequality index $I(\cdot)$ which satisfies properties (A1)–(A5), the aggregate level of inequality may be written:

$$I(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m) = I\left(\mathbf{w}^1 b_1, \mathbf{w}^2 b_2, \dots, \mathbf{w}^m b_m\right) = \hat{I}(\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^m, \mathbf{b}), \tag{5}$$

where $\mathbf{w}^k = \mathbf{y}^k/\mu_k$ is the vector of relative incomes within region k, and $\mathbf{b} = (b^1, b^2, \dots, b^m)$, $b^k = \mu^k/\mu$, denotes the vector of relative mean incomes across regions. Equation (5) makes it clear that aggregate inequality in a country is completely accounted for by differences in relative incomes within regions (as captured in the vectors $\mathbf{w}^k$) and differences in relative mean income between regions (as captured by $\mathbf{b}$). In this context, it is natural to regard the inequality contribution of $\mathbf{w}^k$ as the amount by which aggregate inequality falls if relative incomes in region k are equalized, *ceteris paribus*; and the contribution of $\mathbf{b}$ as the amount by which aggregate inequality falls if regional mean income differences are eliminated, holding constant relative incomes within regions (i.e., by proportionately scaling incomes within each region until each regional mean matches the average for the whole population). For the index E_0, the within- and between-group components (4a) and (4b) conform to these interpretations. Furthermore, the values of the contributions are invariant to the order in which the within- and between-group differences are eliminated.

Other inequality indices drawn from the entropy family (1) satisfy a similar decomposition equation given by

$$E_c(\mathbf{y}) = E_c(\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m) = \sum_{k=1}^{m} \nu_k b_k^c E_c(\mathbf{y}^k) + \sum_{k=1}^{m} \nu_k \varphi_c(b_k) = W + B, \tag{6}$$

which again leads to a natural interpretation in terms of the within- and between-group contributions to inequality:

$$W = \sum_{k=1}^{m} \nu_k b_k^c E_c(\mathbf{y}^k); \tag{7a}$$

$$B = \sum_{k=1}^{m} \nu_k \varphi_c(b_k) = E_c(\bar{\mathbf{y}}^1, \bar{\mathbf{y}}^2, \ldots, \bar{\mathbf{y}}^m). \tag{7b}$$

However, the decomposition provided by (6) is less satisfactory than that given by (3) for two reasons. First, while the within-group term remains a weighted sum of regional inequality values, the weights typically do not sum to one, unless $c = 0$ or 1.[4] So it is usually wrong to interpret W as the *average* level of inequality within regions. Secondly, the within-group component now depends both on within-group differences and (via the weights) on between-group differences. So any attempt to eliminate between-group variation along the lines suggested following Equation (5) now has an indirect effect on the value of the within-group term. As a consequence the quantitative impact of eliminating within- and between-group variations is now sensitive to the order in which the factors are considered.[5]

The main appeal of the decomposition provided by (6) rests on the fact that the inequality indices are *subgroup consistent* in the following sense: holding regional mean incomes and population sizes fixed, an increase in inequality within each region must lead to an increase (or, at least, not a decrease) in inequality in the country as a whole. This property is evidently true for the entropy measures, since the *ceteris paribus* clause implies that the between-group term B is constant in (6), and that the rise in inequality within each region translates into a rise in the weighted sum of regional inequality values captured by the within-group component W.

4 In the latter case (i.e., the Theil coefficient) the weights correspond to the regional income shares.

5 This special property of E_0 corresponds to the 'path independence' property discussed by Foster and Shneyerov (2000). They observe that there are two ways of deriving W and B. As discussed earlier, the first obtains the between-group B contribution as the level of inequality which results after within-group inequality is eliminated by redistributing incomes equally within each region. W is then taken as the residual. The second defines W to be the level of inequality which results when inequalities between groups are eliminated by proportionally scaling each subgroup distribution until it has the same mean as the overall distribution, with the residual now taken to be B. The decomposition is said to be *path-independent* if these two methods produce identical results.

Subgroup consistency is an intuitively appealing and relatively weak property satisfied by the Atkinson class of inequality measures and the Entropy family, but not the Gini coefficient (see Shorrocks, 1988). Although the failure to satisfy subgroup consistency implies that the Gini index is not amenable to a decomposition along the lines of Equation (6), a number of authors have nevertheless proposed alternative forms of Gini decompositions.[6] The method which most closely resembles (6) can be formulated by numbering the regions in order of increasing mean incomes, and by supposing that person i occurs in the ith position when the income distribution is written $\mathbf{y} = (\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m)$, and in position r_i when all incomes are arranged in increasing order. The value of the Gini coefficient is then given by

$$G(\mathbf{y}) = \frac{2}{n^2\mu} \sum_{i \in N} r_i (y_i - \mu) \tag{8}$$

and yields the decomposition equation (see, for example, Lambert and Aronson 1993):

$$\begin{aligned} G = G(\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m) &= \frac{2}{n^2\mu} \sum_{k=1}^{m} \sum_{i \in N_k} r_i (y_i - \mu) \\ &= \frac{2}{n^2\mu} \sum_{k=1}^{m} \left\{ \sum_{i \in N_k} i(y_i - \mu_k) + \sum_{i \in N_k} i(\mu_k - \mu) + \sum_{i \in N_k} (r_i - i) y_i \right\} \\ &= W + B + R, \end{aligned} \tag{9}$$

where

$$W = \frac{2}{n^2\mu} \sum_{k=1}^{m} \sum_{i \in N_k} i(y_i - \mu_k) = \sum_{k=1}^{m} \nu_k^2 b_k G(\mathbf{y}^k) \tag{10a}$$

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

$$B = \frac{2}{n^2\mu} \sum_{k=1}^{m} \sum_{i \in N_k} i(\mu_k - \mu) = \sum_{k=1}^{m} b_k \nu_k \left[\sum_{j=1}^{k} \nu_j - \sum_{j=k}^{m} \nu_j \right] = G(\bar{\mathbf{y}}^1, \bar{\mathbf{y}}^2, \ldots, \bar{\mathbf{y}}^m) \tag{10b}$$

is the ‘between-group component’, representing the value of the Gini coefficient when the income of each individual is replaced by the mean income of the subgroup to which s/he belongs. The final term, R, in Equation (9) is a residual or ‘interaction’ effect which vanishes when the regional income ranges do not overlap (in which case $r_i = i$, for all i), and is otherwise strictly positive.

When the regional income ranges are non-overlapping there is a very clear correspondence between the Gini decomposition (9) and the Entropy formulation (6); the only substantive difference is that the regional inequality weights are given by $\mathrm{v}_k b_k^c$ in (6) and by $\nu_k^2 b_k$ in (9). In this case also, it is natural to regard the ratio B/G as a measure of the proportional contribution of inter-regional income variation to total inequality, mimicking the analogous expression B/E_c used in the context of Entropy indices. The situation becomes more problematic when the regional income ranges overlap, because the interaction term introduces a third, poorly specified, element into the decomposition Equation (9). It is also important to note that (9) is not the only form of Gini

6 Early contributions include Soltow (1960), Bhattacharya and Mahalanobis (1967), Rao (1969), Mangahas (1975), Pyatt (1976), and Zagier (1983).

decomposition on offer; many other specifications have been suggested. The more recent proposals typically retain the division into within-, between-, and interaction terms, but differ in the formulae used for each of the components.[7] In the absence of an obviously superior alternative, we proceed on the assumption that the between-group term B defined in (10b) expressed as a proportion of the overall Gini value captures what we mean by the importance of the contribution of regional income variations to total inequality as measured by the Gini coefficient.[8]

Another departure from the traditional decomposition framework explores the implications of generalising the notion of average regional income to measures other than the mean. The idea dates back to Blackorby et al. (1981) but has recently been explored in greater detail by Foster and Schneyerov (2000). In this framework the between-group term is constructed by replacing the income of each individual with a suitably defined representative income level for the region. Employing representative income levels other than the mean expands the set of inequality indices that have simple and attractive decomposition properties. This opens up some interesting lines for future research, but since no empirical applications have yet seen the light of day they are not pursued further in this paper.

3. Basic theoretical results

Despite the widespread use of decomposition techniques, little attention has been given in the past to establishing general decomposition results. The range of indices considered in the last section, combined with the possibility of alternative decomposition specifications, suggests that it may be difficult to draw general conclusions about the way in which spatial factors affect inequality. This conjecture turns out to be unduly pessimistic. The results given below apply to any inequality index $I(\cdot)$ which satisfies (A1)-(A5), and refer to the properties of a 'between-group term' B defined by

$$B = I(\bar{\mathbf{y}}^1, \bar{\mathbf{y}}^2, \ldots, \bar{\mathbf{y}}^m), \tag{11}$$

in accordance with (4b), (7), and (10b) above.

Consider first the range of values for B, and how the value compares with the overall inequality value $I = I(\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m)$. Intuitively, when only one group is identified (i.e., $m = 1$) then average incomes do not vary across regions and B must be zero. At the other end of the scale, if the number of regions is the same as the size of the population (i.e., $m = n$), then each region contains a single observation and B must equal I. It is also reasonable to expect that these two cases represent the minimum and maximum values that B can take, so that:

Proposition 1: (a) $0 \leq B \leq I$; (b) $B = 0$ if $m = 1$; (c) $B = I$ if $m = n$.

To establish Proposition 1, note that (b) and (c) are both immediate consequences of the definition of B given by (11), because B is the inequality value obtained by replacing the income of each person by the corresponding regional mean. In addition, we have $B \geq 0$

7 See for example Silber (1989), Yitzhaki and Lerman (1991), Yitzhaki (1994), and Sastry and Kelkar (1994).

8 However note that eliminating between-group inequality by scaling incomes within each subgroup until each subgroup mean matches the population average affects both the within-group term (as is the case with most Entropy indices) *and* the interaction component (which has no Entropy counterpart).

because the index $I(\cdot)$ is always non-negative (by A5); and $B \leq I$ because the 'regionally equalized' distribution $(\bar{\mathbf{y}}^1, \bar{\mathbf{y}}^2, \ldots, \bar{\mathbf{y}}^m)$ is obtained from the original distribution $(\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m)$ by applying an equalizing (and hence inequality reducing, by A2) procedure to each region in turn.

The argument in support of Proposition 1 also suggests a non-decreasing relationship between the number of regions and the magnitude of the between-group term. An increase in the number of regions will increase the opportunities for differentiating between the regional mean values used in the calculation of B, thereby causing the value of B to rise. This is most easily seen by reversing the question and asking about the consequences of reducing the number of regions via a merger between two regions. The impact on the value of B is equivalent to that of a mean-preserving equalization of the two subgroup income levels which, by the principle of transfers (A2), cannot increase the value of B. Hence:

Proposition 2: The value of B does not increase if any two regions are combined.[9]

The 'finer partition' characterization in Proposition 2 is one way of capturing the idea that B increases monotonically with m. Another possible interpretation is that the between-group term is larger *on average* when more regions are identified, in other words:

Conjecture 3: The expected value of B increases with m.[10]

This conjecture is not well formulated at present, since it is not clear how the expectation is to be taken over the space of partitions and over the allocation of individuals to subgroups. For example, each of the partitions into m regions may be treated as equally likely, or they may be assigned a probability corresponding to the likelihood that this partition is observed when n individuals are randomly distributed across m categories.

While a formal proof of Conjecture 3 is beyond the scope of this paper, intuition suggests that the result must hold under a variety of interpretations for the following reason. For a fixed-size population, an increase in the number of regions causes the average size to fall, so the distribution shifts towards smaller sized regions. But, as the mean value of smaller samples exhibits greater variability, the net effect is an increase in the expected inequality value captured in the between-group term (11).

The shift towards smaller sized classes can be formalized when n individuals are randomly allocated across m regions, each containing at least one person. The probability that a region contains $r+1$ members ($r \geq 0$) is then given by the multinomial value

$$\pi_{r+1}(m) = \frac{(n-m)!}{r!(n-m-r)!}\left(\frac{1}{m}\right)^r\left(\frac{m-1}{m}\right)^{n-m-r}, \quad r = 0, \ldots, n-m, \tag{12}$$

from which it follows that

$$\begin{aligned} \pi_{r+2}(m)/\pi_{r+1}(m) &= \frac{n-m-r}{r+1}\frac{1}{m-1} = \frac{n-r-1}{(r+1)(m-1)} + \frac{1}{r+1} \\ &\geq \pi_{r+2}(m+1)/\pi_{r+1}(m+1). \end{aligned} \tag{13}$$

9 Note that successive application of Proposition 2 allows Proposition 1(a) to be derived from Proposition 1(b) and 1(c).

10 We are indebted to Ravi Kanbur for this conjecture.

In other words, the frequency of larger regions falls off faster as the number of regions increases. According to Proposition B.1 of Marshall and Olkin (1979, p.129), condition (13) ensures that $\pi(m)$ is majorized by (i.e., Lorenz dominates) $\pi(m+1)$ for all m, so that

$$\sum_{r=0}^{s} \pi_r(m) \geq \sum_{r=0}^{s} \pi_r(m+1), \quad \text{for all } m \text{ and all } s. \tag{14}$$

This is the formal sense in which the distribution of regions shifts towards smaller sizes as the number of regions increases. A similar condition is likely to hold when alternative methods are used to allocate a given population of individuals to a given number of groups.

Let us now fix the partition level (m), the sizes of regions $\{n_1, \ldots, n_m\}$, and the overall income distribution $\mathbf{y} = (y_1, \ldots, y_n)$, and consider what can be said about the way in which B depends on the allocation of individuals (and hence incomes) across the regions. The following two observations follow immediately from the definition (11) of the between-group term.

Proposition 4: (a) if the distribution of income is the same in each region then $B = 0$; (b) if regional mean incomes are all equal then $B = 0$.

Note that the prerequisite in part (b) of Proposition 4 is significantly weaker than the corresponding requirement in part (a).

Proposition 4(a) refers to the situation in which the subgroup distributions overlap to the greatest possible extent. It seems plausible to suppose that a reduction in the degree of overlap will translate into a larger between-group term, but the precise relationship is difficult to formalize given that a reduction in overlap between two subgroups may not necessarily cause the subgroup means to move apart. At the other end of the scale, however, it is possible to establish that if the regional income ranges (strictly) overlap then the between-group term is not a maximum, and hence:

Proposition 5: B is a maximum only if the regional income ranges do not overlap.

The argument is as follows. Suppose that regions k and ℓ have strictly overlapping distributions and that $\mu_k \leq \mu_\ell$. Choose $i \in N_k$ and $j \in N_\ell$ such that $y_i > y_j$. Then swapping the incomes y_i and y_j between the two regions raises the mean income in the 'more affluent' region ℓ, so the switch corresponds to a regressive Pigon-Dalton transfer between the two subgroup income levels which, by appeal to (A2), must increase the inequality value represented by B.

Proposition 5 establishes that for B to be a maximum it is necessary that the income intervals are disjoint. If the subgroup sizes are constrained to be equal, then the non-overlapping condition is also sufficient. In other cases, the magnitude of B will typically depend on the relative sizes of the different subgroups and their positions along the income range. The exact relationship is difficult to predict, but the similar exercise carried out by Davies and Shorrocks (1989) suggests that maximization of B will require the larger (disjoint) groups to be positioned at the centre of the income range, and subgroup sizes to decline monotonically towards each tail.[11]

11 Interestingly, Davies and Shorrocks (1989) show that the between-group component in the Gini decomposition can closely approximate total inequality with a relatively small number of subgroups, as long as the subgroup income ranges are non-overlapping and the group sizes are chosen optimally.

4. Empirical evidence

There is now a large empirical literature on inequality decomposition by population subgroups defined in terms of spatial location. The number of studies which report inequality decompositions using non-spatial elements (education, age, gender, etc.) is even greater. Given the focus of this paper, attention is confined to spatial applications; but the broad conclusions may well apply also in non-spatial contexts.

The types of questions we attempt to address are as follows: do any general patterns or conclusions emerge from the empirical literature? Does the empirical evidence conform with the expectation (via Proposition 2 and Conjecture 3) of a positive relationship between the number of regions and the share of the between-group component? To what extent do the decomposition results depend on the measure of inequality? Are the results sensitive to the 'income' variable used in the analysis?

Most spatial decomposition studies employ either the mean logarithmic deviation index E_0 or the Theil Index E_1. Tables 1A and 1B summarize the results obtained from applying the decomposition of E_0 to many countries and points of time. Given the differences in sample size, choice of income variable, selection of regions, etc., reliable general conclusions are hard to draw. For this reason, it is more appropriate to use the term 'observation' rather than 'finding' or 'conclusion'.

Observation 1: the magnitude of the between-group component

As is typical of most subgroup decompositions, the between-group component is small relative to the within-group component except in the case of rural-urban divide (see Observation 2). This is particularly true when earnings data are used (see Observation 3). Excluding these two sets of circumstances, the share of the between-group component averages 12% with a minimum of 0% and a maximum of 51%.

Some researchers have concluded from this type of evidence that space or location is a relatively unimportant explanation of inequality (see, for example, Cowell and Jenkins, 1995). However, it should be noted that, as a determinant of inequality, space is poorly defined. Spatial location is often not of interest itself, but rather because of its association with many other important influences such as natural resources, weather conditions, infrastructure, cultural traditions, and even institutional arrangements. While some of these factors may contribute positively to the between-group component, nationwide policies such as national wage-setting arrangements are likely to make a negative contribution. Current procedures assign all of these factors to location without trying to disentangle the individual effects. So, until the definition of space is clarified, use of the estimated between-group component as a measure of the spatial contribution should be treated with caution. Caution also needs to be exercised when drawing policy conclusions from the empirical evidence. As noted by Kanbur (2003), if space is related to race or ethnicity, a small between-group component may not accurately reflect the significance of space as a determinant of inequality.

Observation 2: the rural-urban divide

The rural-urban division seems always to produce a much larger between-group component. It ranges from 9% for Greece to as much as 78% for China. This latter result was obtained using regional averages rather than household level data for China, and is therefore not strictly comparable. However, household-level data for

Table 1A. Spatial decomposition of E_0: income or expenditure

Country	No. of groups	Year	Total inequality	Between %	Within %	Category	Reference	Remarks
Canada	9	1991	0.264	1.9	98.1	Province	Gray et al. (2003)	survey data on total household income
	9	1992	0.272	1.5	98.5			
	9	1993	0.273	1.8	98.2			
	9	1994	0.272	1.5	98.5			
	9	1995	0.272	1.8	98.2			
	9	1996	0.281	1.8	98.2			
	9	1997	0.312	1.6	98.4			
China	2	1983	0.079	6.5	93.6	Coast/inland	Kanbur and Zhang (1999)	regional data on per capita consumption expenditure
	2	1984	0.076	6.6	93.5			
	2	1985	0.075	6.0	94.0			
	2	1986	0.080	6.3	93.7			
	2	1987	0.083	6.7	93.4			
	2	1988	0.089	8.0	92.0			
	2	1989	0.088	7.2	92.8			
	2	1990	0.091	7.5	92.5			
	2	1991	0.098	9.1	90.9			
	2	1992	0.108	11.6	88.4			
	2	1993	0.112	12.9	87.1			
	2	1994	0.117	14.7	85.3			
	2	1995	0.120	17.3	82.7			
China	2	1983	0.079	78.1	21.9	Urban/rural		
	2	1984	0.076	75.8	24.2			
	2	1985	0.075	77.0	23.1			
	2	1986	0.080	74.5	25.5			
	2	1987	0.083	74.8	25.2			
	2	1988	0.089	74.7	25.3			
	2	1989	0.088	73.3	26.7			
	2	1990	0.091	74.9	25.1			
	2	1991	0.098	75.5	24.5			
	2	1992	0.108	73.5	26.5			
	2	1993	0.112	75.1	24.9			
	2	1994	0.117	73.3	26.7			
	2	1995	0.120	70.7	29.3			

Continued

Table 1A. (*continued*)

Country	No. of groups	Year	Total inequality	Between %	Within %	Category	Reference	Remarks
China	2	1994	0.330	37.7	62.3	Urban/rural	Lee (2000)	1994 county/city data on per capita consumption
	3	1994	0.330	28.0	72.0	Zone		
	26	1994	0.330	36.8	63.2	Region		
China	2	1994	0.390	25.8	74.2	Urban/rural		1994 county/city data on per capita GVIAO
	3	1994	0.390	39.0	61.0	Zone		
	26	1994	0.390	51.5	48.5	Region		
						County within		
China	4	1994	0.141	24.0	76.0	Jilin	Cheng Y (1996)	1994 household survey data on per capita income
	4	1994	0.070	20.0	80.0	Shandong		
	4	1994	0.075	9.0	91.0	Sichuan		
	4	1994	0.232	9.0	91.0	Guangdong		
	4	1994	0.139	4.0	96.0	Jiangxi		
China	5	1994	0.222	39.0	61.0	Province		
Ecuador	3	1994	N/A	0.0	100.0	Rural region	Elbers et al. (2005)	1994 estimated household data on per capita expenditure
	21	1994	N/A	1.3	98.7	Rural province		
	195	1994	N/A	5.9	94.1	Rural canton		
	915	1994	N/A	14.1	85.9	Rural Parroquia		
	5	1994	N/A	6.6	93.4	Urban region		
	19	1994	N/A	7.3	92.7	Urban province		
	87	1994	N/A	8.6	91.4	Urban Canton		
	664	1994	N/A	23.3	76.7	Urban Zonas		
Finland	4	1971	0.127	12.5	87.5	Region	Loikkanen et al. (2002)	household data on per capita income
	4	1981	0.076	6.3	93.7			
	4	1990	0.069	7.6	92.4			
	4	1993	0.075	7.3	92.7			
	4	1998	0.104	4.4	95.6			
Ghana	3	1996	0.269	29.6	70.4	Residence area	Vanderpuye-Orgle (2002)	household data on per capita consumption expenditure
	6	1996	0.269	29.1	70.9	Residence area		

Greece	2	1974	0.196	10.1	89.9	Urban/rural	Tsakloglou (1993)	household data on per capita consumption expenditure
	2	1982	0.160	9.3	90.7			
	9	1974	0.196	12.4	87.6	Region		
	9	1982	0.159	8.7	91.3			
India	17	1977	0.277	5.0	95.0	Region	Mishra and Parikh (1992)	household data on per capita consumption expenditure
	17	1983	0.182	5.3	94.7			
	17	1977	0.214	10.9	89.1	Rural region		
	17	1983	0.164	8.2	91.8			
	17	1977	0.219	1.8	98.2	Urban region		
	17	1983	0.180	3.4	96.6			
Indonesia	2	1987	0.228	22.3	77.7	Urban/rural	Akita et al. (1999)	household data on per capita expenditure
	2	1990	0.223	22.0	78.0			
	2	1993	0.239	25.2	74.8			
Indonesia	27	1987	0.232	15.1	84.9	Province		
	27	1990	0.227	15.0	85.0			
	27	1993	0.243	17.3	82.7			
	26	1990	0.223	13.0	87.0	Province	Tadjoeddin (2003)	household data on per capita expenditure
	26	1993	0.239	15.0	85.0			
	26	1996	0.216	21.0	79.0			
	26	1998	0.172	22.0	78.0			
	26	1999	0.190	21.0	79.0			
	26	2002	0.233	15.0	85.0			
Madagascar	6	1993	N/A	4.8	95.2	Rural Faritany	Elbers et al. (2005)	1993 estimated household data on per capita expenditure
	104	1993	N/A	15.4	84.6	Rural Fivondrona		
	1117	1993	N/A	18.1	81.9	Rural Firaisana		
	6	1993	N/A	7.7	92.3	Urban Faritany		
	103	1993	N/A	21.7	78.3	Urban Fivondrona		
	131	1993	N/A	23.2	76.8	Urban Firaisana		
Mozambique	424	1996	N/A	22.0	78.0	Administrative post		estimated household data on per capita expenditure
	146	1996	N/A	18.4	81.6	District		
	11	1996	N/A	9.3	90.7	Province		

Continued

Table 1A. (*continued*)

Country	No. of groups	Year	Total inequality	Between %	Within %	Category	Reference	Remarks
Philippines	2	1985	0.282	17.2	82.8	Urban/rural	Balisacan and Fuwa (2003)	family income and expenditure survey data on per capita expenditure
	2	1988	0.264	16.6	83.4			
	2	1991	0.306	16.3	83.7			
	2	1994	0.260	15.6	84.4			
	2	1997	0.303	17.5	82.5			
	13	1985	0.282	15.4	84.6	Region		
	13	1988	0.264	13.0	87.0			
	13	1991	0.306	17.6	82.4			
	13	1994	0.260	13.5	86.5			
	13	1997	0.303	15.1	84.9			
Russia	77	1994	0.297	25.0	75.0	Region	Yemtsov (2005)	household budget survey data on per capita income
	77	1995	0.282	27.0	73.0			
	77	1996	0.316	26.0	74.0			
	77	1997	0.337	23.0	77.0			
	77	1998	0.314	28.0	72.0			
	77	1999	0.329	31.0	69.0			
Switzerland	3	1982	0.136	0.2	99.8	Region	Ernst et al. (2000)	household data on per capita income
	3	1992	0.159	0.6	99.4			

Table 1B. Spatial decomposition of E_0: earnings

Country	No. of groups	Year	Total inequality	Between %	Within %	Category	Reference	Data source
UK	12	1979	0.260	1.0	99.0	Region	Parker (1999)	FES data on employee income
	12	1985	0.310	1.8	98.2			
	12	1991	0.320	2.5	97.5			
	12	1994/5	0.330	2.4	97.6			
UK	12	1979	1.850	2.5	97.5	Region	Parker (2004)	FES data on self-employment income
	12	1985	0.650	3.0	97.0			
	12	1991	0.780	9.8	90.2			
	12	1994/5	1.520	3.0	97.0			
UK	11	1975	0.095	3.2	96.8	Region	Dickey (2001)	New Earnings Survey data for individuals
	11	1980	0.094	4.3	95.7			
	11	1991	0.133	6.8	93.2			
UK	11	1995	0.152	7.2	92.8	Region	Dickey (2001)	British Household Panel Survey data for individuals
	11	1991	0.213	12.2	87.8			
	11	1996	0.286	7.7	92.3			

China still yields a between-group component share of almost 38% (Lee, 2000). Overall, the studies applying a rural-urban split to household data yield an average between-group component of 19.6%, almost 8% higher than the average reported in Observation 1 above.

The between-group component depends on both the number of subgroups and differences in group means (or representative group values). Empirical evidence suggests that differences in means are the more important of these two factors, because the rural-urban distinction involves the minimum number of distinct subgroups. Other spatially defined decompositions often consider a much larger number of groups but produce a smaller within-group component, as is evident from the data for China, Indonesia, and the Philippines reported in Table 1A.

What is the reason for the relatively large between-group component in the rural-urban division? As mentioned earlier, the explanation may well lie in the inability of current decomposition techniques to control for other variables. Lower prices and/or availability of home produced food in rural areas may not be fully reflected in the data on living standards. Furthermore, the rural-urban divide in developing countries is often associated with other differences linked to the provision of infrastructure, employment opportunities, education, health care, access to capital and technology, and so on. In China, where the rural-urban separation has been largely institutionalized, it is not difficult to understand how these factors combine to produce a very large share for the between-group component of inequality. Controlling for the non-spatial factors linked to the rural-urban divide is likely to reveal a smaller between-group component associated with the pure spatial effect.

Observation 3: alternative income concepts

The data in Table 1A refer to income or consumption. Table 1B reports similar data for earnings in one country (the UK). The percentage share of the between-group component turns out to be much lower, ranging from 1% to 12%, with an average of 4.8%. Interestingly, total inequality in earnings is not smaller than inequality of income or consumption, which suggests that there is considerable wage variation across occupations or sectors, but relatively little variation in occupational wages across locations. While it is dangerous to extrapolate from data from a single country, the same result may apply to other market economies where there are no constraints on migration, and where returns to labour and human capital are more or less equalized across space. Collective bargaining, the strength of labour unions and national wage setting policies may also be influential.

Equal factor returns are not sufficient, of course, to produce a negligible between-group component in decompositions of earnings inequality: the employee structure of the workforce must also be similar across space. It would therefore be useful to decompose the within-group component further into a 'returns' effect and a 'workforce structure' effect, the former reflecting market development and migration, and the latter reflecting industrial structure.

Consumption smoothing causes the distribution of consumption to be more equal than the distribution of income, and also leads to the expectation that the share of between-group inequality will be smaller for income compared to consumption. Some support for this prediction is provided by the relatively low values reported for the income observations from Finland and Switzerland in Table 1A; but the small sample size means that the evidence is not conclusive.

Observation 4: alternative measures of inequality

A number of empirical studies report decomposition results based on different inequality indices, allowing the share of the between-group component to be compared across indices. The correlation coefficients presented in Table 2 show that the correlation amongst the various Entropy measures tends to be quite high. The correlation with the Gini values are somewhat lower. Overall, Table 2 suggests that the results obtained using one index should broadly carry over to other indices.

Observation 5: country coverage

Although spatial decompositions exist for the UK, USA, and some other developed countries, results on regional inequality are dominated by developing country evidence. The limited number of studies for developed countries does not imply that spatial inequality is not of interest in the developed world. However, the greater attention to developing and transition economies may reflect the fact that weak market forces, or restrictions on factor mobility, prevent returns to income generating factors from converging. In the search for explanations for the existence of spatial inequality,

Table 2. Correlation coefficient among shares of the between-group component

	E_1	E_2	Gini
E_0	0.98	0.83	0.65
E_1		0.98	0.64
E_2			0.75

it may be useful to compare the values of the between-group component obtained for developed and developing countries.

Observation 6: spatial price variations

The majority of empirical studies reported in Table 1 do not adjust for spatial price differences, although such differences exist and may substantially change the results for both developing countries such as China (see Wan, 2001) and free market economies such as the US (see Ram, 1992). Price levels are often correlated with living standards, so adjusting for spatial price differences will tend to lower the between-group term in the spatial decomposition while not altering inequality within regions (although the within-group component may be affected in an unpredictable fashion due to a change in the weights). Thus, even though the reported share of the between-group component is relatively small (see Observation 1 above), it is likely to be exaggerated, particularly in countries with a big land mass and underdeveloped markets.

Observation 7: the number of subgroups and the size of the between-group component

As discussed in Section 3, the between-group component is expected to rise as the number of groups increases. To examine the empirical evidence, Figures 1–4 present scatter plots of the share of the between-group component against the number of subgroups.[12] The graphs do not show any obvious positive relationship; if anything, the reverse appears to be the case. This apparent conflict with the theoretical predictions is not completely surprising, however, because other factors are not held constant. In particular, comparability is compromized if different criteria are used to group the sample observations. This is easily detected in the results for China, Indonesia, and the Philippines. In the relevant studies, the samples were divided into rural-urban areas as well as into regions (see Table 1A). In moving from the rural-urban division to the regional breakdown, the number of subgroups increases from 2 to 26 or 27 for Indonesia (2 to 13 for the Philippines, 2 to 3 or 26 for China); but the between-group-component falls in most cases. This clearly indicates the dominant impact of differences in living standards between rural and urban residents, which more than offsets the contribution of the number of subgroups.

To examine properly the positive relationship between the number of subgroups and the size of the between-group component requires progressive aggregation of subgroups. This has been done by Elbers et al. (2005) for Ecuador, Madagascar and Mozambique[13] and by Cheng (1996) for China. Results from Elbers et al. indicate small increases in the between-group component, even if the number of groups increases dramatically. Using consumption data, Cheng (1996) reports a rise in the between-group component from 28% to 37% when the number of groups increases from 3 to 26. Using data on the gross value of industrial and agricultural outputs (GVIAO), the corresponding change is from 39% to 51%.

12 For convenience, observations with more than 30 subgroups have been excluded from the graphs. The graphs also exclude the results for the rural-urban division in China because they are not based on household level data.

13 These observations were excluded from Figures 1–4.

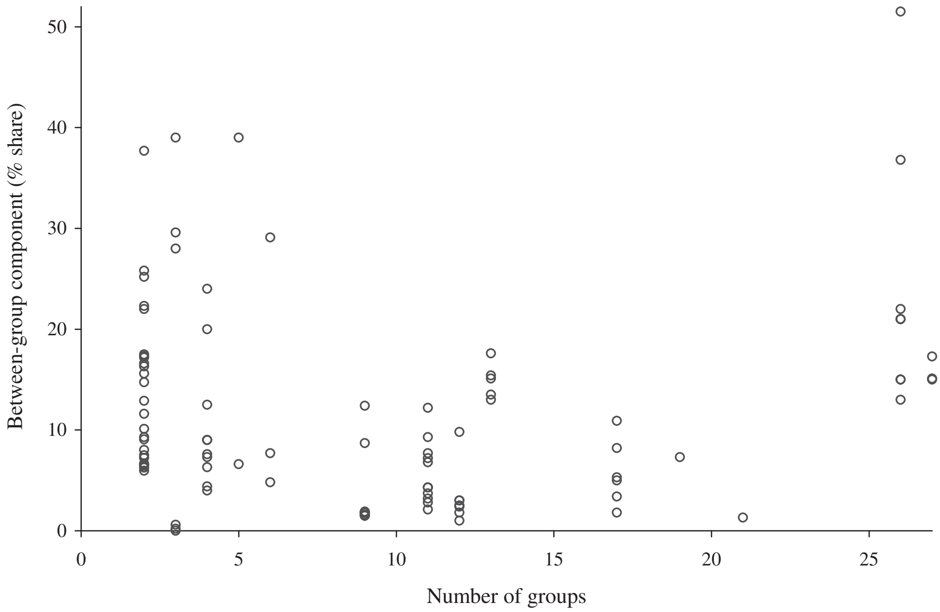

Figure 1. Between-group component and number of groups: E_0.

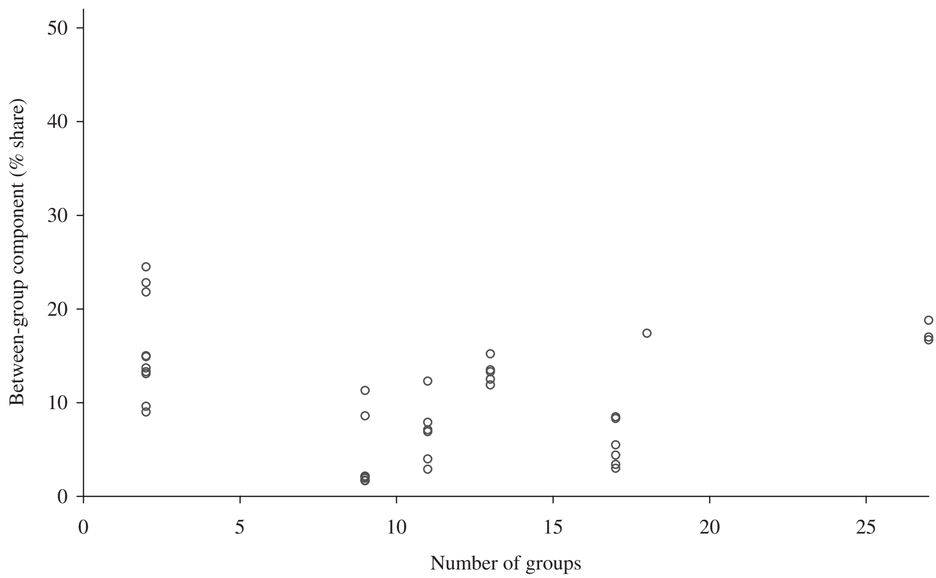

Figure 2. Between-group component and number of groups: E_1.

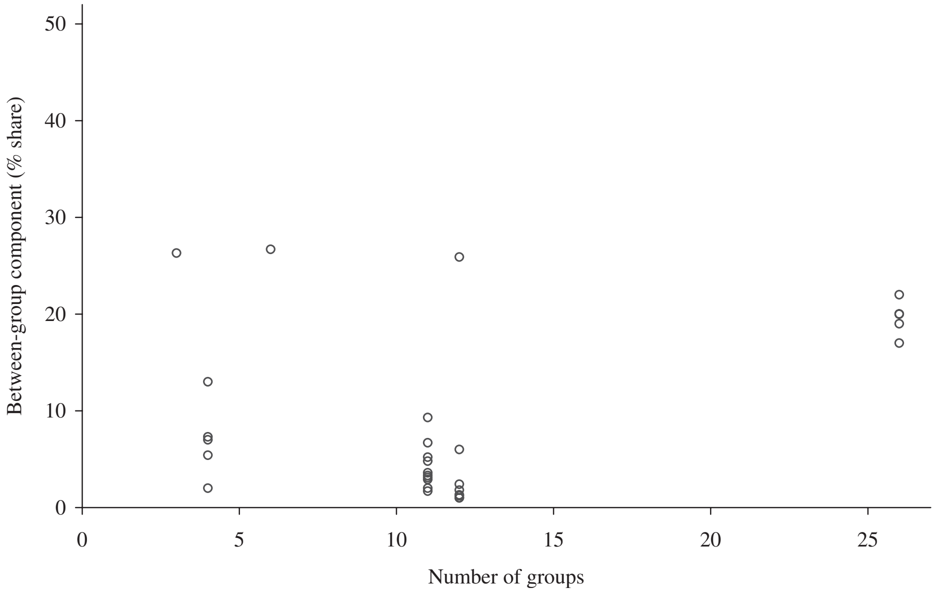

Figure 3. Between-group component and number of groups: E_2.

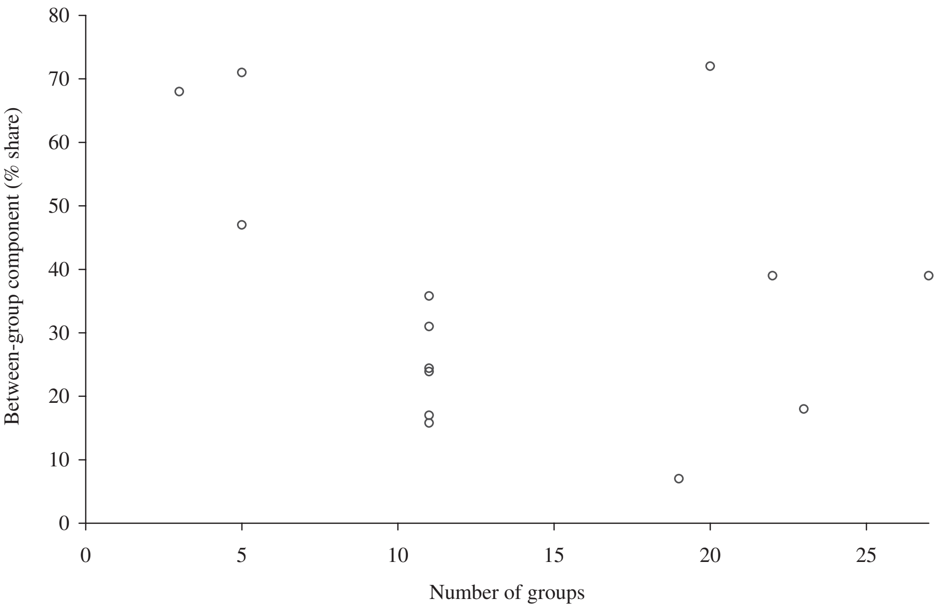

Figure 4. Between-group component and number of groups: Gini.

Table 3. Estimation results

Variable	Coefficient estimate	T-ratio	Level of significance	Box-Cox elasticity
No. of groups	0.13	3.65	0.00	0.17
Dummy for				
E_1	–0.31	–1.12	0.26	–0.03
E_2	–0.39	–0.50	0.62	–0.00
Gini	3.51	8.20	0.00	0.13
Dummy for				
Earnings	–1.50	–5.09	0.00	–0.13
Urban-rural	1.73	4.74	0.00	0.11
CONSTANT	2.63	12.01	0.00	1.24

$R^2 = 0.39$. Sample size = 185.

To explore further the relationship between the number of subgroups and the share of the between-group component, we employed the regression model:

$$S = f(m, D)$$

where S is the share of the between-group component, m is the number of subgroups, and D refers to a set of dummy variables to control for different income concepts, different inequality measures, and the rural-urban division versus other spatial partitions. To allow for possible non-linearities, the model specification takes the Box-Cox form. The standard linear model always produced an insignificant parameter for the core variable m, but a simple χ^2 test suggested preference for the Box-Cox model.

Estimation results for the Box-Cox model are reported in Table 3, with E_0 as the reference index. The results indicate that (a) the size of the between-group component is positively related to the number of subgroups at any conventional significance level; (b) increasing the number of groups by one leads on average to an increase of 0.07 in the percentage share of the between-group component; (c) earnings data tends to yield a smaller between-group component, (d) the rural-urban partition gives a larger between-group component; and (e) the Gini coefficient produces larger shares for the between-group component compared to other indices.

In summary, this section has reported empirical results that will aid future research, both empirical and theoretical. Many questions have been raised that require further attention. Issues of special interest and importance include the appropriate measure of spatial proximity; the relationship between the number of groups and the magnitude of the between-group component; the implications of using measures of representative incomes other than subgroup means; and the choice of inequality index.

5. Concluding remarks

This paper has ranged over a number of theoretical and empirical issues linked to decomposition analysis in a spatial context. Various other issues have yet to be explored. For example, with the exception of Kanbur and Zhang (1999), there is little in the way of empirical literature on the time profile of the within- or between-group component. Availability of data is an obstacle here. Nevertheless, a time profile would enrich the empirical literature by adding a dynamic dimension to the study of spatial inequality decomposition.

More also could be done along the lines suggested by Salas (2002) to link inequality decomposition to the recent literature on growth and convergence. In particular, examining the time trend of the between-group component may offer a better method of studying convergence than the conventional Beta or sigma convergence techniques.

Another set of issues requiring attention concern the underlying factors which ultimately contribute to spatial inequality, factors like economic geography (climate, natural resources), policy regimes, market orientation, and related socio-economic variables. Whether or not spatial differences persist or whither away over time is perhaps influenced most by the freedom to migrate, both within countries and internationally. The extent to which labour migration can help reduce regional disparities is an important question with obvious policy significance.

Acknowledgements

This paper was written for the UNU-WIDER project on *Spatial Disparities in Human Development* and was first presented at the UNU-WIDER conference on *Inequality*, *Poverty and Human Well-being* in May 2003 in Helsinki. We thank Ravi Kanbur, Tony Venables, and an anonymous referee for their valuable comments and suggestions.

References

Akita, T., Lukman, R., Rizal, A., Yamada, Y. (1999) Inequality in the distribution of household expenditures in Indonesia: a Theil decomposition analysis. *Developing Economies*, 37: 197–221.

Balisacan, A., Fuwa, N. (2005) Is spatial inequality increasing in the Philippines. In R. Kanbur, A. J. Venables, and G. H. Wan (eds) *Spatial Disparities in Human Development: Evidence from Asia*. Tokyo: United Nations University Press.

Bhattacharya, N., Mahalanobis, B. (1967) Regional disparities in household consumption in India. *Journal of the American Statistical Association*, 62: 143–161.

Blackorby, C., Donaldson, D., Auersperg, M. (1981) A new procedure for the measurement of inequality within and among population subgroups. *Canadian Journal of Economics*, 14: 665–685.

Bourguignon, F. (1979) Decomposable income inequality measures. *Econometrica*, 47: 901–920.

Bourguignon, F., Morrison, C. (2002) Inequality among world citizens: 1820–1992. *American Economic Review*, 92: 727–744.

Cheng, Y. (1996) A decomposition analysis of income inequality of Chinese rural households. *China Economic Review*, 7: 155–167.

Cowell, F. A., Jenkins, S. P. (1995) How much inequality can we explain? A methodology and an application to the United States. *Economic Journal*, 105: 421–430.

Cowell, F.A., Kuga, K. (1981) Additivity and the entropy concept: an axiomatic approach to inequality measurement. *Journal of Economic Theory*, 25: 131–143.

Davies, J. B., Shorrocks, A. F. (1989) Optimal grouping of income and wealth data. *Journal of Econometrics*, 42: 97–108.

Dickey, H. (2001) Regional earnings inequality in Great Britain: a decomposition analysis. *Regional Studies*, 35: 605–612.

Elbers, C., Lanjouw, P., Mistiase, J., Ozler, B., Simler, K. (2005) Are neighbours equal? Estimating local inequality in three developing countries. In R. Kanbur and A. J. Venables (eds) *Spatial Inequality and Development*. Oxford: Oxford University Press.

Ernst, C., Gerfin, M., Leu, R. E. (2000) Inequality trends in the Swiss income distribution. *Swiss Journal of Economics and Statistics*, 136: 289–305.

Fischer S. (2003) Globalization and its challenges. *American Economic Review*, 93: 1–30.

Foster, J. E., Shneyerov, A. A. (2000) Path independent inequality measures. *Journal of Economic Theory*, 91: 199–222.

Gray, D., Mills, J.A., Zandvakili, S. (2003) New perspectives on Canadian national and provincial income inequality. Paper presented at the UNU-WIDER Conference on Inequality, Poverty and Human Well-Being, Helsinki, 30–31 May.

Kanbur, R. (2003) The policy significance of inequality decompositions. Unpublished paper, Cornell University, August.

Kanbur, R., Zhang, X. (1999) Which regional inequality? The evolution of rural-urban and inland-coastal inequality in China from 1983 to 1995. *Journal of Comparative Economics*, 27: 686–701.

Kanbur, R., Zhang, X. (2005) Fifty years of regional inequality in China: a journey through central planning, reform and openness. In R. Kanbur, A. J. Venables, and G. H. Wan (eds) *Spatial Disparities in Human Development: Evidence from Asia*. Tokyo: United Nations University Press.

Kremer, M., Maskin, E. (2003) Globalization and inequality. Unpublished manuscript, Harvard University.

Lambert, P.J., Aronson, J. R. (1993) Inequality decomposition analysis and the Gini coefficient revisited. *Economic Journal*, 103: 1221–1227.

Lee, J. (2000) Changes in the source of China's regional inequality. *China Economic Review*, 11: 232–245.

Loikkanen, H. A., Riihelä, M., Sullström, R. (2002) Regional income convergence and inequality in boom and bust: results from micro data in Finland 1971–1998, and especially during the 1990s. Paper prepared for the meeting of the ESRC Urban and Regional Economics Seminar Group, 1–3 July, University of Central Lancashire, UK.

Mangahas, M. (1975) Income inequality in the Philippines: a decomposition analysis. World Employment Programme Working Paper No. 12, Geneva, ILO.

Marshall, A. W., Olkin, I. (1979): *Inequalities: Theory of Majorization and its Applications*. New York: Academic Press.

Milanovic, B. (2002) True world income distribution, 1988 and 1993: first calculation based on household surveys alone. *Economic Journal*, 112: 51–92.

Mishra, P., Parikh, A. (1992) Household consumer expenditure inequalities in India: a decomposition analysis. *Review of Income and Wealth*, 38: 225–236.

Pyatt, G.C. (1976) On the interpretation and disaggregation of Gini coefficients. *Economic Journal*, 86: 243–255.

Ram, R. (1992) Interstate income inequality in the United States: measurement, modelling and some characteristics. *Review of Income and Wealth*, 38: 39–48.

Rao, V.M. (1969) Two decompositions of the concentration ratio. *Journal of the Royal Statistical Society, Series A*, 132: 418–425.

Sala-i-Martin, X. (2002) The disturbing 'rise' of global income inequality. Working Paper 8904, NBER, Cambridge, MA, April.

Salas, R. (2002) Multilevel interterritorial convergence and additive multidimensional inequality decomposition. *Social Choice and Welfare*, 19: 207–218.

Sastry, D. V. S., Kelkar, U. R. (1994) Note on the decomposition of Gini inequality. *Review of Economics and Statistics*, 76: 584–586.

Shorrocks, A. F. (1980) The class of additively decomposable inequality measures. *Econometrica*, 48: 613–625.

Shorrocks, A. F. (1984) Inequality decomposition by population subgroups. *Econometrica*, 52: 1369–1385.

Shorrocks, A. F. (1988) Aggregation issues in inequality measurement. In W. Eichhorn (ed.) *Measurement in Economics*. Heidelberg: Physica-Verlag, 429–451.

Silber, J. (1989) Factor components, population subgroups and the computation of the Gini index of inequality. *Review of Economics and Statistics*, 71: 107–115.

Silber, J. (ed.) (1999) *Handbook of Income Inequality Measurement*. Boston, Dordrecht, and London: Kluwer Academic Press.

Soltow, L. (1960) The distribution of income related to changes in the distribution of education, age, and occupation. *Review of Economics and Statistics,* 42: 450–453.

Tadjoeddin, M., Suharyo, W., Mishra, S. (2003) Aspiration to inequality: regional disparity and centre-regional conflicts in Indonesia. Paper presented at the WIDER Conference on Inequality in Asia, Tokyo, 28–29 March.

Theil, H. (1967) *Economics and Information Theory*. Amsterdam: North Holland Publishing Co.

Theil, H. (1972) *Statistical Decomposition Analysis*. Amsterdam: North Holland Publishing Co.

Tsakloglou, P. (1993) Aspects of inequality in Greece: measurement, decomposition and intertemporal change: 1974, 1982. *Journal of Development Economics*, 40: 53–74.

Vanderpuye-Orgle, J. (2002) Spatial Inequality and Polarization in Ghana, 1987–99. Paper presented at the WIDER Conference on Spatial Inequality in Africa, Oxford University, 21–22 September.

Wan, G. H. (2001) Changes in regional inequality in rural China: decomposing the Gini index by income sources. *Australian Journal of Agricultural and Resource Economics*, 45: 361–382.

Yemtsov, R. (2005) Quo vadis? Inequality and poverty dynamics across Russian regions. In R. Kanbur and A. J. Venables (eds) *Spatial Inequality and Development*. Oxford: Oxford University Press.

Yitzhaki, S. (1994) Economic distance and overlapping of distributions. *Journal of Econometrics,* 61: 147–159.

Yitzhaki, S., Lerman, R. I. (1991) Income stratification and income inequality. *Review of Income and Wealth*, 37: 313–329.

Zagier, D. (1983) Inequalities for the Gini coefficient of composite populations. *Journal of Mathematical Economics*, 12: 103–118.