MODELLING RESEARCH GROUP

DEPARTMENT OF ECONOMICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA
90089-0152

MODELLING RESEARCH GROUP

FACTOR COMPONENTS, INCOME CLASSES AND THE COMPUTATION OF THE GINI INDEX OF INEQUALITY

JACQUES SILBER*

MRG WORKING PAPER #M8620

DEPARTMENT OF ECONOMICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA
90089-0152

FACTOR COMPONENTS, INCOME CLASSES AND THE COMPUTATION OF THE GINI INDEX OF INEQUALITY

JACQUES SILBER*

MRG WORKING PAPER #M8620

May 1986

ABSTRACT

The purpose of this note is to propose a simple technique, based on matrix algebra, to compute the Gini Index of Inequality, to obtain a decomposition of this index by factor component, when detailed data on the various components are available, and to derive a breakdown of the inequality into within and between classes inequality, when the income units are grouped by income range.

*On leave from Bar Ilan University, Israel.

I. INTRODUCTION

Several attempts (Lerman and Yitzhaki, 1984; Brown and Mazzarino, 1984; Berrebi and Silber, 1985) have been made recently to derive algorithms which allow a quick computation of the Gini Index of Inequality (Concentration). Other studies (Fei, Ranis and Kuo, 1978; Kakwani, 1980; Pyatt, Chen and Fei, 1980; Shorrocks, 1982; Shorrocks, 1983) attempted to assign inequality contributions to various components of income such as labor or property income.

This note shows how matrix algebra can be used firstly to compute the Gini Index of Inequality, secondly to obtain a decomposition of this index by factor component, when detailed data on the various components are available, and finally to derive a breakdown of the inequality into within, and between classes inequality, when the income units are grouped by income range.

II. MATRIX ALGEBRA AND THE GINI INDEX

Following the work of Sen (1973) and Donaldson and Weymark (1980) it has been shown (Berrebi and Silber, 1986) that the Gini Index of Inequality I_G could be written as:

$$I_G = \sum_{j=1}^{n} s_j \left[\frac{(n-j)}{n} - \frac{(j-1)}{n} \right] \tag{1}$$

where s_j is the proportion of total income earned by the individual whose income has the j^{th} rank in the income distribution, assuming that

$$s_1 \geq s_2 \geq \ldots \geq s_j \geq \ldots \geq s_n$$

Expression (1) may also be written as

$$I_G = \sum_{i=1}^{n} s_i \left[\sum_{j \geq i} \left(\frac{1}{n} \right) - \sum_{j \leq i} \left(\frac{1}{n} \right) \right] \tag{2}$$

or, in matrix notation as

$$I_G = \hat{e}[D - D']\hat{s} \tag{3}$$

where $\hat{e}$ is a column vector of n elements which are each equal to $\frac{1}{n}$ ($\hat{e}'$ being the corresponding line vector), $\hat{s}$ is a column vector of n elements being respectively equal to $s_1, s_2, \ldots, s_n$ and D is a $n \times n$ lower triangular matrix defined as

$$D = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 1 & \ldots & 1 & 0 \\ 1 & 1 & 1 & \ldots & 1 & 1 \end{bmatrix}$$

whereas D' is the transpose of D.

If one defines a new matrix G, henceforth called the G-matrix, as

$$G = D - D' \tag{4}$$

expression (2) may then be written as

$$I_G = \hat{e}' G \hat{s} \tag{5}$$

Since many computer programs have subroutines for matrix multiplications, expression (5) seems to be a very simple way of computing quickly the Gini Index of Inequality (Concentration).

III. FACTOR COMPONENTS AND THE GINI INDEX

Let us call X_{ji} the factor component i of individual j's income and X_j this individual's total income. The share s_j of individual j mentioned in part II will now be written as $s_{j.}$ with

$$s_{j.} = \frac{X_j}{X_T} \tag{6}$$

where $X_T = \sum_{j=1}^{n} X_j$ whereas the share of component i in society's total income will be defined as

$$s_{.i} = \frac{\sum_{j=1}^{n} X_{ji}}{X_T} \tag{7}$$

Let us call s_{ji} the share of the component i of individual j in total income X_T, that is

$$s_{ji} = \frac{X_{ji}}{X_T} \tag{8}$$

and define a n by $(k+1)$ matrix S, whose first column is the vector $\hat{s}$ of the shares $s_{j.}$, whose second column is the vector $\hat{s}_{.1}$ of the shares s_{j1} and whose $(i+1)^{th}$ column is the vector $\hat{s}_{.i}$ of the shares s_{ji}. It should be clear that the product

$$\hat{e}'GS = \hat{z} \tag{9}$$

is a row vector $\hat{z}$ of $(k+1)$ elements, where the first element is the Gini Index of total income inequality and the next k elements are the contributions of the various k components to total inequality.

It should be stressed that the $(i+1)^{th}$ element z_{i+1} of $\hat{z}$, which is equal to $\hat{e}'G\hat{s}_{.i}$ is neither the Gini Inequality Index G_i of the i^{th} component of income, nor what Fei, Ranis and Kuo (1978) called the Pseudo-Gini $\bar{G}_i$ of the i^{th} component and Rao (1969), Kakwani (1980), Pyatt, Chen and Fei (1980) and Shalit (1985) called the concentration ratio of component i. Using the notations of Pyatt, Chen and Fei (1980) and their theorem A.5 (which is Kakwani's (1980) theorem 8.5) it is easily shown that z_{i+1} is equal to $\phi_i\bar{G}_i$.

It is however very simple to compute these Pseudo Gini Indices: call $\hat{v}_{.i}$ the vector of the ratios $\frac{s_{ji}}{s_{.i}}$ so that the element v_{ji} of $\hat{v}_{.i}$ is the share of individual j in the total income derived from component i. Then the Pseudo Gini $\bar{G}_i$ is simply defined as

$$\bar{G}_i = \hat{e}'G\hat{v}_{.i} \tag{10}$$

To obtain the Gini Inequality Index G_i of component i, one has to construct a new vector $\hat{y}_{.i}$, whose elements $\hat{y}_{.i}$ are the shares v_{ji} previously defined, but they are not ordered according to the order of the shares $s_{j.}$ of vector $\hat{s}$ (as were the elements v_{ji} of $\hat{v}_{.i}$) but according to their own rank in the vector $\hat{y}_{.i}$. It is then easy to see that the Gini Index G_i, for component i, is defined as

$$G_i = \hat{e}' G \hat{y}_{.i} \tag{11}$$

In short our procedure to decompose the Gini Inequality Index among all factors that contribute to income is very simple. It does not require the preliminary computation of concentration ratios (pseudo Ginis). It requires simply the construction of a matrix S whose first column refers to the shares of each individual's total income in society's total income whereas the next columns refer to the shares of each individual's income from a specific component in society's total income. All these vectors are ranked according to the (decreasing) rank of the individuals in total income. The product $\hat{e}'GS$ is then a vector $\hat{z}$ whose first element is the Gini Inequality Index I_G and whose other elements are the contributions of each component to the overall inequality I_G. Naturally such a decomposition of inequality remains subject to the serious criticism put forth by Shorrocks (1982). However if one still decides to use the traditional decomposition of the Gini Index used by Pyatt, Chen and Fei (1980) or by Shalit (1985), the procedure proposed here seems to be the simplest one.

IV. INCOME CLASSES AND THE GINI INDEX

When the data on income distribution are grouped by income class (range), it may be of interest to decompose total inequality into two components, the inequality between income classes and that within income classes. Kakwani (1980) has shown that the Gini Index is equal to the sum of the "between classes" Gini Index and of a weighted average of the "within classes" Gini Indices. It will now be shown that, using again matrix algebra, the contribution to total inequality of each of these Gini Indices can be very easily computed.

Let us partition the $n \times n$ G-matrix previously defined into m^2 submatrices where m is the number of income classes (*e.g.,* deciles) used. In each income class h there are n_h

individuals so that $n = \sum_{h=1}^{m} n_h$. The partitioned matrix G will therefore look like

$$G = \begin{bmatrix} G_{11} & G_{12} & \dots & G_{1m} \\ G_{21} & G_{22} & \dots & G_{2m} \\ \vdots & \vdots & & \vdots \\ G_{m1} & G_{m2} & \dots & G_{mm} \end{bmatrix} \tag{12}$$

It is easily seen that the G_{hh} matrices have 0 on their diagonals, (-1)'s in their upper right triangle and (+1)'s in their lower left triangle. The matrices G_{pq} where $q > p$ have all identical elements equal to (-1) whereas the matrices G_{rt}, where $t < r$, have all identical elements equal to (+1).

Let us similarly decompose each of the vectors $\hat{e}'$ and $\hat{s}$ into m components $\hat{e}'_h$ and $\hat{s}_h$, where each subvector $\hat{e}'_h$ or $\hat{s}_h$ has n_h elements. The product $\hat{e}'G\hat{s}$ defined in (5) may now be written, using well known rules on partitioned matrices, as

$$\hat{e}'G\hat{s} = \sum_{p=1}^{m} \hat{e}'_p \left[\sum_{q=1}^{m} G_{pq}\hat{s}_q \right] \tag{13}$$

Expression (13) may also be written as

$$\hat{e}'G\hat{s} = \sum_{p=1}^{m} \hat{e}'_p G_{pp} \hat{s}_p + \sum_{p=1}^{m} \hat{e}'_p \left[\sum_{q \neq p}^{m} G_{pq}\hat{s}_q \right] \tag{14}$$

The first element on the right hand side of (14) gives evidently the "within class" contribution I_w to the Gini Index whereas the second element corresponds to the "between classes" contribution I_B. It is easily seen that $\hat{e}'_p G_{pq} \hat{s}_q$ may be written (assuming $q < p$) as:

$$\underbrace{\left(\frac{1}{n} \cdots \frac{1}{n}\right)}_{n_p \text{ elements}} \underbrace{\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}}_{n_p \times n_q \text{ matrix}} \underbrace{\begin{pmatrix} s_{1q} \\ \vdots \\ s_{n_q q} \end{pmatrix}}_{n_q \text{ elements}} \tag{15}$$

$$\leftrightarrow \hat{e}'_p G_{pq} \hat{s}_q = \frac{n_p}{n} \sum_{j=1}^{n_q} s_{jq} \tag{16}$$

If we call $\bar{s}_q$ the average share $\sum_{j=1}^{n_q} s_{jq}/n_q$, (16) may be written as:

$$\hat{e}'_p G_{pq} \hat{s}_q = \frac{n_p}{n} n_q \bar{s}_q \tag{17}$$

One then verifies easily that the total between classes contribution I_B may be written as

$$I_B = \sum_{p=1}^{m-1} \sum_{q>p} \frac{n_p}{n} n_q \bar{s}_q - \sum_{p=2}^{m} \sum_{q<p} \frac{n_p}{n} n_q \bar{s}_q \tag{18}$$

One may notice that in the specific case where $n_p = n_q = 1 \;\forall\; p$ or q (there is no within inequality because there is only one individual per class) $\bar{s}_q$ is equal to the share s_q defined in part I and I_B will be written as

$$I_B = \frac{1}{n} \sum_{p=1}^{n} \left[\sum_{q>p} s_q - \sum_{q<p} s_q \right] \tag{19}$$

$$\leftrightarrow I_B = \sum_{q=1}^{n/2} (s_q - s_{n-q+1}) \left(\frac{n-q+1}{n} - \frac{q}{n} \right) \tag{20}$$

an algorithm for the computation of the Gini Index which was proposed by Berrebi and Silber (1986).

To summarize, when there are m different income classes, giving m subvectors $\hat{s}_q$ of the shares vector $\hat{s}$, one obtains the within classes contribution I_w to income inequality by computing the sum

$$I_w = \sum_{q=1}^{m} (\hat{e}'_q G_{qq} \hat{s}_q) \tag{21}$$

where $\hat{e}'_q$ and $\hat{s}_q$ were previously defined and G_{qq} is a n_q by n_q G-matrix (n_q is the size of the vectors $\hat{e}'_q$ and $\hat{s}_q$). The "between classes" contribution is naturally obtained by substracting I_w from I_G where I_G is given by expression (5).

V. CONCLUSION

The present note suggested a simple technique based on matrix algebra, for computing the Gini Index of Inequality. This technique seems also to facilitate the decomposition of the Index by factor components when data on the various sources of individual income are given, as well as the breakdown of the Index in between and within classes contributions when income data are classified by income ranges.

REFERENCES

Berrebi, Z.M. and Silber, J. (1985). 'Income Inequality Indices and Deprivation: A Generalization,' *Quarterly Journal of Economics*, Vol. 99, pp. 807-10.

Berrebi, Z.M. and Silber, J. (1986). 'Dispersion, Asymmetry and the Gini Index of Inequality,' mimeo, The University of Southern California.

Brown, T.A. and Mazzarino, G. (1984). 'Drawing the Lorenz Curve and Calculating the Gini Concentration Index from Grouped Data by Computer,' *Oxford Bulletin of Economics and Statistics*, Vol. 47, pp. 273-8.

Donaldson, D. and Weymark, J. (1980). 'A Single Parameter Generalization of the Gini Indices of Inequality,' *Journal of Economic Theory*, Vol. 22, pp. 67-87.

Fei, J.C.H., Ranis G. and Kuo, S.W.Y. (1978). 'Growth and the Family Distribution of Income by Factor Components,' *Quarterly Journal of Economics*, Vol. 92, pp. 17-53.

Kakwani, N.C. (1980). *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. Oxford University Press: London.

Lerman, R.I. and Yitzhaki, S. (1984). 'A Note on the Calculation and Interpretation of the Gini Index,' *Economics Letters*, Vol. 15, pp. 363-8.

Pyatt, G., Chen, C. and Fei, T. (1980). 'The Distribution of Income by Factor Components,' *Quarterly Journal of Economics*, Vol. 94, pp. 451-73.

Rao, V.N. (1969). 'Two Decompositions of Concentration Ratios,' *Journal of the Royal Statistical Society*, series A, Vol. 132, pp. 418-25.

Sen, A. (1973). *On Economic Inequality*. Oxford University Press: London.

Shalit, H. (1985). 'Calculating the Gini Index of Inequality for Individual Data,' *Oxford Bulletin of Economics and Statistics*, Vol. 47, pp. 185-9.

Shorrocks, A.F. (1982). 'Inequality Decomposition by Factor Components,' *Econometrica*, Vol. 50, pp. 193-211.

Shorrocks, A.F. (1983). 'The Impact of Income Components on the Distribution of Family Incomes,' *Quarterly Journal of Economics*, Vol. 97, pp. 311-26.

REPORTS IN THIS SERIES

M8528: REZAUL KHANDKAR: The Effect of Temporary Versus Permanent Job Separation on the Reservation Wage and Job Search Strategy of the Unemployed.

M8529: ROBERT KALABA and LEIGH TESFATSION: Automatic Differentiation of Functions of Derivatives.

M8530: MICHAEL J. P. MAGILL and WAYNE J. SHAFER: Equilibrium and Efficiency in a Canonical Asset Market Model.

M8531: DENNIS J. AIGNER and PIETRO BALESTRA: On the Optimal Experimental Design for Error Commponents Models.

M8532: ROGER GUESNERIE: Stationary Sunspot Equilibria in an N Commodity World.

M8601: JOHN M. VEITCH: Repudiations and Confiscations by the Medieval State.

M8602: MARIO D. TELLO: Imperfect International Markets, Multinational Enterprises and Manufactured Exports from Developing Countries.

M8603: JEFFREY B. NUGENT: Arab Multinationals: Problems, Potential and Policies.

M8604: RICHARD H. DAY and T.-Y. LIN: Irregular Fluctuations and Comparative Policy Dynamics in the Keynesian Regime.

M8605: RICHARD M. GOODWIN: The Economy as an Evolutionary Pulsator.

M8606: Z.M. BERREBI and J. SILBER: Interquantile Differences, Income Inequality Measurement and the Gini Concentration Index.

M8607: JOSEPH G. HIRSCHBERG: A Comparison of Three Flexible Cost Functions Using Establishment Level Electricity Use Data.

M8608: IAN E. NOVOS and MICHAEL WALDMAN: Complementarity and Partial Nonexcludability: An Analysis of the Software/Computer Market.

M8609: ARNOLD ZELLNER: Further Results on Bayesian Minimum Expected Loss (MELO) Estimates and Posterior Distributions for Structural Coefficients.

M8610: ARNOLD ZELLNER: Bayesian Analysis in Econometrics.

M8611: ARNOLD ZELLNER: A Tale of Forecasting 1001 Series: The Bayesian Knight Strikes Again.

M8612: ARNOLD ZELLNER: Biased Predictors, Rationality and the Evaluation of Forecasts.

M8613: ARNOLD ZELLNER: Bayesian Estimation and Prediction Using Asymmetric Loss Functions.

M8614: RICHARD H. DAY, SUDIPTO DASGUPTA, SAMAR K. DATTA, and JEFFREY B. NUGENT: A Note on Instability in Rural-Urban Migration.

M8615: DENNIS J. AIGNER, FRIEDRICH SCHNEIDER and DAMAYANTI GHOSH: Me and My Shadow: Estimating the Size of the U.S. Underground Economy from Time Series Data.

M8616: ROBERT KALABA and LEIGH TESFATSION: Flexible Least Squares.

M8617: JOHN M. VEITCH: Wavering Virgins, the Deficit and Gramm-Rudman.

M8618: TIMUR KURAN: Chameleon Voters and Public Choice.

M8619: RICHARD H. DAY: The Evolving Economy.

M8620: JACQUES SILBER: Factor Components, Income Classes and the Computation of the Gini Index of Inequality.

M8621: MARIO D. TELLO: Manufactured Exports from Developing Countries and Government Policy Under Imperfect International Markets: Corporate Income Taxes and Exchange Rate.

M8622: DANIEL F. SPULBER: The Second Best Core.

M8623: DANIEL F. SPULBER: Variety and Efficiency: Monopolistic Competition with Nonlinear Price Schedules.