Rethinking Inequality Decomposition: Comment

Frank A. Cowell* and Carlo V. Fiorio†

*London School of Economics

†University of Milan and Econpubblica

DARP 82
February 2006

The Toyota Centre
Suntory and Toyota International
Centres for Economics and Related
Disciplines
London School of Economics
Houghton Street
London WC2A 2A

(+44 020) 7955 6674

Distributional Analysis Research Programme

The Distributional Analysis Research Programme was established in 1993 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science. The programme is directed by Frank Cowell. The Discussion Paper series is available free of charge. To subscribe to the DARP paper series, or for further information on the work of the Programme, please contact our Research Secretary, Leila Alberici on:

Telephone: UK+20 7955 6674
Fax: UK+20 7955 6951
Email: l.alberici@lse.ac.uk
Web site: http://sticerd.lse.ac.uk/DARP

Rethinking Inequality Decomposition: Comment

Frank A. Cowell* and Carlo V. Fiorio†

6th December 2005

Decomposition analysis of inequality is important for understanding the main determinants of inequality and for policy analysis. The "traditional" approach to the subject was based purely on the analysis of the mathematical properties of inequality indices and is open to the criticism that the formal requirements for exact decomposition are perhaps too demanding for some practical applications. Recent applied work has reawakened interest in inequality decomposition by focusing on the use of regression-based approaches to avoid some of the restrictions of the traditional methods.

Morduch and Sicular (2002) have suggested a specific regression-based method for decomposing inequality. They claim that their method possesses three main advantages: (a) it yields an exact allocation of contributions to the identified variables, (b) it is general, in that it can be employed with a variety of inequality indices and decomposition rules, (c) it is associated with a simple procedure for deriving standard errors and confidence intervals for estimated components of inequality. In this comment we argue that these points are not

*London School of Economics and STICERD. Address: Houghton Street, London WC2A 2AE, UK. email: f.cowell@lse.ac.uk

†University of Milan and Econpubblica. Address: DEAS, via Conservatorio, 7. 20133 Milan, Italy. email: carlo.fiorio@unimi.it

convincing and, in some respects, wrong. We first summarize the approach taken by Morduch and Sicular (2002) to introduce the regression-based methodology for inequality decomposition; we then examine the three main claims in detail.

1 The Morduch and Sicular (2002) decomposition method

Morduch and Sicular (2002) focus on "natural decomposition rules" for inequality decomposition, which pertain to inequality indices that can be written as a weighted sum of incomes (Shorrocks, 1982):

$$I(\mathbf{y}) = \sum_{i=1}^{n} a_i(\mathbf{y}) y_i \tag{1}$$

where y_i is the income of person i, n is the number of persons in the population, $\mathbf{y} := (y_1, ..., y_n)$, and $a_i(\mathbf{y})$ is a weighting factor. If each y_i is the sum of component incomes y_i^k coming from K different sources (such as pension, employment income, transfers and so on)

$$y_i = \sum_{k=1}^{K} y_i^k$$

then the proportional contribution of source k to overall inequality can be defined as

$$s^k := \frac{\sum_{i=1}^{n} a_i(\mathbf{y}) y_i^k}{I(\mathbf{y})} \tag{2}$$

In the regression-based inequality decomposition proposed by Morduch and Sicular (2002), primary attention is given to the data generating process that led to a particular distribution of income, as in Oaxaca (1973) and Blinder

(1973). Assuming that the income generating process is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}$$

where $\mathbf{y}$ is an $n \times 1$ vector, $\mathbf{X}$ is an $n \times M$ matrix of individual and household characteristics (age, education, household size, residence, etc.), $\boldsymbol{\beta}$ is a $M \times 1$ vector of coefficients and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of residuals, a sample of observations $\{y_i,\, \mathbf{x}_i,\, i = 1, 2, ...n\}$ can be used to estimate the model. The vector $\boldsymbol{\beta}$ can be interpreted as the effect (or "price") of the independent variables on income. Using (3), per capita income of household i is then represented as:

$$y_i = \sum_{m=1}^{M} \widehat{\beta}_m x_i^m + \widehat{\epsilon}_i \tag{4}$$

where $\widehat{\beta}_m$ is the OLS coefficient estimate and $\widehat{\epsilon}_i$ is the OLS residual for household i. By analogy with (2), shares attributable to the characteristic $m = 1, ..., M$ take the form:

$$s^m := \widehat{\beta}_m \left(\frac{\sum_{i=1}^{n} a_i(\mathbf{y}) x_i^m}{I(\mathbf{y})} \right) \tag{5}$$

This decomposition might be applied to any inequality index that can be written as a weighted sum of incomes.

Let us now discuss the three main claims for this model.

2 Exactness

Morduch and Sicular (2002) state that "[this method] yields an exact allocation of contributions to the identified variables." In the traditional inequality decomposition literature a decomposition method is exact if it can express inequality as an additive function of incomes or income shares without residual. For instance, while all members of the class of Generalized Entropy indices are exactly decomposable, the Gini index is not: for arbitrary partitions of the population there will be a residual term that is difficult to interpret unambiguously (among many others, see Cowell, 1980; Shorrocks, 1980; Cowell and Kuga, 1981; Pyatt, 1976; Lambert and Aronson, 1993).

The method is an exact decomposition only of $\widehat{\mathbf{y}}$, not of $\mathbf{y}$, which is of primary interest. In fact, as $\mathbf{y} = \widehat{\boldsymbol{\beta}}\mathbf{x} + \widehat{\boldsymbol{\epsilon}}$, the $\widehat{\boldsymbol{\epsilon}}$ residual is always present, and the decomposition is clearly not exact. Moreover, as the application to Chinese data shows, in some cases the residual term accounts to a very large share of the inequality index: in two cases out of four (the Theil-T and the Alternative CV) the decomposition leaves about 90% of total inequality unexplained![1]

3 Generality

Morduch and Sicular (2002) consider that the method "is general in that it can be employed with different inequality indices and decomposition rules." Although it is true that one can run a regression with whatever random variable available, it is not true that any regression is a good model. In this case the dependent variable is a complex composite (per capita household income, where income comprises both labor and non-labor income) and it is estimated using a linear relationship of household characteristics. It is hardly surprising that the R^2 of this regression is lower than 45% and that all inequality decompositions present very large residuals.

This method also reveals substantial variability across indices. For instance, while the net effect[2] of an additional year of education is to reduce per

[1] See "Residual regression" line in Table 2, p.103.

[2] The net effect of education is given by the sum of the effects of the variables *average*

capita household income inequality by about 50% using the Theil-T index, it increases it using the Gini index. Morduch and Sicular (2002) motivate the difference in outcome between the two indices pointing out the fact that the decomposition of the Gini index does not fulfill the Corollary of the Uniform Addition Property (CUAP), which states that if a component of total income (say, income type k) is positive and equally-distributed its contribution to total income inequality, s^k, must be negative. However, it is not clear whether satisfaction of CUAP is a merit or not.

If it is a merit, and consequently the Theil-T should be preferred to the Gini index in this kind of empirical analysis, then it is difficult to be satisfied with a decomposition method, which leaves over 90% of inequality unexplained. Moreover, this property is not satisfied by the decomposition rule proposed by Shorrocks (1982), which is the unique one to be invariant to the choice of inequality measure. In fact, Shorrocks (1982) showed that the unique decomposition rule, invariant to the inequality index used, is:

$$s^k := \frac{cov(\mathbf{y}^k, \mathbf{y})}{var(\mathbf{y})} \tag{6}$$

where $\mathbf{y}^k = \{y_i^k, i = 1, ...n\}$. Factor contributions can be either positive or negative, depending on the factor providing a disequalizing or equalizing contribution. Using Shorrocks's rule (6) the contribution s^k is negative only if there is a negative correlation between total income and income component k, as it is often the case for taxes or transfers. The suggested decomposition rule is then just one more in a large class of possible decomposition rules. However, unless valid reasons are provided for constraining the set of potential decomposition rules, preferably to the point where inequality index can

education of adults and *education squared*. The same approach is taken also by Fields (2003).

be decomposed in only one way, "the inequality contribution assigned to any income source can vary arbitrarily, depending on the choice of decomposition rule. This turns the calculation of inequality contributions into a meaningless exercise..." (Shorrocks, 1983, p. 315).

Large residuals and highly variable results are most likely due to the inability of a single equation model to explain the complexity of per capita household income. A multiple equation income-generating model such as that of Bourguignon et al. (2001) or semi-parametric methods, such as that proposed by DiNardo et al. (1996), would be more successful although clearly more complex to implement.

4 Estimation

Thirdly Morduch and Sicular (2002) consider that the method "is associated with a simple procedure for deriving standard errors and confidence intervals for the estimated components of inequality." Since the decomposition (5) is linear in the parameters, they claim that the standard error of s^k, $\sigma(s^k)$, is easily estimated using the standard error of $\widehat{\beta}_k$, $\sigma(\widehat{\beta}_k)$, which comes as an output in regression analysis:

$$\sigma(s^m) = \sigma(\widehat{\beta}_m) \times \left(\frac{\sum a_i(\mathbf{y}) x_i^m}{I(\mathbf{y})} \right) \tag{7}$$

However, equation (7) would be correct if and only if $(\sum a_i(\mathbf{y}) x_i^m / I(\mathbf{y}))$ was not stochastic, which is clearly untrue (recall (3)). Hence, the correct standard error is:

$$\sigma(s^m) = \sigma \left(\widehat{\beta}_m \frac{\sum_{i=1}^n a_i(\mathbf{y}) x_i^m}{I(\mathbf{y})} \right) \tag{8}$$

which might be quite complicated to compute, as it requires the use of the bootstrap or non-trivial asymptotic distribution (see for example Cowell, 1989; Cowell and Flachaire, 2002).

References

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Bourguignon, F., Fournier, M., and Gurgand, M. (2001). Fast development with a stable income distribution: Taiwan, 1979-94. *Review of Income and Wealth*, 47(2):139–163.

Cowell, F. and Kuga, K. (1981). Additivity and the Entropy Concept: An Axiomatic Approach to Inequality Measurement. *Journal of Economic Theory*, 25:131–143.

Cowell, F. A. (1980). On the structure of additive inequality measures. *Review of Economic Studies*, 47:521–31.

Cowell, F. A. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics*, 42(1):27–41.

Cowell, F. A. and Flachaire, E. (2002). Sensitivity of inequality measures to extreme values. Distributional Analysis Discussion Paper, 60, STICERD, LSE.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor Market Insititutions and the Distribution of Wages, 1973-1992: a Semiparametric Approach. *Econometrica*, 64:1001–1044.

Fields, G. S. (2003). Accounting for income inequality and its change: a new method with application to distribution of earnings in the United States. In Polachek, S. W., editor, *Research in Labour Economics*, volume 22. Elsevier.

Lambert, P. J. and Aronson, J. R. (1993). Inequality Decomposition Analysis and the Gini Coefficient Revisited. *The Economic Journal*, 103(420):1221–1227.

Morduch, J. and Sicular, T. (2002). Rethinking inequality decomposition, with evidence from rural china. *The Economic Journal*, 112(476):93–106(14).

Oaxaca, R. R. A. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709.

Pyatt, G. (1976). On the Interpretation and Disaggregation of Gini Coefficients. *The Economic Journal*, 86(342):243–255.

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

Shorrocks, A. F. (1982). Inequality Decomposition by Factor Components. *Econometrica*, 50:193–211.

Shorrocks, A. F. (1983). The impact of income components on the distribution of family incomes. *The Quarterly Journal of Economics*, 98(2):311–326.