# Allowing for heterogeneity in the decomposition of measures of inequality in health

*by*

Andrew M. Jones*[a] and Angel López Nicolás[b]

*January 2004*

[a] Department of Economics and Related Studies, University of York, York YO10 5DD, UK

[b] Departament d'Economia i Empresa and CRES, Universitat Pompeu Fabra, 08005-Barcelona, Spain

**Abstract**

This paper shows how recently developed regression-based methods for the decomposition of health inequality can be extended to incorporate heterogeneity in the responses of health to the explanatory variables. We illustrate our method with an application to the GHQ measure of psychological well-being taken from the British Household Panel Survey. The results suggest that there is an important degree of heterogeneity in the association of health to explanatory variables across birth cohorts and genders which, in turn, accounts for a substantial percentage of the inequality in observed health.

JEL classification: D63, I12, C21

Keywords: Health inequalities; heterogeneity; decomposition analysis; panel data.

* Corresponding author. Tel. +44 1904 433790. Fax: +44 1904 433759. E-mail: amj1@york.ac.uk. Data from the *British Household Panel Survey* (BHPS) were supplied by the ESRC Data Archive. Neither the original collectors of the data nor the Archive bear any responsibility for the analysis or interpretations presented here. This chapter derives from the project "The dynamics of income, health and inequality over the lifecycle" (known as the ECuity III Project), which is funded in part by the European Community's Quality of Life and Management of Living Resources programme (contract QLK6-CT-2002-02297). We are grateful to the EC and to the Fundación BBVA and the Spanish DGES (project BEC2002-04294-C02-02) for financial support. We thank Eddy van Doorslaer for support and advice and to participants at the York Seminars in Health Econometrics (YSHE) and two anonymous referees for their comments.

# 1. Introduction

Health economists have adopted the Gini coefficient and concentration indices to provide summary measures of inequalities of health within populations (see e.g. Wagstaff et al, 1989, 1991, 1994, and van Doorslaer et al. 1997). Like the Gini coefficient of income inequality, the Gini and concentration indices for health have the attraction that they can be decomposed by factors (see e.g., Rao, 1969, Kakwani, 1980). A recent contribution by Wagstaff et al. (2003) has used this property to show how a linear regression approach can be used to decompose these indices into the contributions of different explanatory variables. The decomposition treats individual responses to these explanatory variables (the slope coefficients) as homogeneous across individuals. In this paper we show how the decomposition can be extended to allow for heterogeneity in the responses; we illustrate the method with an application to the measurement of inequality in the GHQ measure of psychological well-being using data from the British Household Panel Survey (BHPS) for 1991-99.

The structure of the paper is the following. Section 2 presents our main result and we show how the decomposition of the Gini index and the concentration indices into the contributions of different explanatory variables in a regression model can be modified to incorporate individual heterogeneity in all the coefficients. Section 3 discusses the main features of the BHPS data set and presents and discusses the estimates from regression models for GHQ that allow for heterogeneous responses across age and gender cohorts. Section 4 concludes.

# 2. Regression based decompositions of inequality

The departure point for our methodology is the decomposition of inequality measures into the contributions of different explanatory variables by means of a linear regression model (see e.g., Wagstaff et al., 2003). Suppose we are interested in calculating the Gini coefficient for a measure of health using individual data in a sample from the population of interest. Let $y_i$ denote a measure of health for the $i^{th}$ individual, i=1,2,…N, and $R_i$ denote the cumulative proportion of the population ranked by $y_i$ up to the $i^{th}$ individual (their 'relative rank'). Ignoring, for expositional purposes, the fact that in general sampling weights will be necessary, the Gini coefficient, G, for health is given by (see e.g., Lambert, 1994 p.43),

$$G = \left(\frac{2}{\bar{y}}\right)\operatorname{cov}(y_i, R_i) \tag{1}$$

where $\bar{y} = E(y_i)$. Now let $y_i$ be given by the following linear regression model

$$y_i = \beta_1 + \sum_{k=2}^{K} \beta_k x_{ki} + \varepsilon_i \tag{2}$$

where k is the number of regressors (*x*). By substituting this for $y_i$, the Gini index of y can be written as (see Wagstaff et al., 2003),

$$G = \sum_{k=2}^{K}\left(\beta_k \frac{\bar{x}_k}{\bar{y}}\right)C_k + \left(\frac{2}{\bar{y}}\right)\operatorname{cov}(\varepsilon_i, R_i) \tag{3}$$

The first term in brackets is the elasticity of y with respect to $x_k$ evaluated at the sample means ($\bar{x}_k$ and $\bar{y}$), and $C_k$ is the concentration index of $x_k$ on y. The latter expression can be easily modified to obtain the concentration index of y against another variable of interest. For instance, the concentration index, CI, of health against income would be computed according to

$$CI = \sum_{k=2}^{K}\left(\beta_k \frac{\bar{x}_k}{\bar{y}}\right)C'_k + \left(\frac{2}{\bar{y}}\right)\operatorname{cov}(\varepsilon_i, R'_i) \tag{4}$$

where $C'_k$ denotes the concentration index of $x_k$ against income and $R'_i$ is the cumulative proportion of the population ranked by income up to the $i^{th}$ individual.

In practice, the decomposition formulas, (3) and (4), are applied using OLS estimates of the βs and of the residuals. Thus these inequality measures can be decomposed into an "explained part" and an "unexplained part" (see Wagstaff et al., 2003). The "explained" part can be usefully broken down into the contributions of individual explanatory variables. As for the "unexplained" part, it is a scaled measure of the covariance of the residuals in the regression model with the position of the individual in the distribution of the variable of interest. As such, the unexplained part should be zero if the regression model for the measure of health is specified in a way such that there is no systematic variation in unobserved heterogeneity in health according to the position of the individual in the distribution of the relevant variable. So, the contribution of the residuals should be close to zero for the concentration index but may be quite large for the Gini coefficient.

The presence of heterogeneity in econometric models for microdata, as reflected by low coefficients of determination, would lead to the suspicion that a regression specification that assumes homogeneous responses may not capture all the variation in y. For example, Heckman (2001) argues that:

*"..the most important discovery was the evidence on the pervasiveness of heterogeneity and diversity in economic behaviour…not only were intercepts variable but so were slopes.."*

We now propose a method that deals with heterogeneity, while retaining the useful summary information provided by the regression approach.

A very general way to allow for individual heterogeneity is by means of a regression model for the health variable with heterogeneous parameters. Thus, the regression model can be modified to yield,

$$y_i = \beta_{i1} + \sum_{k=2}^{K} \beta_{ik} x_{ki} + u_i \tag{5}$$

where all the parameters in the model are individual specific. Equation (5) introduces heterogeneity through a linear specification. For ease of interpretation the model can be re-written in the form,

$$\begin{aligned} y_i &= \beta_1 + \sum_{k=2}^{K} \beta_k x_{ki} + (\beta_{i1} - \beta_1) + \sum_{k=2}^{K} (\beta_{ik} - \beta_k) x_{ki} + u_i \\ &= \beta_1 + \sum_{k=2}^{K} \beta_k x_{ki} + v_i \end{aligned}$$

where $\beta_k = E(\beta_{ik})$. Assuming conditional independence of the βs and the elements of $x$, such that $E(\beta_{ik}|x_i)=E(\beta_{ik})= \beta_k$ for all $k=1,\dots K$, gives a random parameters specification and implies linear scale heteroscedasticity in $v_i$ (see, Koenker and Bassett (1982)). This need not be too restrictive as the regressors, $x$, may contain high order polynomials, interactions and other transformations of the variables of interest. Koenker and Bassett (1982, p.45) show that the "linear scale model of heteroscedasticity is an important special case of the general class of models with linear conditional quantile functions. It subsumes many models of systematic heteroscedasticity that have appeared in the econometrics literature". If conditional independence of the βs and the elements of $x$ does not hold, the model leads to a correlated/fixed parameters specification which can be estimated by introducing an individual-specific dummy variable $d_i$, $i=1,\ldots, n$, (by analogy with the least squares dummy variable estimator for the fixed effects model),

$$y_i = \sum_{i=1}^{n} \beta_{i1} d_i + \sum_{i=1}^{n} \sum_{k=2}^{K} \beta_{ik} d_i x_{ki} + u_i$$

So, estimation of the $\beta_i$s would require a long panel data set with sufficient time series variation to estimate a separate regression for each individual. In our empirical application we allow for heterogeneity across age-gender cohorts and estimate separate regressions for each group.

Substituting equation (5) into (1) we obtain,

$$G = \left(\frac{2}{\bar{y}}\right) \operatorname{cov}(y_i, R_i) =$$

$$G = \left(\frac{2}{N\bar{y}}\right) \sum_i (y_i - \bar{y})(R_i - 1/2) =$$

$$G = \left(\frac{2}{N\bar{y}}\right) \sum_i \left(\beta_{i1} + \sum_{k=2}^{K} \beta_{ik} x_{ki} - \bar{y}\right)(R_i - 1/2) + \left(\frac{2}{\bar{y}}\right) Cov(u_i, R_i)$$

To provide a benchmark for our decomposition analysis, consider estimating a "pooled" regression that treats $\beta_i$s as constant and ignores the individual heterogeneity. Since,

$$\bar{y} = \beta_1^{OLS} + \sum_{k=2}^{K} \beta_k^{OLS} \bar{x}_k$$

where $\beta^{OLS}$ denotes the OLS estimates of the coefficients from the pooled model, we can write after some manipulation,

$$G = \left(\frac{2}{N\bar{y}}\right)\left[\sum_i \left(\sum_{k=2}^{K} \beta_k^{OLS}(x_{ik} - \bar{x}_k) + \sum_{k=2}^{K} (\beta_{ik} - \beta_k^{OLS}) x_{ik} + (\beta_{i1} - \beta_1^{OLS})\right)(R_i - 1/2)\right] + \left(\frac{2}{\bar{y}}\right) Cov(u_i, R_i)$$

Collecting terms and changing the order of summation,

$$G=\sum_{k=2}^{K}\left(\frac{2}{\bar{y}N}\right)\beta_k^{OLS}\sum_i\left(x_{ik}-\bar{x}_k\right)\left(R_i-1/2\right)+$$
$$+\left(\frac{2}{\bar{y}N}\right)\sum_{k=2}^{K}\sum_i x_{ik}\left(\beta_{ik}-\beta_k^{OLS}\right)\left(R_i-1/2\right)+$$
$$+\left(\frac{2}{\bar{y}N}\right)\sum_i\left(\beta_{i1}-\beta_k^{OLS}\right)\left(R_i-1/2\right)+\left(\frac{2}{\bar{y}}\right)Cov\left(u_i,R_i\right)$$

Noting that the concentration index of $x_k$ on y is given by,

$$C_k=\frac{2}{\bar{x}_k}\operatorname{cov}(x_{ik},R_i)$$

And considering $\beta_1^{OLS}$ a measure of central tendency for $\beta_{i1}$, we finally obtain,

$$G=\sum_{k=2}^{K}\beta_k^{OLS}\frac{\bar{x}_k}{\bar{y}}C_k+\left(\frac{2}{\bar{y}N}\right)\sum_{k=2}^{K}\sum_i x_{ik}\left(\beta_{ik}-\beta_k^{OLS}\right)\left(R_i-1/2\right)+$$
$$+\left(\frac{2}{\bar{y}}\right)\operatorname{cov}\left(\beta_{i1},R_i\right)+\left(\frac{2}{\bar{y}}\right)Cov\left(u_i,R_i\right)$$

(6)

The first term of this equation is exactly the same as the first term in equation (3) when model (2) is estimated by OLS. The residual term in equation (3) is now split into three components given by the second , third and fourth terms in equation (6). The second term is the contribution to overall inequality of the covariance (weighted by the values of $x_k$) of the slope parameters with the health rank. The third term is simply the covariance of the intercepts (centred at the OLS intercept coefficient) with the health rank. The fourth term corresponds to the "unexplained part" component associated to the model in equation 5. The interpretation of the new decomposition formula has a clear parallel with the use of Oaxaca-type decompositions of changes in inequalities over time or across countries, based on the comparison of two different samples (see Wagstaff *et al.*, 2003). The Oaxaca (1973) decomposition allows an analysis of the extent to which changes in inequality are attributable to changes in inequality in the regressors and to changes in their associated elasticities. Similarly, our new decomposition of the overall level of inequality distinguishes between the contribution of the level of inequality in the regressors, evaluated at the pooled level of the coefficients, and the contribution of heterogeneity in the coefficients around the pooled values.

Like the Gini coefficient, the concentration index for health can be written as,

$$CI = \sum_{k=2}^{K} \beta_k^{OLS} \frac{\bar{x}_k}{\bar{y}} C_k + \left(\frac{2}{\bar{y}N}\right) \sum_{k=2}^{K} \sum_{i} x_{ik} \left(\beta_{ik} - \beta_k^{OLS}\right)\left(R'_i - 1/2\right) +$$
$$+ \left(\frac{2}{\bar{y}}\right) \operatorname{cov}\left(\beta_{i1}, R'_i\right) + \left(\frac{2}{\bar{y}}\right) Cov\left(u_i, R'_i\right)$$

(7)

Each component has a similar interpretation to the Gini coefficient, with health rank, R, replaced by income rank, R'. The first term is identical to the first term in (4) and the second two terms decompose the generalised concentration index of the residual, allowing for heterogeneity.

Implementation of the decomposition formulas in (6) and (7) requires individual-specific estimates of the intercepts and slope coefficients *($\beta_i$).* Panel data, such as the BHPS, provides the greatest scope for identifying these parameters. The *$\beta_i$'s* could be retrieved from a random parameters specification or, given sufficient time series variation, they could be estimated in a fixed effects specification with individual-specific slopes as well as intercepts. In practice, identifying individual-specific *$\beta$'s* from a relatively short time series is problematic. So, for our illustrative empirical application, we allow for heterogeneity in responses across fourteen groups of individuals, defined by age-gender cohorts

## 3. An empirical application

### 3.1 The BHPS data

To illustrate the methods proposed above we estimate a model of psychological well-being based on scores from the General Health Questionnaire (GHQ) as measured in the British Household Panel Survey (BHPS). This application was chosen because the BHPS is a recent panel data set with good quality income and socio-economic variables and because the GHQ measure can be modelled conveniently in a linear regression framework (see e.g., Hauck and Rice, 2003; Wildman, 2003). We use nine waves of data (1991-1999) in order to be able to estimate a health function for different population groups defined by age cohort and gender.

The BHPS is a longitudinal survey of private households in Great Britain (England, Scotland and Wales). It was designed to be a nationally representative survey of over 5,000 households and gives around 10,000 individual interviews of each adult (16+) household member. The BHPS is a repeated panel, with respondents questioned each year. The initial sample, collected in 1991, was selected using a two-stage stratified sampling procedure, designed to give each address an approximately equal probability of selection (Taylor et al., 1998). The first stage consisted of selecting 250 postcode sectors, with probabilities proportional to their size. The second stage selected delivery points. If multiple addresses were found the interviewer selected a particular household. The first wave of the survey was carried out between 1st September 1990 and 30th April 1991. The same individuals are re-interviewed in successive waves and, if new households are formed, the original sample members are interviewed along with all adults in the new household. Information at both the household and the individual level is collected, covering questions on neighbourhood, income, employment, health and caring, demographics, and values and opinions.

For our illustrative application, a subset of individuals who had at least one interview at any of the nine waves, between 1991 and 1999, is used. Selecting the unbalanced cases gives an initial sub-sample of 95,601 observations corresponding to 21,817 individuals. From these we have dropped the observations containing missing values in the GHQ score (5,853 observations) and those whose full household income is reported to be either below £2000 or above £77,000, or is missing (1010 observations). These income thresholds correspond to less than 1% of the observations in the extreme left and right hand tails of the distributions. A further 583 observations with missing values for marital status, social class, job status, number of children or education are dropped from the sample. Thus the estimating samples contain 88155 observations. These observations are grouped according to age cohort and gender resulting in 14 different subsamples as shown in Table 1.

*Table 1. Sub-sample sizes by age cohort and gender*

| Group | Cohort/gender | N |
|---|---|---|
| 1 | Born1976-1983 (women) | 3321 |
| 2 | Born1976-1983 (men) | 2893 |
| 3 | Born1965-1975 (women) | 9642 |
| 4 | Born1965-1975 (men) | 8908 |
| 5 | Born1955-1964 (women) | 9425 |
| 6 | Born1955-1964 (men) | 8323 |
| 7 | Born1945-1954 (women) | 8357 |
| 8 | Born1945-1954 (men) | 7054 |
| 9 | Born1935-1944 (women) | 5901 |
| 10 | Born1935-1944 (men) | 5269 |
| 11 | Born1925-1934 (women) | 5026 |
| 12 | Born1925-1934 (men) | 4341 |
| 13 | Born before 1925 (women) | 5897 |
| 14 | Born before 1925 (men) | 3798 |
| Total | | 88155 |

The BHPS self-completion questionnaire incorporates a reduced version of the General Household Questionnaire (Goldberg and Williams, 1988; Bowling, 1991). The GHQ was developed as a screening instrument for psychiatric illness and is now often used as an indicator of psychological well-being (Hauck and Rice, 2003; Weich et al., 2001; Wildman, 2003). The shortened GHQ includes 12 elements: concentration, sleep loss due to worry, perception of role, capability in decision making, whether constantly under strain, perception of problems in overcoming difficulties, enjoyment of day-to-day activities, ability to face problems, loss of confidence, self-worth, general happiness, and whether suffering depression or unhappiness. Responses are given on a 4-point scale ranging from 0 to 3, with 0 being the best score. For our dependent variable we use the Likert scale, which sums the individual components (Likert, 1952). This gives an overall scale that runs from 0 to 36. To make the interpretation of results more intuitive, we have re-scaled this measure in order to make it increasing in good health. Therefore we use GHQ'=36-GHQ rather than the original GHQ score.

### 3.2 A model with heterogeneous responses for the GHQ score

Next we use a regression model for the level of GHQ' score allowing for heterogeneous responses by age cohort and gender. The intention of the regression model is simply to capture the linear association between the GHQ' score and a range of socio-economic characteristics while allowing for differential responses according to demographic group. It should not be taken as a structural model or used to infer a direction of causality. Table 2 gives the names and definitions of the regressors and their means for the pooled samples of men and women.

*Table 2 Definitions and means for the explanatory variables*

| Var. Name | Definition | Women | Men |
|---|---|---|---|
| Log(income) | Logarithm of equivalised annual real household income | 9.13 | 9.26 |
| Never married | 1 if never married, 0 otherwise | 0.18 | 0.23 |
| Divced., septed., or widwd. | 1 if divorced, separated or widowed, 0 otherwise | 0.20 | 0.08 |
| Children | Number of children in household | 0.63 | 0.59 |
| Professional or managerial | 1 if R. General's SC is prof., manag. or technical occupation, 0 otherwise | 0.17 | 0.26 |
| Skilled | 1 if R. General's SC is skilled manual occupation, 0 otherwise | 0.05 | 0.21 |
| Other SC | 1 if R. General's SC is other, 0 otherwise | 0.54 | 0.42 |
| Unskilled | 1 if R.General's SC is partly skilled or unskilled occupation, 0 otherwise | 0.03 | 0.02 |
| Unemployed | 1 if unemployed, 0 otherwise | 0.03 | 0.06 |
| Student or retired | 1 if student or retired | 0.25 | 0.23 |
| Carer | 1 if family carer, 0 otherwise | 0.15 | 0.01 |
| Degree | 1 if highest academic qualification is degree or higher degree, 0 otherwise | 0.10 | 0.13 |
| HND/HNCT | 1 if highest academic qualification is HND or HNCT, 0 otherwise | 0.06 | 0.07 |
| O/CSE | 1 if highest academic qualification is O level or CSE, 0 otherwise | 0.31 | 0.28 |
| No qualification | 1 if no academic qualifications, 0 otherwise | 0.38 | 0.32 |
| Non-white | 1 if ethnic origin is other than white, 0 otherwise | 0.03 | 0.03 |
| N | | 47569 | 40586 |

Full details of the OLS regression results are given in a table in the Appendix. The coefficients are presented along with Huber-White robust standard errors that are adjusted for clustering within-individuals due to the use of panel data. The first column presents the estimates for a model where all parameters are restricted to be equal across demographic groups (giving the pooled OLS estimates, $\beta^{OLS}$). The following 14 columns (column 1 corresponds to group 1 as defined in Table 1 and so on) present the estimates for the model where all parameters are allowed to vary. While the RESET statistic at the bottom of the pooled model clearly suggests misspecification, the test values for the group-specific models are much lower and in several cases are well below standard critical values. In order to investigate the potential sources of misspecification, a quadratic term in the logarithm of income was included in alternative specifications for both the pooled model and the group specific models but it was not significant in any of the cases. This would favour the view that misspecification is more related to the assumption of common structure than incorrect functional form, suggesting that even finer grouping would lead to more improvements in the RESET statistics. Further evidence against the assumption of common responses is provided in the last column by the F statistic for the null of homogeneity across groups, which is well above standard critical values for the overwhelming majority of regressors.

The estimates reveal a rich pattern of heterogeneity across groups. Notably, while the marginal effect of equivalised household income is statistically insignificant for men and women in both the youngest age cohort (born between 1976-1983) and the oldest age cohorts (born before 1925), it takes a significantly positive value for all other cohorts and, for both men and women, reaches its maximum value for the 1935-1944 and 1945-1954 cohorts. Being "Divorced, separated or widowed" is associated with a lower GHQ' score than the omitted category (living as a couple), and while its effect tends to be greater in absolute value for men in younger cohorts, this event tends to have a greater negative impact on women in cohorts born before 1955. Unskilled women in older cohorts tend to report a greater GHQ' score than skilled non-manual workers (the omitted occupational category) in the same cohorts, but this effect is insignificant for the rest of demographic groups. All individuals within the "other social class" occupational category tend to report a lower GHQ' score than skilled non-manual workers within their group, but the association is stronger in old cohorts, and within the latter the effect is greater in absolute value for males. Unemployed women belonging to the 1945-1954 cohort or younger report a significantly lower GHQ' score than those in employment (the omitted labour status category) in the same cohorts. A similar pattern is found for men belonging to the 1955-1964 cohort or younger. The "student or retired" dummy attempts to capture voluntary inactivity in relation to the employed default category. While this effect is positive and significant for all men born before 1944, it is not statistically significant for women except those belonging to the 1935-1944 cohort. Being a family carer is associated to a lower GHQ' score than being employed in the youngest female cohort, while men in the 1935-1944 cohort in family care tend to report a greater GHQ' score. Women with degrees or no qualifications in the youngest cohort report on average a greater GHQ' score than women with A-levels (the omitted educational category) in the same cohort. But having no qualifications is associated to a lower GHQ' score for women born between 1955-1975 and between 1925-1934 and men born between 1955-1964 and between 1925-1934. Within the 1925-1934 cohort of men, those with A levels tend to report the greatest GHQ' score. On the other hand, within the 1935-1944 cohort, both men and women with O levels tend to report a greater

GHQ' score than, respectively, men and women with A levels. The intercept terms also reveal an interesting pattern of heterogeneity: Intercepts for young cohorts tend to be greater than those of old cohorts and, also, within all but the 1935-1944 cohort, the intercept for males is greater than for females.

### 3. 3 Decomposition analysis

The heterogeneity of the effects discussed above has an impact on the decomposition of the inequality measures, as shown in section 2. We now use these parameter estimates in order to calculate and decompose the Gini coefficient and concentration indices for the GHQ' score in the 1999 wave of the BHPS. Table 3 presents, for each explanatory variable in the model, the results for the decomposition of both inequality indices into:

i) The "homogeneous parameters contribution": i.e. the contribution of the product of the elasticities evaluated at the homogeneous parameters and the concentration indices of the regressors on health rank (or income rank in the case of the CI),

ii) "Heterogeneous parameters contribution": i.e. the contribution of the covariance of the heterogeneous parameters with the health rank (or income rank in the case of the CI),

Recall the expressions for the decompositions of the Gini coefficient and the concentration index in (6) and (7). As demonstrated by Wagstaff et al. (2003), the "homogeneous parameters contributions" depend on the product of the elasticity of health with respect to each explanatory variable and the concentration index for each variable, which in turn depend on the scaled covariance between the variable and the relative rank of health or income. The "homogeneous parameters" component treats the elasticity as homogeneous across individuals as it is evaluated at the OLS estimation of $\beta_k$ and the means of y and $x_k$. In Table 3 the "homogeneous parameters contributions" account for 3.25% of the observed Gini coefficient and 86.46% of the concentration index of health on income. The stark contrast between the ability of the regression model to "explain" total inequality and income-related inequality in the GHQ' score reflects the fact that income-related inequality, measured by C, accounts for only 11 per cent of total inequality, measured by G.

*Table 3. Contribution of explanatory variables to inequality indices*

| GINI INDEX | Hom. contrib. | as % | Het. contrib | as % | Total contrib. | as % |
|---|---|---|---|---|---|---|
| Log(income) | 0.00065 | 0.57% | -0.00418 | -3.67% | -0.00353 | -3.10% |
| Never married | 0.00024 | 0.21% | -0.00039 | -0.34% | -0.00015 | -0.13% |
| Divced., septed., or widwd. | 0.00085 | 0.74% | 0.00005 | 0.04% | 0.00089 | 0.79% |
| Children | 0.00002 | 0.02% | 0.00012 | 0.10% | 0.00014 | 0.12% |
| Professional or managerial | 0.00004 | 0.04% | 0.00009 | 0.08% | 0.00014 | 0.12% |
| Skilled | 0.00051 | 0.45% | -0.00034 | -0.30% | 0.00017 | 0.15% |
| Other SC | 0.00052 | 0.46% | 0.00006 | 0.06% | 0.00058 | 0.51% |
| Unskilled | 0.00003 | 0.03% | -0.00004 | -0.03% | -0.00001 | -0.01% |
| Unemployed | 0.00012 | 0.10% | 0.00012 | 0.11% | 0.00024 | 0.21% |
| Student or retired | 0.00026 | 0.23% | -0.00017 | -0.15% | 0.00009 | 0.08% |
| Carer | 0.00006 | 0.06% | -0.00005 | -0.05% | 0.00001 | 0.01% |
| Degree | 0.00000 | 0.00% | 0.00001 | 0.01% | 0.00001 | 0.01% |
| HND/HNCT | 0.00004 | 0.03% | -0.00002 | -0.02% | 0.00002 | 0.02% |
| O/CSE | 0.00003 | 0.03% | -0.00001 | 0.00% | 0.00003 | 0.02% |
| No qualification | 0.00035 | 0.30% | -0.00025 | -0.22% | 0.00010 | 0.09% |
| Non-white | 0.00000 | 0.00% | 0.00000 | 0.00% | 0.00000 | 0.00% |
| Year 1999 | | | -0.00017 | -0.15% | -0.00017 | -0.15% |
| Intercept | | | 0.00862 | 7.58% | 0.00862 | 7.58% |
| Total | 0.00370 | 3.25% | 0.00346 | 3.04% | 0.00716 | 6.29% |
| Residual | | | | | 0.10665 | 93.71% |
| Actual | | | | | 0.11381 | 100.00% |

| CONC. INDEX | Hom. contrib. | as % | Het. contrib | as % | Total contrib. | as % |
|---|---|---|---|---|---|---|
| Log(income) | 0.00677 | 51.87% | 0.00361 | 27.62% | 0.01038 | 79.49% |
| Never married | -0.00006 | -0.44% | -0.00005 | -0.40% | -0.00011 | -0.84% |
| Divced., septed., or widwd. | 0.00236 | 18.07% | -0.00075 | -5.74% | 0.00161 | 12.34% |
| Children | 0.00015 | 1.17% | -0.00018 | -1.41% | -0.00003 | -0.24% |
| Professional or managerial | 0.00049 | 3.73% | -0.00060 | -4.61% | -0.00012 | -0.89% |
| Skilled | 0.00054 | 4.17% | -0.00025 | -1.95% | 0.00029 | 2.22% |
| Other SC | 0.00266 | 20.41% | 0.00149 | 11.45% | 0.00416 | 31.86% |
| Unskilled | -0.00009 | -0.71% | 0.00004 | 0.29% | -0.00006 | -0.42% |
| Unemployed | 0.00037 | 2.82% | 0.00031 | 2.38% | 0.00068 | 5.20% |
| Student or retired | -0.00351 | -26.87% | 0.00090 | 6.90% | -0.00261 | -19.98% |
| Carer | 0.00016 | 1.20% | 0.00010 | 0.77% | 0.00026 | 1.97% |
| Degree | -0.00004 | -0.27% | -0.00005 | -0.39% | -0.00009 | -0.66% |
| HND/HNCT | 0.00013 | 1.01% | 0.00002 | 0.17% | 0.00015 | 1.18% |
| O/CSE | 0.00002 | 0.12% | 0.00030 | 2.29% | 0.00031 | 2.41% |
| No qualification | 0.00132 | 10.09% | 0.00018 | 1.37% | 0.00150 | 11.46% |
| Non-white | 0.00001 | 0.10% | 0.00001 | 0.08% | 0.00002 | 0.18% |
| Year 1999 | | | 0.00054 | 4.17% | 0.00054 | 4.17% |
| Intercept | | | -0.00490 | -37.53% | -0.00490 | -37.53% |
| Total | 0.01129 | 86.46% | 0.00071 | 5.47% | 0.01200 | 91.93% |
| Residual | | | | | 0.00105 | 8.07% |
| Actual | | | | | 0.01305 | 100.00% |

The "heterogeneous parameters contributions" component of the decomposition shows how heterogeneity in both the intercept and the slope coefficients modifies the contribution to inequality of the regressors. The figures in Table 3 show that heterogeneity in responses increases the Gini coefficient by 3.04% and the concentration index by 5.47%. The impact of income is of particular interest. Heterogeneity in the income effect results in the Gini coefficient being smaller than what it would be if all demographic groups had the average income slope coefficient all else held equal. On the contrary, heterogeneity in the income effect results in the concentration index being greater than what it would be if all demographic groups had the average income slope coefficient all else held equal. In particular, the results in the table suggest that the concentration index would be 27.62% smaller if the marginal effect of income was homogeneous in the population. This empirical finding suggests that the demographic groups, as defined by birth cohort and gender, where income has a greater marginal impact on psychological well-being as measured by the GHQ' score enjoy a higher than average level of income.

An important contributor to income related health inequality is the incidence of divorce, separation or widowhood. Since this event is associated to a lower GHQ' score and there is pro-poor inequality in its distribution, it accounts for a 12.34% of the actual concentration index. However, this contribution would be larger if there was homogeneity in responses. The –5.74% figure in the "heterogeneous contributions" column shows that those individuals for whom the (negative) association between this event and psychological well-being is stronger tend to enjoy a higher than average level of income. Thus the net contribution to income related health inequality is lower than what the 18.07% figure in the "homogeneous contributions" column shows. Another important contributor to the concentration index is the "other social class" indicator. Because belonging to this category is negatively associated to psychological well-being and because there is pro-poor inequality in the distribution of this demographic, this variable accounts for a 31.86% of income related inequality in psychological well-being. A substantial part of this contribution, 11.45%, is due to the heterogeneity of responses, since those individuals for whom the impact is greater in absolute value tend to have a lower than average level of income. A similar pattern is found in the contribution of unemployment, where heterogeneity of responses accounts for nearly one half of the overall 5.2% contribution to the concentration index. The net contribution of the student or retired status is made up of two conflicting components. Its "homogeneous contribution", -26.87% of the overall CI, reflects that this demographic is on average associated to a greater GHQ' score, but there is pro-poor inequality in its distribution of this demographic. However, those for whom the effect is greater than average (men and women in the 1935-1944 cohort) tend to have a greater than average level of income and vice versa. Thus the "heterogeneous contribution" of this explanatory variable is positive.

It is noteworthy that the heterogeneity of the intercepts accounts for a large share, -37.53%, of the concentration index. Young cohorts (those born after 1954) would report, if there were no differences in neither covariates nor responses, a greater GHQ' score than older cohorts (except males in the oldest cohort). But these individuals tend to enjoy a lower than average level of income so this pattern of heterogeneity accounts for a 37.53% reduction in the concentration index.

## 4. Summary and conclusion

In this paper we have shown how the regression based methods for the decomposition of health inequality developed by Wagstaff et al. (2003) can be extended to incorporate individual heterogeneity in the responses of health to the explanatory variables. We have illustrated our proposal with an application to the measure of psychological well-being in the BHPS (waves 1991-1999).

Allowing for heterogeneity in the responses of individuals' health to differences in socio-economic characteristics is one way in which the specification of regression functions may be improved. This improved fit means that a higher proportion of both overall and of income-related inequality can be attributed to the "explained" component in regression-based decompositions of Gini and concentration indices. The new decomposition method presented in this paper extends the work of Wagstaff et al. (2003) to show that the contribution of regressors can be further decomposed; the decomposition of the overall level of inequality distinguishes between the contribution of the level of inequality in the regressors, evaluated at the pooled level of the coefficients, and the contribution of heterogeneity in the coefficients around the pooled values. The interpretation of the new decomposition formula has a clear parallel with the use of Oaxaca (1973) decompositions of changes in inequalities over time or across countries.

The empirical results for the GHQ measure of psychological well-being suggest that there is an important degree of heterogeneity in the association of health to explanatory variables across birth cohorts and genders which, in turn, accounts for a substantial percentage of the inequality in observed health. A particularly interesting finding is that heterogeneity in the marginal effect of income accounts for 27.62% of the observed concentration index of the GHQ' psychological well-being score on income. The evidence presented here cannot provide a causal interpretation for this relationship but it provides a boundary on the reduction in income-related health inequality which might be achieved through policies aimed at eliminating differences in the ability with which individuals are able to transform income in health.

## References


Bowling, A., 1991, *Measuring health. A review of quality of life measurement scales.* OUP, Milton Keynes.

Goldberg, D., Williams, P., 1988, *A user's guide to the General Health Questionnaire.* Nfer-Nelson, Windsor.

Hauck, K., Rice, N., 2003, Health mobility in the UK: a longitudinal analysis of psychological well-being. *University of York,* mimeo.

Heckman, J., 2001. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture, *Journal of Political Economy*, 109, 673-748.

Kakwani, N.C., 1980. *Income Inequality and Poverty. Methods of Estimation and Policy Implications.* Oxford University Press, Oxford.

Koenker, R., and Bassett, G., 1982. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43-61.

Lambert, P., 1993. The distribution and redistribution of income. A mathematical analysis. 2nd Edition. Manchester University Press. Manchester.

Lerman, R.I., Yitzhaki, S., 1989. Improving the accuracy of estimates of Gini coefficients. *Journal of Econometrics* 42, 43-47.

Likert, R., 1952, A technique for the development of attitude scales. *Educational and Psychological Measurement* 12, 313-315.

Oaxaca, R.L., 1973. Male-female wage differentials in urban labor markets. *International Economic Review* 14, 693-709.

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

Taylor, M.F. (ed.) with Brice, J., Buck, N., and Prentice, E., 1998, *British Household Panel Survey User Manual,* University of Essex, Colchester.

Van Doorslaer, E., Wagstaff, A., Bleichrodt, H., *et al*, 1997. Income-related inequalities in health: some international comparisons. *Journal of Health Economics* 16, 93-112.

Wagstaff, A., van Doorslaer, E., 1994. Measuring inequalities in health in the presence of multiple-category morbidity indicators. *Health Economics 3*, 281-291.

Wagstaff, A., van Doorslaer, E., Paci, P., 1989. Equity in the finance and delivery of health care: some tentative cross-country comparisons. *Oxford Review of Economic Policy* 5, 89-112.

Wagstaff, A., Paci, P., van Doorslaer, E., 1991. On the measurement of inequalities in health. *Social Science and Medicine 33*, 545-557.

Wagstaff, A, van Doorslaer, E., Watanabe, N., 2003. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam, *Journal of Econometrics* 112, 207-223..

Weich, S., Lewis, G., Jenkins, S.P., 2001, Income inequality and the prevalence of common mental disorders in Britain. *British Journal of Psychiatry* 178, 222-227.

Wildman, J., 2003. Income related inequalities in mental health in Great Britain: analysing the causes of health inequality over time. *Journal of Health Economics* 22, 295-312

*Table A.1 Full regression estimates (robust standard errors in parentheses)*

| | Hom. model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | Hom. Test F(13,20258) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Log(income) | 0.460 | -0.016 | -0.006 | 0.308 | 0.337 | 0.486 | 0.423 | 0.873 | 0.614 | 0.669 | 0.826 | 0.599 | 0.585 | 0.325 | 0.431 | 2.15 |
| | (0.05) | (0.18) | (0.15) | (0.13) | (0.13) | (0.17) | (0.17) | (0.18) | (0.19) | (0.22) | (0.23) | (0.25) | (0.21) | (0.23) | (0.25) | |
| Never married | 0.421 | 0.127 | -0.380 | 0.202 | -0.111 | -0.359 | -0.537 | -0.403 | 0.398 | -0.282 | 1.739 | -0.147 | -0.493 | 0.543 | -0.639 | 1.51 |
| | (0.08) | (0.28) | (0.31) | (0.17) | (0.16) | (0.39) | (0.33) | (0.47) | (0.55) | (0.72) | (0.63) | (0.61) | (0.60) | (0.41) | (0.58) | |
| Divced., septed., or widwd. | -1.129 | 2.345 | -23.384 | -1.473 | -2.482 | -1.442 | -1.661 | -0.953 | -0.842 | -1.759 | -0.216 | -0.633 | -0.255 | -0.361 | -0.339 | 86.39 |
| | (0.10) | (1.56) | (0.67) | (0.43) | (0.50) | (0.32) | (0.44) | (0.34) | (0.43) | (0.40) | (0.53) | (0.33) | (0.38) | (0.27) | (0.31) | |
| Children | -0.050 | 0.068 | 0.332 | 0.001 | -0.082 | -0.022 | -0.048 | 0.294 | 0.036 | 0.115 | -0.021 | -0.052 | -0.353 | 0.294 | 0.332 | 1.41 |
| | (0.03) | (0.12) | (0.09) | (0.09) | (0.10) | (0.10) | (0.09) | (0.12) | (0.12) | (0.36) | (0.27) | (0.59) | (0.44) | (0.63) | (0.51) | |
| Professional or managerial | 0.136 | 0.547 | -0.279 | 0.031 | 0.145 | -0.250 | 0.006 | -0.365 | 0.380 | -0.797 | -0.118 | -0.317 | -1.154 | -0.240 | 0.402 | 1.12 |
| | (0.09) | (0.44) | (0.42) | (0.20) | (0.21) | (0.24) | (0.29) | (0.27) | (0.39) | (0.36) | (0.38) | (0.70) | (0.60) | (2.83) | (0.69) | |
| Skilled | 0.967 | 0.274 | -0.318 | -0.389 | 0.114 | 0.179 | 0.374 | 0.051 | 0.969 | 0.720 | 0.636 | 1.465 | 0.443 | 1.738 | -0.363 | 1.37 |
| | (0.09) | (0.47) | (0.36) | (0.28) | (0.22) | (0.32) | (0.30) | (0.37) | (0.40) | (0.49) | (0.40) | (0.87) | (0.48) | (1.11) | (0.90) | |
| Other SC | -0.530 | -0.293 | -0.337 | -0.536 | -0.235 | -0.190 | -0.705 | -1.367 | -1.588 | -1.263 | -1.552 | -0.257 | -1.819 | 0.023 | -3.327 | 3.05 |
| | (0.10) | (0.26) | (0.28) | (0.24) | (0.24) | (0.25) | (0.38) | (0.30) | (0.53) | (0.39) | (0.45) | (0.57) | (0.53) | (1.10) | (0.89) | |
| Unskilled | 0.550 | -0.917 | -0.587 | 0.282 | -0.186 | 0.741 | -0.024 | 0.026 | 1.081 | 0.191 | 1.030 | 1.836 | 0.679 | 3.513 | -1.273 | 2.10 |
| | (0.16) | (0.63) | (0.40) | (0.49) | (0.38) | (0.46) | (0.52) | (0.46) | (0.53) | (0.50) | (0.87) | (0.64) | (0.75) | (1.42) | (1.15) | |
| Unemployed | -0.602 | -1.748 | -1.900 | -1.187 | -1.505 | -1.264 | -1.312 | -1.298 | -0.285 | -0.794 | 1.084 | 0.449 | 0.891 | 1.523 | 5.447 | 6.44 |
| | (0.14) | (0.58) | (0.45) | (0.34) | (0.28) | (0.56) | (0.41) | (0.57) | (0.53) | (0.64) | (0.52) | (0.91) | (0.73) | (2.12) | (0.96) | |
| Student or retired | 1.225 | -0.015 | -0.094 | 0.214 | 0.318 | 0.877 | -0.106 | 0.669 | -0.639 | 1.574 | 2.025 | 0.375 | 1.330 | 1.061 | 1.290 | 3.49 |
| | (0.10) | (0.28) | (0.26) | (0.28) | (0.27) | (0.70) | (0.82) | (0.57) | (0.78) | (0.38) | (0.40) | (0.40) | (0.33) | (0.37) | (0.79) | |
| Carer | -0.177 | -1.687 | 0.048 | -0.284 | 0.525 | -0.395 | -0.855 | 0.471 | 0.644 | 0.880 | 3.921 | 0.178 | -1.277 | 0.888 | -4.493 | 5.06 |
| | (0.12) | (0.68) | (1.05) | (0.27) | (0.34) | (0.27) | (1.01) | (0.37) | (1.22) | (0.39) | (0.86) | (0.50) | (1.69) | (0.38) | (1.01) | |
| Degree | -0.020 | 1.223 | -0.928 | 0.234 | -0.137 | 0.255 | -0.209 | -0.065 | -0.680 | -0.349 | 0.955 | -0.461 | -1.002 | 1.550 | -0.770 | 1.82 |
| | (0.12) | (0.42) | (0.53) | (0.25) | (0.23) | (0.34) | (0.33) | (0.47) | (0.42) | (0.80) | (0.56) | (0.85) | (0.76) | (1.24) | (0.85) | |
| HND/HNCT | 0.197 | 0.510 | -0.840 | 0.087 | -0.248 | -0.004 | 0.273 | 0.385 | 0.090 | 0.871 | 1.364 | -0.083 | -1.257 | 1.509 | 0.889 | 1.54 |
| | (0.14) | (0.66) | (0.55) | (0.36) | (0.34) | (0.46) | (0.42) | (0.50) | (0.52) | (0.80) | (0.64) | (1.06) | (0.53) | (0.88) | (0.77) | |
| O/CSE | 0.105 | 0.201 | 0.011 | -0.274 | -0.120 | -0.045 | 0.559 | 0.062 | 0.428 | 1.517 | 0.989 | -0.709 | -1.541 | 1.638 | -0.981 | 2.69 |
| | (0.09) | (0.30) | (0.27) | (0.21) | (0.21) | (0.28) | (0.26) | (0.37) | (0.36) | (0.61) | (0.44) | (0.62) | (0.55) | (0.74) | (0.71) | |
| No qualification | -0.379 | 1.562 | 0.000 | -1.381 | -0.294 | -1.178 | 0.360 | -0.089 | 0.025 | 0.366 | 0.683 | -1.197 | -1.576 | 0.549 | -0.548 | 4.06 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (0.10) | (0.48) | (0.48) | (0.41) | (0.32) | (0.36) | (0.32) | (0.37) | (0.38) | (0.59) | (0.42) | (0.56) | (0.44) | (0.67) | (0.60) | |
| Non-white | -0.326 | -0.809 | -0.797 | -0.061 | -0.269 | -0.377 | -0.348 | 0.900 | 0.443 | -1.515 | -1.581 | 0.584 | -0.978 | -0.549 | -2.751 | 0.84 |
| | (0.19) | (0.71) | (0.63) | (0.42) | (0.36) | (0.48) | (0.55) | (0.64) | (0.52) | (1.15) | (1.20) | (1.01) | (1.01) | (1.39) | (3.26) | |
| 1993-1994 | -0.226 | -0.199 | -0.144 | -0.133 | -0.142 | -0.238 | -0.053 | -0.322 | -0.269 | -0.490 | -0.127 | -0.501 | -0.138 | -0.567 | -0.003 | 1.10 |
| | (0.05) | (0.49) | (0.48) | (0.16) | (0.14) | (0.15) | (0.14) | (0.15) | (0.16) | (0.18) | (0.17) | (0.17) | (0.16) | (0.16) | (0.18) | |
| 1995-1998 | -0.409 | -1.174 | -0.628 | -0.552 | -0.581 | -0.599 | -0.648 | -0.560 | -0.069 | -0.410 | -0.082 | -0.634 | 0.185 | -0.500 | 0.039 | 2.70 |
| | (0.05) | (0.46) | (0.49) | (0.16) | (0.14) | (0.16) | (0.15) | (0.16) | (0.17) | (0.20) | (0.18) | (0.20) | (0.18) | (0.17) | (0.19) | |
| 1999 | -0.243 | -0.846 | -0.589 | -0.296 | -0.410 | -0.320 | -0.855 | 0.089 | 0.119 | -0.108 | -0.040 | -0.469 | 0.104 | -0.679 | -0.075 | 2.11 |
| | (0.06) | (0.48) | (0.51) | (0.19) | (0.17) | (0.18) | (0.18) | (0.21) | (0.21) | (0.24) | (0.24) | (0.25) | (0.23) | (0.23) | (0.26) | |
| Constant | 20.981 | 25.993 | 28.256 | 22.698 | 23.867 | 20.444 | 21.975 | 16.476 | 19.225 | 18.167 | 16.850 | 21.001 | 22.729 | 20.158 | 24.193 | 3.58 |
| | (0.49) | (1.73) | (1.53) | (1.29) | (1.24) | (1.64) | (1.70) | (1.80) | (1.93) | (2.21) | (2.27) | (2.50) | (2.07) | (2.43) | (2.48) | |
| Reset. F(3,N-23) | 27.40 | 0.68 | 0.35 | 5.14 | 0.31 | 10.18 | 0.36 | 19.00 | 6.82 | 5.42 | 2.38 | 2.89 | 3.17 | 2.47 | 2.16 | |
| N | 88155 | 3321 | 2893 | 9642 | 8908 | 9425 | 8323 | 8357 | 7054 | 5901 | 5269 | 5026 | 4341 | 5897 | 3798 | |