GENERATING INTERNATIONALLY COMPARABLE INCOME DISTRIBUTION DATA: EVIDENCE FROM THE FEDERAL REPUBLIC OF GERMANY (1974), MEXICO (1968) AND THE UNITED KINGDOM (1979).

BY WOUTER VAN GINNEKEN*

International Labour Office, Geneva

In this article an attempt is made to generate internationally comparable income distribution data for the Federal Republic of Germany (1974), Mexico (1968) and the United Kingdom (1979). To that end, the same income concept and income unit were adopted for each country, i.e. respectively household available income and the household. Moreover, incomes from various sources were adjusted for inconsistency with National Accounts according to Altimir's methodology. The paper finds that the distribution of persons by household income per equivalent unit is probably the best way of looking at the distribution of economic welfare. It further demonstrates that the distribution of persons by household available income per capita is much closer to this 'ideal' distribution than the distribution of households by household available income. Finally, the paper discusses some of the problems arising from the fact that one normally works with grouped data. It is found that in the case of the three countries under study, grouping is likely to have had only a small impact on the results.

1. INTRODUCTION

The question of income distribution has become a matter of major importance in the last development decade. This is explained by the fact that not all groups in society have equally benefited from economic growth and that in some developing countries the number of absolute poor has not decreased. As a result, various attempts have been made to produce a set of internationally comparable income data. Given the fragility of underlying data most of these attempts have been criticised.[1] The attempt described in this paper (which is part of a World Bank/ILO project) represents a new effort to improve international comparability—an effort that will hopefully be followed by new improvements in the future.

Collection of income distribution data started with the efforts of Adelman and Morris,[2] which included 43 distributions from developing countries, with different income concepts and income units. Their work was supplemented by Paukert[3] who achieved better comparability for 56 developed and developing countries, since he chose primary income as the income concept and the

*This article is written "à titre personnel'. It is an outcome of work carried out within ILO's World Employment Programme. I would like to thank Christian Grootaert, Jong-goo Park and Felix Paukert for their useful comments and Mr Garzuel for computer assistance.

[1]See, for example, D. McGranahan, "International Comparability of Statistics on Income Distribution," United Nations Research Institute for Social Development, Geneva, 1979.

[2]Adelman, I. and Morris, C. F., *Economic Growth and Social Equity in Developing Countries*, Stanford, 1973.

[3]Paukert, F. "Income distribution at different levels of development: a survey of evidence", *International Labour Review* 108 (2–3), August–September 1973, pp. 97–125.

household or the family as the income unit. The collection of income distribution data was further extended by S. Jain[4], who included information on 83 developed and developing countries (often more than one distribution per country). The collection contains distributions with different income concepts, income units and geographical coverage. M. Sawyer[5] produced comparable pre-tax and post-tax income distribution data on a per household and a per capital basis for about 12 OECD countries. P. Visaria[6] reviewed the income distribution data of six Asian countries (or regions) and experimented in particular with various income units (households, household member and equivalent unit) and their impact on income distribution and poverty analysis. O. Altimir[7] produced comparable income distribution data for seven Latin American countries. He used the same methodology for adjusting various income sources for inconsistency with National Accounts. The income unit he employed was the household while the income concept was total household income (including primary income, property income received, current transfers and other benefits received minus employers' contributions to Social Security). C. Morrison[8] generated a series of income distribution data including 40 developing countries. The income concept employed is primary income while the income unit is household, person or income earner. Income data are in some cases adjusted for inconsistency with National Accounts. Finally, the UN Statistical office[9] has produced a new compilation of income distribution data. Each of the sixty country chapters contains information on concepts, sample design, and sampling errors as well as on tabulations undertaken and selected actual results.

2. Income Concept, Income Unit and Welfare Unit

The WB/ILO Income Distribution Statistics Project[10] aims at generating about 30 new estimates, in both developed and developing countries. The approach taken in this paper will not be uniformly applied to all countries, because data availability is often a constraining factor. However, for the country estimates shown in this paper (Federal Republic of Germany, Mexico and U.K.) there was no major data constraint.

[4]Jain, S., *Size Distribution of Income. A complilation of data*, The World Bank, 1975.

[5]Sawyer, M., *Income Distribution in OECD Countries*, Occasional Studies, OECD Economic Outlook, Paris, July 1976.

[6]Visaria, P., "Poverty and Living Standards in Asia: an overview of the main results and lessons of selected household surveys," Development Research Center, World Bank (mimeo), 1980.

[7]Altimir, O. and Sourrouille, J. V., "Measuring levels of living in Latin America: an overview of main problems." Paper presented to the Expert Group Meeting on the Measurement of Living Standards, Washington D.C., February 1980.

[8]Paukert, F., Lecaillon, J., Morrison, C., and Germidis, D., "Repartition, Emploi et Développment Économoque: un essai de synthèse (to be published), ILO, Geneva, 1980.

[9]UNSO, "A Survey of National Sources of Income Distribution Statistics," Department of International Economic and Social Affairs, United Nations, New York, 1981.

[10]van Ginneken, W., "A Background Paper for the WB/ILO Income Distribution Statistics Project" (mimeo) ILO, Geneva, 1980. Country estimates have so far been completed on Malawi, Tanzania, India, Indonesia, Philippines, Nepal, Dominican Republic, El Salvador, Panama, Trinidad and Tobago, Bangladesh, Yugoslavia, Federal Republic of Germany, Mexico, United Kingdom, France, Spain and Sweden.

The *income concept* employed in the project is a slight modification of total available household income as defined by the UN Provisional Guidelines.[11] It includes wages and salaries—in cash and in kind—(excluding social security and private insurance contributions both by employers and employees), net income from self-employment (including consumption of own produce), income from personal property and investment (including imputed rental income of owner-occupied housing), social security and private (insurance) transfers, minus personal income and property taxes. This type of definition covers only part of total welfare, i.e. "that part of total welfare which is attributable to the consumption of goods and services of the kinds which are normally sold on the market."[12] In other words, welfare derived from government services is not included in this definition.

The *income unit* which corresponds with our income concept is undoubtedly the household. The household concept adopted here will be as close as possible to the definition of the 1980 World Population Census Programme[13]:

> "a household may be either (a) a one-person household . . . or (b) a multi-person household, that is, a group of two or more persons who make common provision for food or other essentials for living."

In most previous studies, the distribution of households has been analysed by household income. Although the household is the central unit within which production and consumption decisions are taken, it is not certain whether it is an appropriate welfare unit. If one wishes to analyse the primary income distribution, it seems that the income earner or tax-unit is more appropriate. Depending on the aim of the analysis one could also defend using the household, if one is interested in the earning power of a household. It is, however, not completely consistent to consider the household as the *welfare unit* for available household income (i.e. income after tax and transfers). The reason is that households of different size and composition need a different amount of available income to reach the same level of economic welfare.

As a result, more and more authors propose to look at the distribution of persons by household income per capita. Danziger and Taussig[14] for example therefore proposed that the person is a more appropriate income-receiving unit than the household. This is confirmed by Sen,[15] who states that—for poverty analysis[16]—the person is the appropriate unit because—from a social welfare point of view—one should weigh each individual's welfare equally.

[11]U.N., *Provisional Guidelines on Statistics of the Distribution of Income, Consumption and Accumulation of Households*, Studies in Methods, Series M, No. 61, New York 1977. See also ILO, *International Recommendations on Labour Statistics*, Geneva, 1975.

[12]Lydall, H., "Effects of alternative measurement techniques on the estimation of the inequality of income," World Employment Research Working Paper, No. 2-23/100, ILO, Geneva, June 1981, p. 11.

[13]U.N., "Draft Principles and Recommendations for Population and Housing Censuses", E/CN.3/515/add. 2, 1978, para. 73–74.

[14]Danziger, S. and Taussig, M. K., "The income unit and the anatomy of income distribution," *Review of Income and Wealth*, Series 25, No. 4, December 1979, pp. 365–375.

[15]Sen, A. K., "Three notes on the concept of poverty," Income Distribution and Employment Programme. Working Paper WEP 2-23/WP 65, ILO, Geneva, p. 17.

[16]For an application of this principle see W. van Ginneken, "Some methods of poverty analysis: an application to Iranian data 1975–1976," *World Development*, Vol. 8, No. 9, September 1980, pp. 639–646.

If one can derive equivalence scales then it is clear that household income per equivalent unit is a better criterion of economic welfare than household income or household income per capita. One would therefore come to the conclusion that the distribution of persons according to household income per equivalent unit is the best way of looking at the distribution of economic welfare.

Although we believe that this is indeed the case, it presents us with an aggregation problem. If one looks at the distribution of households by household income and of persons by household income per capita, then the aggregated income is equal to total income in the population. The third distribution consistent with the aggregation criterion is the distribution of equivalent units by household income per equivalent unit. We reject this distribution, however, since the number of equivalent units is not an appropriate welfare unit. As Sen points out,[17] the adoption of this unit would imply that the economic welfare of three two-member households is more important than two three-member households.[18]

In diagram 1 we show the nine combinations of welfare units and measures of economic welfare (ranking criteria). It shows that of the nine possible distributions, the distribution in cell 6 is considered optimal, those in cells 1 and 5 are sub-optimal, while those in the other cells are considered unimportant.

Welfare Unit / Economic Welfare (Ranking Criterion)	Household Income	Household Income Per Capita	Household Income per Equivalent Unit
Household	1	2	3
Person	4	5	6
Equivalent unit	7	8	9

It is likely that the distribution of persons by household income per capita (cell 5) is fairly close to that of cell 6, and we shall try to measure this by means of transition matrices (section 6). In section 5, where the main results are presented, we shall also show the distribution of households by household income, since this is the distribution most commonly used.

3. A Simple Methodology to Estimate Equivalence Scales

There are various ways to estimate equivalence scales.[19] For our purposes, we have chosen a simple methodology which can be applied to most available

[17]Sen, A. K., *op.cit.*

[18]It is pointed out by Deaton and Muellbauer, (*Economics and Consumer Behaviour*, Cambridge, 1980, p. 227) that one really needs a theory of allocation within the household. The assumption of the equivalent unit estimation assumes that the available household income is spent in proportion to the equivalent value that each household member represents. This may not be realistic given the fact that in many LDCs women and children receive less than their equivalent unit share.

[19]See for example, Goedhart, T., Halberstadt, V., Kapteyn, A. and van Praag, B., "The poverty line: concept and measurement," *Journal of Human Resources*, Vol. XII, No. 4, Fall 1977, pp. 503–520 and J. L. Nicholson, "Appraisal of different methods of estimating equivalent scales and their results," *Review of Income and Wealth*, Series 22, No. 1, March 1976, pp. 1–11.

data sets, both in developed and developing countries. It is based on a double-log food expenditure function of the following shape:

$$\log F = a + b \log Y + c \log N \tag{1}$$

where F = food expenditure per household

Y = available household income

N = household size.

If one assumes that a household's level of living varies inversely with the proportion of its expenditure that it devotes to food[20] (F/Y), one can mathematically derive that the economies of scale factor e (where $e = \partial \log Y / \partial \log N$, for $d(F/Y) = 0$) equals $(c/1-b)$. By reformulating the food expenditure function it is also possible to estimate directly the scale factor t which is related to e:

$$\log (F/N) = a + b \log (Y/N) - t \log N \tag{2}$$

If $b + c + t = 1$, one can derive that

$$e = 1 - \frac{t}{1-b} \tag{3}$$

since $b < 1$ in the case of food expenditure. One can further derive that if

$t > 0$ $e < 1$ there are economies of scale

$t = 0$ $e = 1$ there are no economies of scale

$t < 0$ $e > 1$ there are diseconomies of scale.

This means that if t is significantly different from zero, the economies of scale factor e is significantly different from zero.

In general there is a problem with estimating functions (1) and (2) because there is collinearity between $\log Y$ and $\log N$ and even more so between $\log (Y/N)$ and $\log N$. As a result we shall use the estimates of e derived from equation (1).

The estimates shown in table 1 are based on sources[21] which we have also used for estimating the distribution of income (see section 5). This means that the income data have been adjusted for National Accounts.

The table shows that both types of food expenditure functions produce about the same economies of scale factors. The economies of scale factors for

[20] A recent, more sophisticated, way to estimate equivalent scales on the basis of food shares is presented by A. Deaton and J. Muellbauer, *op. cit.*, pp. 205–212. They propose to consider the food share as a function of the logarithm of total expenditure. This has the advantage that the sum of expenditures on all items adds up to total expenditure and that equivalence scales are a function of total expenditures. The methodology was applied to Sri Lanka data. See A. Deaton, "Inequality and needs: some experimental results for Sri Lanka". (mimeo), Living Standards Measurement Study, World Bank, 1980.

[21] Göseke, G. and Bedau, K.-D., *Einkommens-und Verbrauchsschichtung für die grösseren Verwendungsbereiche des privaten Verbrauchs und die privaten Ersparnisse in der Bundersrepublik Deutschland, 1955 bis 1974*, Deutsches Institut für Wirtschaftsforschung, Berlin, 1978. Dirección General Coordinadora de la Programmación Económica y Social, *Estudios de ingresos y gastos de las familias*. Mexico, D.F., 1974. Department of Employment, *Family Expenditure Survey, 1979*. HMSO. London, 1980.

TABLE 1

ESTIMATES OF DOUBLE-LOG FOOD EXPENDITURE FUNCTIONS

Country	$\log F = a$	$+ \; b \log Y$	$+ \; c \log N$	R^2	$e^* = \frac{c}{1-b}$	Number of observations
Federal Republic of Germany (1974)	1.995	0.611 (0.008)	0.246 (0.009)	0.99	0.632	100
Mexico (1968)	1.136	0.433 (0.006)	0.319 (0.014)	0.56	0.563	5,608
United Kingdom (1979)	1.252	0.426 (0.002)	0.497 (0.002)	0.98	0.866	120
	$\log \frac{F}{N} = a$	$+ \; b \log \frac{Y}{N}$	$- \; t \log N$	R^2	$e^* = 1 - \frac{t}{1-b}$	Number of observations
Federal Republic of Germany (1974)	2.068	0.601 (0.007)	0.127 (0.008)	0.99	0.682	100
Mexico (1968)	1.045	0.464 (0.002)	0.230 (0.006)	0.61	0.571	5,608
United Kingdom (1979)	1.267	0.424 (0.001)	0.099 (0.001)	0.93	0.828	120

* = Economies of scale factor.
Figures between brackets represent standard deviations.
Income data are adjusted for inconsistencies with National Accounts.

the Federal Republic of Germany and Mexico are somewhat lower than for the United Kingdom. This indicates that the per capita income distribution in the U.K. is closer to the "true" distribution of economic welfare than in the other two countries because if $e = 1$ both distributions are exactly the same.

4. CONSISTENCY WITH NATIONAL ACCOUNTS AND POPULATION ESTIMATES

In most cases income distribution data are derived from household income and/or expenditure surveys and are, therefore, based on a sample. As a result, the quality of the data on incomes and households is dependent on sampling and non-sampling errors of those surveys. There is not really a theory about how income and population data should be adjusted. In this paper we shall therefore take a pragmatic approach. The basic fact underlying the adjustment is that aggregated income—as measured by household surveys—underestimates National Accounts income by between 15 to 30 percent. It can generally be assumed that National Accounts data are more accurate than aggregate household income estimates, because the former data combine multiple data sources and have been subject to several rounds of consistency checking.

There are many reasons for this discrepancy. Altimir,[22] after a thorough survey of Latin American data, comes to the conclusion that "biases in sample structure and in responses to income questions appear to be less important

[22] Altimir, O., "Income distribution estimates from household surveys and population censuses in Latin America: an assessment of reliability" (mimeo), World Bank, Development Research Center, Washington D.C. 1977.

sources of discrepancies than underestimation of income." He further finds that[23] "estimates of property income (housing in particular) are higher in surveys than in the National Accounts, which tend to undervalue this item. Survey figures for wages and salaries are in most cases consistent with National Accounts total.... Entrepreneurial income in the survey data is usually 25–50 percent below the comparative macro-economic estimates." It seems, therefore, necessary to adjust survey data in order to provide a more realistic picture of income inequality and poverty.

Many authors have proposed ways to make these adjustments. Ojha and Bhatt[24] and Ranadive[25] proposed to take the expenditure distribution and to estimate separately savings ratios per expenditure class. The rationale of this methodology is that expenditure gives a more accurate picture of economic welfare (in particular of low income groups), because expenditure in kind is normally valued at retail prices and because expenditure is usually more closely related to permanent income. Mrs de Navarrete[26] proposed a similar methodology for Mexico, but she adjusted the incomes of higher income groups somewhat differently. In fact, for all income classes where measured income exceeds measured expenditure, incomes were proportionally increased in order to arrive at consistency with National Accounts. For income classes where measured expenditure exceeds measured income, expenditure is taken as a proxy for income.[27] Altimir proposes to adjust the original household income distribution by type of income. In other words, he corrects separately wages and salaries, income from self-employment, property income and transfers for inconsistencies with National Accounts. The three methodologies normally lead to a higher estimate of income inequality, but to a lower estimate of poverty.

There is in principle, however, no limit to the number of adjustments that can be made. If a Social Accounting Matrix[28] is available, then the original survey data have been made consistent with various sources of data, such as population estimates, data from tax and social security surveys, population and industrial censuses.[29] The most advanced way of adjusting is a so-called merge-file,[30] which includes adjusted income and other data for individual household

[23] *Ibid.*, p. 95.

[24] Ojha, P. D. and Bhatt, V. V., "Pattern of income distribution in India: 1953–55 to 1963–65", *Sankhya*, Vol. 36, Series C, 1974, pp. 163–166.

[25] Ranadive, K. R., "Distribution of Income Trends Since Planning," paper presented at the ISI Seminar on Income Distribution, February 1973.

[26] Navarrete, de, "La distribución del ingreso en México: tendencias y perspectivas" in *El Perfil de México en 1980*, Vol. 1, 3rd edition (México, D.F., 1971), pp. 15–62.

[27] See also W. van Ginneken, "Socio-economic groups and income distribution in Mexico," Crown Helm, London, 1980.

[28] See for example, G. Pyatt and E. Thorbecke, "Planning techniques for a better future," ILO, Geneva, 1976.

[29] A good example is the income distribution estimates by the Deutsches Institut für Wirtschaftsforschung; see G. Göseke and K. D. Bedau, *Verteilung und Schichtung der Einkommen der privaten Haushalte in der Bundesrepublik Deutschland 1950 bis 1975* (Distribution and classification of private household incomes in the Federal Republic of Germany, 1950 to 1975), Berlin, 1974.

[30] Kortmann, K., "Die Generierung einer geschlossen Mikrodatenbasis für die Bundesrepublik Deutschland" (The generation of a closed microdata file for the Federal Republic of Germany) in H.-J. Krupp and W. Glatzer (ed.), *Umverteilung in Sozialstaat* (Redistribution in the Welfare State) Frankfurt, 1978, pp. 193–236. For a description of an American merge-file, see J. A. Pechman and B. A. Okner, *Who bears the tax burden?*, The Brookings Institution, Washington, D.C. 1974, Appendix A. This file is regularly updated and the latest available data refer to 1975.

observations. It is only this latter type of information which permits estimates of the income distribution without statistical bias (see section 7).

It was not possible to make the last two types of adjustment. In addition, we consider the second type of adjustment less appropriate, because it is based on a rather limiting assumption. It is difficult to choose between the first and the third type of adjustment, but considering the availability of data, we have opted for Altimir's correction methodology. Since we also wish to estimate the *per capita* and per equivalent distributions, we therefore need a number of tables which cross-classify average incomes (by type of income) and food expenditure (for estimating equivalence scales) by household available income and household size.

As is commonly observed in developed as well as in developing countries, the aggregate household survey estimates are lower than the corresponding totals of National Accounts. The under-estimation is normally highest for net income from self-employment and capital income. The adjustment which Altimir[31] proposes is the following. For income types where aggregate survey estimates are lower than the corresponding National Accounts totals, one adjusts survey income proportionally according to the under-estimation with National Accounts. If the survey total is higher, then one retains the survey data. This can happen (and frequently happens!) in the case of wages and salaries and particularly of imputed rent. The assumption of proportional correction is likely to have little effect for the following sources of income: wages and salaries and transfer income, because they are normally well-captured by the survey. In the case of net income from self-employment, the assumption is reasonable because the under-estimation is the result of the long recall period (normally one year) and some voluntary under-reporting out of fear of taxation. For capital income the difference between National Accounts and aggregated survey data is completely attributed to the highest quintile of the household available income distribution—a fixed proportion of household available income. The reason for the different treatment of capital (or investment) income is that the under-estimation is mainly due to deliberate under-reporting of the highest income groups, while this is not the case for lower income classes. The choice of the highest quintile of households is, however, somewhat arbitrary. It would be better to attribute the under-estimated income to the richest primary income earners and to determine what proportion of them is likely to under-report for each individual country. However, the data are lacking to undertake this type of adjustment and, as a result, we applied the standard procedure to all countries.

The income distribution estimates for the Federal Republic of Germany which are reported in this paper are based on a Social Accounting Matrix constructed by the German Institute of Economic Research (DIW).[32]

The estimates for Mexico and the United Kingdom which are presented in this paper have only been adjusted for inconsistencies with National Accounts. There is reason to believe that the 1968 Mexican Survey under-represents small households. But since the household definition of the 1968 Survey is quite

[31] Altimir, O., op. cit.
[32] Göseke, G. and Bedau, K. D., op. cit.

different from that of the 1970 Population Census,[33] we decided not to adjust the distribution of households by household size. We did, however, "blow-up" the number of persons included in the survey in order to make it match with a recent population estimate. One needs this adjustment in order to compute aggregated household income.

The 1968 Mexican Survey underestimates household available income—as measured by National Accounts—by about 30 percent, while this percentage is about 15 percent for the 1979 Family Expenditure Survey. Wages and salaries and transfers are relatively well measured in both countries (respectively 102.4 and 77.8 percent in Mexico and respectively 94.4 and 84.4 percent in the U.K.). Income from self-employment and from capital, on the other hand, are not well measured (respectively 45.9 and 48.1 percent in Mexico and respectively 50.1 and about 70 percent in the U.K.). The capacity of both surveys to capture these types of income is therefore about the same. However, since in Mexico the proportion of income from self-employment and from capital is much higher than in the U.K., it is likely that the adjustment has a more important effect on measured inequality in Mexico.

5. Main Results

Table 2 shows the main results of this article, i.e. comparable income distribution data for three countries. They are comparable in the sense that the same adjustment procedure was applied to the original data. They do not, therefore, show the best estimates for each individual country. In the case of Mexico, for example, the tape of the 1968 survey was available which enabled us to make any desired tabulations. The estimates shown in table 2 are based on tables which cross-classify household available income (Federal Republic of Germany and Mexico), household gross income (UK) and household size. (1, 2, 3, 4, 5, 6+ for the Federal Republic of Germany and the UK and 1, 2, 3, 4, 5, 6, 7, 8, 9+ for Mexico). Although the tables included about 100 cells, measured income inequality is lower than its "true" value as the value of the highest decile tends to be somewhat underestimated while that of the lowest decile is somewhat overestimated. In the next section we shall analyse the difference in measured inequality based on individual household data and grouped data. A second bias is included in the estimation of the Gini-index. The Gini-index shown in table 2 is estimated on the basis of deciles in order to ensure comparability with decile distributions (to be) generated for other countries. The theoretically correct way would be to calculate the upper and lower bounds of the Gini-index[34] based on the original table and calculate the appropriate average.[35]

[33]For more information, see W. van Ginneken (in collaboration with M. Garzuel), "Comparable Income Distribution Data for Mexico (1968), United Kingdom (1979) and Federal Republic of Germany (1974)," World Employment Research Working Paper, WEP 2-23/105, ILO, Geneva, September 1981.

[34]Mehran, F., "Bounds on the Gini-index based on observed points on the Lorentz curve", *Journal of the American Statistical Association* (Washington), March 1975, Application Section, pp. 64–66.

[35]The correct estimate is the sum of one-third of the lower bound and two-thirds of the upper bound. See F. A. Cowell, *Measuring Inequality*, Oxford, 1977, p. 129., and F. A. Cowell and F. Mehta, "The estimation and interpolation of inequality measures," paper presented at the EADI symposium entitled From Income Distribution Research to Income Distribution Policy in LDCs, Paderborn, April 1981.

TABLE 2

GINI INDICES AND DECILE DISTRIBUTIONS (IN PERCENTAGES). ORIGINAL DATA ADJUSTED FOR NATIONAL ACCOUNTS. FEDERAL REPUBLIC OF GERMANY (1974), MEXICO (1968), UNITED KINGDOM (1979)

	Federal Republic of Germany (1974)			Mexico (1968)						United Kingdom (1979)					
Type of data	Adjusted			Original data			Adjusted			Original data			Adjusted		
Income concept*	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
Welfare unit	Household	Person	Person	Household	Person	Person	Household	Person	Person	Household	Person	Person	Household	Person	Person
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Deciles															
I	2.9	3.6	3.7	1.0	1.0	1.1	0.7	0.8	0.8	2.9	3.9	4.1	2.8	3.9	4.2
II	4.0	4.8	5.0	2.1	2.1	2.2	2.0	1.8	1.9	4.1	5.6	5.7	4.5	5.6	5.7
III	5.0	5.7	5.9	3.0	2.9	3.1	2.7	2.5	2.8	5.4	6.4	6.5	5.5	6.3	6.5
IV	6.0	6.7	6.7	4.1	3.8	4.1	3.5	3.4	3.6	6.9	7.2	7.3	6.9	7.2	7.2
V	7.1	7.4	7.6	5.1	4.9	5.1	4.4	4.3	4.5	8.4	8.1	8.2	8.2	8.0	8.1
VI	8.3	8.7	8.7	6.5	6.3	6.6	5.6	5.8	5.8	9.9	9.3	9.3	9.5	9.1	9.1
VII	9.9	9.9	10.0	8.3	8.2	8.4	7.4	7.6	7.9	11.4	10.4	10.7	10.9	10.5	10.5
VIII	12.0	11.6	11.8	11.2	10.6	11.3	10.5	9.8	10.5	13.1	12.3	12.4	12.5	12.0	12.1
IX	16.0	14.8	14.5	16.4	17.7	16.9	15.9	17.0	16.2	15.6	15.2	14.9	15.5	14.8	14.5
X	28.8	26.8	26.0	42.2	42.4	41.2	47.2	47.1	46.0	22.4	21.5	20.7	23.8	22.6	22.0
Gini-index[1]	0.365	0.320	0.308	0.526	0.535	0.518	0.569	0.574	0.560	0.310	0.267	0.254	0.315	0.273	0.261
Gini-index[2]	0.373	0.328	0.316	0.543	0.551	0.534	0.589	0.594	0.580	0.315	0.271	0.258	0.320	0.278	0.266

[1]based on deciles: lower bound.

[2]based on deciles: one-third of lower bound + two-thirds of upper bound.

*I = Household available income; II = Household available income *per capita*; III = Household available income per equivalent unit.

Note: The decile distributions are calculated on the basis of tables cross-classifying income and household size (about 100 cells).

For comparison, these theoretically more correct estimates have been included in the table.[36] The decile distributions have been estimated with the so-called "portable" method developed by Mehran.[37]

Looking at table 2, one is struck by the fact that the adjustment for National Accounts has fairly little impact on measured inequality. This is particularly so for the U.K. where underestimation towards National Accounts amounts to about 15 percent and where underestimated investment income represents only a small portion of total household available income. In the case of Mexico underestimated income from capital respresents a larger part so that the Gini index of adjusted income is about four percentage points higher than that of unadjusted incomes. The same is true for the highest deciles since the proportion of capital income in that decile is particularly high. For the Federal Republic of Germany the original data are not available.

An unexpected result of table 2 is that the distribution of persons by income *per capita* in Mexico is more unequal than the distribution of households by household income. One would have expected the opposite as is the case for the Federal Republic of Germany and the United Kingdom since the concept of household income *per capita* (and per equivalent unit) seems to imply a certain averaging of income variation. The result for Mexico may be explained by the fact that average income per household is about the same for each household size while it increases strongly with household size for the Federal Republic of Germany and the U.K.

As expected, it is clear that inequality of income in Mexico is significantly higher than it is in the two European countries and that this inequality is somewhat lower in the U.K. than in the Federal Republic of Germany.

6. Transition Matrices

In section 2 we argued that the person is a more appropriate welfare unit than the household. In the previous section we saw that in some cases measured inequality of the distribution of household income is closer to the distribution of persons by equivalent income than the distribution by *per capita* income. With the help of so-called transition matrices we would like to find out in this section which distribution could be considered as second best.

In this paper we define the transition matrix as the distribution of persons (or households)[38] ranked in deciles according to two measures of economic welfare. Given the fact that we consider three such measures (household available income per household, *per capita* and per equivalent unit) there are three transition matrices for each data set.

We are particularly interested to know whether persons stay in the same decile irrespective of the criterion of economic welfare. A convenient way of

[36]In the table we show the average Gini index based on the decile distribution instead of on the original table. Although theoretically incorrect the difference is normally not more than 0.1 of a percentage point and therefore negligible.

[37]Mehran, F., "Dealing with grouped income distribution data," Income Distribution and Employment Working Paper, WEP 2-23/WP 20, ILO, Geneva, August 1975.

[38]We shall not show the transition matrices for households which are normally quite similar to those for persons.

measuring this is the sum of the diagonal elements ("trace") which indicates the percentage of persons that remain in the same decile (see table 3). The trace can be considered as a simplified ranking correlation coefficient.

TABLE 3

SUM OF DIAGONAL ELEMENTS (TRACE) OF VARIOUS TRANSITION MATRICES IN DECILES OF PERSONS (PERCENTAGES)

	Original Data			Adjusted Data		
Income Concept*	I/II	I/III	II/III	I/II	I/III	II/III
Federal Republic of Germany (1974)	n.a.	n.a.	n.a.	17.0	23.8	41.0
Mexico (1968)	34.3	52.0	54.6	41.4	54.8	62.5
United Kingdom (1979)	14.0	18.1	60.6	13.6	17.0	69.3

*Household available income I = per household;

II = per capita;

III = per equivalent unit.

Note: The matrices are calculated on the basis of tables cross-classifying household income and household size (about 100 cells).

Table 3, which is based on grouped data,[39] shows that the trace is highest for the transition matrices combining household available income *per capita* and per equivalent unit. It is therefore justified to conclude that the *per capita* distribution is a "second-best" distribution for the countries under study in this paper. This conclusion depends, however, on two assumptions, first, that the methodology proposed in this paper for estimating equivalence scales is acceptable and, secondly, that the economies of scale factor does not approach zero. As to the first assumption, we have shown that the economies of scale factor calculated for the Federal Republic of Germany, for example,[40] leads to equivalence scales which are remarkably similar to the OECD scales for developed countries. The second assumption is also reasonable because it is very unlikely that household income decreases with increasing household size.

7. STATISTICAL BIAS AS A RESULT OF GROUPED DATA

In the two previous sections we have mentioned in passing that there may be an important difference between inequality measured on the basis of grouped data (in our case tables cross-classifying households and persons by household income and household size) and inequality measured on the basis of individual household observations. We were fortunate to have available the individual household observations of the 1968 Mexican survey, so that we can analyse

[39]For Mexico we also calculated transition matrices on the basis of individual household data. The "traces" calculated on this basis are not much different from those of table 3 (not more than two percentage points).

[40]W. van Ginneken, *op. cit.*

the differences between the two data sets. There are two types of bias that we would like to investigate here. The first relates to the distribution of income *per capita* and per equivalent unit and the second relates to the adjustment of income.

The first type of bias arises when one derives the distribution of income *per capita* and per equivalent unit on the basis of grouped data (see table 2). If one compares these distributions with those based on individual household data one finds a remarkably small difference. The Gini index (lower bound based on deciles) calculated with the individual household data is 0.540 for the *per capita* distribution (compared with 0.535, see column (5) of table 2) and is 0.521 for the per equivalent distribution (compared with 0.518, see column (6) of table 2).

The second type of bias arises when we adjust incomes for inconsistencies with National Accounts without being able to rerank welfare units according to adjusted income. This creates a bias which was recently investigated—in a somewhat different context—by Pyatt, Chen and Fei.[41] If we define adjusted income (y_i^*) as the sum of unadjusted income (y_i) and a correction factor (u_i), i.e. $y_i^* = y_i + u_i$, we are interested to know whether the ranking of adjusted income $r(y_i^*)$ equals that of unadjusted income $r(y_i)$, or whether

$$f(u_i) = r(y_i).$$

In most cases the adjustment is proportional to observed income so that

$$y_{ik}^* = (1 + c_k) y_{ik}$$

where k refers to type of income. The question is now: under what conditions is it true that $r(u_i) = r(y_i)$? Three exceptional cases exist: first, if $u_i = 0$, in other words, if no adjustment takes place; secondly, if $u_i = Cy_i$, i.e. if the adjustment is the same for each factor income; and thirdly, if $r(y_{ik}) = r(y_i)$, i.e. if the ranking of each type of factor income is the same as that for total income.

In the case of Mexico, the last condition is broadly fulfilled for wages and salaries and income from self-employment (looking at deciles). Together these two income sources represent about three-quarters of household income. Since we allocate underestimated income from capital to the highest quintile of households, this is not likely to disturb the ranking either. Finally, the ranking of transfer income is almost opposite to that of total household income and is, therefore, likely to disturb the ranking. However, since transfer incomes represent only five percent of household income, it is not likely to have a great impact (table 4).

If one compares table 4 with columns (7) to (9) of table 2, one sees that the two ways of grouping have a fairly small impact on measured inequality. In table 2 the welfare units have been reranked according to adjusted income, but within a table including about 100 observations. In table 4 (columns (1) to (3)) the welfare units were not reranked and the adjustments were applied to the distribution of households by unadjusted income.

In the case of the United Kingdom, we had two ranking problems. First, the tables available to us show the distribution of households by gross household income. Given the available grouped data it is impossible to estimate correctly

[41] Pyatt, G., Chen, C., Fei, J. "The Distribution of Income by Factor Components," *The Quarterly Journal of Economics*, November 1980, pp. 451–473.

TABLE 4

GINI INDICES AND DECILE DISTRIBUTIONS OF INCOMES ADJUSTED FOR NATIONAL ACOUNTS. TWO RANKING CRITERIA: UNADJUSTED INCOME AND ADJUSTED INCOME. MEXICO (1968)

Ranking Criterion	Unadjusted Income			Adjusted Income		
Income concept*	I	II	III	I	II	III
Welfare Unit	House-hold (1)	Person (2)	Person (3)	House-hold (4)	Person (5)	Person (6)
Deciles						
I	0.8	0.8	0.9	0.7	0.7	0.7
II	2.2	2.0	2.3	2.0	1.9	2.0
III	2.9	2.9	3.0	2.8	2.6	2.8
IV	3.8	3.6	3.9	3.6	3.5	3.6
V	4.7	4.5	4.7	4.5	4.4	4.5
VI	5.9	5.9	6.0	5.7	5.7	5.9
VII	7.4	7.6	7.5	7.5	7.4	7.8
VIII	10.6	9.7	10.4	10.6	9.9	10.3
IX	16.1	16.1	15.7	15.9	15.5	15.7
X	45.5	47.0	45.5	46.7	48.5	46.7
Gini-index	0.551	0.562	0.544	0.564	0.575	0.562

*I: Household available income.
II: Household available income per capita.
III: Household available income per equivalent unit.

the rank correlation between gross and available household income. However, by computing the transition matrix of deciles in gross and in available household income, one is able to estimate the rank correlation between the two types of income approximately. The sum of the diagonal elements (in households) is 96.1 percent which indicates a high correlation. If one computes the Gini indices of available household income ranked by gross and available income, one finds only slightly different Gini indices, i.e. respectively 0.309 and 0.310. One can therefore conclude that the ranking of households by gross income did not significantly disturb the results of table 2.

Secondly, there is a bias which arises when one adjusts various factor incomes for inconsistency with National Accounts. Average factor incomes from most sources (wages and salaries, income from self-employment, investment income, occupational pensions and other income) increase monotonically with household available income (measured in deciles). One factor income (state benefits representing one-eighth of gross income) and three deduction posts (income tax, Social Security and private insurance contributions, representing one-fifth of gross income) would in principle be able to significantly disturb the rank correlation between adjusted and unadjusted income. However, adjusted income still increases sharply with every decile. This is due to the fact that the correlation factor for the three deduction posts and State Benefits are relatively small and that this weight in gross income is significantly lower than that of the five "well-behaving" factor incomes.

8. Conclusion

In this article we tried to generate internationally comparable income distribution data for the Federal Republic of Germany (1974), Mexico (1968) and the United Kingdom (1979). To that end, the same income concept and income unit were adopted for each country, i.e. respectively household available income and the household. Moreover, incomes from various sources were adjusted for inconsistency with National Accounts according to Altimir's methodology. The paper finds that the distribution of persons by household income per equivalent unit is probably the best way of looking at the distribution of economic welfare. It further demonstrated that the distribution of persons by household available income *per capita* is much closer to this "ideal" distribution than the distribution of households by household available income. Finally, the paper discusses some of the problems arising from the fact that one normally works with grouped data. It is found that in the case of the three countries under study, grouping is likely to have had only a small impact on the results.