The Economic and Social Review, Vol. 16, No. 3, April 1985, pp. 225-249

Calculation of Gini and Theil Inequality Coefficients for Irish Household Incomes in 1973 and 1980

D.C. MURPHY*
Central Statistics Office, Dublin

Abstract: The exact values of Gini and Theil inequality coefficients calculated on the basis of discrete data are provided for direct, gross, disposable and final (i.e., after all taxes and State benefits) Irish household incomes in 1973 and 1980. These exact values are used to quantify the errors inherent in the grouped data approximations customarily used in most analyses. It is shown that quantile income classifications in general offer the most consistent approximation basis. A procedure is presented for selecting income ranges to improve the precision of grouped data approximations. Exact Gini and Theil coefficients for direct and final income are also provided for a wide range of different types of households in 1973 and 1980.

I INTRODUCTION

Cost analyses of income distribution in Ireland and abroad use Gini and Theil inequality measures approximated from published data classified by income ranges. Recent Irish studies by Nolan (1978 and 1981) and O'Connell (1982) based on the Household Budget Survey (HBS) and related income redistribution analyses published by the Central Statistics Office (CSO) used inequality measures approximated from decile classifications interpolated from the published standard income ranges. Murphy (1984) provided measures estimated from previously unavailable decile ranges.

These grouped data approximations provide *lower bounds* for the exact

*Appreciation is extended to a referee for comments and to Tom Keane and Damien Malone of the ADP Division in the CSO for undertaking the computer work required for the compilation of the exact Gini and Theil coefficients based on discrete observations. The views expressed are those of the author.

Gini and Theil coefficients. They are based on the assumption that all income observations in each range distinguished are equal to the published average income of the range. This means that they under-estimate the exact coefficient by ignoring the income inequality arising within income ranges.

Upper bounds for Gini and Theil inequality coefficients may also be derived by allowing for the maximum possible inequality within income ranges. This would occur if all income observations in a range (a_{i-1}, a_i) are located only at the end points a_{i-1} and a_i. Using the published average income $(\overline{x}_i)$ in the range the proportion of observations lying at each end point can be calculated, e.g., the proportion of observations at a_{i-1} is $(a_i - \overline{x}_i) \div (a_i - a_{i-1})$.

These maximum upper and lower bounds have been used, for example, by Theil (1967, p. 132) and Gastwirth (1972 for the Gini coefficient, 1975 for the Theil coefficient). Gastwirth also derived sharper bounds by making some allowances for the distribution of income within ranges. The derivation of these sharper bounds is somewhat laborious, particularly for the Theil coefficient, and is not generally practical unless the calculations are automated.

The main purpose of this paper is to quantify the estimation errors inherent in the use of these upper and lower bounds. Murphy (1984) presented limited results comparing exact Gini and Theil inequality values for gross household income with lower bound estimates derived using different numbers of income ranges for the 1973 and 1980 HBS. This paper presents much more detailed results for different income concepts and for the decomposition of inequality across different types of households.

The paper first looks in Section II at the Gini inequality coefficient and sets out the findings for the different income concepts. The Theil coefficient is similarly considered in Section III where results are also provided on the decomposition of inequality for five types of household on both the exact and approximate basis. Section IV develops a procedure for selecting income ranges to minimise the errors inherent in grouped data approximations. The findings are summarised in Section V.

II GINI INEQUALITY COEFFICIENT

Definition

The Gini coefficient is mathematically defined as:

$$G = \frac{\text{Mean difference of income observations}}{2\ (\text{Arithmetic mean})}$$

$$= \frac{1}{2n^2\overline{x}} \sum_{h=1}^{n} \sum_{i=1}^{n} |x_h - x_i| \qquad (1)$$

for a discrete set of income observations x_i ($i = 1$ to n) with mean $\overline{x}$. It is a measure of relative inequality, i.e., the ratio of a measure of dispersion to the average value $\overline{x}$. Since the mean absolute difference is used as the measure of dispersion the Gini coefficient is dependent on the spread of income values amongst themselves and not from a central value.

In the case of income classified into groups (i.e., n_j observations and average income $\overline{x}_j$ for j = 1 to g groups) the Gini coefficient may, as originally shown by Bhattacharya and Mahalanobis (1967), be disaggregated into three separate components, namely:

$$G = \underbrace{\frac{1}{2\overline{x}} \sum_{j=1}^{g} \sum_{K=1}^{g} P_j P_K |\overline{x}_j - \overline{x}_K|}_{\text{(between-group)}} + \underbrace{\sum_{j=1}^{g} P_j S_j G_j}_{\text{(within-group)}} + \underbrace{\text{3rd component}}_{\text{(interaction)}} \qquad (2)$$

where $P_j = n_j \div n$ is the proportion of income values and $S_j = n_j\overline{x}_j \div n\overline{x}$ is the proportion of total income in the jth group. The 3rd term depends on the overlapping of incomes distributions between the groups and, therefore, disappears when the income observations are grouped by *income ranges*. In this latter situation the overall Gini coefficient based on discrete incomes decomposes completely into *between-group* (G_{BG}) and *within-group* (G_{WG}) inequality components. If all the income observations in each range are equal the within-group inequality disappears (i.e., $G_{WG} = 0$) and the overall Gini coefficient reduces to the between group component (i.e., G_{BG}).

Upper and Lower Gini Bounds Derived from Grouped Data

When income observations are grouped into income ranges (j = 1 to R) the between-group inequality component G_{BG} corresponds, as already explained, to the general lower bound (i.e., G_L) of the exact Gini coefficient which assumes that all n_j income observations within each income range (a_{j-1}, a_j) equal the mean income ($\overline{x}_j$) of the range (i.e., $G_{WG} = 0$). This is, in fact, the standard grouped data estimate of the Gini coefficient.

In the case of the general upper bound (G_U) the Gini coefficient G_j(j = 1 to R) for the maximum possible inequality of income observations within each range, obtained by assuming that they are located at the end-point a_{j-1} and a_j, is calculated as:

$$G_j = \frac{\text{Mean difference of observations in jth range}}{2\overline{x}_j}$$

$$= \frac{1}{2\overline{x}_j} \cdot \frac{2(a_j - \overline{x}_j)}{(a_j - a_{j-1})} \cdot \frac{(\overline{x}_j - a_{j-1})}{(a_j - a_{j-1})} \cdot (a_j - a_{j-1})$$

$$= \frac{(a_j - \overline{x}_j)(\overline{x}_j - a_{j-1})}{\overline{x}_j(a_j - a_{j-1})} \qquad (2)$$

From (2) the general upper bound grouped data Gini estimate then becomes

$$G_U = G_{BG} + \text{Max } G_{WG} \tag{3}$$

where

$$\text{Max } G = \sum_{j=1}^{R} \frac{n_j}{n} \cdot \frac{n_j \bar{x}_j}{n\bar{x}} \cdot \frac{(a_j - \bar{x}_j)(\bar{x}_j - a_{j-1})}{\bar{x}_j(a_j - a_{j-1})}$$

$$= \frac{1}{\bar{x}} \sum_{j=1}^{R} \left(\frac{n_j}{n}\right)^2 \cdot \frac{(a_j - \bar{x}_j)(\bar{x}_j - a_{j-1})}{(a_j - a_{j-1})}$$

which is identical to the formula used by Gastwirth (1972). His *sharper* bounds for G_{WG} are applied for intervals (a_{j-1}, a_j) satisfying the conditions:

$$\bar{x}_j < \tfrac{1}{2}(a_{j-1} + a_j)$$

$$n_{j-1}(a_{j-1} - a_{j-2})^{-1} > n_j(a_j - a_{j-1})^{-1} > n_{j+1}(a_{j+1} - a_j)^{-1}$$

for decreasing income density.

Calculations Based on HBS data[1]

HBS household income data incorporates reweighting to allow for differential non-response. The allowance which must be made for this in Gini calculations is described in Appendix 1. Since the HBS is a sample survey the derived Gini coefficients are estimates subject to sampling variations. This should be small for estimates based on the full HBS samples. It is higher for sub-classifications where its level is directly related to sub-sample size.

Gini Coefficient Estimates

Table 1 compares the upper and lower Gini coefficient bounds for the State as a whole based on different numbers of income ranges with the exact Gini values calculated using the discrete HBS sample observations in 1973 and 1980. Four different household income concepts are distinguished, namely:

(1) *Direct income:* All regular income from employment, property, investments, pensions, etc., accruing to households;

(2) *Gross income:* Defined as direct income *plus* State transfer pay-

1. The published average household incomes and the reweighted sub-sample sizes in the HBS are rounded to 3 decimal places. Because of this the weighted average of sub-classification average incomes does not correspond exactly with the published combined average income. To ensure overall consistency the former was used in all calculations for this paper.

ments (i.e., unemployment pensions, old age pensions, children's allowances, etc.);

(3) *Disposable income:* Defined as gross income *less* direct taxes (i.e., income tax and employee share of social insurance contributions);

available directly from the HBS, and

(4) *Final income:* Calculated as disposable income *plus* estimated value of non-cash State benefits *less* indirect taxes (i.e., motor tax, licences, estimated VAT and duty content of household expenditure) in the HBS-based income redistribution analyses.

Zero direct incomes arising, for example, in the case of households completely dependent on State transfers as well as negative disposable (due to payment of back-tax) and final (because of expenditure with high tax content) incomes were incorporated in all calculations. Maximum weekly household incomes of £1,000 (1973) and £2,000 (1980) were arbitrarily taken in calculating the maximum inequality *within* the corresponding top income ranges.

The precision with which the upper and lower bounds bracket the exact Gini cofficient improves considerably as the number of income ranges on which they are based increases. For example, in the case of direct household income in 1980 the width of the bracketing interval for the standard bounds fell from 0.0131 using decile ranges, to 0.0038 with the 20-quantile ranges, and to 0.0007 when the maximum number of 60 ranges was used. The sharper bounds proposed by Gastwirth do not improve the precision of the bounds very substantially. The upper and lower bounds estimated on the basis of 60 income ranges for disposable income in 1980 do not bracket the exact Gini value. The very precise estimation in these circumstances was probably sensitive to the fact that the lowest income range used did not reflect the existence of some negative disposable income observations. These were, however, taken into account in calculating the exact Gini value which, therefore, registered a slightly higher level of inequality.

Table 1 shows that for all four income concepts the accuracy of the standard lower bound grouped data Gini estimate improves as the number of ranges used increases. The under-estimation is of the order of 1-2 per cent when decile or published (i.e., 11 or 13) ranges are used. This reduces to less than 1 per cent using 20-quantiles and to negligible proportions when based on 60 ranges. In absolute terms the degree of understatement when published income ranges are used varies considerably between income concepts and this distorts any comparisons based on them. This is not the case for either the decile and 20-quantile based estimates where the absolute level of understatement is relatively constant for each income concept.

Table 1: *Grouped and discrete HBS data estimates of Gini inequality coefficients for household income, 1973 and 1980*

Gini calculation basis	*1973* Direct income	*1973* Gross income	*1973* Disposable income	*1973* Final income	*1980* Direct income	*1980* Gross income	*1980* Disposable income	*1980* Final income
1. *Published Intervals* (11 in 1973, 13 in 1980)								
(a) Standard bounds								
– Lower	0.4480	0.3800	0.3699	0.3811	0.4685	0.3853	0.3625	0.3507
(error*)	(-0.0073)	(-0.0071)	(-0.0058)	(-0.0058)	(-0.0079)	(-0.0073)	(-0.0042)	(-0.0043)
– Upper	0.4583	0.3906	0.3778	0.3889	0.4812	0.3979	0.3691	0.3573
(b) Gastwirth's sharper bounds								
– Lower	0.4497	0.3820	0.3717	0.3826	0.4694	0.3859	0.3636	0.3519
– Upper	0.4576	0.3896	0.3770	0.3882	0.4808	0.3976	0.3685	0.3568
2. *Decile Intervals*								
(a) Standard bounds								
– Lower	0.4465	0.3790	0.3674	0.3782	0.4683	0.3855	0.3597	0.3479
(error*)	(-0.0088)	(-0.0081)	(-0.0083)	(-0.0087)	(-0.0081)	(-0.0071)	(-0.0070)	(-0.0071)
– Upper	0.4600	0.3912	0.3796	0.3907	0.4814	0.3972	0.3711	0.3591
(b) Gastwirth's sharper bounds								
– Lower	0.4483	0.3807	0.3689	0.3798	0.4702	0.3872	0.3613	0.3495
– Upper	0.4591	0.3904	0.3789	0.3899	0.4805	0.3965	0.3704	0.3584
3. *20 Quantile Intervals*								
(a) Standard bounds								
– Lower	0.4523	0.3842	0.3726	0.3836	0.4740	0.3905	0.3645	0.3528
(error*)	(-0.0030)	(-0.0029)	(-0.0031)	(-0.0033)	(-0.0024)	(-0.0021)	(-0.0022)	(-0.0022)
– Upper	0.4563	0.3877	0.3763	0.3874	0.4778	0.3940	0.3678	0.3561
(b) Gastwirth's sharper bounds								
– Lower	0.4528	0.3845	0.3729	0.3840	0.4744	0.3910	0.3649	0.3531
– Upper	0.4560	0.3876	0.3761	0.3873	0.4776	0.3937	0.3676	0.3560
4. *60 Intervals* (1980 only)	60 interval income classifications not distinguished in 1973 HBS							
(a) Standard bounds								
– Lower					0.4757	0.3923	0.3661	0.3547
(error*)					(-0.0007)	(-0.0003)	(-0.0006)	(-0.0003)
– Upper					0.4764	0.3928	0.3666	0.3552
(b) Gastwirth's sharper bounds								
– Lower					0.4758	0.3924	0.3663	0.3548
– Upper					0.4764	0.3927	0.3665	0.3552
5. EXACT VALUE based on discrete data	0.4553	0.3871	0.3757	0.3869	0.4764	0.3926	0.3667	0.3550

*Under-estimation = exact value minus lower bound.

Quantile Gini approximations, therefore, provide a consistent basis for estimating the absolute change in inequality levels resulting from the State tax/benefit redistribution process. This is a significant result.

Exact Gini coefficient values for direct and final household incomes in 1980 and 1973 are shown in Appendices 2 and 3, respectively, for households of different composition, life cycle, etc., calculated using the discrete sample observations (incorporating reweighting to allow for differential HBS response). The standard lower bound grouped data Gini approximations, based in all cases on decile classifications for the whole population (i.e., not published ranges), are also shown for comparison purposes. The degree to which these understate the exact values vary considerably, e.g., for direct income in 1980 the Gini coefficient for the State is under-estimated by 1.7 per cent, households with employee heads by 3.3 per cent and, exceptionally, households with three or more earners by 25.5 per cent. This is partly due to differences in sampling variability since the sub-samples differ widely in size. However, a considerable proportion of it also arises because of the variations in the under-estimation of the standard lower bound Gini value due to the use of the same income ranges for each sub-population distinguished. This error is highest for sub-populations with income distributions skewing extremely from that of the overall population (to which the decile ranges used relate), i.e., for predominantly low income (e.g., retired life cycle) and, more particularly, high income (e.g., three or more earners) household categories. These results illustrate the danger of using the same income ranges for estimating Gini coefficients for different groupings of households.

III THEIL INEQUALITY COEFFICIENT

Definition

The Theil inequality coefficient is defined as:

$$T = \frac{1}{n\overline{x}} \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{\overline{x}}\right) \qquad (4)$$

for a discrete set of income observations x_i(i = 1 to n) with mean $\overline{x}$. Common logarithms are used for calculation purposes in this paper. The popularity of the Theil coefficient for income inequality analysis purposes stems from the fact that when income observations are classified into any set of mutually exclusive groups (j = 1 to g) it can, as shown, for example, by Shorrocks (1980) and Bourguignan (1979), be naturally decomposed (i.e., without an interactive term as in the case of the Gini coefficient) into between-group (T_{BG}) and within-group (T_{WG}) components:

Table 2: *Grouped and discrete HBS data estimates of Theil inequality coefficients for household income, 1973 and 1980*

Theil calculation basis	*1973*				*1980*			
	Direct income	*Gross income*	*Disposable income*	*Final income*	*Direct income*	*Gross income*	*Disposable income*	*Final income*
1. *Published Intervals* (11 in 1973, 13 in 1980)								
(a) Standard bounds								
– Lower	0.1526	0.1033	0.0990	0.1060	0.1684	0.1053	0.0944	0.0897
(error*)	(-0.0112)	(-0.0081)	(-0.0074)	(-0.0054)	(-0.0109)	(-0.0076)	(-0.0049)	(-0.0026)
– Upper	0.2093	0.1545	0.1391	0.1446	0.2247	0.1556	0.1241	0.1183
(b) Gastwirth's sharper bounds								
– Lower	0.1531	0.1038	0.0994	0.1064	0.1686	0.1054	0.0947	0.0900
– Upper	0.2083	0.1536	0.1382	0.1438	0.2243	0.1553	0.1236	0.1178
2. *Decile Intervals*								
(a) Standard bounds								
– Lower	0.1533	0.1026	0.0965	0.1024	0.1703	0.1056	0.0920	0.0858
(error*)	(-0.0105)	(-0.0088)	(-0.0099)	(-0.0090)	(-0.0090)	(-0.0073)	(-0.0073)	(-0.0065)
– Upper	0.2211	0.1613	0.1611	0.1690	0.2218	0.1493	0.1404	0.1340
(b) Gastwirth's sharper bounds								
– Lower	0.1538	0.1029	0.0968	0.1028	0.1708	0.1060	0.0924	0.0861
– Upper	0.2202	0.1605	0.1604	0.1682	0.2207	0.1485	0.1397	0.1333
3. *20 Quantile Intervals*								
(a) Standard bounds								
– Lower	0.1578	0.1060	0.1002	0.1071	0.1747	0.1088	0.0952	0.0896
(error*)	(-0.0060)	(-0.0054)	(-0.0062)	(-0.0043)	(-0.0046)	(-0.0041)	(-0.0041)	(-0.0027)
– Upper	0.1915	0.1354	0.1335	0.1427	0.1989	0.1300	0.1196	0.1140
(b) Gastwirth's sharper bounds								
– Lower	0.1580	0.1061	0.1003	0.1072	0.1749	0.1089	0.0954	0.0897
– Upper	0.1911	0.1352	0.1333	0.1425	0.1986	0.1297	0.1193	0.1138
4. *60 Intervals* (1980 only)								
(a) Standard bounds								
– Lower					0.1774	0.1122	0.0989	0.0920
(error*)	60 interval income classifications not distinguished in 1973 HBS				(-0.0019)	(-0.0007)	(-0.0004)	(-0.0003)
– Upper					0.1809	0.1139	0.1002	0.0932
(b) Gastwirth's sharper bounds								
– Lower					0.1774	0.1122	0.0990	0.0920
– Upper					0.1809	0.1139	0.1001	0.0931
5. EXACT VALUE based on discrete data	0.1638	0.1114	0.1064	0.1114	0.1793	0.1129	0.0993	0.0923

*Under-estimation = exact value minus lower bound.

Quantile Gini approximations, therefore, provide a consistent basis for estimating the absolute change in inequality levels resulting from the State tax/benefit redistribution process. This is a significant result.

Exact Gini coefficient values for direct and final household incomes in 1980 and 1973 are shown in Appendices 2 and 3, respectively, for households of different composition, life cycle, etc., calculated using the discrete sample observations (incorporating reweighting to allow for differential HBS response). The standard lower bound grouped data Gini approximations, based in all cases on decile classifications for the whole population (i.e., not published ranges), are also shown for comparison purposes. The degree to which these understate the exact values vary considerably, e.g., for direct income in 1980 the Gini coefficient for the State is under-estimated by 1.7 per cent, households with employee heads by 3.3 per cent and, exceptionally, households with three or more earners by 25.5 per cent. This is partly due to differences in sampling variability since the sub-samples differ widely in size. However, a considerable proportion of it also arises because of the variations in the under-estimation of the standard lower bound Gini value due to the use of the same income ranges for each sub-population distinguished. This error is highest for sub-populations with income distributions skewing extremely from that of the overall population (to which the decile ranges used relate), i.e., for predominantly low income (e.g., retired life cycle) and, more particularly, high income (e.g., three or more earners) household categories. These results illustrate the danger of using the same income ranges for estimating Gini coefficients for different groupings of households.

III THEIL INEQUALITY COEFFICIENT

Definition

The Theil inequality coefficient is defined as:

$$T = \frac{1}{n\overline{x}} \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{\overline{x}}\right) \qquad (4)$$

for a discrete set of income observations x_i(i = 1 to n) with mean $\overline{x}$. Common logarithms are used for calculation purposes in this paper. The popularity of the Theil coefficient for income inequality analysis purposes stems from the fact that when income observations are classified into any set of mutually exclusive groups (j = 1 to g) it can, as shown, for example, by Shorrocks (1980) and Bourguignan (1979), be naturally decomposed (i.e., without an interactive term as in the case of the Gini coefficient) into between-group (T_{BG}) and within-group (T_{WG}) components:

$$T = \sum_{j=1}^{g} \frac{n_j \overline{x}_j}{n\overline{x}} \log(\frac{\overline{x}_j}{\overline{x}}) + \sum_{j=1}^{g} \frac{n_j \overline{x}_j}{n\overline{x}} T_j = T_{BG} + T_{WG} \tag{5}$$

where T_j (j = 1 to g) is the Theil inequality coefficient for the income observations in the jth group:

$$T_j = \frac{1}{n_j \overline{x}_j} \sum_{i=1}^{n_j} x_{ij} \log(\frac{x_{ij}}{\overline{x}_j})$$

Again, if all income observations with each group are equal, the within-group inequality disappears and the overall Theil coefficient reduces to the between-group component T_{BG}.

Upper and Lower Theil Bounds Derived from Grouped Data

As in the case of Gini, when income observations are grouped by income ranges (j = 1 to R), the value of the between-group Theil inequality component (T_{BG}) corresponds the general lower Theil coefficient (i.e., T_L) estimate obtained by assuming that income observations in each income range equal the mean income of the range. This is again the standard grouped data estimate of the Theil coefficient. It understates the exact value by ignoring the income inequality arising *within* the income ranges distinguished (i.e., T_{WG}) and is directly equivalent to:

$$T_L = \sum_{j=1}^{R} \frac{n_j}{n\overline{x}} N - \log \overline{x} \tag{6}$$

where $N = \overline{x}_j \log \overline{x}_j$

which is the formulation given by Gastwirth (1975).

The general upper bound T_U for Theil coefficient is again obtained by allowing for the maximum possible within-group inequality which arises when the observations within each income range (a_{j-1}, a_j) are located at either end point such that their average value equals the mean income $\overline{x}_j$. On incorporating the Theil coefficient for these observations:

$$T_j = \frac{1}{n_j \overline{x}_j} \left[n_j \frac{a_j - \overline{x}_j}{a_j - a_{j-1}} a_{j-1} \log(\frac{a_{j-1}}{\overline{x}_j}) + n_j (\frac{\overline{x}_j - a_{j-1}}{a_j - a_{j-1}}) a_j \log(\frac{a_j}{\overline{x}_j}) \right]$$

in T_{WG} formula (5) readily reduces to:

$$T_U = \sum_{j=1}^{R} \frac{n_j}{n\overline{x}} M - \log \overline{x} \tag{7}$$

where

$$M = \left(\frac{a_j - \overline{x}_j}{a_j - a_{j-1}}\right)a_{j-1}\log a_{j-1} + \left(\frac{x_j - a_{j-1}}{a_j - a_{j-1}}\right) a_j \log a_j$$

which is the formulation given by Gastwirth (1975).

The sharper lower and upper bounds proposed by Gastwirth are again applied for T_{WG} in intervals with decreasing income density (tested for as before) with adjustment for the use of common logarithms instead of natural logarithms (as used by Gastwirth).

Calculation of Theil Coefficient using HBS Data

Because of the use of logarithms (base 10) in the formulation of the Theil coefficient £0.01 (i.e., one penny) was taken in place of any zero or negative incomes in calculations using discrete observations. This introduced a slight inconsistency since the grouped-data Theil estimation used the average of actual discrete observations in the lowest income range distinguished and, therefore, reflected the occurrence of any negative values. There were only a few instances[2] where this average income was negative; taking account of the substituted £0.01 amounts, an appropriate positive value was approximated for Theil estimation purposes in these cases. Maximum weekly household incomes of £1,000 (1973) and £2,000 (1980) were again used for calculating the maximum inequality *within* the corresponding top income ranges.

The results for the Theil coefficient given in Table 2 exhibit the same general patterns as already discussed for the Gini coefficient. One major difference, however, is the far greater sensitivity of the Theil coefficient to income variations within ranges. As a result of this the standard lower bound grouped data Theil estimate is far less precise than the corresponding Gini values. The under-estimation is generally of the order of 3-9 per cent for all four income concepts when the estimation is based on the decile or published income ranges. It only reduces to 3-6 per cent when the 20-quantile classification is used and is, of course, negligible when based on 60 ranges. The sensitivity of the Theil coefficient to within-range income variations is also exemplified by the exceedingly high upper bounds even when 20-quantile intervals are used. The sharper Gastwirth bounds improve the precision of the bracketing intervals only marginally.

The understatement of the standard lower-bound Theil approximation based on published ranges is significantly greater in magnitude and varies more considerably between the four income concepts than the corresponding

2. 1980 final income with 60 intervals distinguished and the 1973 overall decile classifications of final income for two life cycle (i.e., early school and pre-adolescent) and three household composition (i.e., 2 adults with 1 child, 2 and 4+ children) categories.

Table 2: *Grouped and discrete HBS data estimates of Theil inequality coefficients for household income, 1973 and 1980*

Theil calculation basis	*1973*				*1980*			
	Direct income	*Gross income*	*Disposable income*	*Final income*	*Direct income*	*Gross income*	*Disposable income*	*Final income*
1. *Published Intervals* (11 in 1973, 13 in 1980)								
(a) Standard bounds								
– Lower	0.1526	0.1033	0.0990	0.1060	0.1684	0.1053	0.0944	0.0897
(error*)	(-0.0112)	(-0.0081)	(-0.0074)	(-0.0054)	(-0.0109)	(-0.0076)	(-0.0049)	(-0.0026)
– Upper	0.2093	0.1545	0.1391	0.1446	0.2247	0.1556	0.1241	0.1183
(b) Gastwirth's sharper bounds								
– Lower	0.1531	0.1038	0.0994	0.1064	0.1686	0.1054	0.0947	0.0900
– Upper	0.2083	0.1536	0.1382	0.1438	0.2243	0.1553	0.1236	0.1178
2. *Decile Intervals*								
(a) Standard bounds								
– Lower	0.1533	0.1026	0.0965	0.1024	0.1703	0.1056	0.0920	0.0858
(error*)	(-0.0105)	(-0.0088)	(-0.0099)	(-0.0090)	(-0.0090)	(-0.0073)	(-0.0073)	(-0.0065)
– Upper	0.2211	0.1613	0.1611	0.1690	0.2218	0.1493	0.1404	0.1340
(b) Gastwirth's sharper bounds								
– Lower	0.1538	0.1029	0.0968	0.1028	0.1708	0.1060	0.0924	0.0861
– Upper	0.2202	0.1605	0.1604	0.1682	0.2207	0.1485	0.1397	0.1333
3. *20 Quantile Intervals*								
(a) Standard bounds								
– Lower	0.1578	0.1060	0.1002	0.1071	0.1747	0.1088	0.0952	0.0896
(error*)	(-0.0060)	(-0.0054)	(-0.0062)	(-0.0043)	(-0.0046)	(-0.0041)	(-0.0041)	(-0.0027)
– Upper	0.1915	0.1354	0.1335	0.1427	0.1989	0.1300	0.1196	0.1140
(b) Gastwirth's sharper bounds								
– Lower	0.1580	0.1061	0.1003	0.1072	0.1749	0.1089	0.0954	0.0897
– Upper	0.1911	0.1352	0.1333	0.1425	0.1986	0.1297	0.1193	0.1138
4. *60 Intervals* (1980 only)								
(a) Standard bounds								
– Lower	60 interval income classifications not distinguished in 1973 HBS				0.1774	0.1122	0.0989	0.0920
(error*)					(-0.0019)	(-0.0007)	(-0.0004)	(-0.0003)
– Upper					0.1809	0.1139	0.1002	0.0932
(b) Gastwirth's sharper bounds								
– Lower					0.1774	0.1122	0.0990	0.0920
– Upper					0.1809	0.1139	0.1001	0.0931
5. EXACT VALUE based on discrete data	0.1638	0.1114	0.1064	0.1114	0.1793	0.1129	0.0993	0.0923

*Under-estimation = exact value minus lower bound.

Gini estimates. This means that income inequality analyses are subject to even greater distortion when using these approximate Theil measures. However, the decile and 20-quantile estimates again provide a consistent basis for assessing absolute changes in inequality levels although there is a slight tendency for the degree of under-estimation for final income to be less than that for the other three income concepts.

Contribution of Sub-Populations to Overall Income Inequality

Exact Theil direct and final income inequality coefficients for different types of households in 1973 and 1980 are presented in Appendices 2 and 3, respectively. Standard lower bound Theil estimates based on overall State decile classifications in each instance are also shown for comparison purposes. The varying precision of these bounds is, as was the case for the equivalent Gini particulars, due to under-estimation and sampling variability. The former was because of the use of the same overall decile ranges in all instances and the latter because of different sub-sample sizes.

The Theil coefficients for different groupings of households may be combined using formula (5) to assess how much of the overall income inequality in the community is due to differentials in income levels *between* the different groups and how much arises because of inequality existing *within* the groups. The results based on the standard lower-bound decile Theil approximations have already been provided by Murphy (1984).[3]

These results are summarised in Table 3 and compared with the exact results for the Theil coefficients calculated on the basis of discrete observations. This type of presentation is not possible on the Gini basis because of the interactive inequality component resulting from the overlap of income distributions between groups.

The between-group inequality component, which reflects differences in group average incomes, is exactly calculated on both basis. In the case of the exact discrete data calculations the within-group inequality component may be determined simply by subtraction if the exact overall Theil coefficient is available. The percentage within-group inequality contribution was, of course, under-estimated using the decile-based approximations. However, the maximum level of understatement was only 3 per cent (i.e., 63.73 per cent *vis-à-vis* 60.77) in the case of 1973 final income for households of differing compositions.

Applying the standard lower bound approximation approach, under-estimation arises from the exclusion of the within-decile inequality in each

3. The equivalent results given in this paper differ marginally in some instances because the weighted average of decile incomes was used instead of the corresponding published average to avoid any inconsistencies due to the rounding of data in the CSO reports.

Table 3: *Decomposition of Theil inequality coefficient for different household groupings, 1973 and 1978*

Houshold groupings[1]	*1973*				*1980*			
	Direct income		*Final income*		*Direct income*		*Final income*	
	Exact[2]	*Approx.*[3]	*Exact*	*Approx.*	*Exact*	*Approx.*	*Exact*	*Approx.*
Household composition								
Within-group	0.1249	0.1145	0.0710	0.0626	0.1407	0.1320	0.0550	0.0490
Between-group	0.0389	0.0389	0.0404	0.0404	0.0386	0.0386	0.0373	0.0373
Total	0.1638	0.1534	0.1114	0.1030	0.1793	0.1706	0.0923	0.0863
Life cycle of head of household								
Within-group	0.1285	0.1182	0.0841	0.0758	0.1338	0.1252	0.0642	0.0584
Between-group	0.0353	0.0353	0.0273	0.0273	0.0455	0.0455	0.0281	0.0281
Total	0.1638	0.1535	0.1114	0.1031	0.1793	0.1707	0.0923	0.0865
Age of head of household								
Within-group	0.1483	0.1377	0.1020	0.0934	0.1528	0.1438	0.0816	0.0755
Between-group	0.0155	0.0155	0.0094	0.0094	0.0265	0.0265	0.0107	0.0107
Total	0.1638	0.1532	0.1114	0.1028	0.1793	0.1703	0.0923	0.0862
Number of earners in household								
Within-group	0.0976	0.0871	0.0812	0.0727	0.0912	0.0823	0.0607	0.0547
Between-group	0.0661	0.0662	0.0302	0.0302	0.0881	0.0881	0.0316	0.0316
Total	0.1638	0.1533	0.1114	0.1029	0.1793	0.1704	0.0923	0.0863
Livelihood status of head								
Within-group	0.1318	0.1217	0.0997	0.0914	0.1315	0.1228	0.0816	0.0756
Between-group	0.0320	0.0320	0.0117	0.0117	0.0478	0.0478	0.0107	0.0107
Total	0.1638	0.1537	0.1114	0.1031	0.1793	0.1706	0.0923	0.0863
STATE	0.1638	0.1533	0.1114	0.1024	0.1793	0.1703	0.0923	0.0858

Notes: (1) Defined in Appendices 2 and 3.
(2) Calculated on the basis of discrete data.
(3) Standard lower bound Theil coefficient approximations using overall State decile income ranges in all cases.

group. This situation may be summarised algebraicly as follows:

$$T = \sum_{K=1}^{g} \frac{n_K \overline{x}_K}{n\overline{x}} \log(\frac{\overline{x}_K}{\overline{x}}) + \sum_{K=1}^{g} \frac{n_K \overline{x}_K}{n\overline{x}} \cdot T_K \tag{8}$$

(between-group) (within group)

where T_K, the Theil inequality coefficient for the Kth group (K = 1 to g), is further decomposed to distinguish the between and within-decile contribution within the group:

i.e.,

$$T_K = \sum_{j=1}^{d} \frac{n_{jK} \overline{x}_{jK}}{n_K \overline{x}_K} \log(\frac{\overline{x}_{jK}}{\overline{x}_K}) + \sum_{j=1}^{d} \frac{n_{jK} \overline{x}_{jK}}{n_K \overline{x}_K} \cdot T_{jK} \tag{9}$$

(between decile in Kth group) (within decile in Kth group)

Where T_{jK} is the Theil coefficient for income observations in jth decile (j = 1 to d) in the Kth group. The within-decile inequality is ignored in determining the standard lower bound Theil estimate at group level. This means that, when aggregated over all groups, the resulting overall within-group inequality component is under-estimated by:

$$\sum_{K=1}^{g} \sum_{j=1}^{d} \frac{n_{jK} \overline{x}_{jK}}{n\overline{x}} \cdot T_{jK}$$

This represents the inequality arising within deciles in each group (or within groups in each decile). It is interesting to note that when this aggregated within-group lower bound estimate is added to the between-group inequality contribution:

i.e.,

$$T_L = \sum_{K=1}^{g} \frac{n_K \overline{x}_K}{n\overline{x}} \cdot \log(\frac{\overline{x}_K}{\overline{x}}) + \sum_{K=1}^{g} \sum_{j=1}^{d} \frac{n_{jK} \overline{x}_{jK}}{n_K \overline{x}_K} \cdot \log(\frac{\overline{x}_{jK}}{\overline{x}_K})$$

$$= \sum_{K=1}^{g} \sum_{j=1}^{d} \frac{n_{jK} \overline{x}_{jK}}{n\overline{x}} \cdot \log(\frac{\overline{x}_{jK}}{\overline{x}}) \tag{10}$$

the resulting global Theil estimate incorporates the within-group inequality in each decile. It, therefore, provides a *more precise* lower bound for the exact Theil value than the standard estimate based on overall decile ranges which excludes *all* inequality arising within deciles. However, as can be seen from Table 3, the gain in precision is relatively small. It varies between the different groupings of households distinguished, but tends to be higher for final income than for direct income.

IV SELECTION OF INCOME RANGES

General Points

The standard income ranges used in the published CSO reports were not selected for the purpose of optimising the estimation of Gini and Theil coefficients. One standard set of income ranges could not achieve this purpose for different income concepts.

In general, researchers are dependent on the standard set of published income ranges. As was seen from Tables 1 and 2, these ranges are more suited to some income concepts than others. The resulting under-estimation variations in the standard Gini and Theil inequality derived measures can result in misleading conclusions.

Other than increasing the number of intervals distinguished (which would be impractical in published reports) it has been shown that one way of avoiding this problem is the general use of decile classifications. Another possible approach is the selection of income ranges specifically to minimise the under-estimation of the standard grouped data Gini and Theil values.

Improved Selection of Income Ranges

The under-estimation of the standard lower bound approximation of the Gini and Theil inequality coefficients is reduced by choosing income ranges which increase the differentials between the average income of ranges and also reduce the income variability within ranges. This is quite akin to the similar problem in stratified sampling where the variance of the estimate is reduced by choosing strata which maximise differences between strata averages and minimise data variability within strata. When the stratifying characteristic is the variable being estimated a number of approximate methods have been developed to choose strata boundaries to minimise the overall variance. One such method, described in Cochran (1963, p. 130), involves the aggregation of $\sqrt{f_i}$ (i = 1 to n) in a detailed frequency distribution of n intervals containing f_i observations and the choice of strata boundaries to create equal intervals on the cumulative $\sqrt{f_i}$ scale.

This method is used to select 13 direct, gross, disposable and final 1980 household income ranges. The resultant lower and upper bounds for the Gini and Theil coefficients are compared with those obtained from the 13 standard published intervals to see if there is any gain in precision. The detailed frequency distributions (incorporating reweighting) based on the 60 standard income ranges:

i.e., Less than £10; £5 intervals to £100; £10 intervals to £400; £20 intervals to £500; £600 and over

are used as the basis for selecting the 13 ranges. Allowance is made for variable interval widths of £5c_i by cumulating $\sqrt{c_i f_i}$. The end point of the upper interval £600 and over is taken as £2,000 for consistency with the earlier calculations; this far exceeds the actual maximum incomes in all four cases. The income ranges resulting from this approach are summarised in Table 4. Two different sets of ranges essentially emerge. There is close similarity between the direct and gross income ranges and also between those for disposable and final income. The direct/gross income ranges differ quite radically from the standard published intervals; fewer (and wider) ranges are distinguished at the lower income levels with a corresponding greater number of high income ranges. This more detailed breakdown of higher income levels is reduced for disposable and final income, but is still much greater than that of the published standard intervals. This is the reason why the under-estimation of the standard (i.e., lower bound) Gini and Theil values based on published income ranges was lower for disposable and final income than for direct and gross income. It also explains why the under-estimation of overall decile based values was particularly high for sub-populations of high income households.

The Gini and Theil upper and lower bounds based on these new sets of 13 income ranges are given in Table 5. As can be seen, there is a considerable gain in precision compared with the results calculated with the published data. The increased precision is particularly large in the case of the Theil coefficient where the under-estimation of the standard lower bound estimate based on these new sets of 13 income ranges is less (and in the case of gross, disposable and final income substantially less) than the corresponding errors obtained using the 20-quantiles. The reduced under-estimation of the standard lower bound Gini measure is of the same order of a magnitude for each of the four income concepts and so can be used to provide consistent measures of absolute changes in inequality levels resulting from the payment of taxes and the receipt of benefits. The Theil under-estimation still varies but, since its magnitude is considerably reduced, the dangers of misinterpretation is correspondingly lessened.

It would clearly be impractical to use variable sets of ranges for different income concepts in published reports. However, this method of selecting income ranges provides far more precise Gini and Theil approximations than obtained using decile classifications. It could, therefore, be of practical use for special analysis purposes when exact Gini and Theil coefficient values are not available.

V SUMMARY

The traditional lower bound Gini and Theil inequality approximations based on the standard HBS published income ranges for different income

Table 4: *Improved selection of 13 direct, gross, disposable and final weekly household income ranges, 1980*

Ranges in published 1980 reports	*Direct income*		*Gross income*		*Disposable income*		*Final income*	
	Improved ranges	*Households (adjusted)*	*Improved ranges*	*Households (adjusted)*	*Improved ranges*	*Households (adjusted)*	*Improved ranges*	*Households (adjusted)*
Under £20	Under £15	1,392	Under £30	777	Under £20	311	Under £30	509
£20 & under £30	£15 & under £40	573	£30 & under £50	791	£20 & under £45	1,119	£30 & under £50	885
£30 " " £40	£40 " " £70	726	£50 " " £75	881	£45 " " £65	847	£50 " " £65	784
£40 " " £60	£70 " " £90	711	£75 " " £95	759	£65 " " £80	744	£65 " " £85	1,000
£60 " " £80	£90 " " £110	730	£95 " " £120	938	£80 " " £100	991	£85 " " £100	723
£80 " " £90	£110 " " £140	867	£120 " " £140	628	£100 " " £120	778	£100 " " £120	816
£90 " " £100	£140 " " £170	610	£140 " " £170	683	£120 " " £140	643	£120 " " £140	631
£100 " " £120	£170 " " £210	593	£170 " " £210	655	£140 " " £170	657	£140 " " £160	481
£120 " " £140	£210 " " £250	351	£210 " " £260	464	£170 " " £210	502	£160 " " £200	649
£140 " " £170	£250 " " £320	365	£260 " " £320	323	£210 " " £260	314	£200 " " £250	378
£170 " " £200	£320 " " £420	183	£320 " " £420	186	£260 " " £340	184	£250 " " £340	226
£200 " " £230	£420 " " £600	64	£420 " " £600	76	£340 " " £580	82	£340 " " £560	89
£230 & over	£600 & over	21	£600 & over	23	£580 & over	13	£560 & over	15

Table 5: *Grouped HBS data estimates of Gini and Theil inequality coefficients based on improved selection of 13 ranges, 1980*

Calculation basis	*Direct income*	*Gross income*	*Disposable income*	*Final income*
Gini Coefficient				
(a) Standard bounds				
– Lower	0.4724	0.3890	0.3627	0.3516
(error* with selected 13 ranges)	(-0.0040)	(-0.0036)	(-0.0040)	(-0.0034)
(error* with published 13 ranges)	(-0.0079)	(-0.0073)	(-0.0042)	(-0.0043)
– Upper	0.4779	0.3945	0.3685	0.3568
(b) Gastwirth's sharper bounds				
– Lower	0.4741	0.3908	0.3648	0.3539
– Upper	0.4772	0.3937	0.3676	0.3557
(c) Exact value	0.4764	0.3926	0.3667	0.3550
Theil Coefficient				
(a) Standard bounds				
– Lower	0.1750	0.1105	0.0971	0.0912
(error* with selected 13 ranges)	(-0.0043)	(-0.0024)	(-0.0022)	(-0.0011)
(error* with published 13 ranges)	(-0.0109)	(-0.0076)	(-0.0049)	(-0.0026)
– Upper	0.1843	0.1182	0.1037	0.0984
(b) Gastwirth's sharper bounds				
– Lower	0.1756	0.1111	0.0977	0.0918
– Upper	0.1830	0.1171	0.1024	0.0969
(c) Exact value	0.1793	0.1129	0.0993	0.0923

*Under-estimation = exact value minus lower bound.

concepts understate the exact values in varying degrees. The magnitude and variability of understatement is substantially higher for the Theil approximation. The use of these estimates, therefore, can distort analyses of the inequality changes resulting from the State taxes/benefits redistribution process.

Gini and Theil inequality approximations based on decile ranges each understate the exact values for different income concepts by approximately the same absolute amounts and, therefore, provide a consistent basis for analysis purposes. It would, therefore, appear appropriate for the CSO to use decile rather than standard ranges in future HBS and income redistribution reports now that it has developed a decile classification procedure incorporating the reweighting adjustments for differential response. The method developed for selecting income ranges provides more precise approximations, but this may be of practical use only for special analysis purposes.

REFERENCES

BHATTACHARYA, N. and B. MAHALANOBIS, 1967. "Regional Disparities in Household Consumption in India", *Journal of American Statistical Association*, Vol. 62, March, pp. 143-161.

BOURGUIGNON, FRANCOIS, 1979. "Decomposable Income Inequality Measures", *Econometrica*, Vol. 47, No. 4, July, pp. 901-920.

CENTRAL STATISTICS OFFICE, 1980. *Redistributive Effects of State Taxes and Benefits on Household Incomes in 1973*, Dublin: Stationery Office.

CENTRAL STATICS OFFICE, 1982. *Household Budget Survey 1980, Volume I*, Dublin: Stationery Office.

CENTRAL STATISTICS OFFICE, 1983. *Redistributive Effects of State Taxes and Benefits on Household Incomes in 1980*, Dublin: Stationery Office.

COCHRAN, W.G., 1963. *Sampling Techniques*, Second Edition, John Wiley & Sons Ltd.

GASTWIRTH, JOSEPH L., 1972. "The Estimation of the Lorenz Curve and Gini Index", *Review of Economics and Statistics*, Vol. 63, pp. 306-316.

GASTWIRTH, JOSEPH L., 1975. "The Estimation of a Family of Measures of Inequality", *Journal of Econometrics*, Vol. 3, pp. 61-70.

MORGAN, JAMES, 1962. "The Anatomy of Income Distribution", *Review of Economics and Statistics*, Vol. 44, August, pp. 270-282.

MURPHY, D.C., 1984. "The Impact of State Taxes and Benefits on Irish Household Incomes", *Statistical and Social Inquiry Society of Ireland*, 23 February.

NOLAN, BRIAN, 1978. "The Personal Distribution of Income in the Republic of Ireland", *Journal of the Statistical and Social Inquiry Society of Ireland*, Vol. XXIII, Part V.

NOLAN, BRIAN, 1981. "Redistribution of Household Incomes in Ireland by Taxes and Benefits", *The Economic and Social Review*, Vol. 13, No. 1, October.

O'CONNELL, Philip J., 1982. "The Distribution and Redistribution of Income in the Republic of Ireland", *The Economic and Social Review*, Vol. 13, No. 4, July.

SEN, AMARTYA, 1973. *On Economic Inequality*, Oxford: Clarendon Press.

SHORROCKS, A.F., 1980. "The Class of Additively Decomposable Inequality Measures", *Econometrica*, Vol. 48, No. 3, April, pp. 613-625.

THEIL, HENRI, 1967. *Economics and Information Theory*, Amsterdam: North-Holland Publishing Company.

APPENDIX 1

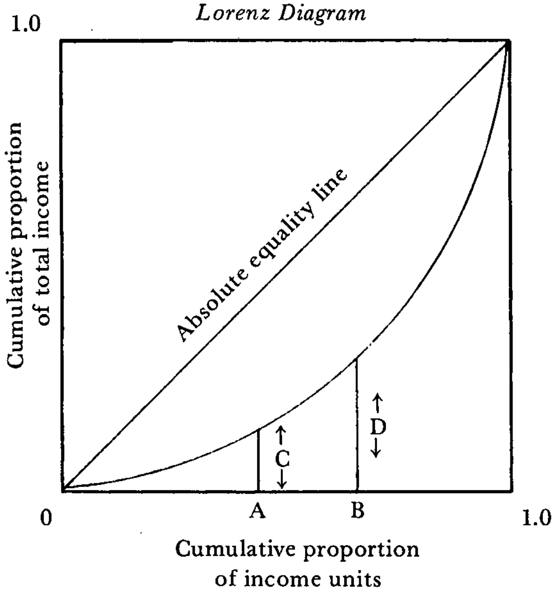
Calculation of Gini and Theil Coefficients

Gini Coefficient (Discrete Data)

The Gini coefficient defined at (1) is also presented using the *Lorenz diagram* (which plots the cumulative proportion of income observations ordered in increasing magnitude along the horizontal axis against their corresponding cumulative proportionate share of total income) as:

$$\frac{\text{Area between Lorenz Curve and diagonal}}{\text{Area under diagonal}}$$

Lorenz Diagram



The coefficient is generally calculated on this basis using the method described by Morgan (1962) with the Lorenz curve obtained by joining plotted points with straight lines:

$$\text{i.e.}\quad G = \frac{\tfrac{1}{2} - \text{Area under Lorenz Curve}}{\tfrac{1}{2}}$$

$$= 1 - 2 \times \text{Area under Lorenz Curve}$$

$$= 1 - \Sigma\,(B - A)\,(C + D)$$

In the case of unweighted *discrete data*:

$$A = \frac{i-1}{n}\,; \qquad B = \frac{i}{n}\,; \qquad B - A = \frac{1}{n}$$

$$C = \sum_{h=1}^{i-1} x_h \div \sum_{h=1}^{n} x_h = T_{i-1} \div T_n$$

$$D = \sum_{h=1}^{i} x_h \div \sum_{h=1}^{n} x_h = T_i \div T_n$$

it follows that:

$$G = 1 - \sum_{i=1}^{n} \frac{1}{n} \left[\sum_{h=1}^{i-1} x_h + \sum_{h=1}^{i} x_h \right] \div \sum_{h=1}^{n} x_h$$

$$= 1 - \frac{1}{nT_n} \sum_{i=1}^{n} (T_{i-1} + T_i)$$

This formula is exactly equivalent to that given by Sen (1973, p. 31)

$$\text{i.e.,} \quad G = 1 + \frac{1}{n} - \frac{2}{n^2 \overline{x}} \left[nx_1 + (n-1)x_2 + \ldots + x_n \right]$$

where $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$.

In the case of grouped data, where the n observations in the jth income range (j = 1 to R) are assumed to be equal to the mean income $\overline{x}_j$, it follows that:

$$B - A = \frac{n_j}{n} \quad \text{and} \quad T_j = \sum_{K=1}^{j} n_K \overline{x}_K$$

so that the standard grouped Gini estimate (i.e., the general lower Gini bound) is:

$$G_L = 1 - \sum_{j=1}^{R} \frac{n_j}{n} \left(\frac{T_{j-1} + T_j}{T_R} \right)$$

$$= 1 - \frac{1}{nT_R} \sum_{j=1}^{R} n_j (T_{j-1} + T_j)$$

In the HBS average weekly household income is derived (see, for example, Appendix of 1980 HBS Report, Volume 1) as:

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{n} g_i x_i$$

$$= \frac{1}{n} \sum_{i=1}^{n} (g_i \div \frac{N}{n}) \; x_i$$

where n = number of households in HBS sample;
N = number of households in the population;
g_i = grossing factor applicable to the ith household which allows for differential non-response related to such factors as household size, social group, regional location, farm size, etc.

This HBS reweighting is equivalent to the proportional adjustment:

$$a_i = g_i \div \frac{N}{n}$$

which, in calculating the Gini coefficient, must be incorporated in *both* the household and income distributions from which the horizontal and vertical axes, respectively, of the Lorenz diagram are derived. Using Morgan's approach, this results in the following Gini formulation for reweighted HBS discrete data:

$$G = 1 - \sum_{i=1}^{n} \left[\frac{\sum_{h=1}^{i} a_h - \sum_{h=1}^{i-1} a_h}{n} \right] \cdot \left[\frac{\sum_{h=1}^{i-1} a_h x_h + \sum_{h=1}^{i} a_h x_h}{\sum_{h=1}^{n} a_h x_h} \right]$$

$$= 1 - \sum_{i=1}^{n} \frac{a_i}{n} \left(\frac{T_{i-1} + T_i}{T_n} \right)$$

$$= 1 - \frac{1}{NT_n} \sum_{i=1}^{n} g_i (T_{i-1} + T_i)$$

where

$$T_i = \sum_{h=1}^{i} g_h x_h$$

Theil Coefficient (Discrete Data)

The Theil coefficient for unweighted discrete data is essentially complied as the simple arithmetic average of the function:

$$\frac{x_i}{\bar{x}} \log\left(\frac{x_i}{\bar{x}}\right)$$

calculated for each observation x_i ($i = 1$ to n).

In the case of the HBS data allowance is made as follows in this averaging process for the reweighting made to correct for differential non-response:

$$T = \frac{1}{n\bar{x}} \sum_{i=1}^{n} a_i x_i \log\left(\frac{x_i}{\bar{x}}\right)$$

$$= \frac{1}{T_n} \sum_{i=1}^{n} g_i x_i \log\left(\frac{N x_i}{T_n}\right)$$

where

$$T_n = \sum_{i=1}^{n} g_i x_i .$$

Grouped Data Estimates

The standard grouped data Gini and Theil coefficients formulations may be used directly using the n_j and $\bar{x}_j$ values given in the published HBS reports since they already incorporate the reweighting for differential responses.

APPENDIX 2

Sub-population Gini and Theil inequality coefficients for direct and final household income, 1980

Sub-populations	No. of households (adjusted)	1980 Gini Coefficient				1980 Theil Coefficient			
		Direct income		Final income		Direct income		Final income	
		Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data
Household composition									
1 Adult	1,181	0.6360	0.6458	0.3140	0.3422	0.3268	0.3367	0.0807	0.0883
2 Adults	1,451	0.5415	0.5482	0.2984	0.3042	0.2255	0.2356	0.0679	0.0735
2 Adults with 1 child	443	0.3084	0.3139	0.2527	0.2562	0.0774	0.0822	0.0493	0.0516
2 " " 2 children	733	0.3368	0.3441	0.2761	0.2803	0.0925	0.1009	0.0607	0.0684
2 " " 3 "	532	0.3390	0.3437	0.2344	0.2398	0.0945	0.0983	0.0426	0.0466
2 " " 4+ "	464	0.3757	0.3797	0.2120	0.2163	0.1190	0.1214	0.0337	0.0361
3 Adults	517	0.4457	0.4533	0.2775	0.2803	0.1489	0.1568	0.0579	0.0593
3 " with children	505	0.3840	0.3908	0.2488	0.2509	0.1166	0.1228	0.0452	0.0510
4 Adults	292	0.3371	0.3577	0.2353	0.2476	0.0861	0.1021	0.0404	0.0521
4 " with children	308	0.3561	0.3655	0.1921	0.2010	0.0966	0.1012	0.0279	0.0305
Other households without children	234	0.2592	0.3205	0.1912	0.2474	0.0626	0.0781	0.0330	0.0450
" " with children	525	0.3475	0.3749	0.2044	0.2430	0.0996	0.1089	0.0348	0.0430
Life cycle of head of household (HOH)*									
Without wife and children:									
Young	474	0.3870	0.3950	0.3749	0.3816	0.1153	0.1260	0.1099	0.1131
Middle aged	517	0.5714	0.5818	0.4104	0.4327	0.2583	0.2678	0.1279	0.1383
Retired	702	0.7230	0.7437	0.2781	0.2937	0.4507	0.4707	0.0652	0.0690
With wife and/or children:									
Pre-family	246	0.3019	0.3070	0.2639	0.2699	0.0711	0.0746	0.0530	0.0542
Pre-school	776	0.3212	0.3289	0.2743	0.2790	0.0866	0.0915	0.0600	0.0633
Early school	917	0.3576	0.3643	0.2432	0.2490	0.1072	0.1135	0.0464	0.0525
Pre-adolescent	823	0.3845	0.3909	0.2414	0.2482	0.1180	0.1243	0.0432	0.0486
Adolescent	855	0.3708	0.3782	0.2383	0.2500	0.1035	0.1098	0.0405	0.0467
Adult	1,184	0.3773	0.4029	0.2919	0.3127	0.1124	0.1246	0.0619	0.0698
Empty nest	345	0.5597	0.5667	0.3469	0.3532	0.2383	0.2464	0.0894	0.0919
Retired	345	0.6319	0.6486	0.1861	0.1926	0.3215	0.3344	0.0303	0.0314

†Standard lower bound grouped data estimates based on overall State decile income ranges in all cases.
*Defined in Murphy (1984) and 1980 Household Budget Survey, Volume 4.

Appendix 2 (cont'd)

Sub-populations	No. of households (adjusted)	1980 Gini Coefficient				1980 Theil Coefficient			
		Direct income		Final income		Direct income		Final income	
		Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data
Age of Head of Household (HOH)									
Under 30 years	1,002	0.3586	0.3656	0.2890	0.2930	0.1069	0.1156	0.0664	0.0741
30-44 years	2,147	0.3462	0.3537	0.2693	0.2755	0.0951	0.1013	0.0544	0.0590
45-64 "	2,407	0.4439	0.4572	0.3580	0.3712	0.1489	0.1592	0.0923	0.0999
65 and over	1,628	0.6689	0.6800	0.3393	0.3471	0.3620	0.3742	0.0859	0.0884
Number of earners									
0 Earners	1,608	0.7589	0.7832	0.2885	0.3332	0.5146	0.5428	0.0629	0.0831
1 Earner	4,018	0.3411	0.3472	0.2968	0.3223	0.0885	0.0941	0.0656	0.0820
2 Earners	1,023	0.2549	0.2726	0.2409	0.2655	0.0486	0.0572	0.0451	0.0568
3+ Earners	536	0.1714	0.2300	0.1584	0.2230	0.0258	0.0409	0.0223	0.0392
Livelihood status of HOH									
Self-employed	1,647	0.4781	0.4892	0.4007	0.4139	0.1674	0.1805	0.1164	0.1257
Employee	3,387	0.2848	0.2945	0.2698	0.2767	0.0582	0.0638	0.0517	0.0558
Out of work	460	0.7621	0.7697	0.3383	0.3437	0.5125	0.5237	0.0803	0.0833
Retired	983	0.6799	0.6902	0.3225	0.3294	0.3757	0.3886	0.0778	0.0810
Other	708	0.6996	0.7069	0.3944	0.4035	0.4035	0.4249	0.1170	0.1270
STATE	7,185	0.4683	0.4764	0.3479	0.3550	0.1703	0.1793	0.0858	0.0923

† Standard lower bound grouped data estimates based on overall State decile income ranges in all cases.
* Defined in Murphy (1984) and 1980 Household Budget Survey, Volume 4.

APPENDIX 3

Sub-population Gini and Theil inequality estimates for direct and final household income, 1973

Sub-populations	No. of households (adjusted)	1973 Gini Coefficient				1973 Theil Coefficient			
		Direct income		Final income		Direct income		Final income	
		Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data
Household composition									
1 Adult	1,084	0.6503	0.6665	0.3326	0.3597	0.3496	0.3627	0.0971	0.1059
2 Adults	1,537	0.5175	0.5253	0.3860	0.3937	0.2091	0.2205	0.1180	0.1254
2 Adults with 1 child	375	0.3679	0.3768	0.3588	0.3658	0.1089	0.1341	0.1137	0.1326
2 " " 2 children	511	0.2855	0.2908	0.2676	0.2727	0.0628	0.0661	0.0556	0.0556
2 " " 3 "	389	0.3068	0.3119	0.2419	0.2461	0.0766	0.0810	0.0467	0.0496
2 " " 4+ "	583	0.3492	0.3551	0.2437	0.2527	0.0997	0.1085	0.0468	0.0557
3 Adults	782	0.4176	0.4245	0.3156	0.3205	0.1301	0.1359	0.0729	0.0765
3 " with children	624	0.3406	0.3502	0.2573	0.2648	0.0890	0.1010	0.0484	0.0594
4 Adults	460	0.3282	0.3416	0.2692	0.2757	0.0783	0.0850	0.0509	0.0572
4 " with children	497	0.2907	0.2988	0.2226	0.2333	0.0646	0.0687	0.0372	0.0414
Other households without children	280	0.2983	0.3265	0.2609	0.2832	0.0661	0.0780	0.0484	0.0604
" " with children	620	0.3188	0.3432	0.2105	0.2463	0.0810	0.0914	0.0367	0.0463
Life cycle of head of household (HOH)*									
Without wife and children:									
Young	336	0.4032	0.4129	0.3899	0.3995	0.1193	0.1253	0.1119	0.1162
Middle aged	664	0.5520	0.5608	0.4295	0.4457	0.2365	0.2436	0.1380	0.1466
Retired	704	0.7078	0.7298	0.3394	0.3576	0.4289	0.4464	0.1002	0.1047
With wife and/or children:									
Pre-family	195	0.3109	0.3217	0.3365	0.3379	0.0749	0.0807	0.0819	0.0879
Pre-school	670	0.3249	0.3340	0.3372	0.3455	0.0860	0.1002	0.0974	0.1086
Early school	760	0.3224	0.3299	0.2657	0.2712	0.0828	0.0897	0.0554	0.0603
Pre-adolescent	787	0.3557	0.3623	0.2657	0.2737	0.1018	0.1070	0.0522	0.0579
Adolescent	1,146	0.3582	0.3670	0.2754	0.2883	0.0967	0.1055	0.0544	0.0639
Adult	1,709	0.3671	0.3824	0.3151	0.3243	0.1008	0.1095	0.0706	0.0777
Empty nest	468	0.4997	0.5099	0.4423	0.4545	0.1981	0.2150	0.1593	0.1693
Retired	299	0.6693	0.6838	0.3214	0.3331	0.3782	0.3937	0.0988	0.1032

†Standard lower bound grouped data estimates based on overall State decile income ranges in all cases.
*Defined in Murphy (1984) and 1980 Household Budget Survey, Volume 4.

Appendix 3 (cont'd)

Sub-populations	No. of households (adjusted)	1973 Gini Coefficient				1973 Theil Coefficient			
		Direct income		Final income		Direct income		Final income	
		Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data	Grouped† data	Discrete data
Age of head of household (HOH)									
Under 30 years	592	0.3298	0.3379	0.3163	0.3216	0.0885	0.1057	0.0839	0.1001
30-44 years	1,951	0.3289	0.3365	0.2857	0.2931	0.0839	0.0895	0.0615	0.0660
45-64 "	3,320	0.4261	0.4373	0.3790	0.3903	0.1367	0.1471	0.1033	0.1124
65 and over	1,876	0.6083	0.6165	0.4148	0.4220	0.2905	0.2999	0.1261	0.1311
Number of earners									
0 Earners	1,380	0.7688	0.7926	0.3467	0.4008	0.5445	0.5702	0.1002	0.1192
1 Earner	4,297	0.3607	0.3677	0.3454	0.3662	0.0984	0.1078	0.0893	0.1080
2 Earners	1,311	0.2772	0.2886	0.2691	0.2835	0.0565	0.0630	0.0535	0.0646
3+ Earners	750	0.1970	0.2359	0.2066	0.2452	0.0307	0.0413	0.0329	0.0465
Livelihood status of HOH									
Self-employed	2,241	0.4543	0.4671	0.4262	0.4419	0.1502	0.1635	0.1321	0.1430
Employee	3,280	0.2791	0.2890	0.2722	0.2791	0.0557	0.0626	0.0531	0.0594
Out of Work	380	0.7024	0.7087	0.3808	0.3850	0.4123	0.4194	0.1024	0.1058
Retired	818	0.6089	0.6171	0.3932	0.4000	0.2924	0.3019	0.1124	0.1152
Other	1,020	0.6575	0.6630	0.4301	0.4378	0.3495	0.3568	0.1356	0.1399
STATE	7,739	0.4465	0.4553	0.3782	0.3869	0.1533	0.1638	0.1024	0.1114

†Standard lower bound grouped data estimates based on overall State decile income ranges in all cases.
*Defined in Murphy (1984) and 1980 Household Budget Survey, Volume 4.