SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Geographic Decomposition of Inequality in Health and Wealth: Evidence from Cambodia

Tomoki Fujii

December 2007

Paper No. 24-2007

Geographic Decomposition of Inequality in Health and Wealth: Evidence from Cambodia

Tomoki Fujii*

School of Economics
Singapore Management University

Email: tfujii@smu.edu.sg

1. Introduction

The small-area estimation developed by Elbers, Lanjouw and Lanjouw (2002, 2003), in which a census and a survey are combined to produce the estimates of welfare measures for small geographic areas, has become a standard tool for poverty analysis in developing countries. The small-area estimates are typically plotted on a map, which are commonly called a poverty map. Poverty maps proved useful for policy analysis and formulation, and have become increasingly popular among policy-makers and researchers. In Cambodia, poverty maps have been used by various international organizations, ministries and non-governmental organizations for analyzing the poverty situations for their operation areas, selecting the potential locations for their projects and programs, and educating students in classrooms (Fujii, 2007).

Besides creating poverty maps, the small-area estimation has been used for a wide array of purposes. For example, it has been used to analyze geographic targeting (Elbers et al., 2007 and Fujii, 2008), consumption inequality (Elbers et al., 2004), local inequality and crime (Demombynes and Özler, 2005), and impacts of trade liberalization (Fujii and Roland-Holst, 2008). In this paper, we offer another new

* Earlier stages of this research were supported by the World Bank Research Group, the Government of Japan under the Millennium PHRD grant, and the SMU Research Grant (05-C208-SMU-030). I thank Alain de Janvry, Chris Elbers, Jean Olson Lanjouw, Peter Lanjouw, Livia Montana, Mahadevan Ramachandran, Martin Ravallion, H.E. Kim Saysamalen, Elisabeth Sadoulet, Boreak Sik and H.E. San Sythan for comments and help I received at various stages of my research.

application of the small-area estimation; We use the small-area estimation to look at whether poverty is more spatially unequally distributed than child undernutrition. More precisely, we decompose inequality of consumption and child nutrition status into the within-group and between-group inequalities at various levels of spatial aggregation, and compare the decomposition results.

While it is widely known that the health and wealth are positively correlated, it is not clear whether the spatial inequality in health and wealth necessarily exhibits a similar pattern. The significance of this point can be easily understood in a simple example. Suppose that the wealthy people in a country only live in the north and poor people only in the south, and suppose further that mosquitoes carrying malaria parasites exist uniformly across the country. Since wealthy people have better knowledge to cope with malaria, and resources to prevent the infection (such as mosquito repellants and mosquito nets), they are less likely to get infection than poor people. However, since there is no perfect preventive measure, the incidence of malaria would be less unequally distributed than poverty across the country. This example is extreme, of course. But it is of interest to see how different the spatial patterns of inequality in poverty and undernutrition are.

The knowledge of spatial inequality in consumption and health is valuable for geographic targeting, because the spatial inequality prescribes the potential gains from geographic targeting. In the example given above, the resources for anti-poverty programs can be fully efficiently used if they are delivered to the south because everyone is poor and thus the resources all go to poor people. However, if we deliver all the resources (say, malaria tablets) to the south, the outcome may not be fully efficient. We would be giving the tablets to some in the south who are less vulnerable to malaria while not giving to others in the north who are more vulnerable to malaria. If geographic information is the only information available to the policy-maker, geographic targeting is still useful (and efficient given the available information), but the extent to which one may gain from geographic targeting is determined by the pre-existing spatial inequality.

This paper is organized as follows. In the next section, we review the related literature. In Section 3, we shall discuss the small-area estimation methods for consumption and child nutrition status. We shall develop a unified framework for the standard small-area estimation developed by Elbers, Lanjouw and Lanjouw (2002, 2003) and its extension for the estimation of the prevalence of malnutrition by Fujii (2005). In Section 4, we shall discuss the method of inequality decomposition. In Section 5, we shall discuss the data we use. We then present the decomposition results in Section 6. Section 7 provides concludes.

2. Literature Review

There is a large body of the literature on the (positive) relationship between health and wealth. There are three types of explanations: (i) health causes wealth, (ii) wealth causes health and (iii) there is a tertiary factor that is correlated with both health and wealth. Let us briefly look at each type of explanation.

Prichett and Summers (1996) confirm that wealth causes health, using instrumental variables to isolate the reverse causation or incidental association. The long-run income elasticity of infant and child mortality in developing countries are estimated at between –0.2 and –0.4. The results are intuitive as one would expect wealthier populations tend to enjoy better health care and safer food and water.

Bloom and Canning (2000), on the other hand, points out the several mechanisms that could account for this relation. First, healthier populations tend to have higher productivity. Second, healthier population have stronger incentives to save and invest in human and physical capital. Third, demographic dividends are also a possible mechanism. That is, when the decline in infant mortality initiates the decline in fertility, the proportion of working age increases. As a result, the income per capita in the country also increases.

The positive association between health and wealth can also be explained by the existence of the third factor that is correlated with both. For example, better educated population may well have higher income as well as better knowledge of health that help them stay healthy. Indeed, Elo and Preston (1996) show that the percentage reduction of mortality associated with one year increase in education for persons aged 35-54 is around 7-9% for males and 2-8% for females in various rich countries. Besides education, health care and behavior may also be the third factor that is correlated both health and wealth.

The three types of relationships mentioned above are not mutually exclusive and may be at work simultaneously. They have, however, very different policy implications. For example, suppose that education is the important factor for determining health. Then, redistribution of income will not help reduce health inequality. Reducing the health inequality requires educating the poor.

Another important and controversial point in the relevant literature is whether the absolute or relative income matters for the health outcome. Some researchers, including Wilkinson (1992, 1996, 1997) and Kawachi et al. (1997), have noted that poor health in developed countries is strongly related to income inequality. Kawachi et al. argue that income inequality is related to reduction in social cohesion, which in turn is associated with poor health (as measured by mortality). Marmot and Wilkinson (2001) point out that the psychosocial effects of relative deprivation, including control, anxiety, insecurity, depression and social afflictions, are negatively correlated with health. These studies would suggest that the inequality in health outcome may be caused by income inequality. This, in turn, provides the rationale for redistribution of income as a means to reduce health inequality.

Economists have been skeptical about this argument. Wagstaff and van Doorslaer (2000) point out that the empirical evidence in earlier studies based on aggregate data is often insufficient to discriminate one hypothesis from another. Based on the analysis of individual-level data in the United States, they found strong support for the absolute-income hypothesis (health is determined by the level of income) and no or little support for the relative-income hypothesis (health is determined by the deviation in income from the mean) and income-inequality hypothesis (health is determined by the level of income and the income inequality). Deaton (2003) provides a more comprehensive critique of the relative-income and income-inequality hypotheses.

This study also investigates into the relationship between health and wealth, but we look at a different aspect, namely the spatial inequality in health and wealth. Only a few studies carried that address the spatial inequality. For example, Pradhan, Sahn and Younger (2003) decomposed the world health inequality into within-country inequality and between-country inequality, where the latter accounts only for 31% of the total world inequality. This contrasts with similar exercises for income inequality, where the various empirical studies agree that between-country inequality accounts for the total inequality more than the within-country inequality (Firebough, 2000). Using a cross-country regression analysis, van Doorslaer et al. (1997) directly relate a health inequality index to an income inequality index, and find the estimated coefficient is positive and significant.

While these cross-country comparisons are interesting, the inequality decomposition within a country is arguably more important because it is more closely related to the geographic targeting of health programs, anti-poverty programs or combination of both. In this sense, our study is closely related to Wagstaff (2005), in which he asks, "How far are income-related inequalities in the health sector due to gaps *between* poor and less poor areas, rather than due to differences between poor and less poor people *within* areas." He sets out a method of answering this question and apply to the geographically decomposition of the concentration index of health subsidies in Vietnam and insurance coverage in rural China.

This study is different from Wagstaff (2005) in four aspects. First, we use various decomposable measures that are not considered in his study, as we shall discuss in Section 4. Second, instead of "health access" type indicators, we use the individual health outcome. Third, by applying the small-area estimation, we were able to decompose at various levels of decomposition covering virtually all over the country. In contrast, the number of geographic areas in Wagstaff (2005) was relatively small (58 provinces in Vietnam and 225 villages in China) and covers only a fraction of geographic areas covered in the survey.[i] Fourth, Wagstaff (2005) uses the concentration index, which can only be decomposed into the between-area component, within-area components and the residual. The residual can be interpreted, but the interpretation is not straightforward, a feature that makes the concentration index unattractive. Hence, we only choose the measures that allow us to decompose neatly.

3. Small-Area Estimation

Let us now turn to the two small-area estimation methods we used for consumption and the status of the child nutrition. The former is based on Elbers, Lanjouw and Lanjouw (2002, 2003) and the latter on Fujii (2005). These estimation methods share the same basic procedures. That is, model parameters are first estimated with survey data. Estimated parameters are then used to impute the welfare indicator for each unit record in the census. Finally, the imputed welfare indicator is aggregated so that the standard errors are at an acceptable level. We carry out the imputation repeatedly by Monte-Carlo simulation in order to take into account of the various sources of error.

Thus, the fundamental idea of the small-area estimation is straightforward. However, the details of these two methods of small-area estimation differ substantially. We shall summarize the two methods in a unified framework, and highlight their differences.

Suppose that there are K welfare indicators of interest. For consumption-based small-area estimation, we use the *per capita* logarithmic consumption as a measure of welfare as with the standard practice of “poverty mapping.” Obviously, we have $K=1$ in this case.

For the “nutrition mapping”, or the small-area estimation for the status of the child nutrition, we have $K=2$. We use measures based on the height-for-age and weight-for-age Z-scores, which represent the number of standard deviations between an individual’s value of these anthropometric indicators and the median for the reference group of health population of the same sex and age. Because height changes more slowly than weight, the former is considered to reflect the nutrition status of children in the longer run than the latter.[ii]

We convert these Z-scores into the corresponding height and weight for the 24-months old year girls as with Pradhan, Sahn and Younger (2003), which we shall call standardized height and weight.[iii] Since we adopt affine transformation for this conversion, the regression results remain unaffected and there are two additional merits. First, for practically possible values of the Z-scores, the standardized height and weight take a positive value so that we can employ conventionally used inequality measures. Second, we can compare our results with their results so that we can see whether the within-country inequality, which is a major source of the world health inequality, comes from the inequality within areas or between areas in the same country.

Now, let l be the identifier at the unit-record level. For example, the consumption measure is typically recorded at the household level, and thus l is the household identifier. For the child nutrition status, l is the child identifier. Let us further denote the k-th welfare indicator for the observation unit l by $y_l^{(k)}$, where $k \in \{1, \ldots, K\}$. We assume that the $y_l^{(k)}$ is related to a $d^{(k)}$-vector of characteristics $\mathbf{x}_l^{(k)}$ in the following manner:

$$\begin{aligned} y_l^{(k)} &= E[y_l^{(k)} \mid \mathbf{x}_l^{(k)}] + u_l^{(k)} \\ &= \mathbf{x}_l^{(k)} \cdot \beta^{(k)} + u_l^{(k)}, \end{aligned}$$

where $\cdot$ is the inner-product operator, $\beta^{(k)}$ a $d^{(k)}$-vector of parameters, and $u_l^{(k)}$ the residual term that is not explained by $\mathbf{x}_l^{(k)}$.

The structure of the error term is what distinguishes between the small-area estimation for consumption and that for the status of child nutrition. In the standard poverty mapping exercise, we can omit (k) as $K = 1$, and write the structure of u_l as follows:

$$u_l = \eta_{c(l)} + \varepsilon_l,$$

where $c(l)$ is a function that maps each household to the cluster (which is usually a community or village) it belongs to. In other words, the error term is decomposed into the cluster-specific random component $\eta_{c(l)}$ and household-specific random component ε_l. This structure would be reasonable because we often have many cluster-level variables that are not recorded in the survey but correlated with the consumption. Variables like the distance to the market and average land quality of the cluster are a typical example.

The error component for the nutrition mapping is decomposed as follows:

$$u_l^{(k)} = \eta_{c(l)}^{(k)} + \varepsilon_{h(l)}^{(k)} + \delta_l^{(k)},$$

where $h(l)$ and $c(l)$ are functions that map from each child to the household and cluster the child belongs to. Two remarks are in order. First, note that we must include (k) in the notation because $K > 1$ for nutrition mapping in general. Second, we have additional term $\delta_l^{(k)}$ in the expression, which denotes the individual-specific random component. This additional term is important because every child has different levels of nutritional status even if the children are from the same household.

In the poverty mapping, $\eta_{c(l)}$ and ε_l are assumed to be uncorrelated, and $\eta_{c(l)}$ is independently and identically distributed. The household-specific error component ε_l is allowed to be heteroskedastic. In the nutrition mapping, $\eta_{c(l)}^{(k)}$, $\varepsilon_{h(l)}^{(k)}$ and $\delta_l^{(k)}$ are assumed to be uncorrelated piecewise for any given combination of k and l. Further, $\eta_{c(l)}^{(k)}$ and $\delta_l^{(k)}$ are respectively assumed to be independently and identically distributed for any given k. As with the poverty mapping, the household-specific error component $\varepsilon_l^{(k)}$ is allowed to be heteroskedastic. Also, its is assumed that $\eta_{c(l)}^{(k)}$ and $\varepsilon_{h(l)}^{(k)}$ are uncorrelated across k. That is $E[\eta_{c(l)}^{(k)}\eta_{c(l)}^{(k')}] = 0$ and $E[\varepsilon_{h(l)}^{(k)}\varepsilon_{h(l)}^{(k')}] = 0$ for any $k \neq k'$ and any l. However, $\delta_l^{(k)}$ is allowed to be correlated across k.

While these assumptions are arguably somewhat arbitrary, they are designed to capture the important aspects of the error components while keeping the model parsimonious and estimable given the practical constraints imposed by the data availability. For example, in principle, cluster-specific random components may be heteroskedastic in the poverty mapping. However, since the number of clusters is typically small and it is difficult to distinguish between the heteroskedasticity at the household level and heteroskedasticity at the cluster level, we choose to allow for the heteroskedasticity only at the household level. This choice is also justifiable on the ground that the cluster specific effect is usually small relative to the household-specific random component.

In the nutrition mapping, heteroskedasticity is allowed for at the household level but not at the individual level. This is because there are only a limited number of variables observed at the household level. One could argue that the correlation across welfare indicators may occur at the cluster and household levels. However, the effect is most important at the individual because the individual-level component account for a majority of variations of the random component $u_l^{(k)}$. This would not be difficult to imagine. For example, the various indicators of the status of child nutrition are simultaneously affected by how the child is taken care of and how well the child is fed relative to other children in the household, both of which are important determinants of the nutrition status of children but not observed in the survey.

Let us now turn to the implementation of the estimation. We first specify the variables to be included in $\mathbf{x}_l^{(k)}$. They must be shared by the census and the survey. We then run the ordinary least square (OLS) regression to obtain the OLS estimate of the parameter $\hat{\beta}_{OLS}^{(k)}$. We let the regression residual be $\hat{u}_l^{(k)} = y_l^{(k)} - \mathbf{x}_l^{(k)} \cdot \beta_{OLS}^{(k)}$ for each k. The OLS regression must be weighted by the population expansion factor. The estimations described in the subsequent paragraphs must also be appropriately weighted in practice. However, we shall assume hereafter that the survey data has a unit weight for the sake of the simplicity of the presentation.

We then estimate the distribution of the error terms. In the standard poverty mapping, we approximate the distribution of η_c and ε_l by their empirical analogues in the following manner:

$$\hat{\eta}_c \equiv \frac{1}{\#\{L_c\}} \sum_{l \in L_c} \hat{u}_l \text{ and } \hat{\varepsilon}_l \equiv \hat{u}_l - \hat{\eta}_c,$$

where $L_c \equiv \{l \mid c(l) = c\}$ is the set of households that belong to cluster c (with a little abuse of notation) and $\#\{\bullet\}$ the counting measure. In other words, we take the cluster average of the residuals as the estimate of the cluster-specific random component, and regard as the household-specific random component the difference between the regression residual and the cluster-specific random component. We can use this approximation because each cluster typically contains sufficient number of households. We then estimate the regression parameters for the heteroskedastic model for the household-specific random component.

$$\ln \frac{\hat{\varepsilon}_l^2}{A^* - \hat{\varepsilon}_l^2} = \mathbf{z}_l \cdot \alpha + \gamma_l$$

where the maximum bound is conventionally set at $A^* \equiv 1.05 \max_l \{\hat{\varepsilon}_l^2\}$. $\mathbf{z}_l$ is the regressors for the heteroskedastic regression, α the model parameter and γ the residual for this model.

Nutrition mapping requires somewhat more complicated treatment because we cannot use the empirical analogue of $\delta_l^{(k)}$. That is, we cannot approximate $\delta_l^{(k)}$ by taking the difference between $\hat{u}_l^{(k)}$ and its household-level mean, because the number of children in a household is typically too small to justify such an operation. Therefore, we need to correct for the finite sample. For example, we can estimate the variance of the individual-level random components $\left(\sigma_\varepsilon^{(k)}\right)^2 = E\left[\left(\delta_l^{(k)}\right)^2\right]$ in the following manner:

$$\left(\hat{\sigma}_\varepsilon^{(k)}\right)^2 = \sum_{c \in C} \sum_{h \in \tilde{H}_c} \frac{1}{\#\{\tilde{H}_c\}} \sum_{l \in L_h} \left(\#\{L_h\} - 1\right)^{-1} \cdot \left(\hat{u}_l^{(k)} - \overline{u}_{h(l)}^{(k)}\right)^2$$

where C is the set of all clusters, $L_h \equiv \{l \mid h(l) = h\}$ is the set of individuals in household h, $\tilde{H}_c \equiv \{h \mid c(l^{-1}(h)) = c, \#\{L_h\} > 1\}$ is the set of households in cluster c that have more than one children, and $\overline{u}_h^{(k)} \equiv (\#\{L_h\})^{-1} \sum_{l \in L_h} \hat{u}_l^{(k)}$ is the household-level average of the regression residual. Similarly, we estimate the covariance between the individual-specific random components for two different indicators $\sigma_\varepsilon^{(k,k')} = E\left[\delta_l^{(k)} \delta_l^{(k')}\right]$ as follows:

$$\hat{\sigma}_\varepsilon^{(k,k')} = \sum_{c \in C} \sum_{h \in \tilde{H}_c} \frac{1}{\#\{\tilde{H}_c\}} \sum_{l \in L_h} \left(\#\{L_h\} - 1\right)^{-1} \cdot \left(\hat{u}_l^{(k)} - \overline{u}_{h(l)}^{(k)}\right) \cdot \left(\hat{u}_l^{(k')} - \overline{u}_{h(l)}^{(k')}\right).$$

While the finite-sample correction for the cluster-specific effect is less straightforward, Fujii (2005) showed that the variance of the cluster-specific random component $\left(\sigma_\eta^{(k)}\right)^2 = E\left[\left(\eta_l^{(k)}\right)^2\right]$ can be estimated as follows:

$$\left(\hat{\sigma}_{\eta}^{(k)}\right)^{2}=\left(\sum_{c\in C}(\#\{H_{c}\}-1)\right)^{-1}\cdot\left(\sum_{c\in C}\#\{H_{c}\}\left(\overline{u}_{c}^{(k)}\right)^{2}-\sum_{c\in C}\frac{1}{\#\{H_{c}\}}\sum_{h\in H_{c}}\left(\overline{u}_{h}^{(k)}\right)^{2}\right),$$

where $\overline{u}_{c}^{(k)}$ is the cluster level average of the regression residual. The estimation of the heteroskedastic model in the nutrition mapping is carried out in a manner similar to that for the poverty mapping. However, because it is difficult to separate the individual-specific and household-specific random components, we work with the sum of the two, noting that the heteroskedasticity of this sum comes only from the heteroskedasticity in the household-specific random component.

Once we specify the error structure, we can find the generalized least square estimate $\hat{\beta}_{GLS}$ of the model parameters as well as the estimate $\widehat{VC}(\hat{\beta}_{GLS})$ of its associated variance-covariance matrix. We also use the empirical distribution of $\hat{u}_{l}^{(k)}$ to estimate the population distribution of $u_{l}^{(k)}$. Notice that the model parameters are jointly estimated across welfare indicators in this procedure.

Once all the distributional and model parameters of interest are estimated, we can impute the welfare measure to each record in the census. The prediction error at the unit-record level are subject to the model error (error associated with the estimation of β) and the idiosyncratic error (error due to the disturbance term u_{l}). The latter tend to decrease by aggregation as the idiosyncratic errors tend to cancel out with each other, but the former does not systematically change by aggregation. We estimate the standard error associated with aggregate welfare measures by repeatedly simulating these two sources of errors .

Suppose we carry out R rounds of the simulation and let $r\in\{1,\ldots,R\}$. The estimate $\tilde{y}_{l,(r)}^{(k)}$ of the welfare indicator for record l and indicator k in round r is given by:

$$\tilde{y}_{l,(r)}^{(k)}=\mathbf{x}_{l}^{(k)}\cdot\tilde{\beta}_{(r)}^{(k)}+\tilde{u}_{l,(r)}^{(k)},$$

$\tilde{\beta}_{(r)}$ is randomly drawn from a normal distribution with mean $\hat{\beta}_{GLS}$ and variance-covariance matrix $\widehat{VC}(\hat{\beta}_{GLS})$. The idiosyncratic error component $\tilde{u}_{l,(r)}^{(k)}$ is drawn to retain the original error structure. For example, in the nutrition mapping,

$$\tilde{u}_{l}^{(k)}=\tilde{\eta}_{c(l)}^{(k)}\tilde{\sigma}_{\eta,(r)}^{(k)}+\tilde{\varepsilon}_{h(l)}^{(k)}\tilde{\sigma}_{\varepsilon,(r)}^{(k)}+\tilde{\delta}_{l}^{(k)}\tilde{\sigma}_{\delta,(r)}^{(k)},$$

$\tilde{\eta}^{(k)}_{c(l)}$, $\tilde{\varepsilon}^{(k)}_{h(l)}$ and $\tilde{\delta}^{(k)}_{l}$ are random draws from the empirical distribution standardized to have a mean zero and a unit standard deviation. The r-th round estimates of $\tilde{\sigma}^{(k)}_{\eta,(r)}$, $\tilde{\sigma}^{(k)}_{\varepsilon,(r)}$ and $\tilde{\sigma}^{(k)}_{\delta,(r)}$ are $\hat{\sigma}^{(k)}_{\eta}$, $\hat{\sigma}^{(k)}_{\varepsilon}$ and $\hat{\sigma}^{(k)}_{\delta}$ calculated with a bootstrapped sample.

Once we have the imputed welfare measure $\tilde{y}^{(k)}_{l,(r)}$, we can aggregate at any level of aggregation. Let the estimate of the welfare indicator of interest for the r-th round aggregated for a certain geographic group G be $\tilde{\mathbf{W}}^{(k)}_{(r),G} = \mathbf{W}(\{\tilde{y}^{(k)}_{l,(r)}\}_{l\in G})$.[iv] We take the average and standard deviation of $\tilde{\mathbf{W}}^{(k)}_{(r),G}$ over r to arrive at the point estimate and its associated standard error of the estimate of the welfare measure $\tilde{\mathbf{W}}^{(k)}_{G}$ for group G.

For example, the point estimate of the FGT measure with parameter α (Foster, Greer and Thorbecke, 1984) for the k-th indicator is as follows:

$$P^{(k)}_{G}(\alpha) = \frac{1}{R\cdot \#\{G\}} \sum_{r=1}^{R} \sum_{l\in G} (1-\tilde{y}^{(k)}_{l,(r)} z^{-1})^{\alpha} \cdot \mathrm{Ind}(z > \tilde{y}^{(k)}_{l,(r)}),$$

where $\mathrm{Ind}(\bullet)$ is the indicator function and z is the cut-off level. For poverty mapping, z corresponds to the poverty line. For the nutrition mapping, we take the standardized height and weigh corresponding to the conventional cut-off point of the Z-score of negative two, below which the child is deemed undernourished (stunted for the case of height and underweight for the case weight respectively). In the poverty mapping, $P(0)$, or the poverty rate, is often plotted on a map to arrive at a poverty map. Similarly, we can plot the prevalence $P(0)$ for the standardized height and weight, or the prevalence of stunting and underweight on a map to create nutrition maps.

While the FGT measure is the most frequently employed measure of welfare in poverty mapping, we can also choose other welfare measures, including inequality measures. For the purpose of the decomposition analysis, we shall employ two alternative measures, which we shall discuss in the next section.

4. Decomposition Analysis

The results for the decomposition analysis are at least in part driven by the choice of the decomposable inequality index. Thus, we use two alternative measures that allow for a neat decomposition in order to see if the choice of inequality measure matters. The first decomposable measure we consider is the generalized entropy measure with parameter $\alpha(\neq 0,1)$, which is defined as follows:

$$GE_G(\alpha) \equiv \frac{1}{\alpha(\alpha-1)}\left(\frac{1}{\#\{G\}}\sum_{l\in G}\left(\frac{y_l}{\bar{y}}\right)^{\alpha} - 1\right), \text{ where } \bar{y} \equiv \frac{1}{\#\{G\}}\sum_{l\in G} y_l .$$

When α is 0 or 1, the generalized entropy measure is defined as follows:

$$GE_G(0) = \frac{1}{\#\{G\}}\sum_{l\in G}\ln\frac{\bar{y}}{y_l} \text{ and } GE_V(1) = \frac{1}{\#\{G\}}\sum_{l\in G}\frac{y_l}{\bar{y}}\ln\frac{y_l}{\bar{y}} .$$

It is well known that the generalized entropy measure is additively decomposable and satisfies desirable characteristics of inequality measures, such as the transfer principle, scale independence, population-replication independence, and anonymity (Shorrocks, 1980; Shorrocks 1984). Pradhan, Sahn and Younger (2003) implicitly set the parameter value at $\alpha = 0$, which means that everyone is equally weighted. While this choice is sensible as there is no other obvious choice, we varied the parameter values to see if our results are sensitive to the choice of the parameter.

When we calculate the health inequality or inequality in standardized height and weight, we need to take into account the inequality arising from the genetic variations. When calculating the proportion of the between-country inequality to the total world inequality, Pradhan, Sahn and Younger (2003) adjusted the denominator (the total inequality in the world) by subtracting the natural inequality. . That is, instead of using $GE_{World}(0)$, they used $GE_{World}(0) - GE_{Natural}(0)$ in the denominator. This adjustment can be justified on the ground that the natural inequality does not exist between areas but within each area. We adjust the inequality in an analogous way. That is, we take the denominator (total inequality in Cambodia) to be $GE_{Cambodia}(\alpha) - GE_{Natural}(\alpha)$.

While the descriptive statistics of the share of the between-area inequality is interesting, we may be more interested in the spatial inequality of certain types of population. For example, if we are interested in the spatial concentration of the poor or the undernourished, statistics focused on the lower tail of the distribution is more appropriate. Even if the between-group inequality in the general entropy measure is low relative to the total inequality, spatial targeting is not necessarily ineffective. This is because low or zero between-group inequality and heterogeneous levels of malnourishment or poverty across areas can happen at the same time.

This point may be more clearly understood with a simple numerical example of income distribution. Suppose that there are two villages A and B in a small country and each village has 100 people. In Village A, everyone earns 10. In Village B, there is one rich person whose income is 901 and the remaining 99 people earn only 1. Let's suppose that the poverty line is 5. In this case, the between-group inequality measured by the generalized entropy measures is zero because the average income is 10 for both villages. However, geographic targeting is obviously very useful because all the poor people live in Village B. This shows that what really matters for the targeting of poverty alleviation programs is not the inequality of income but the "inequality of poverty" across areas. Similarly argument holds for the targeting of child nutrition programs.

To focus on undernutrition and poverty, we also carry out a simple variance decomposition analysis of the FGT measure. That is, let us suppose that the country C consists of J areas, and that each individual or household belongs to one and only one area of $G_1,\ldots,G_J$. Then, the total variance in the country V_C can be decomposed into the within-area variance V_W and between-area variance V_B in the following manner:

$$V_C \equiv \sum_{j=1}^{J}\sum_{l\in G_j}(P_l - P_C)^2 = \sum_{j=1}^{J}\sum_{l\in G_j}(P_l - P_{G_j})^2 + \sum_{j=1}^{J}\sum_{l\in G_j}\#\{G_j\}(P_{G_j} - P_C)^2 = V_W + V_B$$

Four remarks are in order. First, the FGT measure can be defined for each household or individual. Hence, taking the FGT measure at the unit-record level as the welfare indicator of interest, we can simply use the variance decomposition. Second, the proportion of the between-group variance, $V_C^{-1}V_W$, does not depend on the (arbitrary) choice of the reference population. This is because a change in the reference population translates into an affine transformation of the FGT measure, of which the proportion of the between-group variance is independent.

Third, we can readily compare the spatial inequality in poverty and undernutrition when $\alpha = 0$, because both poverty and undernutrition indices are taken as a discrete variable. Hence, we can see whether poverty is more spatially concentrated than undernourishment. Note that this does not hold other values of α. For example, there is no clear comparison between the poverty gap and the undernutrition gap (*i.e.* how below y is from z) when $\alpha = 1$ because the nutrition gap in part depends on the choice of the reference group.

Fourth, this decomposition analysis (rightly) ignores what happens in the upper tail. This asymmetry between the lower and upper tails is important for the purpose of policy analysis and formulation. For those implementing anti-poverty programs, transfer of income from the poorest to the second poorest would be a concern, but the transfer from the second richest man to the richest man would not.

5. Data

Both the poverty mapping and nutrition mapping require a survey and a census. Ideally, if we had a survey that includes consumption and nutrition indicators at the individual level in a single survey, we are in an ideal situation in which we can estimate consumption and nutrition indicators jointly. However, it is not the case in Cambodia, and thus we conducted poverty mapping and nutrition mapping separately.

Both the poverty mapping and nutrition mapping use the Cambodian National Population Census for 1998 (For details, see National Institute of Statistics (2000)). The census covers virtually all persons staying in Cambodia at the time of census, and includes variables for housing characteristics, conditions and facilities as well as individual variables for sex, age, relation to the head of household, marital status, migration, literacy, education, and employment. After excluding the records with missing values, the census contains about 2.1 million records of households and 1.4 million records of children under five.

In poverty mapping, we use the Cambodia Socio-Economic Survey (CSES) for 1997 (For details, see National Institute of Statistics (1998)). The CSES covers 6,010 households from 474 villages and village was taken as the primary sampling unit. The CSES 1997 includes consumption indicator as well as various other indicators both at the household and individual levels. It is representative at the level of three strata:

Phnom Penh, Other Urban and Rural.

In the nutrition mapping, we used the Cambodia Demographic and Health Survey (CDHS) for 2000 (National Institute of Statistic, Directorate General for Health and ORC Macro (2001) for details). It was designed to collect health and demographic information for the Cambodian population with a particular focus on women of childbearing age and young children. The sample covered 12,236 households in 17 strata across the country. In addition to detailed information about each household, its members, and housing characteristics, one-quarter of these households were systematically selected to participate in the anthropometric data collection. All children under 60 months of age in the sub-sampled households were weighed and measured. After excluding children for which information on height or weight is missing or implausible, 3,596 observations were used for this analysis.

Because Cambodia has a rich collection of geographic data, indicators on a range of characteristics could be merged into the census and the survey in both poverty mapping and nutrition mapping. These indicators include distance calculations, land use and land cover information, climate indicators, vegetation, agricultural production and flooding as well as the village-level means generated from the census. Inclusion of these geographic variables and their cross terms with other individual-level and household-level variables has improved substantially the ability to explain the variation of consumption and anthropometric indicators.

Using these data, we estimated the FGT measures and general entropy measures, among other things, for consumption, standardized height and standardized weight. A consumption model was constructed for each of the three strata for CSES 1997, while anthropometric models (height and weight) were constructed for each of the following five zones ("ecozones"): Urban, Plain, Tonlesap, Coastal and Plateau. These ecozones are a combination of provinces that have similar agro-climatic and socio-cultural characteristics. We aggregated provinces because some of the strata for DHS 2000 had too few observations to carry out meaningful analysis. In each model, we checked the robustness of the regression coefficients by randomly dropping some households or clusters as was done in Elbers, Lanjouw and Lanjouw (2002).

In the next section, we shall present the estimation results. We shall, however, only present the results relevant to the inequality decomposition. Readers are referred to World Food Programme (2002) and Fujii (2006) for the details of the standard poverty mapping exercise, and Fujii (2005) for the nutrition mapping exercise.

6. Results

Before carrying out the decomposition exercise, we plotted the $GE(0)$ inequality index for per capita consumption, standardized height and standardized weight in the form of a map. Because the comparison of $GE(0)$ in absolute terms is not meaningful, we simply looked at the quartiles. Figures 1, 2, and 3 are the maps of GE(0) at the commune-level, where a commune is the third largest administrative unit after province and commune, and before village.[v] The darker areas represent more unequal communes. For example, Q1 is the top one-quarter of most equal communes.

Three cautions are in order. First, the area of the commune varies substantially across the country. In general, the area of a commune is larger in more remote and sparsely populated areas. This in turn means that the remote communes tend to be overrepresented in the map. In particular, the northeastern part of the country have communes with larger areas. Second, the map does not take into account the standard errors associated with the estimates. Thus, even if we chose one commune from one quarter and another commune from another quarter, the difference in inequality may not be significant. Third, since $GE(0)$ is sensitive to the lower tail distribution, it is responsive to the worst-off children or households.

With these points in mind, let us look at the maps. One can immediately see that the maps for consumption and standardized height are similar. Indeed, there is a moderately positive correlation between the consumption inequality and height inequality; their Spearman rank-correlation is 0.24. The rank-correlations of weight inequality with consumption or height inequality are much weaker and negative (-0.14 for consumption and -0.04 for height).

While negative correlation may be unexpected, it is not surprising. Weight can fall very rapidly. Thus negative idiosyncratic shocks to worse-off children or household could substantially increase inequality. On the other hand, consumption and height do not change so quickly and thus consumption and height inequality are much less sensitive to short-run negative shocks. This also explains why there may be negative correlation; if the commune is more equal in the long-run, then the impact of the negative shock on inequality is larger.

Another point to note is that those communes with highest weight inequality are concentrated in the southeastern as one can see from Figure 3. While we cannot conclusively find the reason for this concentration, it is worth pointing out that the southeastern areas approximately correspond to the areas most vulnerable to flood. Flood tends to affect most severely those who are worse-off, because they live in lower land and have less protection from flood. This in turn means that areas affected by flood are likely to have higher local inequality other things being equal.

Let us now look at the decomposition results shown in Table 1. The figures for the generalized entropy measures for the standardized height and weight are adjusted for the natural inequality. We calculate the proportion for each round of the simulation and take the mean and standard deviation over the simulation to find the point estimate and the standard error as discussed in Section 3.

By construction, the proportion of between-group inequality goes down as we take more aggregated groups. In particular, the proportion of between-group to the total inequality is equal to zero at the level of Cambodia (because there is only one group) and one at the level of unit records (because there is no inequality within group).

For most of the inequality measures, the between-group component is highest for consumption followed by standardized height and standardized weight. The spatial inequality of poverty can be explained for the most part by the geographic variations (at the village level) in the distribution of poor people. However, this is not the case of standardized height and weight. Thus, geographic targeting is likely to be much more effective for anti-poverty programs than for nutrition programs.

How does the numbers in Table 1 compare with cross-country decomposition? While this study is a result for only one country and thus cannot be generalized, it does seem to indicate that health inequality is a much more local phenomenon than consumption inequality. As noted earlier, the between-country component accounts for only 31% of the total world health inequality (Pradhan, Sahn and Younger, 2003) as measured by $GE(0)$ for standardized height adjusted for natural inequality. In Cambodia, only 19% of the total health inequality can be explained by between-village component of the inequality using the same measure. The comparable figures consumption would be much higher. More than half of the total income inequality in the world is due to the between-country inequality. Similarly, more than half of the total consumption inequality in the world is due to between-village inequality in Cambodia.

One could object to a comparison of this sort on the ground that the choice of reference age and sex group—which we need to construct the standardized height and weight—is arbitrary. We argued that this problem could be avoided by conducting variance decomposition of $P(0)$. The qualitative nature of the results remain the same. The between-group component is much larger for the consumption measure than the standardized height and weight measures. Additional advantage of using $P(0)$ is that the standard errors associated with the decomposition results are much smaller than that for $GE(0)$. This stems from the fact that $P(0)$ is insensitive to the tails.

So far, we have only considered $GE(\bullet)$ and $P(\bullet)$ when their parameter value is equal to zero. While the parameter value of zero has some advantages over other values as we have argued, it is important to verify how much our results are driven by the choice of the parameter. As we can see from Table 1, the choice of the parameter for the generalized entropy measure does not matter much for the range between -1 and 1. When the parameter value is larger than one (not reported), the choice of parameter matters more significantly. This is because the generalized entropy measure tends to become more sensitive to extreme values. Similarly, the choice of the parameter for the FGT measure does not affect the results so much.

7. Discussions

We have conducted decomposition analyses of consumption poverty and health inequality for children by applying the small-area estimation. We presented the small-area estimation for consumption by Elbers, Lanjouw and Lanjouw (2002, 2003), standardized height and weight by Fujii (2005) in a unified framework, highlighting their similarities and differences. While this study is purely descriptive, it is useful for elucidating the significance of the geographic information in explaining overall inequality. The magnitude of the ratio of between-location health inequality to the overall health inequality in a country was not previously known.

We first decomposed the inequality in health and wealth following the approach by Pradhan, Sahn and Younger (2003). By comparing the proportion of the between-country inequality to the total world inequality and the proportion of the between-village inequality to the total inequality in Cambodia, we found that the sizable proportion of wealth (consumption) inequality is determined by geography whereas health inequality is intrinsically a local phenomenon.

This comparison may not convincing, because the choice of reference group is arbitrary. In order to overcome this problem, we argued that we can decomposed the health and wealth inequality by looking at the variance of $P(0)$. This measure allows for a valid comparison because it does not depend on the reference group and the left-hand-side variable that is used to produce $P(0)$ is essentially a discrete variable. This approach could be used to compare different kinds of other key indicators, including mortality, literacy, and prevalence of diseases.

Our conclusions are robust with respect to the choice of parameters for $GE(\bullet)$ and $P(\bullet)$; We consistently found that the between-group component is smaller for health inequality than consumption inequality regardless of the choice of the decomposable index. We also consistently found that the proportion of between-country inequality in the world is larger than the proportion of between-group inequality in Cambodia, even when the group is taken to be the smallest administrative unit, village. We cannot over-extrapolate from a study in one country, but this indicates that the inequality between the countries in the world is indeed important than the inequality between communities that exist within a country.

References

Bloom, D.E., and Canning, D. "The Health and Wealth of Nations." *Science.* 287(5456): 1207-1208.

Deaton, A. (2003) "Health, Inequality, and Economic Development." *Journal of Economic Literature*. 41(1): 113-159.

Demombynes, G., and Özler, B. (2005) "Crime and Local Inequality in South Africa." *Journal of Development Economics*. 76(2): 265-292.

Dibley, M.M., Gloldby, J.B., Staehling, N.W., and Trowbridge, F.L. (1987a) "Development of Normalized Curves for the International Growth Reference: Historical and Technical Considerations." *American Journal of Clinical Nutrition.* 46(5): 736-748.

Dibley, M.M., Staehling, N.W., Nieburg, P., and Trowbridge, F.L. (1987b) "Interpretation of Z-score Anthropometric Indicators Derived from the International Growth Reference." *American Journal of Clinical Nutrition.* 46(5): 749-762.

Elbers, C., Fujii, T., Lanjouw, P., Özler, B., and Yin, W. (2007) "Poverty Alleviation through Geographic Targeting." *Journal of Development Economics.* 83(1): 198-213.

Elbers, C., Lajouw, P., Mistiaen, A., Özler, B., and Simler, K. (2004) "On the Unequal Inequality of Poor Communities." *World Bank Economic Review*. 18(3): 401-421.

Elbers, C., Lanjouw. J.O., and Lanjouw, P. (2002) "Welfare in Villages and Towns: Micro-level estimation of poverty and inequality." Policy Research Working Paper 2911. Washington D.C., USA: The World Bank.

Elbers, C., Lanjouw. J.O., and Lanjouw, P. (2003) “Micro-level Estimation of Poverty and Inequality.” *Econometrica.* 71(1): 355-364.

Elo, I., and Preston, S.H. (1996) “Educational Differentials in Mortality: United States, 1979-85,” *Social Science and Medicine.* 42(1): 47-57.

Firebaugh, G. (2000) “The Trend in Between-Nation Income Inequality.” *Annual Review of Sociology*. 26: 323-339.

Foster, J., Greer, J., and Thorbecke, E. (1984) “A Class of Decomposable Poverty Measures.” *Econometrica.* 52(3): 761-766.

Fujii, T. (2005) “Micro-level Estimation of Child Malnutrition Indicators and its Application in Cambodia.” Policy Research Working Paper 3662. Washington D.C., USA: The World Bank.

Fujii, T. (2006) “Commune-level Estimation of Poverty Measure and its Application in Cambodia.” In Kanbur R., Venables A.J., and Wan, G. (eds) *Spatial Disparities in Human Development: Perspectives from Asia* pp. 289-314. United Nations University Press.

Fujii, T. (2007) “To Use or Not to Use?: Poverty Mapping in Cambodia.” In Bedi, T., Coudouel, A. and Simler, K. (eds) *More than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions.* pp. 125-142. The World Bank.

Fujii, T. (2008) “How Well Can we Target Aid with Rapidly Collected Data? Empirical Results for Poverty Mapping from Cambodia.” *World Development*. Forthcoming.

Fujii, T., and Roland-Holst, D.W. (2008) “How does Viet Nam’s Accession to the World Trade Organization Change the Spatial Incidence of Poverty?” In Nissanke, M., and Thorbecke E. (eds) *Globalization and the Poor in Asia: Can Shared Growth be Sustained?* Palgrave Macmillan. Forthcoming.

Kawachi, I., Kennedy, B.P., Lochner, K., and Prothrow-Stith, D. (1997) “Social Capital, Income Inequality and Mortality.” *American Journal of Public Health*. 87(9): 1491-1498.

Marmot, M., and Wilkinson, R.G. (2001) “Psychosocial and Material Pathways in the Relation between Income and Health: A Response to Lynch et al.” *British Medical Journal*. 322: 1233-1236.

National Institute of Statistic (1998) *Cambodia Socio-Economic Survey 1997: Technical Report on Survey Design and Implementation.* Phnom Penh, Cambodia: National Institute of Statistic, Ministry of Planning, Royal Government of Cambodia.

National Institute of Statistic (2000) *General Population Census of Cambodia 1998: Village Gazetteer*. Phnom Penh, Cambodia: National Institute of Statistic, Ministry of Planning, Royal Government of Cambodia.

National Institute of Statistics, Directorate General for Health, and ORC Macro (2001) *Cambodia Demographics and Health Survey 2000*. Phnom Penh, Cambodia: National Institute of Statistics, Ministry of Planning, Royal Government of Cambodia.

Pradhan, M., Sahn, D.E., and Younger, S.D. (2003) “Decomposing World Health Inequality.” *Journal of Health Economics*. 22: 271-293.

Pritchett, L., and Summers, L.H. (1996) “Wealthier is Healthier.” *Journal of Human Resources.* 31(4): 841-868.

Shorrocks, A.F. (1980) “The Class of Additively Decomposable Inequality Measures.” *Econometrica.* 48(3): 613-625.

Shorrocks, A.F. (1984) “Inequality Decomposition by Population Subgroups.” *Econometrica.* 52(6):1369-1385.

WHO Working Group (1986) “Use and Interpretation of Anthropometric Indicators of Nutritional Status.” *Bulletin of the World Health Organization*. 64(6): 929-941.

WHO Working Group (1995) “An Evaluation of Infant Growth: the Use and Interpretation of Anthropometry in Infants.” *Bulletin of the World Health Organization*. 73(2): 165-174.

Wagstaff, A. (2005) “Inequality Decomposition and Geographic Targeting with Applications to China and Vietnam.” *Health Economics.* 14: 649-653.

Wagstaff, A., and van Doorslaer, E. (2000) “Income Inequality and Health: What does Literature Tell Us?” *Annual Review of Public Health*. 21: 543-567.

Waterlow, J.C., Buznina, R., Keller, W., Lane, J.M., Nichaman, M.Z. and Tanner, J.M. (1977) “The Presentation and Use of Height and Weight Data for Comparing the Nutritional Status of Groups of Children under the Age of 10 Years.” *Bulletin of the World Health Organization*. 55(4): 489-498.

Wilkinson, R.G. (1992) “Income Distribution and Life Expectancy.” *British Medical Journal*. 304: 165-8.

Wilkinson, R.G. (1996) *Unhealthy Societies: The Affliction of Inequality*. Routledge.

Wilkinson, R.G. (1997) “Income Inequality Summarizes the Health Burden of Individual Relative Deprivation.” *British Medical Journal*. 314: 1727-1728.

van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U-G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., Leu, R.E., O’Donnell, O., Propper, C., Puffer, F., Rodr í guez, M., Sundberg, G., and Winkelhake, O. (1997) “Income-related inequalities in health: some international comparisons.” *Journal of Health Economics.* 16(1): 93-112.

World Food Programme (2002) *Estimation of Poverty Rate at Commune Level in Cambodia: Using the Small-Area Estimation Technique to Obtain Reliable Estimates*. Phnom Penh, Cambodia: World Food Programme.

Tables and Figures

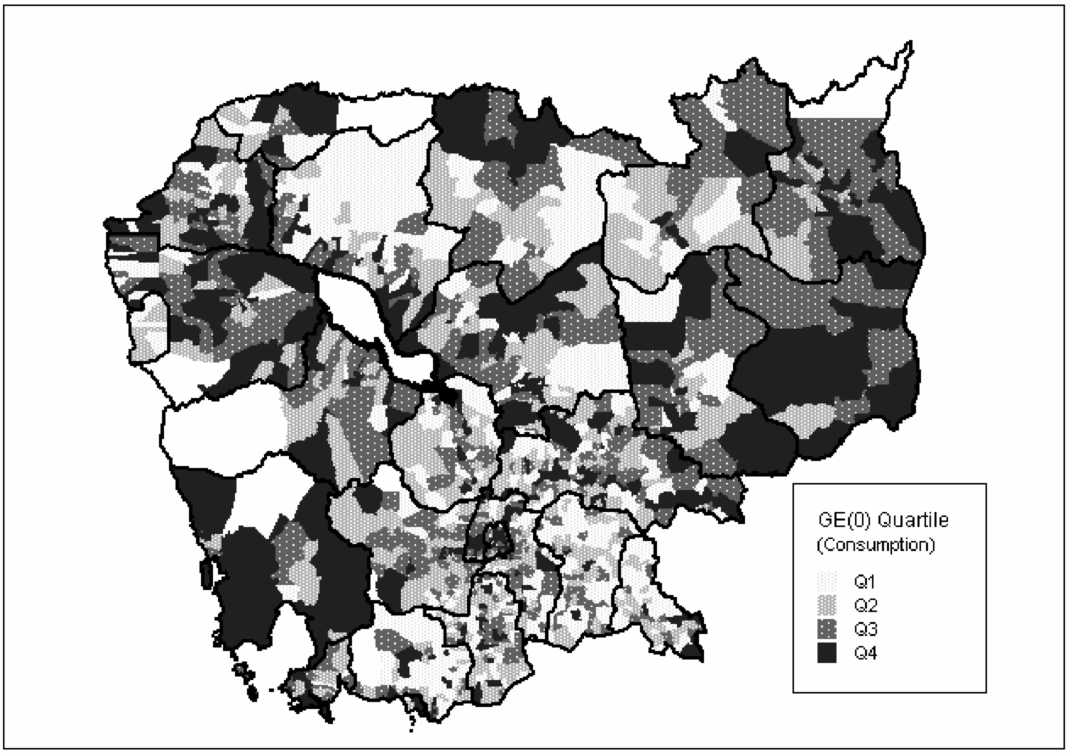

Figure 1. Map of inequality (GE(0)) in consumption at the commune level.

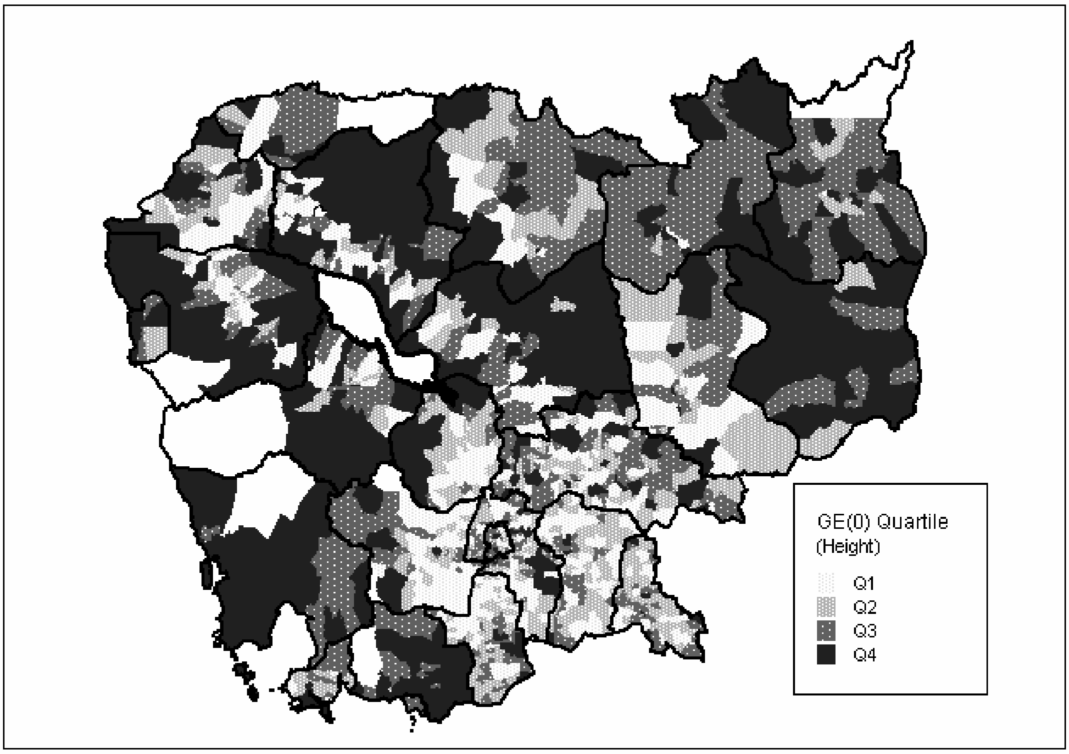

Figure 2. Map of inequality (GE(0)) in standardized height at the commune level.

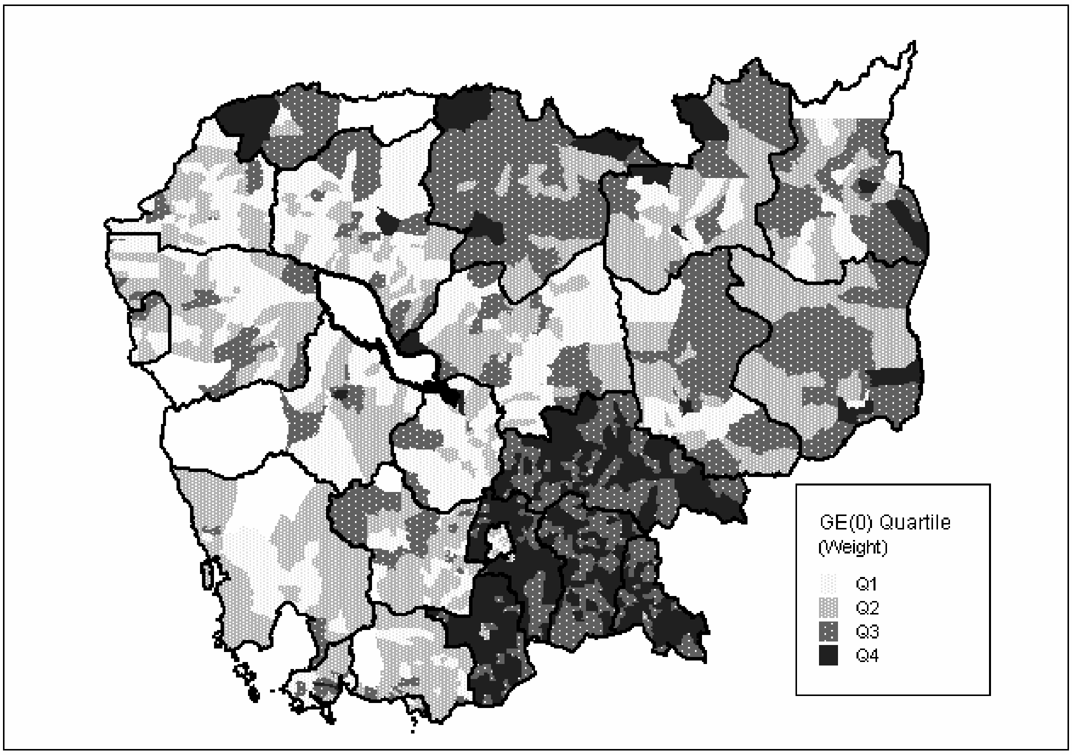

Figure 3. Map of inequality (GE(0)) in standardized weight at the commune level.

Table 1. Proportion of between-group inequality for various indicators at various levels.[vi]

		Cambodia		Province		District		Commune		Village		Unit Record	
Standardized Height	P(0)	0.00	(0.00)	1.31	(0.30)	2.73	(0.35)	6.32	(0.54)	12.22	(0.90)	100.00	(0.00)
	P(1)	0.00	(0.00)	1.24	(0.34)	3.06	(0.50)	7.82	(1.12)	14.19	(1.48)	100.00	(0.00)
	P(2)	0.00	(0.00)	0.90	(0.27)	2.45	(0.47)	6.89	(1.47)	12.02	(1.74)	100.00	(0.00)
	GE(-1)	0.00	(0.00)	2.30	(0.55)	5.13	(0.64)	13.62	(1.35)	23.94	(1.86)	100.00	(0.00)
	GE(-0.5)	0.00	(0.00)	2.28	(0.54)	5.07	(0.63)	13.51	(1.35)	23.74	(1.84)	100.00	(0.00)
	GE(0)	0.00	(0.00)	2.23	(0.53)	4.95	(0.61)	13.23	(1.33)	23.24	(1.81)	100.00	(0.00)
	GE(0.5)	0.00	(0.00)	2.08	(0.49)	4.62	(0.57)	12.39	(1.25)	21.76	(1.69)	100.00	(0.00)
	GE(1)	0.00	(0.00)	2.59	(0.61)	5.74	(0.70)	15.45	(1.58)	27.12	(2.12)	100.00	(0.00)
Standardized Weight	P(0)	0.00	(0.00)	0.88	(0.23)	2.04	(0.27)	4.62	(0.40)	10.76	(0.83)	100.00	(0.00)
	P(1)	0.00	(0.00)	1.99	(0.56)	3.17	(0.61)	5.70	(0.77)	11.76	(1.09)	100.00	(0.00)
	P(2)	0.00	(0.00)	2.03	(0.56)	2.86	(0.61)	4.66	(0.76)	9.06	(0.94)	100.00	(0.00)
	GE(-1)	0.00	(0.00)	1.33	(0.36)	3.29	(0.50)	7.83	(0.88)	17.94	(1.76)	100.00	(0.00)
	GE(-0.5)	0.00	(0.00)	1.38	(0.38)	3.41	(0.51)	8.12	(0.90)	18.57	(1.80)	100.00	(0.00)
	GE(0)	0.00	(0.00)	1.41	(0.39)	3.48	(0.52)	8.29	(0.91)	18.94	(1.81)	100.00	(0.00)
	GE(0.5)	0.00	(0.00)	1.40	(0.38)	3.44	(0.52)	8.20	(0.89)	18.72	(1.78)	100.00	(0.00)
	GE(1)	0.00	(0.00)	1.59	(0.44)	3.92	(0.59)	9.35	(1.01)	21.33	(2.01)	100.00	(0.00)
Consumption	P(0)	0.00	(0.00)	7.10	(0.50)	12.00	(0.71)	19.06	(0.88)	37.14	(1.16)	100.00	(0.00)
	P(1)	0.00	(0.00)	8.20	(0.86)	14.29	(1.36)	23.15	(1.76)	46.72	(2.01)	100.00	(0.00)
	P(2)	0.00	(0.00)	7.20	(1.00)	12.98	(1.81)	21.53	(2.36)	45.87	(2.82)	100.00	(0.00)
	GE(-1)	0.00	(0.00)	15.20	(11.07)	24.22	(10.47)	36.56	(9.26)	57.87	(7.22)	100.00	(0.00)
	GE(-0.5)	0.00	(0.00)	17.64	(14.30)	27.49	(13.48)	41.01	(11.33)	62.17	(7.55)	100.00	(0.00)
	GE(0)	0.00	(0.00)	18.69	(16.32)	28.84	(15.84)	43.19	(13.16)	64.14	(8.68)	100.00	(0.00)
	GE(0.5)	0.00	(0.00)	17.04	(12.66)	27.18	(14.47)	42.25	(12.99)	63.66	(9.09)	100.00	(0.00)
	GE(1)	0.00	(0.00)	11.71	(2.59)	19.91	(3.54)	34.42	(4.06)	57.24	(6.12)	100.00	(0.00)
Number of Observations		1		24		180		1594		13320/13233		1424907/2130544	

[i] Further, the survey may not be representative at the level of decomposition. For example, the 1998 Vietnam Living Standard Survey used in Wagstaff (2005) is not representative at the provincial level. Hence, the within-province concentration index is unreliable at best.

[ii] See Dibley et al. (1987a, 1987b), Waterlow et al. (1977), and WHO Working Group (1986, 1995) for further discussion of the Z-score.

[iii] Pradhan, Sahn and Younger (2003) only use height because too much weight is obviously not healthy. However, this is not a concern in Cambodia as only less than 1 percent of children under five is overweight in Cambodia.

[iv] In principle, G does not have to be a group defined by geographic location. However, we only consider geographic aggregation in this study.

[v] There are about 1,600 communes in Cambodia, and each commune contains around 1,300 households.

[vi] The number of villages for standardized height and weight are 13320 whereas that for the consumption is 13233. This discrepancy comes from the fact that there are some villages without children under five. The number of unit records corresponds to the number of individuals for standardized height and weight (1424907) whereas it corresponds to the number of households for consumption (2130544).