Fei, John C. H.; Ranis, Gustav

**Working Paper**

# Income Inequality by Additive Factor Components

Center Discussion Paper, No. 207

**Provided in Cooperation with:**
Yale University, Economic Growth Center (EGC)

*Suggested Citation:* Fei, John C. H.; Ranis, Gustav (1974) : Income Inequality by Additive Factor Components, Center Discussion Paper, No. 207, Yale University, Economic Growth Center, New Haven, CT

This Version is available at:
https://hdl.handle.net/10419/160135

ECONOMIC GROWTH CENTER

YALE UNIVERSITY

Box 1987, Yale Station
New Haven, Connecticut

CENTER DISCUSSION PAPER No. 207

# INCOME INEQUALITY BY ADDITIVE FACTOR COMPONENTS

John C.H. Fei and Gustav Ranis

June 1974

# INCOME INEQUALITY BY ADDITIVE FACTOR COMPONENTS

## Introduction

The inequality of family income distribution has long been recognized as an important social problem in both the rich and poor countries. But only recently has there been an increasing awareness that our basic understanding of the relationship between growth and income distribution must be substantially improved if both LDC and donor policies are to be in a position to minimize the conflict between these objectives. Our aim is a fairly ambitious deterministic theory of income distribution, but we must try to approach it one step at a time. The first step, focused on in this paper, is the evolution of a framework of measurement which will facilitate the pursuit of a causal analysis underlying the determination of the distribution of income.

As is well known, economists have long learned to measure the degree of income distribution inequality, with the use of the Lorenz Curve and the Gini Coefficient, at least since the beginning of this century. And yet while intuitive notions abound, there exists, at the present time, no satisfactory theory which explains how this inequality is being determined. In short, our knowledge on this subject is still largely in the state characterized by "measurement without theory". What is missing is a framework of data processing based on preconceived theoretical ideas, so that the empirical evidence (i.e., the statistical data) can be scrutinized more efficiently and be made more conducive to the requirements of a positive theory of the determination of income distribution inequality. The present paper aims to construct such a framework.

A respectable theory of income distribution inequality must give prominent recognition to three facets of the problem. First, family income consists of different types of factor income (e.g., wage, property and transfer income); thus, the inequality of total family income is attributable to the inequality of factor incomes, for example, property income. Second, factor income can, in turn, be traced to the distribution of factor ownership (e.g., family ownership of capital and labor force) and factor prices (e.g., wage rates and rates of return to capital). Finally, factor ownership and factor prices change mainly as a result of economic development. This chain of reasoning suggests that income distribution in LDCs is essentially a growth-relevant and growth-sensitive issue. The recent empirical findings of Kuznets and Adelman, to the effect that "income distribution must get worse before it gets better" lend support to this viewpoint.

Two types of questions may thus be raised. First, how, or, in what way do the factor distribution and factor prices affect the total family income distribution? Second, how does economic development affect factor distribution and factor prices? Our paper addresses mainly the first question--for unless we have a firm answer to this question, it is obviously premature even to raise the second question. It should be noted that the second question belong properly to the domain of theories of economic development. This paper aims to build a bridge between income distribution theory and development theory.

An exploration of the first question could start with the intuitively obvious point that the very unequal distribution of a particular factor

income (e.g., wage income) can contribute to total inequality especially when that factor income also accounts for a large distributive share of national income (e.g., 65%). However, imagine a case when the property income is concentrated very heavily in the hands of a few very wealthy families,(i.e., very unequally distributed) and where government welfare payments are made mainly to a small number of very poor families, thus, also very unequally distributed. Thus, the unequal factor distribution in the former (i.e., property income) contributes to the overall inequality while the opposite is true in the latter case. This unequal distribution of welfare payments favoring the poor can contribute to overall equality. To put it differently, the unequal distribution of factor income can contribute to inequality or equality depending upon the degree of correlation between the factor income and the total income. The more precise meaning of this relationship will be explored below.

In order to fully assess the effect of a particular factor income distribution on total family income distribution, at least three dimensions are involved: the relative size of the factor's distributive share, the degree of inequality of factor income distribution, and the degree of correlation between factor income and total income. This paper will design a statistical method for decomposing these effects. This method will allow us to trace the total family income inequality to the various factor components (e.g., wage, property, and transfer income), and will be applied to the case of Taiwan as a pilot empirical effort.

Turning to the second issue, the impact of economic development on income distribution can be analyzed in terms of the above (three) effects.

The economist is familiar with the impact of growth on distributive shares, e.g., through analysis of the elasticity of substitution and, to this extent, our analysis is directly tied to growth theory. The impact of growth on factor distribution inequality and the correlation of factor with total income distribution are less familiar. As we shall see, many of the ambivalent and intuitive notions of how "growth" affects "income distribution" really allude to "these" unfamiliar effects. Our paper finally tries to indicate the direction of future research on these relatively unfamiliar issues.

In Section I, we introduce the familiar Gini Coefficient as a measure of total income distribution inequality. We also introduce two less familiar, but useful concepts, one the so-called Pseudo-Gini Coefficient, and the other the Gini Error term. In Section II, we formally introduce the various components of total family income and analyze the contribution of various factor incomes to total income inequality using the concepts of Section I. In Section III and IV, we analyze the nature as well as the magnitude of the inequality effect and the correlation effect of factor components on total income distribution. The empirical application of these methods to the case of Taiwan is presented in Section IV. Finally, the significance of our analysis for linking income distribution theory to growth theory is introduced in Section VI. The conclusion is presented in Section VII. Some of the technical details are relegated to the Appendices.

## SECTION I

## Gini Coefficient and Gini Error

Let there be n-families, a pattern of family income distribution is denoted by a non-negative row vector Y:

$$Y = (Y_1, Y_2, \ldots Y_n), \quad Y_i \geq 0, \; i = 1, 2, \ldots n \tag{1.1}$$

When Y is normalized, we have

$$\text{a)} \quad y = (y_1, y_2, \ldots y_n) \quad \text{where} \tag{1.2}$$

$$\text{b)} \quad y_i = Y_i / S_y \quad i = 1, 2, \ldots n \quad \text{and}$$

$$\text{c)} \quad S_y = Y_1 + Y_2 + \ldots + Y_n$$

$$\text{d)} \quad y_1 + y_2 + \ldots + y_n = 1$$

In (1.2) $S_y$ is the national income and $y_i$ is the fraction of national income received by the i-th family.

When we use the Gini Coefficient as a measure of the degree of inequality of Y, the fractions $(y_i)$ must be rearranged in a monotonically non-decreasing order. In other words, there is a permutation $(i_1, i_2, \ldots i_n)$ of the first n-integers (1, 2, ... n) such that the following condition is satisfied:

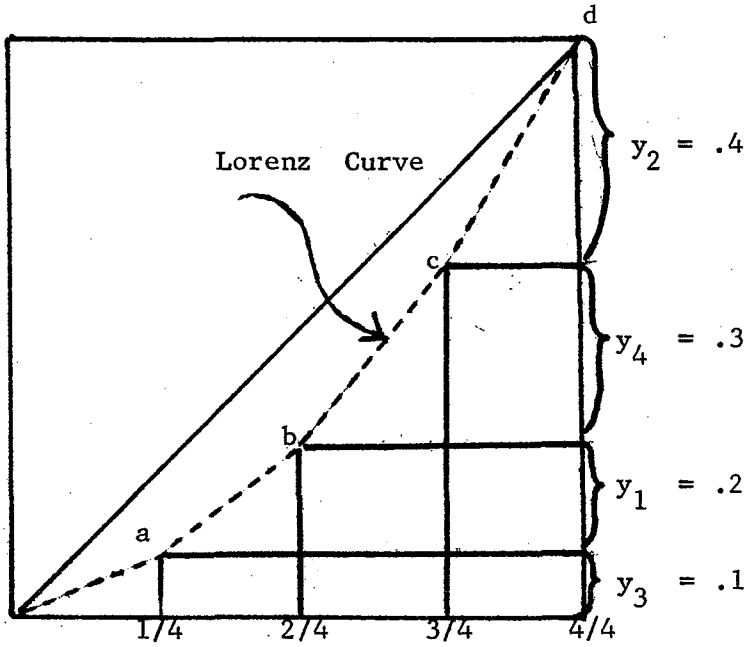
$$y_{i_1} \leq y_{i_2} \leq \ldots \leq y_{i_n} \tag{1.3}$$

A Lorenz Curve is a real-valued function, defined on (1/n, 2/n, ... n/n) such that

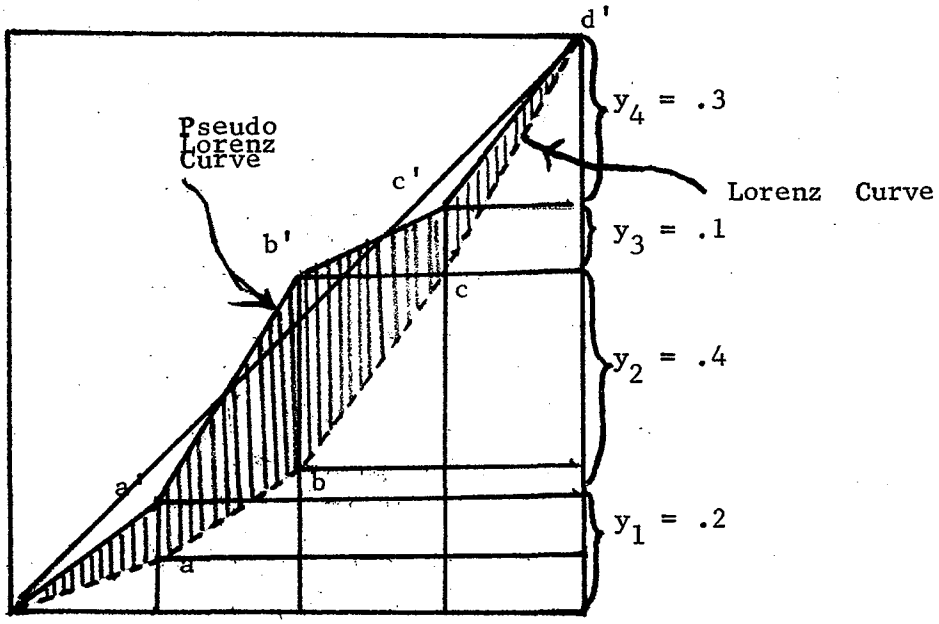
$$L_Y(j/n) = y_{i_1} + y_{i_2} + \ldots + y_{i_j} \quad j = 1, 2, \ldots n \tag{1.4}$$

This is illustrated by the following example.

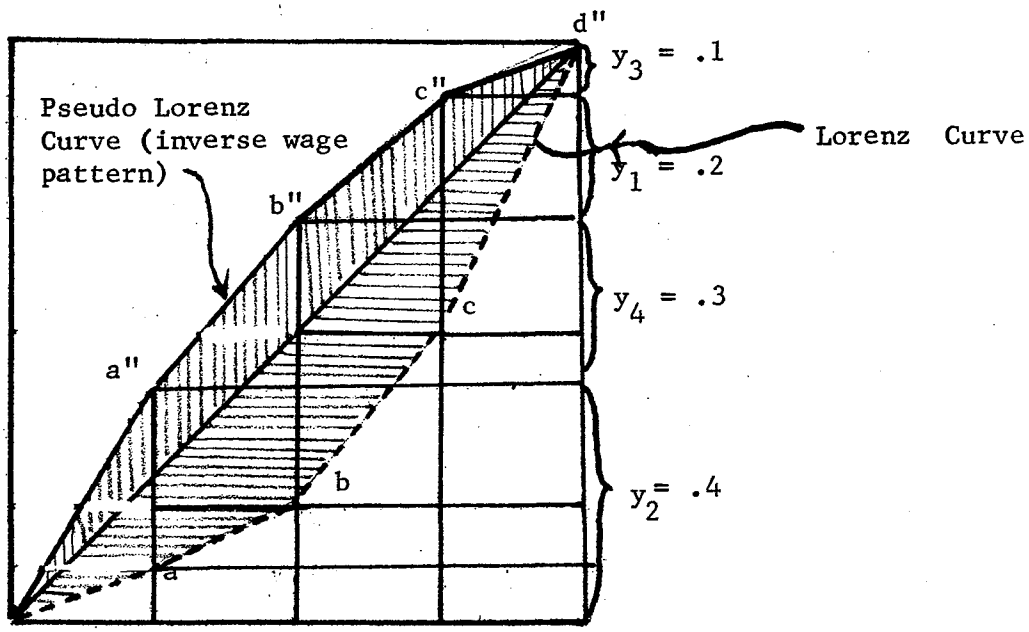
(1a)


d
Lorenz Curve
c
b
a
$y_2$ = .4
$y_4$ = .3
$y_1$ = .2
$y_3$ = .1
1/4
2/4
3/4
4/4

(1b)


d'
Pseudo Lorenz Curve
c'
b'
c
a'
b
a
$y_4$ = .3
Lorenz Curve
$y_3$ = .1
$y_2$ = .4
$y_1$ = .2

(1c)


d"
Pseudo Lorenz Curve (inverse wage pattern)
c"
b"
c
a"
b
a
$y_3$ = .1
Lorenz Curve
$y_1$ = .2
$y_4$ = .3
$y_2$ = .4

Example 1

Let there be four families (i.e., n = 4); and, let Y = (20, 40, 10, 30) with y = ($y_1$ = .2, $y_2$ = .4, $y_3$ = .1, $y_4$ = .3). Then, in a monotonically non-decreasing order,

$$(y_{i_1}, y_{i_2}, y_{i_3}, y_{i_4}) = (y_3 = .1, y_1 = .2, y_4 = .3, y_2 = .4).$$

In diagram 1a, these income fractions are marked off on the vertical axis of a unit square with the **positive** fractions, 1/4, 2/4, 3/4, 4/4, marked off on the horizontal axis. The Lorenz Curve is abcd.

Let the area under the Lorenz Curve be denoted by B. The Gini Coefficient is defined as 1 - 2B. Thus:

$$G_y = 1 - 2B \tag{1.5}$$

where B is the area under the Lorenz Curve of Y.

**We then have the following theorem:**

Theorem 1.1 The Gini Coefficient of Y (1.1) is

(a) $G_y = au_y - b$, where (1.6)

(b) $a = 2/n$; $b = (n + 1)/n$, and

(c) $u_y = \lambda_1 y_{i_1} + \lambda_2 y_{i_2} + \ldots + \lambda_n y_{i_n}$, where

(d) $\lambda_1 = 1, \lambda_2 = 2, \ldots \lambda_n = n$

Proof: The area above the Lorenz Curve is

$$1 - B = (y_1/2 + y_2 + \ldots + y_n)/n + (y_2/2 + y_3 + \ldots + y_n)/n + \ldots (y_n/2)/$$

$$= (y_1 + y_2 + \ldots + y_n)/n + (y_2 + y_3 + \ldots + y_n)/n + \ldots + y_n/n -$$

$$(y_1 + y_2 + \ldots y_n)/2n$$

$$= (1y_1 + 2y_2 + \ldots + ny_n)/n - 1/2n \quad \text{by (1.2d)}$$

$$G_y = 1 - 2B = 1 + 2\left[\sum_{i=1}^{n} iy_i)/n - \frac{1}{2n} - 1\right] \text{ Q.E.D.}$$

There is a perfectly natural economic interpretation of $u_y$ in (1.6c), namely, that it is the **weighted average of the ranks of families ($\lambda_i$),** weighted by the income fractions ($y_{i_j}$). Thus $u_y$ will be called the rank index of Y. Theorem 1.1 states that the Gini **Coefficient is a** linear transformation of the rank index of Y. Thus a higher value of the rank index implies a higher value of $G_y$, i.e., a more unequal pattern of income distribution.

When a Y (1.1) is given, we may also define a Pseudo-Lorenz Curve which is similar to the formal definition of a Lorenz Curve (1.4) except that the income fractions $y_i$ in (1.2a) are not arranged in a monotonically non-decreasing order. Thus, formally, the Pseudo-Lorenz Curve is defined as:

$$L_Y^S (j/n) = y_1 + y_2 + \ldots + y_j \quad j = 1, 2, \ldots n \qquad (1.7)$$

This is illustrated in the following example:

Example 2

For the pattern of income distribution given in Example 1, the Pseudo-Lorenz Curve is a'b'c'd' in diagram 1b.

**When we denote the area under the Pseudo-Lorenz Curve as $\bar{B}$, a Pseudo-Gini Coefficient can also be defined in a way similar to (1.5), namely:**

$$\bar{G}_y = 1 - 2\bar{B} \qquad (1.8)$$

where $\bar{B}$ is the area under the Pseudo-Lorenz Curve. **We can then deduce:**

Theorem 1.2, **i.e., the Pseudo-Gini Coefficient of Y (1.8) is**

$$\text{(a)}\quad \bar{G}_y = a\bar{u}_y - b, \text{ where} \tag{1.9}$$

$$\text{(b)}\quad \bar{u}_y = \lambda_1 y_1 + \lambda_2 y_2 + \ldots + \lambda_n y_n$$

and where a, b and $\lambda_i$ are defined in (1.6b - d).

The Gini-Error of Y is simply defined as the difference between the Gini Coefficient and the Pseudo-Gini Coefficient:

$$\varepsilon_y = G_y - \bar{G}_y = 2(u_y - \bar{u}_y)/n \ \ldots \text{ by 1.6a and 1.9a} \tag{1.10}$$

In Section IV, we shall prove that the Gini Error is always non-negative; hence, the Gini Coefficient is always larger than the Pseudo-Gini Coefficient. It follows from this definition that:

$$\varepsilon_y = (1 - 2B) - (1 - 2\bar{B}) = 2(\bar{B} - B) \tag{1.11}$$

which leads to the following geometric interpretation of the Gini Error:

Theorem 1.3 The area between the Pseudo-Lorenz Curve and the Lorenz Curve is one-half the value of the Gini Error.

In diagram 1b, the dotted Lorenz Curve of diagram 1a is reproduced, and the shaded area is one-half the value of the Gini Error. Notice that when condition (1.3) is satisfied to begin with, we have the following corollary:

Corollary 1.1 If the income distribution pattern Y(1.1) satisfies the monotonic condition (1.3), then the Pseudo-Gini Coefficient equals the Gini Coefficient ($G_y = \bar{G}_y$) and the Gini Error **equals 0, ($\varepsilon_y = 0$)**

**In this section we shall state one more very obvious theorem:**

Theorem 1.4 If $X = kY$, $(k > 0)$ (i.e., X is a scalar multiple of Y) then $G_x = G_y$, $\bar{G}_x = \bar{G}_y$ and $\varepsilon_x = \varepsilon_y$

This theorem states that the Gini coefficient, the Pseudo-Gini Coefficient and the Gini Error depend only on the proportionality of the family income distribution, y(1.2b) and that the overall level of income is irrelevant. We shall make use of this theorem in a later section.

We shall now explore the economic significance of the Pseudo-Gini Coefficient and the Gini Error when the income pattern (Y) has several components.

## SECTION II

## Factor Components of Total Income

Suppose income $Y_i$(1.1) is classified into several types such as Wage Income, $W_i$, Property Income, $R_i$, and Transfer Income, $T_i$. Then:

$$Y_i = W_i + R_i + T_i, \qquad i = 1, 2, \ldots n \tag{2.1}$$

The distributive shares of the total income $S_Y$ may be denoted by

$$\begin{aligned} &\text{(a)}\quad \phi_w = S_w/S_y \text{ (wage share)};\ \phi_r = S_r/S_y \text{ (property share)};\\ &\quad\ \text{and}\ \ \phi_t = S_t/S_y \text{ (transfer share)}\\ &\text{(b)}\quad S_w = \sum_{i=1}^{n} W_i;\ S_r = \sum_{i=1}^{n} R_i;\ S_t = \sum_{i=1}^{n} T_i\\ &\text{(c)}\quad \phi_w + \phi_r + \phi_t = 1 \end{aligned} \tag{2.2}$$

Then, the i-th family receives the following shares of income for each factor:

$$w_i = W_i/S_w;\quad r_i = R_i/S_r;\quad t_i = T_i/S_t,\quad i = 1, 2, \ldots n \tag{2.3}$$

We then readily see the following:

Theorem 2.1 The total income share of the i-th family is the weighted average of the factor shares, i.e.,

$$y_i = \phi_w w_i + \phi_r r_i + \phi_t t_i,\quad i = 1, 2, \ldots n \tag{2.4}$$

where the weights are the national factor distributive shares, (2.2a)

Proof: $y_i = (W_i + R_i + T_i)/S_y$

$= (S_w/S_y)(W_i/S_w) + (S_r/S_y)(R_i/S_r) + (S_t/S_y)(T_i/S_t)$ Q.E.D.

In the rest of this paper, we shall assume that the total family income, $Y_i$(1.1), is arranged in a monotonically non-decreasing order, i.e.:

$$Y_1 \leq Y_2 \leq \ldots Y_n \tag{2.5}$$

so that the first family is the poorest and the last family is the wealthiest. Under this convention, the factor incomes in (2.1) do not always satisfy the following monotonic conditions, i.e.:

$$\text{(a)}\quad W_1 \leq W_2 \leq \ldots \leq W_n \tag{2.6}$$

$$\text{(b)}\quad R_1 \leq R_2 \leq \ldots \leq R_n$$

$$\text{(c)}\quad T_1 \leq T_2 \leq \ldots \leq T_n$$

In general, property income is probably more likely to satisfy (2.6b), i.e., property income is more likely to be concentrated among the wealthier families, while transfer income is the least likely to satisfy (2.6c). Thus, for each factor component, we can define its Gini Coefficient, Pseudo-Gini Coefficient and Gini Error. Using the wage income as an illustration, we have, from (1.6)(1.9) and (1.10):

$$\text{(a)}\quad G_w = au_w - b \quad \text{(Gini Coefficient of W)} \tag{2.7}$$

$$\text{(b)}\quad \bar{G}_w = a\bar{u}_w - b \quad \text{(Pseudo-Gini Coefficient of W)}$$

$$\text{(c)}\quad \epsilon_w = G_w - \bar{G}_w \quad \text{or} \quad \bar{G}_w = G_w - \epsilon_w \quad \text{(Gini Error of W)}$$

$$\text{(d)}\quad u_w = \lambda_1 w_{i_1} + \lambda_2 w_{i_2} + \ldots + \lambda_n w_{i_n} \quad (\text{where } w_{i_1} \leq w_2 \ldots \leq w_{i_n})$$

$$\text{(e)}\quad \bar{u}_w = \lambda_1 w_1 + \lambda_2 w_2 + \ldots + \lambda_n w_n$$

The economic interpretation of the Gini Coefficient, $G_w$, is that it measures the degree of the inequality of the distribution of wage income. The economic significance of the Pseudo-Gini Coefficient is seen from the following theorem:

Theorem 2.2 The Gini Coefficient of the total family income ($G_y$) is the weighted sum of factor Pseudo-Gini Coefficients

$$G_y = \phi_w \bar{G}_w + \phi_r \bar{G}_r + \phi_t \bar{G}_t \tag{2.8}$$

where the national distributive shares are the weights.

Proof: By (1.6a)

$$u_y = \sum_{i=1}^{n} \lambda_i y_i = \sum_{i=1}^{n} \lambda_i (\phi_w w_i + \phi_r r_i + \phi_t t_i) \text{ . . . by (2.4)}$$

$$= \phi_w \bar{u}_w + \phi_r \bar{u}_r + \phi_t \bar{u}_t \text{ . . . by (1.9b)}$$

Then, by (1.9a)

$$G_y = au_y - b = a(\phi_w \bar{u}_w + \phi_r \bar{u}_r + \phi_t \bar{u}_t) - b(\phi_w + \phi_r + \phi_t) \text{ . . . by (2.2c)}$$

$$= \phi_w(a\bar{u}_w - b) + \phi_r(a\bar{u}_r - b) + \phi_t(a\bar{u}_t - b)$$

$$= \phi_w \bar{G}_w + \phi_r \bar{G}_r + \phi_t \bar{G}_t \text{ . . . by (1.9)}$$

Q.E.D.

It follows from this theorem that we can estimate the Gini Coefficient of the total income as the sum of two terms:

$$\text{(a)} \quad G_y = H - E \text{ where,} \tag{2.9}$$

$$\text{(b)} \quad H = \phi_w G_w + \phi_r G_r + \phi_t G_t$$

$$\text{(c)} \quad E = \phi_w \varepsilon_w + \phi_r \varepsilon_r + \phi_t \varepsilon_t$$

The term H (2.9b) is the weighted sum of the factor Gini Coefficients and E (2.9c) is the weighted sum of the factor Gini Errors. Notice that when the errors are small, i.e., E is negligible, H can be taken as an approximation of $G_y$. When this is the case, we can use this approximation (H) and say that a factor income contributes to inequality of total family income distribution when the factor is very unequally distributed, i.e., $G_w$, $G_r$, $G_t$ are large, and/or when a factor distributive share is large, **(as indicated earlier).** However, when the Gini Errors are large, i.e., when E is not negligible, this approximation cannot be taken for granted. In the following sections, we shall analyze the forces that determine the direction of change as well as the magnitude of the Gini Error.

The ideas in this section can be illustrated by the following numerical example.

Example 3

In Table 2.1, we illustrate the computational procedure. Rows 1-4 of column I show the total and factor incomes of four families which are the primary data (2.1). Condition (2.5) is satisfied by total income. Total incomes and factor distributive shares (2.2) are in column II. The total and factor income fractions (1.2b and 2.3) are shown in rows 5-8 of column I. Since condition (2.6) is not satisfied, the factor shares are rearranged into a monotonic order as shown in column II. The computation of the $\bar{u}$-index (1.9b) **is** shown in rows 9-13 of column I, while the u-index (rank index of Y, 1.5c) **is** calculated in column II. The Pseudo-Gini Coefficients (1.8a), the Gini Coefficients (1.6a) and the Gini Errors (1.10) are computed in rows 14-18. Equation (2.9) is **calculated as follows:**

## Table 2.1

| | (I) Family Incomes | (II) Total Income | (II) National Factor Shares | (III) |
|---|---|---|---|---|
| 1. | $Y_1 = 10$ $Y_2 = 20$ $Y_3 = 30$ $Y_4 = 40$ | $S_y = 100$ | | |
| 2. | $W_1 = 2$ $W_2 = 5$ $W_3 = 3$ $W_4 = 20$ | $S_w = 30$ | $\phi_w = .30$ | |
| 3. | $R_1 = 3$ $R_2 = 7$ $R_3 = 20$ $R_4 = 15$ | $S_r = 45$ | $\phi_r = .45$ | |
| 4. | $T_1 = 5$ $T_2 = 8$ $T_3 = 7$ $T_4 = 5$ | $S_t = 25$ | $\phi_t = .25$ | |
| 5. | $y_1 = .10$ $y_2 = .20$ $y_3 = .30$ $y_4 = .40$ | Factor Fractions in natural order | | |
| 6. | $w_1 = .07$ $w_2 = .16$ $w_3 = .10$ $w_4 = .67$ | $w_1 = .07$ $W_3 = .10$ $W_2 = .16$ $w_4 = .67$ | | |
| 7. | $r_1 = .07$ $r_2 = .16$ $r_3 = .44$ $r_4 = .33$ | $r_1 = .07$ $r_2 = .15$ $r_4 = .33$ $r_3 = .44$ | | |
| 8. | $t_1 = .20$ $t_2 = .32$ $t_3 = .28$ $t_4 = .20$ | $t_1 = .20$ $t_4 = .20$ $t_3 = .28$ $t_2 = .32$ | | |
| 9. | $\lambda_1 = 1,\ \lambda_2 = 2,\ \lambda_3 = 3,\ \lambda_4 = 4$ | $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3, \lambda_4 = 4$ | | |
| 10. | $\bar{u}_y = 1(.10) + 2(.20) + 3(.30) + 4(.40) = 3.00$ | $u_y = \bar{u}_y = 3.00$ | | |
| 11. | $\bar{u}_w = 1(.07) + 2(.16) + 3(.10) + 4(.67) = 3.37$ | $u_w = 1(.07) + 2(.10) + 3(.16) + 4(.67) = 3.43$ | | |
| 12. | $\bar{u}_r = 1(.07) + 2(.16) + 3(.44) + 4(.33) = 3.03$ | $u_r = 1(.07) + 2(.16) + 3(.33) + 4(.44) = 3.14$ | | |
| 13. | $\bar{u}_t = 1(.20) + 2(.32) + 3(.28) + 4(.20) = 2.48$ | $u_r = 1(.20) + 2(.20) + 3(.28) + 4(.32) = 2.72$ | | |
| | Pseudo-Gini Coefficient | Gini Coefficient | | Gini Error |
| 14. | $\bar{G}_x = a\bar{u}_x - b$, $a = 2/4$, $b = 5/4$ | $G_x = au_x b$, $a = 2/4$, $b = 5/4$ | | $\varepsilon_x = G_x - \bar{G}_x$ |
| 15. | $\bar{G}_y = .5(3.00 - 1.25 = .250$ | $G_y = .5(3.00) - 1.25 = .250$ | | $\varepsilon_y = 0$ |
| 16. | $\bar{G}_w = .5\ (3.37) - 1.25 = .435$ | $G_w = .5(3.43) - 1.25 = .465$ | | $\varepsilon_w = .030$ |
| 17. | $\bar{G}_r = .5(3.03) - 1.25 = .265$ | $G_r = .5(3.14) - 1.25 = .320$ | | $\varepsilon_r = .055$ |
| 18. | $\bar{G}_t = .5(2.48) - 1.25 = -.010$ | $G_t = .5(2.72) - 1.25 = .110$ | | $\varepsilon_t = .120$ |

(a) $H = .30(.465) + .45(.320) + .25(.110) = .311$ (2.10)

(b) $E = .30(.030) + .45(.055) + .25(.120) = .061$

(c) $G_y = H - E = .250$

The value of $G_y$ computed this way is seen to have the same value as those shown in row 15 of Column II in Table 2.1.

## SECTION III

## Gini Error and Rank Correlation

For any particular type of factor income, e.g., wage income, we can distinguish two effects of the impact of factor income distribution on total income distribution inequality because $G_y$ is the weighted sum of the component Pseudo-Gini Coefficients, $\bar{G}_w$, which is the difference of two terms, the factor Gini Coefficient, $G_w$, and the Gini Error, $\varepsilon_w$ (see 2.7c and 2.8). The factor Gini Coefficient term measures the inequality effect on total family income, while the Gini Error measures the correlation effect. In this section, we will show that the Gini Error measures the degree of correlation between the factor income pattern and the total family income pattern.

Definition: Given a wage pattern $w = (w_{j_1}, w_{j_2}, \ldots w_{j_n})$ (2.3), the wage pattern space contains all the wage patterns formed of permutations of $w_1, w_2, \ldots w_n$.

Formally, let $\Omega_i = (j_1, j_2, \ldots j_n)$ be a **permutation of the integers** $(1, 2, \ldots n)$. The $n!$ **permutations of (1, 2, ... n) and the corresponding** wage patterns in the WPS may be denoted as:

$$\begin{aligned} &\text{(a)} \quad \Omega_i = (j_1, j_2, \ldots j_n) \quad i = 1, 2, \ldots n! \\ &\text{(b)} \quad w(\Omega_i) = (w_{j_1}, w_{j_2}, \ldots w_{j_n}) \quad i = 1, 2, \ldots n! \end{aligned} \tag{3.1}$$

Notice that all the wage patterns in the WPS obviously have the same factor Gini Coefficient ($G_w$). They differ from each other **only in respect to** the order in which the wage income fractions of different magnitudes appear.

Thus the wage pattern space contains all conceivable patterns of family wage income which have the same inequality effect (i.e., same value for wage Gini Coefficient, $G_w$, as defined in 2.7a), but, different Gini errors (i.e., different values for $\bar{G}_w$ as defined in 2.7b). Thus the WPS allows us to concentrate on the analysis of the variation of the Gini Error term while isolating the inequality effect. Our immediate task in this section is to show that the Gini Error term reflects a correlation effect between the wage income and the total income of the families. The result of our analysis in this section allows us to interpret $E = \phi_w\varepsilon_w + \phi_r\varepsilon_r + \phi_t\varepsilon_t$ (2.9c) as a weighted average of the "correlation effects" just as the term $H = \phi_w G_w + \phi_r G_r + \phi_t G_t$ (2.9b) can be interpreted as the weighted average of the inequality effects. The terms H and E which affect total income distribution inequality (as seen in 2.9a) are not statistical artifacts but can be interpreted in real economic terms which will be explored in the latter part of this paper.

Two wage patterns in the WPS deserve special emphasis: i.e., the wage patterns associated with the natural and the inverse permutation:

$$\begin{aligned} &\text{(a)}\quad \text{the natural permutation } \Omega_1 = (1, 2, \ldots n) \\ &\text{(b)}\quad \text{the inverse permutation } \Omega_{-1} = (n, n-1, \ldots 1) \end{aligned} \tag{3.2}$$

$$\begin{aligned} &\text{(a)}\quad \text{the natural wage pattern } w(\Omega_1) = (w_1, w_2, \ldots w_n) \\ &\text{(b)}\quad \text{the inverse wage pattern } w(\Omega_{-1}) = (w_{n-1}, w_{n-2}, \ldots w_1) \end{aligned} \tag{3.3}$$

Intuitively, these two wage patterns lie at the "extreme ends" in the "scale of correlation." All the other wage patterns in WPS lie between these two extremes.

Example 3.1

Let n = 4; the 4! = 24 permutations of (1, 2, ... n) are written in the brackets of diagram 3.1. The natural permutation (1, 2, 3, 4) and the inverse permutation (4, 3, 2, 1) are written at the very top and

bottom, respectively. (A bracket in this chart represents a particular permutation (3.1a) as well as a particular wage pattern (3.1b) associated with it, although the latter is not shown explicitly).

Let us adopt the convention defined in (2.5) and assume that the first (last) family is the poorest (wealthiest) from the viewpoint of total income. **Second, we will adopt the convention that the monotonic condition defined in (2.6a) is satisfied by the natural wage pattern (3.3a). Given these conventions, in the inverse wage pattern of (3.3b), the first (last) family receives the highest (lowest) wage income.**

Since it is **assumed that the total family income satisfies the monotonic** condition (2.5), the total income ranks of the n-families are indicated by the natural permutation $\Omega_1$, (3.2a). The natural wage pattern $w(\Omega_1)$ satisfies the monotonic condition, (2.6a). Thus the wage income rank for the wage pattern $w(\Omega_1)$ is $\Omega_1$, and the wage income rank for a wage pattern $w(\Omega_i)$ is $\Omega_i$. Thus, using the Spearman Formula of rank correlation we have:

Definition: The rank correlation between total family income (2.1) and wage pattern $w(\Omega_i)$ is

$$\text{(a)} \quad R' = (\Omega_1, w(\Omega_i)) = 1 - (6 \Sigma d_i^2)/n^3 - n) \text{ where}$$

where (3.4)

$$\text{(b)} \quad d_i^2 = (1 - j_1)^2 + (2 - j_2)^2 + \ldots (n - j_n)^2$$

The ordinary correlation between the wage pattern $w(\Omega_i)$ and the natural wage pattern $w(\Omega_1)$ **is:**

(a) $R = (w(\Omega_1), w(\Omega_i))$

$$\sum_{i=1}^{n} (w_i - \bar{w}_i)(w_{j_i} - \bar{w}_{j_i}) \Big/ \sqrt{\Sigma(w_i - \bar{w}_i)^2 \, \Sigma(w_{j_i} - \bar{w}_{j_i})^2} \quad \text{where} \tag{3.5}$$

(b) $\bar{w}_i = \bar{w}_{j_i} = \frac{1}{n}$

In (3.4), the rank correlation is between the rank of total family income and the rank of wage income, while, in (3.5), the ordinary correlation is between the wage fractions of a particular wage pattern in the WPS and the fractions of **the natural wage pattern. The above will** be **illustrated in the next example.**

Example 4 Referring to diagram 3.1, of example 3, the rank correlation for each wage pattern is indicated by the first number in parentheses next to each permutation. For the computation of the ordinary correlation, the magnitudes of the wage income fractions for this numberical example are $w(\Omega_1) = (w_1 = .05; w_2 = .20; w_3 = .30; w_4 = .45)$. The ordinary correlation coefficient is indicated by the numbers below the rank correlation.

Referring to diagram 3.1, as we move upward in this diagram from the reverse wage pattern to the natural wage pattern, the value of R and R' **generally show an** increasing tendency. The arrows in the diagram are intended to show that the wage pattern in the WPS are incompletely ordered according to the following definition.

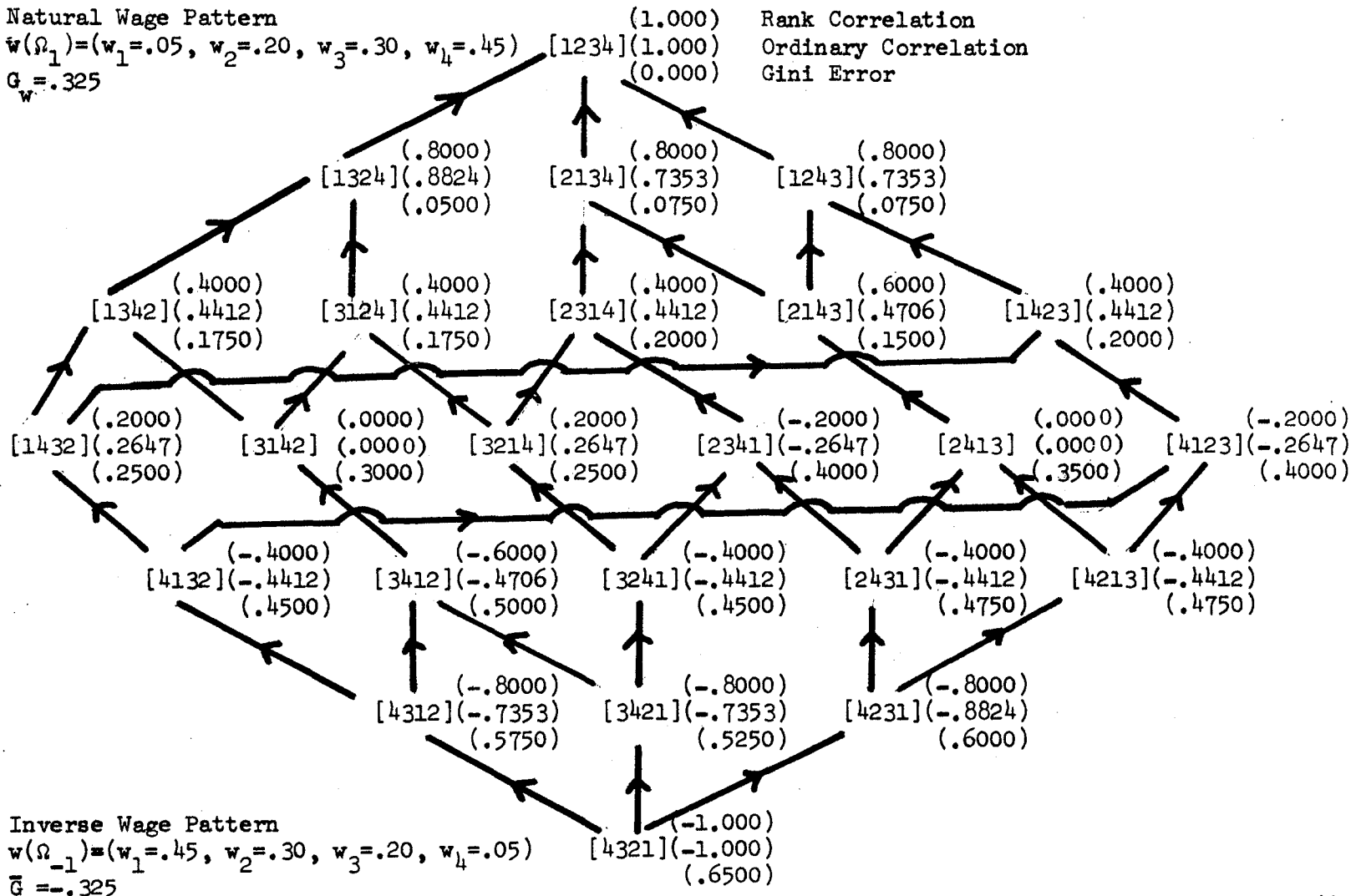
Definition: A permutation $X = (x_1, x_2, \ldots x_n)$ of $(1, 2, \ldots n)$ is a successor of $Z = (z_1, z_2, \ldots z_n)$ in notation $X = S(Z)$ if X can be obtained from Z by interchanging two adjacent integers such that the larger integer is moved to the right in that exchange.

This definition is illustrated by the following example:

Diagram 3.1

## Causal Order Chart

| Order | # of Vertices | # of Directed Edges |
|---|---|---|
| I | II | III |
| 0 | 1 | 0 |
| 1 | 3 | 3 |
| 2 | 5 | 6 |
| 3 | 6 | 9 |
| 4 | 5 | 9 |
| 5 | 3 | 6 |
| 6 | 1 | 3 |
| Total | 24 | 36 |



Example 5

Let Z = (1, 4, 3, 6, 2, 5). Interchange the two underlined adjacent integers by moving the larger integer "6" to the right to obtain X = (1, 4, 3, 2, 6, 5). Thus X is a successor of Z, or X succeeds Z, i.e., X = S(Z).

The successor relation X = S(Z) is shown diagramatically in diagram 3.1 by an arrow which initiates from "Z" and terminates at "X", i.e., the arrows point at the successor of a permutation. When X and Z in the above definition are interpreted as the factor income rank for the wage pattern space, the economic interpretation of X = S(Z) is that the rank of wage pattern w(X) is closer to the rank of the natural wage pattern $w(\Omega_1)$ (which is the same as the rank of the total income pattern) than w(Z). This is illustrated in the following example.

Example 6

Referring to example 4, suppose $w(Z) = (w_1, w_4, w_3, w_6, w_2, w_5) = (.05, .19, .15, .30, .10, .21)$ and $w(X) = (.05, .19, .15, .10, .30, .21)$. The natural wage pattern is $(.05, .10, .15, .19, .21, .30)$. Thus by moving the larger wage fraction ($w_6 = .30$) to the right, the wage pattern w(X) is closer to the natural pattern than w(Z).

Let $\varepsilon_X$ and $\varepsilon_Z$ denote the Gini Error of the wage pattern of X and Z, respectively. We have the following theorem:

Theorem 3.1 If X = S(Z) (i.e., X is a successor of Z) then:

a) The Rank Correlation $R'(\Omega_1, w(X)) \geq R'(\Omega_1, w(Z))$

b) The Ordinary Correlation Coefficient $R(w(\Omega_1), w(X)) \geq R(w(\Omega_1), w(Z))$

c) The Gini Error of $\varepsilon_X = \varepsilon_Z$

The proof of this theorem is given in Appendix I.

This theorem states that if X is a successor of Z, i.e., if the wage pattern w(X) is closer to the natural wage pattern than w(Z), the correlation coefficients (R' and R in 3.4 and 3.5) for X are higher and the value of the Gini Error, $G_w$ is lower than for Z. In this sense, a larger Gini Error corresponds to a lower degree of correlation between the factor income rank and the total income rank.

Since the wage patterns in the wage pattern space (WPS) have the same wage Gini Coefficient ($G_w$), i.e., the same degree of inequality of the distribution of wage income, the economic interpretation of the above theorem is that the wage Gini Error can exercise an additional influence on total income distribution inequality. This influence, which is attributable to the degree of rank correlation between the factor income, (i.e., in this case the wage income) and total income, can either contribute to inequality or contribute to equality. Although property income and welfare (i.e., transfer) income may be both very unequally distributed, i.e., $G_r$ and $G_t$ are large, the property income contributes to the inequality of total income distribution while the transfer income contributes to equality (see above). This is because property income usually has a high positive correlation with total income and hence a low value of Gini Error, (i.e., $\varepsilon_r$ is low) while the transfer income usually has a low, and possibly negative, correlation with that of total income and hence a large Gini Error term (i.e., $\varepsilon_t$ is high). This will be fully explored in the next section and may be summarized here as follows:

Summary: The distribution of factor income (e.g., wage or property or transfer income) exercises two influences on the inequality of family income distribution: (a) the inequality effect as measured by the factor Gini Coefficient, $G_x$ and (b) the correlation effect as measured by the Gini Error, $\varepsilon_x$. The net impact is measured by the Pseudo-Gini Coefficient, $\bar{G}_x = G_x - \varepsilon_x$, which is

the difference between these two effects.

The above is illustrated by another example:

Example 6: In diagram 3.1, the Gini Errors are indicated by the third number beside each permutation. As we move upward along the arrows in this chart, we see that the Gini Error decreases while both the Rank and Ordinary Correlation Coefficients increase.

## SECTION IV

## The Magnitude of the Gini Error and the Pseudo-Gini Coefficient

In the last section, we identified two types of effects, i.e., the inequality effect and the correlation effect, that a factor income distribution may have on total income distribution inequality. In this section, we want to analyze the magnitude, i.e., the absolute value and the sign of the Gini Error and the Pseudo-Gini Coefficient. Analysis of the above will show that factor income distribution can contribute "heavily" or "lightly" to income distribution inequality. It may, in fact, contribute to equality rather than inequality as a result of the magnitude of the Gini Error and the Pseudo-Gini Coefficient. The analysis of this section then provides a rational economic framework for interpreting the decomposition of the empirical data presented in the next section.

Referring to diagram 3.1, we see that as we move upward along a path formed by consecutive arrows, the value of the Gini Error decreases. Since by (corollary 1.1) the magnitude of the Gini Error associated with the natural wage pattern is zero, we have:[1]

Theorem 4.1: The value of the Gini Error of a factor pattern (e.g., $\varepsilon_w$) is always non-negative. The value is zero when the factor pattern is the natural one ($w(\Omega_1)$), i.e., when the factor pattern satisfies the monotonic condition (2.6a).

This theorem is verified by the following example:

---

[1] A rigorous proof of this theorem requires that, in diagram 3.1, every permutation lie on a path linking the inverse permutation with the natural permutation. This will be proved in Appendix I.

**DIAGRAM 4.1**

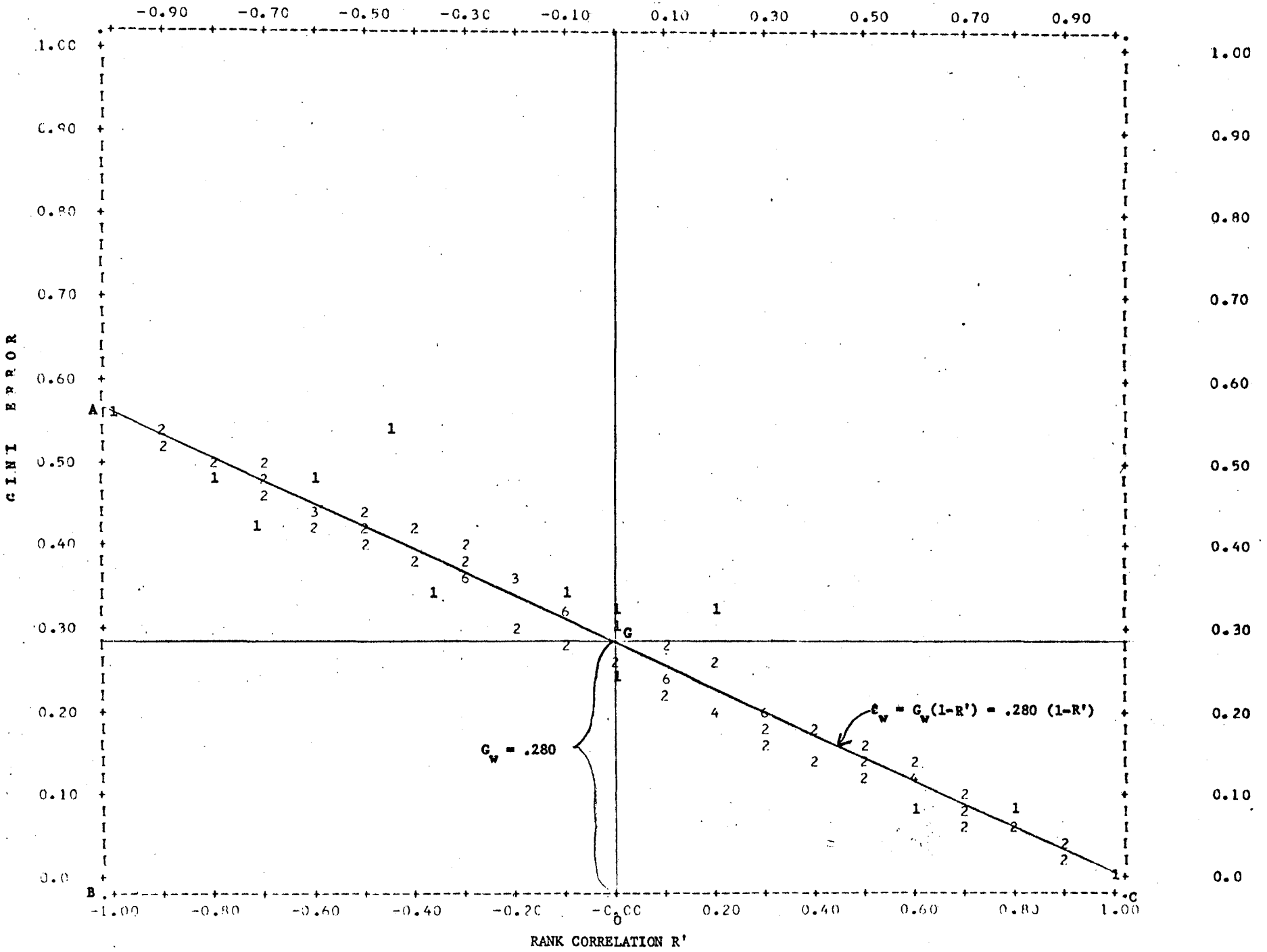
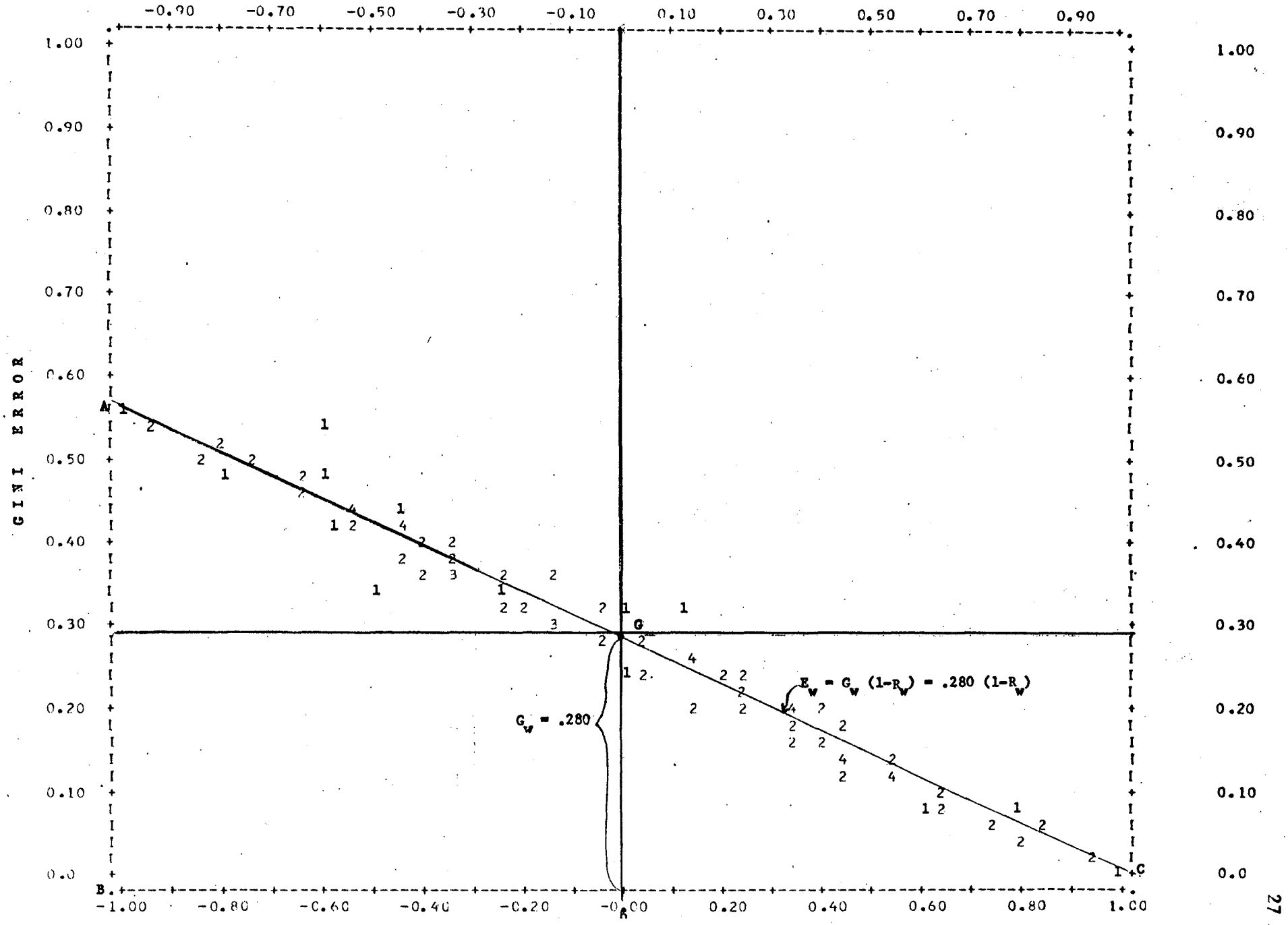

-0.90
-0.70
-0.50
-0.30
-0.10
0.10
0.30
0.50
0.70
0.90
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.0
GINI ERROR
A
B
C
G
O
$G_w = .280$
$e_w = G_w(1-R') = .280\ (1-R')$
-1.00
-0.80
-0.60
-0.40
-0.20
0.00
0.20
0.40
0.60
0.80
1.00
RANK CORRELATION R'

DIAGRAM 4.2


-0.90 -0.70 -0.50 -0.30 -0.10 0.10 0.30 0.50 0.70 0.90
1.00 0.90 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0.0
GINI ERROR
A
B
C
G
$G_w$ = .280
$E_w = G_w (1-R_w) = .280 (1-R_w)$
-1.00 -0.80 -0.60 -0.40 -0.20 0.00 0.20 0.40 0.60 0.80 1.00
R

Example 7:

Referring to table 2.1 and example 3, we see that all the Gini Errors shown in rows16-18 (column III) take on positive values.

Using 2.7c, equation 2.8 can be rewritten as:

$$G_y = \phi_w \bar{G}_w + \phi_r \bar{G}_r + \phi_t \bar{G}_t$$
$$= \phi_w(G_w - \varepsilon_w) + \phi_r(G_r - \varepsilon_r) + \phi_t(G_t - \varepsilon_t) \qquad (4.1)$$

Thus, in estimating $G_y$ (the family income distribution inequality) according to this equation, we have to make a downward adjustment of the factor Gini Coefficients ($G_w$, $G_r$, $G_t$) by corresponding magnitudes of the Gini Error before the weighted average can be taken. This downward adjustment is determined by the degree of correlation between the factor income distribution and the total income distribution. The less the factor income is correlated with the total family income (e.g., the transfer income), the more such an adjustment needs to be made. This may be summarized as follows:

Corollary 4.2 The value of $H = \phi_w G_w + \phi_r G_r + \phi_t G_t$ (2.9b) always overestimates $G_y$.

This is verified by the numerical example summarized in equation 2.10.

To explore the magnitude of the Gini Error further, the scatter diagram between the Rank Correlation Coefficient, R', (3.4a) and the Gini Errors in (2.7c) is shown in diagram 4.1. R', varying between +1 and -1, is measured on the horizontal axis and the Gini Error is measured on the vertical axis. This diagram is constructed for the case when there

are five families, i.e., n = 5, and hence there are 5! = 120 points in the wage pattern space. The numerical values assumed for the natural wage pattern is $(w_1, w_2, w_3, w_4, w_5) = (.05, .15, .20, .25, .35)$.

Notice that as $R'_w$ decreases the value of the Gini Error increases; however, the relation is not an exact one. Instead, there appears to be a linear regression line between $R'_w$ and $\varepsilon_w$. The linear regression in this numerical example has a good fit, with a correlation coefficient between $R'_w$ and $\varepsilon_w$ of .969.[1] Thus, for each value of Rank Correlation R', there is an expected value of the Gini Error satisfying the regression result:

$$\hat{\varepsilon}_w = c + dR'_w \quad , \quad \text{where} \quad R'_w = R'(W(\Omega_1), W(\Omega_i)) \tag{4.2}$$

To find the value of "c" and "d" in this regression equation, we have the following theorem:

Theorem 4.2

a) $\varepsilon_w = 0$ when $R'_w = 1$

b) $\varepsilon_w = 2G_w$ when $R'_w = -1$

The validity of (4.2a) follows directly from Theorem 4.1, namely, for the natural wage pattern, the rank correlation is +1 and the Gini Error is zero. The validity of (4.2b) can be seen in the following way. In diagram 1c, the Pseudo-Lorenz Curve for the inverse wage pattern is the curve a"b"c"d". Since the shaded area above the 45 degree line obviously equals

[1] In Appendix I we shall explain why this is so.

the area below the 45 degree line, it follows from theorem 1.3 that $\varepsilon_w = 2(2A) = 2G_w$ where A is the value of the shaded area below the 45 degree line. This is a rigorous proof of (4.2b).

Using Theorem 4.2ab, and equation 4.2, the values "c" and "d" can be determined as $c = G_w$ and $d = -G_w$, leading to the following regression line:

$$\hat{\varepsilon}_w = G_w (1 - R'_w) \tag{4.3}$$

In diagram 4.1, this equation is represented by the straight line AGC where the distance OG is the Gini Coefficient ($G_w$) and the distance BA is twice the value of $G_w$. In the remainder of this section, we shall assume that the approximation equation (4.3) is valid in the expectations sense. Using this approximation, we have

$$\bar{G}_w = G_w - \epsilon_w = G_w(1 - \varepsilon_w/G_w) = G_w R' \tag{4.4}$$

which shows that the Pseudo Gini Coefficient is the product of the Gini Coefficient $G_w$ and the Rank Correlation Coefficient (R'). Thus, using the notation "≃" to stand for "approximately true", we have:

$$G_y \simeq \phi_w(G_w R'_w) + \phi_r(G_r R'_r) + \phi_t(G_t R'_t) \tag{4.5}$$

From this we readily see that there are two types of cases depending upon the sign of the rank correlation, R', i.e., positive or negative. The conclusion may be summarized as follows:

Conclusion

a) The Pseudo Gini Coefficient is approximately the product of the factor Gini Coefficient and the factor Rank Correlation.

b) The Pseudo-Gini Coefficient is positive (negative) when the rank correlation, R', is positive (negative).

Later, we shall use the approximation equation (4.5) to help us examine the relation between growth and income distribution, with the understanding that all such inferences will only be approximately true.

Up to now in this section we have only made use of the Rank Correlation Coefficient R' as defined in (3.4). All the results are still valid if we use the Ordinary Correlation Coefficient, R, as defined in (3.5), in place of R'. In diagram 4.2, the scatter diagram between $R_w$ and $\varepsilon_w$ is shown. (The same numerical data are used in this diagram as in diagram 4.1.) The same regression equation line (4.3) is fitted for this scatter diagram, which also appears to be a good fit. The correlation coefficient is .974 in this case.

## SECTION V

## EMPIRICAL APPLICATION

The framework which we have developed can now be used to analyze empirical data for total income ($Y_i$) and for factor income ($W_i$, $R_i$, $T_i$) as defined in (2.1). The empirical analysis can provide numerical answers to several types of questions. The first question we can raise is: how good is H (2.9b), i.e., the weighted average of factor Gini Coefficients as an estimator of the family income distribution, $G_y$? Since H always overestimates $G_y$ (Corollary 4.2), we can define the degree of overestimation by:

$$d = (H - G_y)/G_y \quad \text{(degree of overestimation)} \tag{5.1}$$

If "d" is small, we can, for practical purposes, accept the validity of the estimator equation (2.9b). This implies that the correlation effects of factor income distribution can be neglected as a first approximation.

The second question we can raise is: which factor is more responsible for the inequality of income distribution? We can divide equation (4.1) by $G_y$ to obtain:

$$\text{a)} \quad 1 = F_w + F_r + F_t \quad \text{where}$$

$$\text{b)} \quad F_w = \phi_w \bar{G}_w/G_y; \quad F_r = \phi_r \bar{G}_r/G_y; \quad F_t = \phi_t \bar{G}_t/G_y \tag{5.2}$$

$$\text{c)} \quad F_w \simeq \phi_w R'_w G_w/G_y; \quad F_r \simeq \phi_r R'_r G_r/G_y; \quad F_t \simeq \phi_t R'_t G_t/G_y$$

$F_w$, $F_r$, $F_t$ will be referred to as **factor inequality weights**, FIW.

The FIW are the product of the Pseudo-Gini Coefficients ($\bar{G}_x$) and the relative share of national income ($\phi_x$), expressed as a fraction of the total Gini Coefficient. They correspond to the fraction of inequality (i.e., fraction of $G_y$) accounted for by the various factor income components. Thus a factor contributes heavily to total income distribution inequality when its distributive share is large and/or when the Pseudo-Gini Coefficient is large relative to the total Gini Coefficient. Notice that when the Rank Correlation Coefficient, $R'_x$, is negative in (5.2c), the FIW is negative. This means that this factor contributes to equality rather than inequality.

Using equation (2.7c) we can decompose the FIW into an inequality effect and a correlation effect:

$$F_w = \frac{\phi_w G_w}{G_y} - \frac{\phi_w \varepsilon_w}{G_y} \; ; \; F_r = \frac{\phi_r G_r}{G_y} - \frac{\phi_r \varepsilon_r}{G_y} \; ; \; F_t = \frac{\phi_t G_t}{G_y} - \frac{\phi_t \varepsilon_t}{G_y} \qquad (5.3)$$

where the inequality effect is $\phi_x G_x / G_y$ and the correlation effect is $\phi_x \varepsilon_x / G_y$. Notice that the inequality effect and the correlation effect are both non-negative. When the correlation coefficient, R', is positive, the inequality effect outweighs the correlation effect, and both the Pseudo-Gini Coefficient and the FIW are positive. Conversely, (i.e., when $\varepsilon_x > G_x$ and R' is negative), the inequality effect is outweighed by the correlation effect, and both the Pseudo-Gini Coefficient and the FIW become negative.

The third question we can raise is the causation of change of the magnitude of income distribution inequality ($G_y$) in a statistical sense. Using the notation $\eta_x$ to denote the rate of change of the time variable, x, we have from (4.5):

$$(a)\quad \eta_{G_y} = F_w V_w + F_r V_r + F_t V_t \quad \text{where}$$

$$(b)\quad V_w = \eta_{\phi_w} + \eta_{G_w} + \eta_{R'_w}\ ;\ V_r = \eta_{\phi_r} + \eta_{G_r} + \eta_{R'_r}\ ;\ V_t = \eta_{\phi_t} + \eta_{G_t} + \eta_{R'_t} \qquad (5.4)$$

which states that the rate of growth of the Gini Coefficient, $G_y$, is the weighted average of the sum of the rates of growth of $\phi_x$, $G_x$ and $R'_x$. This statistical technique may be first illustrated by a numerical example:

Example 8

In the following numerical example, there are five families. Property income is assumed to concentrate very heavily in the hands of the wealthiest families, while transfer income is assumed to concentrate in the hands of the poorest families.

| | Family 1 | Family 2 | Family 3 | Family 4 | Family 5 | Total |
|---|---|---|---|---|---|---|
| Wage Income | 3 | 1 | 17 | 15 | 9 | 45 |
| (rank) | (2) | (1) | (5) | (4) | (3) | |
| Property Income | 0 | 0 | 2 | 8 | 25 | 35 |
| (rank) | (1) | (2) | (3) | (4) | (5) | |
| Transfer Income | 8 | 12 | 0 | 0 | 0 | 20 |
| (rank) | (4) | (5) | (3) | (2) | (1) | |
| **Total Income** | 11 | 13 | 19 | 23 | 34 | 100 |

The total income is \$100 of which the wage share is 45% ($\phi_w$), the property share is 35% ($\phi_r$) and the transfer share is 20% ($\phi_t$). The factor income rank for each family is indicated by the numbers in brackets.

The **computational results** are shown in Table 5.1 and are self-explanatory. The estimated Gini Coefficient, H, according to (2.9b) is in

Table 5.1

| | | | | Wage (1) | Property (2) | Transfer (3) | Total (4) |
|---|---|---|---|---|---|---|---|
| 1. | | Factor Share | $\phi_x$ | .4500 | .3500 | .2000 | 1.0000 |
| 2. | Gini Coefficient | Factor Gini | $G_x$ | .3912 | .6628 | .6400 | |
| 3. | Gini Coefficient | Weighted Gini | $\phi_x G_x$ | .1760 | .2320 | .1280 | H = .5360 |
| 4. | Gini Coefficient | Factor Inequality Effect | $\phi_x G_x / G_y$ | .7862 | 1.0362 | .5717 | 2.3940 |
| 5. | Pseudo-Gini Coefficient | Pseudo-Gini | $\bar{G}_x$ | .2308 | .6628 | -.5600 | |
| 6. | Pseudo-Gini Coefficient | Weighted Pseudo-Gini | $\phi_x \bar{G}_x$ | .1039 | .2320 | -.1120 | $G_y$ = .2239 |
| 7. | Pseudo-Gini Coefficient | Factor Inequality Weight | $\phi_x \bar{G}_x / G_y$ | .4640 | 1.0362 | -.5002 | FIW=1.0000 |
| 8. | Gini Error | Gini Error | $\epsilon_x$ | .1604 | .0000 | 1.2000 | |
| 9. | Gini Error | Weighted Error | $\phi_x \epsilon_x$ | .0722 | .0000 | .2400 | E = .3122 |
| 10. | Gini Error | Correlation Effect | $\phi_x \epsilon_x / G_y$ | .3225 | .0000 | 1.0719 | d =1.3940 |
| 11. | Rank Correlation R' | Rank Correlation | R' | .5000 | 1.0000 | -.8240 | |
| 12. | Rank Correlation R' | 1-R' | 1-R' | .5000 | .0000 | 1.8240 | |
| 13. | Rank Correlation R' | Gini Error Fraction | $\epsilon_x / G_x$ | .4100 | .0000 | 1.8750 | |
| 14. | Rank Correlation R' | Approximate Error | $1-\left(\frac{\epsilon_x / G_y}{1-R'}\right)$ | .1800 | .0000 | -.0280 | |

row 3. The Pseudo-Gini Coefficients are shown in row 5 and the Gini Coefficient for total family income, $G_y$, as defined in (2.8), is shown in row 6. The Factor Gini Errors, as defined in (2.7c), are shown in row 8. The E term, as defined in (2.9c) is shown in row 9. The factor inequality weights, FIW, as defined in (5.2b), are shown in row 7. The inequality effects and the correlation effects, as defined in (5.3), are shown in rows 4 and 10, respectively.

The degree of over-estimation, (5.1), is:

$$d = (.5360-.2239)/.2239 = 1.39 \qquad (5.5)$$

This means that the over-estimation of $G_y$ is more than 100%! Thus, in this example, the estimator, H, (2.9b), cannot be accepted as it neglects all the correlation effects.

Referring to the FIW in row 7, we see that the property FIW ($F_r = 1.036$) accounts for a major share of the inequality. Its weight is due to a compounding (i.e., product) of the large property income as a distributive share ($\phi_r = .35$) and a Pseudo-Gini Coefficient, ($\bar{G}_r = .6628$), which is the largest. Notice that the FIW for transfer income is not only the lowest but is in fact negative (i.e., $F_t = -.500$). Thus, transfer income contributes to equality.

The decomposition of FIW into the inequality effects and the correlation effects yields the following:

(5.6)

| | Wage | Property | Transfer | Total |
|---|---|---|---|---|
| Inequality Effect | .7862 | 1.0362 | .5717 | 2.394 |
| Correlation Effect | .3222 | 0.0000 | 1.0719 | 1.394 |
| FIW | .4637 | 1.0362 | - .5002 | 1.000 |

From these results we see that total income inequality is due mainly to the inequality effect of property income ($F_r$). On the other hand, transfer income contributes to equality mainly because of its negative correlation effect. When all three types of factors are considered together, the total inequality effect is 2.394 versus a correlation effect of 1.394.

The Rank Correlation coefficients (R' as defined in 3.4) are shown in row 11. To verify equation (4.3), the terms $\varepsilon_x/G_y$ are shown in row 13 which should be approximately the same as row 12. The approximate errors are indicated in row 14. To verify the approximation equation (4.5) we have:

(5.7)

$$G_y = (.45)\ (.3912)\ (.5) + (.35)\ (.6628)\ (1.0) + (.2)\ (.64)\ (-.824)$$
$$= .215$$

As compared with the true value of $G_y = .2239$, there is an underestimation of 4%.

The above analysis may now be applied to empirical data collected for Taiwan. The year we have chosen is 1972. For that year, the primary economic data are listed in Table 5.2.[1]

In Column 1 we show the 23 income classes. The range of income of each class is listed in Column 2 (in units of 1000 New Taiwan dollars, N.T.).

---

[1]Source: Report on the Survey of Family Income and Expenditure, Taiwan Province, Republic of China, 1972, Table 5: Composition of Total Current Receipts and Expenditures of All Households by Income Size.

TABLE 5.2

PRIMARY INCOME DATA FOR TAIWAN, 1972, IN N.T. $1000

| INCOME | | NUMBER OF FAMILIES | INCOME | | | | | |
|---|---|---|---|---|---|---|---|---|
| CLASS (1) | RANGE (2) | (3) | TOTAL (4) | WAGE (5) | PROPERTY (6) | MIXED (7) | TRANSFER (8) | OTHER (9) |
| 1 | 000-010 | 11100 | 90417 | 12135 | 8778 | 13412 | 54948 | 1144 |
| 2 | 010-015 | 23621 | 303579 | 115751 | 30656 | 105537 | 47853 | 3782 |
| 3 | 015-020 | 61497 | 1089263 | 605541 | 86888 | 270127 | 119697 | 7010 |
| 4 | 020-025 | 100868 | 2308013 | 1275522 | 172150 | 719988 | 128969 | 11384 |
| 5 | 025-030 | 149334 | 4095893 | 2534357 | 268412 | 1050196 | 212978 | 29950 |
| 6 | 030-035 | 202522 | 6587184 | 4256089 | 446599 | 1549750 | 300102 | 34644 |
| 7 | 035-040 | 231484 | 8723384 | 5371777 | 601425 | 2400515 | 320362 | 29305 |
| 8 | 040-045 | 226934 | 9640168 | 5923378 | 675624 | 2762163 | 243403 | 35600 |
| 9 | 045-050 | 221916 | 10540494 | 6730770 | 731348 | 2711221 | 322272 | 44883 |
| 10 | 050-055 | 173608 | 9135277 | 5588705 | 710248 | 2575133 | 206359 | 54832 |
| 11 | 055-050 | 161963 | 9322386 | 5528434 | 811122 | 2628247 | 329825 | 24758 |
| 12 | 060-065 | 118830 | 7431800 | 4532223 | 658666 | 1977842 | 240404 | 22665 |
| 13 | 065-070 | 112929 | 7608642 | 4610944 | 697747 | 1912834 | 379949 | 7168 |
| 14 | 070-075 | 102613 | 7413269 | 4376691 | 660617 | 2053589 | 295669 | 26703 |
| 15 | 075-080 | 78217 | 6050068 | 3629718 | 599124 | 1602002 | 210896 | 8328 |
| 16 | 080-085 | 62331 | 5146635 | 2648751 | 603883 | 1543047 | 322255 | 28699 |
| 17 | 085-090 | 54551 | 4778202 | 2708502 | 437059 | 1453364 | 168475 | 10802 |
| 18 | 090-095 | 42690 | 3937332 | 2346247 | 436580 | 1049223 | 101123 | 4159 |
| 19 | 095-100 | 26277 | 2563560 | 1417494 | 229765 | 726530 | 177265 | 12506 |
| 20 | 100-150 | 164348 | 19266070 | 10611106 | 2330238 | 5111696 | 1130002 | 83028 |
| 21 | 150-200 | 27558 | 4721174 | 1950944 | 821743 | 1618439 | 328907 | 1141 |
| 22 | 200-300 | 14330 | 3308482 | 1280859 | 350143 | 1153805 | 513192 | 10483 |
| 23 | 300+ | 1046 | 341586 | 177419 | 134305 | 1555 | 27099 | 1208 |
| GRAND TOTAL | | 2370567 | 134402878 | 78233357 | 12503120 | 36990215 | 6182004 | 494182 |

Table 5.3

| | | | Wage (1) | Mixed (2) | Property (3) | Transfer (4) | Other (5) | Total (6) |
|---|---|---|---|---|---|---|---|---|
| | Factor Share | $\phi_x$ | .582 | .275 | .093 | .046 | .004 | 1.0000 |
| Gini Coefficient | Factor Gini | $G_x$ | .2518 | .2968 | .4020 | .3965 | .2925 | |
| | Weighted Gini | $\phi_x G_x$ | .1466 | .0816 | .0374 | .0182 | .0012 | H = .2850 |
| | Factor Inequality Effect | $\phi_x G_x / G_y$ | .5191 | .2891 | .1324 | .0646 | .0042 | 1.0100 |
| Pseudo-Gini Coefficient | Pseudo-Gini | $\bar{G}_x$ | .2516 | .2958 | .4013 | .3584 | .1668 | |
| | Weighted Pseudo-Gini | $\phi_x \bar{G}_x$ | .1464 | .0814 | .0373 | .0165 | .0007 | $G_y$ = .2823 |
| | Factor Inequality Weight | $\phi_x \bar{G}_x / G_y$ | .5187 | .2882 | .1322 | .0584 | .0024 | FIW=1.0000 |
| Gini Error | Gini Error | $\varepsilon_x$ | .0002 | .0010 | .0007 | .0381 | .1257 | E = .0028 |
| | Weighted Error | $\phi_x \varepsilon_x$ | .0001 | .0003 | .0001 | .0018 | .0005 | d = .0100 |
| | Correlation Effect | $\phi_x \varepsilon_x / G_y$ | .0004 | .0010 | .0003 | .0062 | .0018 | |
| Rank Correlation R' | Rank Correlation | $R'$ | .9987 | .9953 | .9996 | .6803 | .3159 | |
| | 1-R' | $1-R'$ | .0021 | .0047 | .0004 | .3197 | .6841 | |
| | Gini Error Fraction | $\varepsilon_x / G_x$ | .0008 | .0035 | .0018 | .0961 | .4298 | |
| | Approximate Error | $1-\left(\frac{\varepsilon_x / G_y}{1-R'}\right)$ | .6190 | .2553 | -3.5000 | .6994 | .3717 | |

The number of families for each income class is listed in Column 3. The total income for each income class is shown in Column 4.[1] These incomes are divided into five components, namely wage income (Column 5), profit income (Column 6), mixed income[2] (Column 7), transfer income[3] (Column 8), and other income (Column 9). The total income of each type is indicated at the bottom of the table.

The results of our computation are shown in Table 5.3 which has the same structure as Table 5.1. The degree of overestimation (5.1) is:

$$d = (.285 - .2823)/.2823 = .009 \qquad (5.8)$$

Thus the degree of over-estimation is less than 1%. This implies that there is a large positive correlation between factor incomes and total income for all major components, i.e., those with large distributive shares, and hence the correlation effect can be neglected.[4] (See row 1 and row 11).

From rows (7) and (1), we see that those factors which, together, contribute about 94% of total income inequality are the three major components which account for 95% of national income. The comparative magnitudes of FIW for these major components are:

$$\text{FIW (wage)} = .5187 > \text{FIW (mixed)} = .2882 > \text{FIW (property)} = .1322 \qquad (5.9)$$

---

[1] This conventional way of listing the primary data involves the specification of family classes, class income range and class frequency, and requires a redesign of the computational procedure for the Gini Coefficient, the Pseudo-Gini coefficient and the Rank Correlation, which will be explained in Appendix II.

[2] Mixed Income includes Net Operation Surplus, Net Professional Income and Net Agricultural Income.

[3] Transfer Income includes Gifts, Transfers and other Receipts from Government, Households, Enterprises, and Abroad.

[4] Thus, for the year 1972 we picked, the analysis of the correlation effect, e.g., the approximation error in row 14, in this section is negligible and is undertaken only to illustrate the methodology involved.

From this evidence, we see that the wage income appears to have contributed the most to total inequality while profit income does not contribute as much as we would have intuitively expected. However, when the distributive shares and the Pseudo-Gini Coefficients of these components are listed separately below, we see:

| | Wage | | Mixed | | Property | (5.10) |
|---|---|---|---|---|---|---|
| Distributive Share | .582 | > | .275 | > | .093 | |
| Pseudo Gini | .2516 | < | .2958 | < | .4013 | |

From this, we see that the FIW for wage (profit) is the largest (smallest) primarily because of its largest (smallest) distributive share, but, the Pseudo-Gini Coefficient for wage (profit) is the smallest (largest). The fact that the Pseudo-Gini Coefficient for property income is nearly 1.6 times the value of that for wage income, represents an important empirical finding, as we will see below.

A decomposition of the FIW into the inequality effect and the correlation effect shows that the correlation effects for all the major components are negligible, and thus, supports our previous assertion (5.8).

| | Wage | Mixed | Property | Total | (5.11) |
|---|---|---|---|---|---|
| Inequality Effect | .5191 | .2891 | .1324 | 1.0100 | |
| Correlation Effect | .0004 | .0010 | .0003 | .0100 | |
| FIW | .5187 | .2882 | .1321 | 1.0000 | |

In applying our method of analysis to the case of Taiwan, three remarks should be added. First, Taiwan had a peculiar pattern of post-War growth and the year picked (1972) probably non-optimal.[1] Hence, the empirical results can be interpreted as typical for Taiwan during its entire labor surplus period,

---

[1]It fell in the post-turning point era. It is interesting and relevant from a growth-theoretical point of view (see below).

or relevant for other LDC's. Second, the primary economic data, i.e., the Taiwan authorities' classification of income into these five categories, was not guided by any clear theoretical notions. For example, the category of "mixed income," which includes a mixture of "agricultural income, net operational surplus of small enterprises and professional income," is a mixture of property and wage income which ideally should have been separated.[1] "Transfer income" which includes receipts other than government transfer payments does not meet the theoretical requirement which allows us to assess the pure impact of government policy on income distribution. Thus, to increase the usefulness of our future empirical research, we need to refine the data and orient it more to our theoretical requirements.

We have to date limited our empirical application to the statistical data for one random year just to test out the methodology.[2] This precludes a more important analysis of the causation of the change of income distribution inequality through time as traced to factor components. Since changes in income distribution through time are growth relevant issues, we shall now turn to the investigation of the relation between income distribution and growth in the next section.

---

[1]Since "mixed" income is a mixture of "wage and property" income we expect that it shows those characteristics that lie between wage and property income. Our empirical findings in this section confirms this expectation. For example, the Pseudo-Gini Coefficient for mixed income, as shown in 5.10, lies between those for property and wage income.

[2]The empirical analysis in this section should thus be regarded only as a preliminary pilot study. A larger scale study based on cross-sectional, cross-country, and time-series data, beyond the scope of the present paper, can be easily carried out given the framework developed in this paper.

## SECTION VI

## The Impact of Growth on Income Distribution

In the last section we have shown the potential usefulness of our framework for an empirical analysis of the impact on total income distribution inequality through factor income distribution patterns. In this section, we want to show the usefulness of our framework as a conceptual tool for the analysis of the impact of economic development on income distribution. The wage and property income pattern can be traced to factor ownerships, factor prices and the degree of employment so that:

$$\text{a)}\quad W = (W_1, W_2, \ldots W_n) = (v_1 m_1 L_1, v_2 m_2 L_2, \ldots v_n m_n L_n)$$

$$R + (R_1, R_1, \ldots R_n) = (p_1 k_1, p_2 k_2, \ldots p_n k_n) \quad \text{where}$$

$$\text{b)}\quad L = (L_1, L_2, \ldots L_n) \quad \ldots\ldots\text{family ownership of labor} \qquad (6.1)$$

$$K = (K_1, K_2, \ldots K_n) \quad \ldots\ldots\text{family ownership of capital}$$

$$\text{c)}\quad v = (v_1, v_2, \ldots v_n) \quad \ldots\ldots\text{wage structure}$$

$$p = (p_1, p_2, \ldots p_n) \quad \ldots\ldots\text{structure of the rates of return to capital}$$

$$\text{d)}\quad m = (m_1, m_2, \ldots m_n) \quad \ldots\ldots\text{structure of labor employment rates}$$

Having identified factor ownership (6.1b), factor prices (6.1c), and employment rates (6.1d) as those factors that determine factor incomes (6.1), we can easily see that whatever impact economic development has on income distribution is really brought about by its impact on these underlying factors. At the present time, the economic literature is full of intuitive statements of this type.

The inequality of the family ownership of capital stock (K in 6.1b) and of the differentiated wage structure (v in 6.1c), traced to differences in education and skills, are probably the two most important causes of family income inequality. Economic development, through the accumulation of physical and human capital, will obviously affect income distribution. When the capital market is imperfect, as in most LDC's, there is a differentiated structure in the rates of return to capital (p in 6.1c)[1] which is known to gradually modify as the financial institutions develop in the growth process. Changing female participation and the variation of the age composition of the labor force also change the structure of labor ownership (L, 6.1b). Finally, there may be a differentiated pattern of unemployment rates (6.1d) because of "premature"unionization[2] and/or "over-education"[3] of the labor force. All these dimensions can be, and indeed have been, identified as relevant to both growth and income distribution.

---

[1] e.g., in many LDC's, the wealthy are benefitted by political forces which allow them to have a higher rate of return to capital.

[2] Non-union labor tends to have a higher unemployment rate.

[3] It is well known that many LDC's have an "educated unemployed"class.

There is another kind of linkage between growth and income distribution and that is with the theory of development in an aggregate or macro theoretic construct. In conforming with this type of theory, the proportionality of the variables in K,L,v,p,m are assumed to be fixed over time while their magnitudes increase. Denoting the total labor force (capital stock) by $S_L$ ($S_k$) and the absolute real wage (rate of return to capital) by $v_a$ ($p_a$) and denoting the average employment rate by $m_a$, we can rewrite 6.1a as:

(6.2)

a) $W(t) = (W_1(t), W_2(t), \ldots W_n(t)) = S_L(t)\ m_a(t)\ v_a(t)\ (g_1, g_2, \ldots g_n)$

b) $R(t) = (R_1(t), R_2(t), \ldots R_n(t)) = S_k(t)\ p_a(t)\ (h_1, h_2, \ldots h_n)$

where the row vectors $(g_1, g_2, \ldots g_n)$ and $(h_1, h_2, \ldots h_n)$ (See Appendix III) are assumed to be constant (i.e., maintaining constant proportionality) while the increase of wage and property income levels of the families over time (i.e., $W_i(t)$ and $R_i(t)$) are determined by total capital stock ($S_k(t)$), total labor force ($S_L(t)$), absolute factor prices ($v_a(t)$) and $p_a(t)$) and the aggregate employment rate ($m_a(t)$).

Traditional aggregate growth theory is essentially concerned with such macro economic variables as the growth of total capital stock ($S_k(t)$), population growth ($S_L(t)$), the wage rate ($v_a(t)$), the rate of return to capital ($p_a(t)$), and the average unemployment rate ($m_a(t)$).

Economists have had a long standing interest in the question of whether or not labor's distributive share

$\phi_L$ is likely to rise or to fall at the expense of the capital share $\phi_K$ over time.[1] Inspite of the interest in this issue on functional (or social) income distribution grounds, there is still no theory that links the behavior of factor shares ($\phi_L$, $\phi_K$) to family income distribution inequality, i.e., there is still no link between the "functional" and the "size" distribution of income.

To see how our framework may provide an answer, we see that under the special assumption of constant proportionality (6.2), the Pseudo-Gini Coefficient remains constant over time (see Theorem 1.4). Equation 4.1 then shows that an improvement of labor's distributive share will also improve income distribution inequality provided the following condition is satisfied, i.e., provided that

$$\bar{G}_w < \bar{G}_r \tag{6.3}$$

i.e., provided the Pseudo-Gini Coefficient for wage income is smaller than the Pseudo-Gini Coefficient for property income. However, whether or not this holds can only be answered on empirical grounds. Our empirical analysis in Section V implies that this is indeed the case in Taiwan.

The above analysis indicates that, during the process of growth, changes in the values of the factor Pseudo-Gini Coefficient are brought

---

[1] This interest has been manifested both on the theoretical front and on an empirical front. With respect to the former, analytical concepts employed include the elasticity of substitution and the degree of factor bias of innovation.

about by variations in the proportionality of the variables in L,K,v,p,m, i.e, variations in the proportions of the capital stock (and labor force) owned by families, and variations in relative factor prices, and in employment rates. Our paper suggests that such variations will affect income distribution inequality through an inequality effect and a correlation effect, the two component parts of the Pseudo-Gini Coefficient.

These two effects may be illustrated via variations in the relative capital ownership $(K_1/S_k, K_2/S_k)$ as an example. There are two types of variations, corresponding to the two effects. In the first case, the rankings of these ratios are not disturbed[1] which means that capital owning families may own more (or less) capital than before and hence the ownership of the capital stock can become more (or less) unequal. However, the correlation effect does not change. In the second case, the rankings of these ratios are disturbed (i.e, reversed) which means that some relatively poor families which previously owned little or no capital are successful in acquiring more capital leading to an improvement in their capital-owning ranking. In this type of variation both effects will be present. The feasibility of rapid capital augmentation in the course of development not only increases the overall size of the capital stock (as envisioned in aggregate growth theory), but also opens up opportunities for the participation of families to improve their relative status in respect to the ownership of capital stock. It is conceivable that the impact of growth on income distribution may be brought about through this type of upward family "mobility." A similar type of reasoning lies behind the variation of the proportionality of other variables in K,v,p, and m. This type of problem really eludes the traditional aggregate theory and requires further study.

---

[1] For example the inequality of $(K_1/S_{k_1} \leq K_2/S_{k_2} \leq \ldots\ldots \leq K_n/S_{k_n})$ is preserved through time.

## Section VII

## Conclusion

We have tried, in this paper, to provide a conceptual framework for disaggregating total family income inequality into its factor components. The impact of each component on overall inequality can be traced to three separable effects, the impact of that factor's distributive share, the impact of that factor's own income inequality, and the impact of the correlation between that factor's income and total income. We can say that if a factor's income is positively correlated with total income, its contribution to total inequality increases as (i) its distributive share increases and/or (ii) its income inequality increases and/or (iii) the degree of positive correlation increases. The precise opposite is true when the correlation is negative. This conceptual apparatus permits us to assess the quantitative importance of various factor contributions to overall inequality. Moreover, it paves the way for the next important step, i.e. tracing various factor income distributions in turn directly to the impact of such basic growth relevant forces as those determining factor ownership, factor prices, and factor employment rates. Thus the overall distribution of income will ultimately be seen to be determined by the same basic forces which determine the growth performance of an economic system over time. Our future research efforts will in large part be directed here.

Two types of policy implications are indicated by the results of this paper. The first follows directly from our decomposition analysis and permits us to examine the quantitative impact of government tax and transfer payments on overall family income distribution. The regressiveness or pro-

gressiveness of any system of taxes and transfers is customarily considered an important potential tool of government redistributive policy. The actual impact of such policies on the overall size distribution of income can now be analyzed in terms of the share effect, the inequality effect and the correlation effect for T, as discussed above. For example, a larger government welfare budget unequally distributed (and concentrated in the hands of the poorest families) would serve to make the overall distribution of income more equal. Our framework thus allows us to measure the actual historical impact of government tax and transfer activity on the overall distribution of income in a given country and to assess its quantitative importance as a potential tool of redistributive policy in the future.

More important is the examination of the indirect effects of government policy on income distribution via growth. An analysis of these effects can be undertaken by means of the type of structural analysis we have merely begun to outline in Section VI of this paper and which we intend to pursue more fully in our future work. By understanding the basic forces of growth which affect factor rewards, factor distribution, factor employment rates, etc., in different phases of a country's growth process, and how these forces can be affected through the whole arsenal of public policies, we are at the same time understanding--and learning how to affect--the overall distribution of income. The major brunt of our further theoretical and empirical work will be directed at the construction of such a deterministic theory of income distribution.

# APPENDIX I

In this appendix, we shall first prove theorem (3.1) in the text. We assume $X = S(Z)$, i.e., the wage pattern X is a successor of the wage pattern Z. Let wage pattern $W_1 \leq W_2 \leq \ldots \leq W_n$ (arranged in the monotonic order) be rearranged to become $X = (W_{i_1}, W_{i_2}, \ldots W_{i_n})$ and $Z = (W_{k_1}, W_{k_2}, \ldots W_{k_n})$ where $\Omega_i = (i_1, i_2, \ldots i_n)$ and $\Omega_k = (k_1, K_2, \ldots k_n)$ are two permutations of $(1, 2, \ldots n)$. From (1.10) the Gini Error of a wage pattern is:

(A.1)

a) $G_w = (2/n)\, d_w$ where

b) $d_w = u_w - \bar{u}_w = \lambda_1 w_1 + \lambda_2 w_2 + \ldots + \lambda_n w_n - \lambda_1 w_{i_1} - \lambda_2 w_{i_2} \ldots + \lambda_n w_{i_n}$

Apply (A.1b) to X and Z, separately, we have:

(A.2)

a) $d(X) = \lambda_1 (w_1 - w_{i_1}) + \lambda_2 (w_2 - w_{i_2}) + \ldots + \lambda_n (w_n - w_{i_n})$

b) $d(Z) = \lambda_1 (w_1 - w_{k_1}) + \lambda_2 (w_2 - w_{k_2}) + \ldots + \lambda_n (w_n - w_{k_n})$

We want to shown $d(X) \leq d(Z)$ which obviously implies that $\varepsilon_x \leq \varepsilon_z$ by (A.1a). Since $X = S(Z)$ the permutation $(i_1, i_2, \ldots i_n)$ is obtained from $(k_1, k_2, \ldots k_n)$ by exchanging two adjacent integers such that

(A.3)

a) $i_p = k_{p+1}$; $i_{p+1} = k_p$

b) $k_p > k_{p+1}$ or $w_{k_p} \geq w_{k_{p+1}}$ or $w_{i_{p+1}} \geq w_{i_p}$ (by A.3a)

c) $i_q = k_q$ for $q \neq p$ and $q \neq p+1$

Let the difference between d(Z) and d(X) be denoted by:

$$S = d(Z) - d(X) = \lambda_p(w_p - w_{k_p}) + \lambda_{p+1}(w_{p+1} - w_{k_{p+1}})$$

$$- \lambda_p(w_p - w_{i_p}) - \lambda_{p+1}(w_{p+1} - w_{i_{p+1}}) \text{ ---- by (A.3c)}$$

$$= \lambda_p(w_{i_p} - w_{k_p}) + \lambda_{p+1}(w_{i_{p+1}} - w_{k_{p+1}}) \text{ -----A}$$

$$= \lambda_p(w_{i_p} - w_{i_{p+1}}) + \lambda_{p+1}(w_{i_{p+1}} - w_{i_p}) \text{ ----by A.3a}$$

$$= w_{i_p}(\lambda_p - \lambda_{p+1}) - w_{i_{p+1}}(\lambda_p - \lambda_{p+1})$$

Thus $S \geq 0$ if and only if

$$w_{i_p}(\lambda_p - \lambda_{p+1}) \geq w_{i_{p+1}}(\lambda_p - \lambda_{p+1}) \text{ or}$$

$$w_{i_{p+1}}(\lambda_{p+1} - \lambda_p) \geq w_{i_p}(\lambda_{p+1} - \lambda_p) \text{ or}$$

$$w_{i_{p+1}} \geq w_{i_p} \text{ ---------- because } \lambda_{p+1} > \lambda_p$$

The last inequality is seen to be true by (A.3b). This proves $\varepsilon_X \leq \varepsilon_Z$. Next we want to prove that the ordinary correlation coefficient $R(X, w(\Omega_1)) - R(Z, w(\Omega_1))$ where $w(\Omega_1) = (w_1, w_2, \ldots w_3)$ is the natural wage pattern (i.e., wage pattern is in monotonic order). By (3.5ab), the correlation coefficient can be written as:

$$\text{a)} \quad R(X, w(\Omega_1)) = \sum_{k=1}^{n}(w_k - 1/n)(w_{i_k} - 1/n) \Big/ \sum_{k=1}^{n}(w_k - 1/n)^2 \text{ [1/]} \qquad \text{(A.4)}$$

---

[1/] Notice that $(w_{i_1}, w_{i_2}, \ldots w_{i_n})$ is merely a permutation of $(w_1, w_2, \ldots w_n)$ so that $\sum_k (w_k - 1/n)^2 = \sum_k (w_{i_k} - 1/n)^2$.

$$= \left(\sum_{k=1}^{n} w_k w_{i_k} - 1/n\right) \Big/ \left(\sum_{k=1}^{n} w_k^2 - 1/n\right)$$

$$= \left[n(w_1 w_{i_1} + w_2 w_{i_2} + \ldots + w_n w_{i_n}) - 1\right] \Big/ \left(n\sum_{k=1}^{n} w_k^2 - 1\right)$$

b) $R(Z, w(\Omega_1)) = \left[n(w_1 w_{k_1} + w_2 w_{k_2} \ldots + w_n w_{k_n}) - 1\right] \Big/ \left(n \sum_{k=1}^{n} w_k^2 - 1\right)$

We see that $S = (R(X, w(\Omega_1)) - R(Z, w(\Omega_1)) \geq$ if, and only if,

$w_1 w_{i_1} + w_2 w_{i_2} + \ldots + w_n w_{i_n} \geq w_1 w_{k_1} + w_2 w_{k_2} + \ldots + w_n w_{k_n}$ or

$w_1(w_{i_1} - w_{k_1}) + w_2(w_{i_2} - w_{k_2}) + \ldots + w_n(w_{i_n} - w_{k_n}) \geq 0$ or

$w_p (w_{i_p} - w_{k_p}) + w_{p+1}(w_{i_p} - w_{k_p}) \geq 0$ ----- by A.3c

Compare the above inequality with expression A in the early proof, we see that, since $w_p \leq w_{p+1}$ they can take the place of $\lambda_p \leq \lambda_{p+1}$ in A. This proves that $S \geq 0$.

Finally, we want to prove that the Rank Correlation coefficient $R'(\Omega_1, X) \geq R'(\Omega_1, Z)$ where $\Omega_1 = (1,2,\ldots.n)$ is the natural permutation. It follows from the definition of the rank correlation in (3.4)"

(A.5)

a) $R'(\Omega_1, \Omega_i) = 1 - (6/(n^3 - n)) \quad ((1 - i_1)^2 + (2 - i_2)^2 + \ldots + (n - i_n)^2)$

b) $R'(\Omega_1, \Omega_k) = 1 - (6/(n^3 - n)) \quad ((1 - k_1)^2 + (2 - k_2)^2 + \ldots + (n - k_n)^2)$

Hence $S = R'(\Omega_1, \Omega_i) - R'(\Omega_1, \Omega_k) = 0$ if, and only if,

$$(1 - i_1)^2 + (2 - i_2)^2 + \ldots + (n - i_n)^2 \leqq (1 - k_1)^2 + (2 - k_2)^2 + (n - k_n)^2 \text{ or}$$

$$(p - i_p)^2 + (p+1 - i_{p+1})^2 \leqq (p-k_p)^2 + (p+1 - k_{p+1})^2 \text{ by (A.3c) or}$$

$$(p - i_p)^2 + (p+1 - i_{p+1})^2 \leqq (p - i_{p+1})^2 + (p+1 - i_p)^2 \text{ by (A.3a) or}$$

$$i_p \leqq i_{p+1}$$

which is seen to be true by (A.3ab). This completes the proof of theorem (3.1) in the text.

In the rest of this appendix we want to investigate the reasons that lie behind the linear regression relations shown in the scatter diagram 4.1. The successor relation $X = S(Z)$ is indicated in caused order chart (Diagram 3.1) by the arrows initiating from Z and terminating at X. A number of such consecutive arrows form a <u>directed path</u>, $P(Z,V)$, leading from a permutation Z to a permutation V. The above theorem immediately implies that compared with Z, the Gini Error at V is smaller while the correlation coefficient is higher. A rigorous formulation of the idea of a directed path, $P(Z,V)$, is that an <u>incomplete ordering</u> is introduced on the wage pattern space according to the following definition of "Dominance":

Definition: A permutation V dominates Z, in notation, $V = D(Z)$, when

$$V = S(S\ (S\ \ldots.\ S(Z)))$$

Thus when "V dominates Z" we can obtain V from Z by interchanging adjacent integers, in a finite number of steps, always moving a larger integer to

the right. This can be illustrated by the following example for

$Z = (4,3,5,1,2)$ and $V = (3,4,1,2,5)$, we see:

$$Z = (\underline{4},\underline{3},5,1,2) \longrightarrow (3,4,\underline{5},\underline{1},2) \longrightarrow (3,4,1,\underline{5},\underline{2}) \longrightarrow (3,4,1,2,5) = V$$

So it takes three steps to transform Z into V and $V = D(Z)$ (i.e., there is a directed path $P(Z,V)$). A direct corollary of theorem 3.1 is:

Corollary A1 When $V = D(Z)$

a) $R'(\Omega_1, V) \geq R'(\Omega_1 Z)$

b) $R(W(\Omega_1), W(V)) \geq R(W(\Omega_1), W(Z)$

c) $\varepsilon_V \leq \varepsilon_Z$

It follows from this corollary that we can say, unambiguously, that the Gini Error decreases (and correlation coefficient increases) only when this relation of dominance exists. If, for a pair of permutations (E,F) such a path does not exists, then (a) (b) and (c) in the above corollary may not be valid. The relation of dominance, is, then, an incomplete ordering of the points in the wage pattern space.

When there are n-integers (i.e., n-families) we can define an order of a permutation according to the following definition:

Definition: The order of a permutation $(i_1, i_2, \ldots i_n)$ is the number of "adjacent integer exchange" (moving a larger integer to the right) which transforms it into the natural order $(1,2, \ldots n)$.

To compute the order for (5,3,2,1,6,4,) we see:

1) It takes 1 step to move 6 into its natural place (5,3,2,1,4,6)
2) " " 4 steps " " 5 " " " " (3,2,1,4,5,6)
3) " " 0 " " " 4 " " " " (3,2,1,4,5,6)
4) " " 2 " " " 3 " " " " (2,1,3,4,5,6)
5) " " 1 step " " 2 " " " " (1,2,3,4,5,6)

Thus, the order is 8 (= 1 + 4 + 9 + 2 + 1). If $m_i$ is the number of integers smaller than "i" that occupy positions to the right of "i" in a given permutation, then the order of the permutation is:

$$\text{Order} = m_1 + m_2 + \ldots\ldots + m_n \tag{A.6}$$

We readily see that the natural (reverse) permutation has the minimum (maximum) order given by the formula

(A.7)

a) Order of $(1,2,\ldots.n) = 0$

b) Order of $(n, n-1,\ldots.1) = n(n-1)/2$

In the causal order chart (Diagram 3.1), the order of the permutations is indicated at the left-hand margin where the number of permutations and

directed edges initated from vertices of a given order are also shown. The economic interpretation of a permutation having a lower order is that the wage income ranking conforms more closely to the total income rank, and hence, the wage pattern tends to have a higher correlation coefficient in the sense of theorem 3.1.

For the case n = 5, the scatter diagram between the order and the Rank Correlation Coefficient $R'((1,2,3,4,5), (i_1, i_2, i_3, i_4, i_5))$ is given in diagram A1. The scatter diagram between the order and the Gini Error[1] is given in diagram A2. We see that the "order" is positively (negatively) correlated with $\varepsilon(R')$. This is the reason that lies behond the negative correlation between R' and $\varepsilon$ as shown in diagram 4.1 in the text. From the causal order chart, diagram 3.1, we see that the incompleteness of the relation of dominance really lies behind the linear regressions in these scatter diagrams. The parameters in the linear regression equation (4.3) in the text and the correlation coefficient implied therein have not been deduced rigorously in this paper. A formal derivation of these parameter values necessitates a fuller understanding of the relation of dominance. A few remarks may be added on this issue for future reference.

When there are n-families the wage pattern space has n!.points (or vertices). One can easily verify that the total number of directed edges (i.e., total number of successor relations) as:

$$\text{Number of successor relations} = (n-1)\, n!/2 \qquad \text{(A.8)}$$

---

[1]For this numerical example we assume ($w_1 = .05$, $w_2 = .15$, $w_3 = .20$, $w_4 = .25$, $w_5 = .35$).

**DIAGRAM A1**

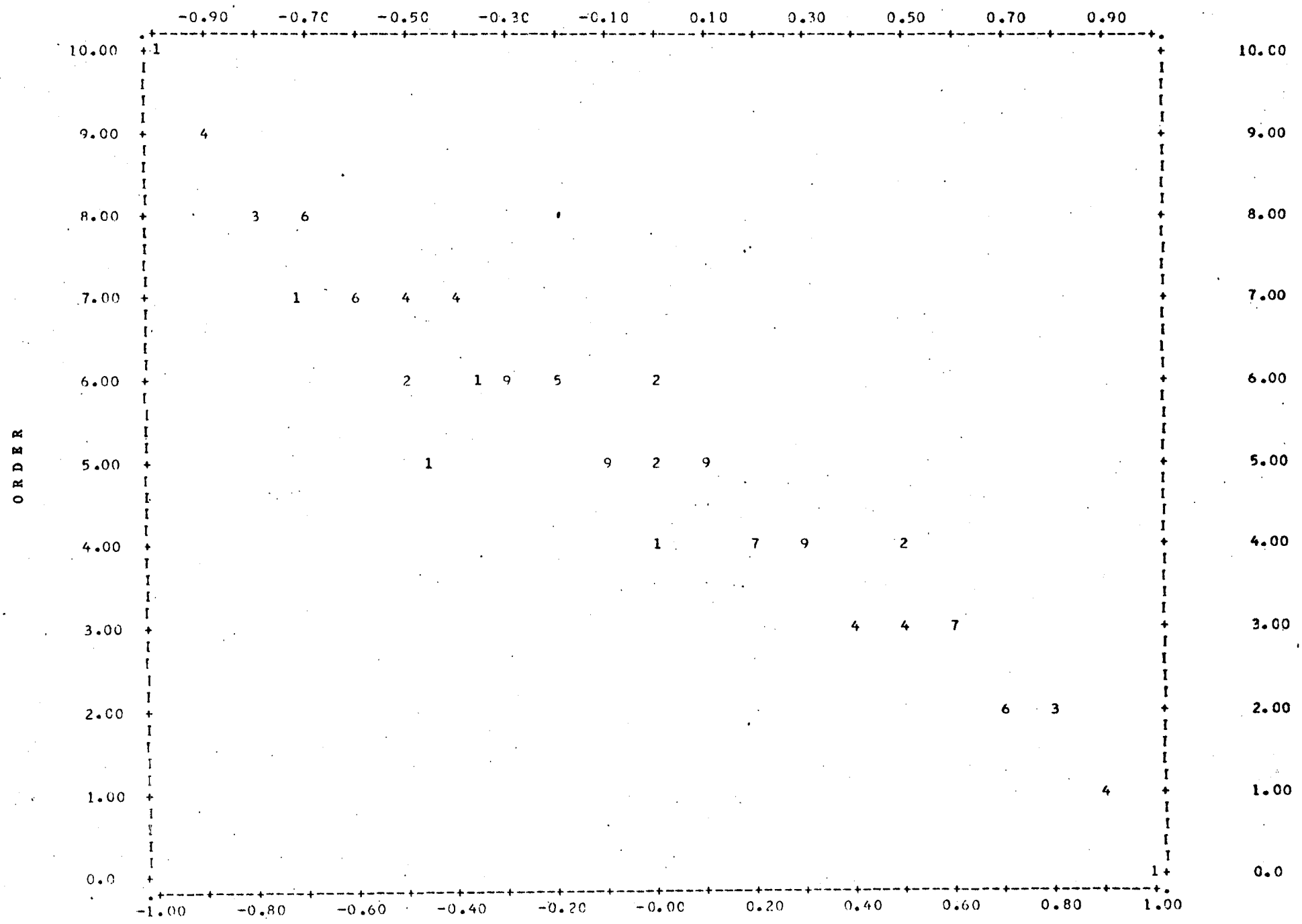
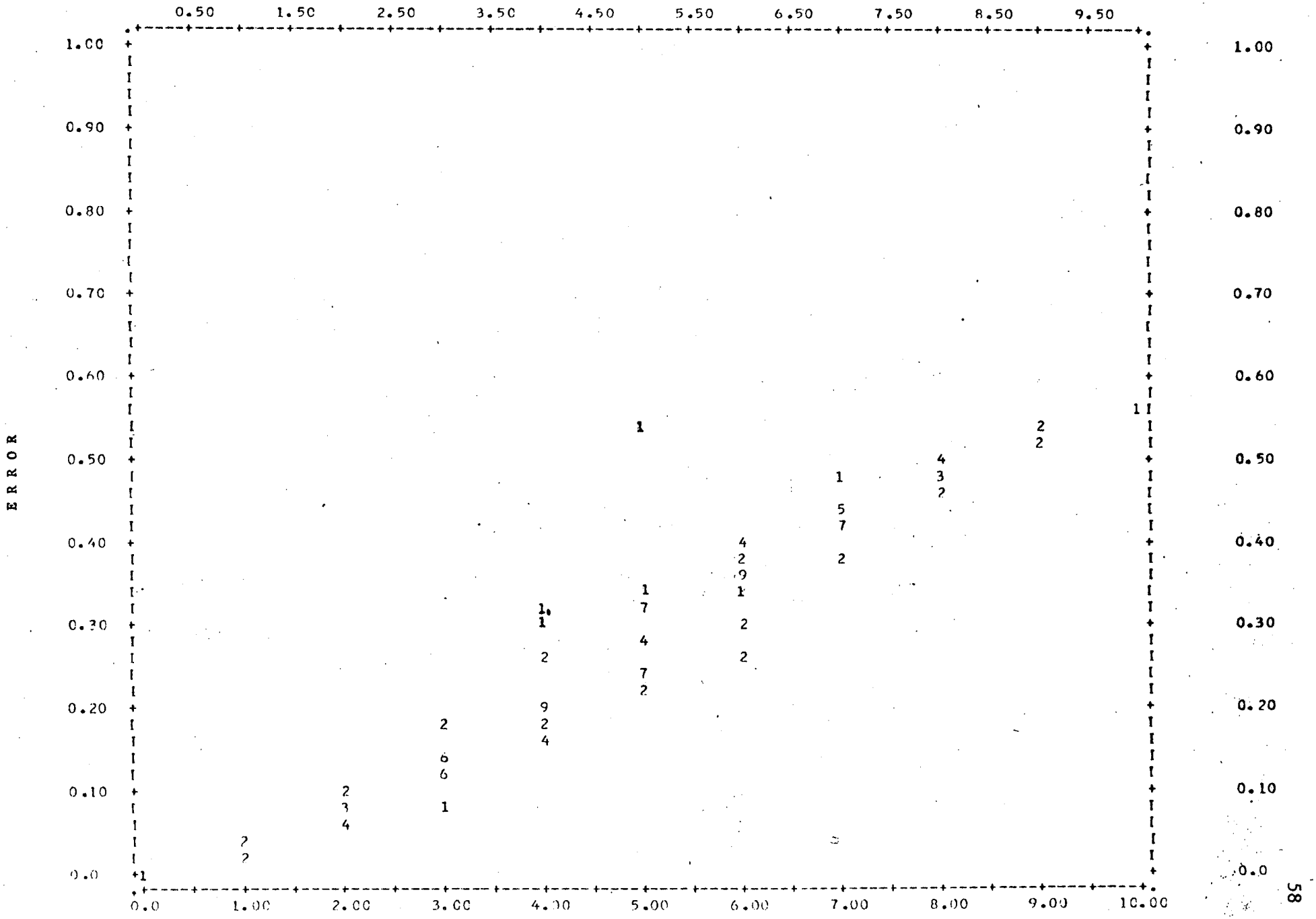

-0.90 -0.70 -0.50 -0.30 -0.10 0.10 0.30 0.50 0.70 0.90
10.00 9.00 8.00 7.00 6.00 5.00 4.00 3.00 2.00 1.00 0.0
ORDER
-1.00 -0.80 -0.60 -0.40 -0.20 -0.00 0.20 0.40 0.60 0.80 1.00

DIAGRAM A2


0.50
1.50
2.50
3.50
4.50
5.50
6.50
7.50
8.50
9.50
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.0
ERROR
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

forming a directed linear graph. This graph is connected and circuit free (i.e., there is no directed closed path). The natural (reverse) permutation is a sink (source), i.e., a vertex which is not an initiating (a terminating) vertex of any edge. The maximum length of the directed path is the maximum order (A.7b). All paths with the maximum length connect the sink and the source. All paths leading from a vertex to the sink have the same length which means that the order of a vertex is uniquely determined. Every vertex lies on a path of the maximum length which is the reason that lies behond theorem 4.1 in the text. When the direction of all edges are reversed, a linear graph is formed which is isomorphic with the given linear graph. Under this isomorphism, the sink (source) becomes the souree (sink). A rigorous investigation of the relation between the Gini Error and the Rank Correlation Coefficient would have to be based on further exploration of these properities.

# APPENDIX II

## Computation Procedure

In this Appendix, we will discuss the computational procedure for the Gini Coefficient, the Pseudo-Gini Coefficient and the Rank Correlation Coefficient when the primary economic data is in the form of a frequency distribution table, e.g., Table 5.3 in the text for Taiwan. Abstractly, the primary data is shown in Columns (1), (2), (3) in Table AII.1, i.e., the class rank (Column 1), the class frequence, $n_i$, (Column 2) and total class income, $z_i$ (Column 3).

The conventional method for computing the Gini Coefficient from the frequency tables begin with the computation of the class population share, $\Theta_i = n_i / N$, and the class income share, $\beta_i = z_i/Z$, where:

$$\text{a)}\quad N = n_1 + n_2 + \ldots + n_m \qquad \text{(AII.1)}$$

$$\text{b)}\quad Z = z_1 + z_2 + \ldots + z_m$$

are the total frequences (i.e., the total number of families) and the total income (i.e., the national income), respectively. The cumulative class income and the weighted cumulative class income are shown in Columns (6 and 7), the sum of the latter is denoted by B, i.e.;

$$B = \Theta_1(\beta_1/2) + \Theta_2(\beta_1 + \beta_2/2) + \ldots\ldots + \Theta_m(\beta_1 + \beta_2 + \ldots + \beta_{m-1} + \beta_m/2) \qquad \text{(AII.2)}$$

and the Gini Coefficient is computed as:

**TABLE AII.1**

| | Class Rank (1) | Class Frequency (2) | Total Class Income (3) | Class Population Share (4) | Class Income Share (5) | For Computation of the Gini Coefficient: Cumulative income (6) | For Computation of the Gini Coefficient: Weighted Cumulative income (7) | For a Typical Family: Income (8) | For a Typical Family: Income Share (9) | Computation of V-index (10) | Computation of V-index (11) | $S_i$ (12) | $H_i$ (13) | $(S_o-H_i)^2$ (14) | $n_i(S_i-H_i)^2$ (15) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | $n_1$ | $z_1$ | $\theta_1=n_1/n$ | $\beta_1=z_1/Z$ | $1/2\beta_1$ | $\theta_1(1/2\beta_1)$ | $Y_1=z_1/n_1$ | $Y_1/Z=z_1/n_1Z$ | $V_1$ | $V_1Y_1/Z=V_1\beta_1/n_1=V_1\beta_1/\theta_1N$ | | | | |
| | 2 | $n_2$ | $z_2$ | $\theta_2=n_2/n$ | $\beta_2=z_2/Z$ | $\beta_1+1/2\beta_2$ | $\theta_2(\beta_1+1/2\beta_2)$ | $Y_2=z_2/n_2$ | $Y_2/Z=z_2/n_2Z$ | $V_2$ | $V_2Y_2/Z=V_2\beta_2/n_2=V_2\beta_2/\theta_2N$ | | | | |
| | 3 | $n_3$ | $z_3$ | $\theta_3=n_3/n$ | $\beta_3=z_3/Z$ | $\beta_1+\beta_2+1/2\beta_3$ | $\theta_3(\beta_1+\beta_2+1/2\beta_3)$ | $Y_3=z_3/n_3$ | $Y_3/Z=z_3/n_3Z$ | $V_3$ | $V_3Y_3/Z=V_3\beta_3/n_3=V_3\beta_3/\theta_3N$ | | | | |
| | m | $n_m$ | $z_m$ | $\theta_m=n_m/n$ | $\beta_m=z_m/Z$ | $\beta_1+\beta_2+\beta_3+\ldots1/2\beta_m$ | $\theta_m(\beta_1+\beta_2+\beta_3+\ldots1/2\beta_m)$ | $Y_m=z_m/n_m$ | $Y_m/Z=z_m/n_mZ$ | $V_m$ | $V_mY_m/Z=V_m\beta_m/n_m=V_m\beta_m/\theta_mN$ | | | | |
| Total | | N | Z | 1 | 1 | | B | | | | u | | | | |
| | 1 | 10<br>(20) | 20<br>(60) | .1<br>(.2) | .05<br>(.15) | .025<br>(.075) | .0025<br>(.0150) | 2<br>(3) | .0050<br>(.0075) | 55<br>(210) | .275<br>(1.575) | 0 | 10 | 100 | 2000 |
| | 2 | 20<br>(10) | 60<br>(20) | .2<br>(.1) | .15<br>(.05) | .125<br>(.175) | .0250<br>(.0175) | 3<br>(2) | .0075<br>(.0050) | 410<br>(255) | 3.075<br>(1.275) | 20 | 0 | 400 | 4000 |
| | 3 | 30<br>(40) | 120<br>(200) | .3<br>(.4) | .30<br>(.50) | .350<br>(.450) | .1050<br>(.1800) | 4<br>(5) | .0100<br>(.0125) | 365<br>(2020) | 13.650<br>(25.25) | 30 | 60 | 900 | 36000 |
| | 4 | 40<br>(30) | 200<br>(120) | .4<br>(.3) | .50<br>(.30) | .750<br>(.850) | .3000<br>(.2550) | 5<br>(4) | .0125<br>(.0100) | 3220<br>(2565) | 40.250<br>(25.65) | 70 | 30 | 1600 | 48000 |
| Total | | 100<br>(100) | 400<br>(400) | 1<br>(1) | 1<br>(1) | | B = .4325<br>($\bar{B}$ = .4675) | | | | u = 57.25<br>($\bar{u}$ = 53.75 | | | | $\sum d_i^2$ = 90,000 |

$G_y = 1-2B$, where (AII.3)

B is the area under the Lorenz curve which is conventionally constructed from the frequency distribution data.

In the theory presented in the text, the N families are <u>not</u> stratified into income classes. Furthermore, for the non-stratified data, the Gini Coefficient is defined by (eq. 1.6):

a) $g = 2(u)/N - (N+1)/N$ where (AII.4)

b) $u = 1h_1 + 2h_2 + .... + Nh_N$

c) $h_1 \leq h_2 \leq .... \leq h_N$

In (AII.4b), $h_i$ is the <u>income share of the i-th family</u> arranged in a monotically non-decreasing order (AII.4c).

To compute the Gini Coefficient according to (AII.4a), the u-index (AII.4b) is computed as follows. The income, $Y_i$, and the income share, $Y_i/Z$, of a <u>typical family</u> in the i-th income class is shown in Columns (8) and (9). Conceptually, the n families in the i-th income class receive the weights from $n_i + n_2 + .... + n_{i-1} + 1$ to $n_1 + n_2 + .... + n_i$. Since all the families in this class receives the same income share, $Y_i/Z$, the u-index is computed as:

(AII.5)

a) $u = V_1(Y_1/Z) + V_2(Y_2/Z) + .... + V_m(Y_m/Z)$ where

b) $Y_i = z_i/n_i = \beta_i/N\Theta_i$; $\quad i = 1,2,....m$

c) $V_i = (S_i + 1) + (S_i + 2) + .... + (S_i + n_i)$; $S_i = n_1 + n_2 + .... + n_{i-1}$

Conceptually, the computation of the u-index is indicated in Column (10) and (11). To show that the conventional method of computation is relevant to our paper, we have the following theorem:

Theorem AII.1 <u>The Gini Coefficient conventionally computed (i.e., by AII.2 and AII.3) is exactly the same as that theoretically defined (i.e., by AII.4 and AII.5).</u>

Proof: From (AII.2) and (AII.3):

$$G_y = 1-2B = \beta_1 + \beta_2 + .... + \beta_m - 2B$$

$$= \beta_1 + \beta_2 + .... + \beta_m - 2\,[\beta_1(\Theta_1/2 + \Theta_2 + .... \Theta_m) ..... + \beta_2(\Theta_2/2 + \Theta_3 + .... + \Theta_m) + .... + \beta_m(\Theta_m/2)]$$

$$= \beta_1 j_1 + \beta_2 j_2 + .... + \beta_m j_m \text{ where}$$

$$j_i = 1-\Theta_i - 2(\Theta_{i+1} + \Theta_{i+2} ....+ \Theta_m) \text{ ----------(a)}$$

On the other hand, $V_i$ in (AII.5c) can be rewritten as:

$$V_i = n_i s_i + (1+2 + \ldots + n_i) = n s_i + n_i(n_i + 1)/2 = n_i(2s_i + n_i + 1)/2$$

$$= n_i [2(n_1 + n_2 + \ldots + n_{i-1}) + n_i + 1)/2 \text{ ----------(b)}$$

By (AII.4) and (AII.5):

$$g = (2/N) [V_i\beta_i/\theta_1 N + V_2\beta_2/\theta_2 N + \ldots + V_m\beta_m/\theta_m N] \ldots\ldots\ldots$$
$$- (N + 1)(\beta_1 + \beta_2 + \ldots + \beta_m)/N$$

$$= k_1\beta_1 + k_2\beta_2 + \ldots + k_m\beta_m \text{ where}$$

$$k_i = 2V_i/N^2\theta_i - (N + 1)/N = (2V_i - (N + 1) N\theta_i)/N^2\theta_i$$

$$= [n_i [2(n_1 + n_2 + \ldots + n_{i-1}) + n_i + 1] - (N + 1) N\theta_i]/N^2\theta_i$$

$$= [\theta_i [2(\theta_1 + \theta_2 + \ldots + \theta_{i-1}) + \theta_i + 1] - \theta_i(N+1)/N]/\theta_i$$

$$= 2(\theta_1 + \theta_2 + \ldots + \theta_{i-1}) + \theta_i + 1 - 1 - i/N$$

$$= 2 V_i/N^2\theta_i - (N + 1)/N$$

$$= n_i [2(n_1 + n_2 + \ldots + n_{i-1}) + n_i + 1]/N^2\theta_i - (N + 1)/N$$

$$= \theta_i (2(\theta_1 + \theta_2 + \ldots + \theta_{i-1}) + \theta_i + 1/N)/\theta_i - (N + 1)/N$$

$$= 2 (1-(\theta_i + \theta_{i+1} + \ldots + \theta_m)) + \theta_i + 1/N - 1 - 1/N$$

$$k_i = 1 - 2(\theta_i + \theta_{i+1} + \ldots + \theta_m) + \theta_i$$

$$= 1 - 2(\theta_{i+1} \ldots + \theta_m) - \theta_i = j_i \quad \text{Q.E.D.}$$

The Pseudo-Gini Coefficient in the text (1.9) is defined by (AII.4ab) where (AII.4c) is <u>not</u> satisfied, i.e., when the income share of the families $(h_1, h_2, \ldots h_n)$ appear in the given order:

(AII.6)

$$\bar{G} = 2(u)/N - (N + 1)/N \text{ where}$$

$$\bar{u} = 1\, h_1 + 2\, h_2 + \ldots + N\, h_N$$

We can use the same table (AII.1) to compute the $\bar{u}$-index with the understanding that the typical family income, $Y_i$ in Column (8), now appear in the given order not satisfying:

$$Y_1 \leq Y_2 \leq \ldots \leq Y_n \tag{AII.7}$$

The sum B in Column (7), now denoted by $\bar{B}$ is the area under the Pseudo-Lorenz Curve. The Pseudo-Gini Coefficient is computed as

(AII.8)

$$\bar{G} = 1-2\bar{B}$$

Since, in the proof of theorem (AII.1), the ordering of $Y_i$ (i.e., whether or not (AII.7) is satisfied) is irrelevant, (AII.8) will lead to the same

value of the Pseudo-Gini Coefficient as theoretically defined in (AII.6).

The computational procedure for the Gini Coefficient and for the Pseudo-Gini Coefficient are illustrated by the two numerical examples indicated at the bottom of Table AII.1. The numbers written without a bracket in the table is the example in which the typical family incomes do appear in the natural order. The values of B = .4325 and u = 57.25 lead to alternative methods for the computation of the Gini Coefficient which lead to the same result, i.e.:

(AII.9)

a) $G = 1-2B = 1 - 2 \times .4325 = .135$

b) $G = 2u/N - (N+1)/N = 2(57.25)/100 - 101/100 = .135$

The numerical example in the brackets shows that the typical family incomes (Column 8) do not appear in a monotonic order. This leads to $\bar{B} = .4675$ and $\bar{u} = 53.75$ for the two alternative methods for the computation for the Pseudo-Gini Coefficient and both yield the same result:

(AII.10)

a) $\bar{G} = 1 - 2\bar{B} = 1 - 2\ (.4675) = .07$

b) $\bar{G} = 2\bar{u}/N - (N+1)/N = 2(53.75) - 101/100 = .07$

In actual practice, we recommend the use of the conventional method. Notice that in this numerical example, the number <u>without</u> the brackets is merely a reordering of the families for the numbers <u>within</u> the brackets.

(See Columns 2 and 3). Thus the true Gini Coefficient for the latter is given by AII.9. Hence the Gini Error is:

$$\varepsilon = G - \bar{G} = .135 - .07 = .065 \tag{AII.11}$$

Finally, for the computation of the Rank Correlation Coefficient from this frequency table, let us assume that (AII.7) is not satisfied. Let us denote the factor income rank of the first family in the i-th class by $H_i + 1$, the factor income rank of $n_i$ families in this class are then:

$$H_i + 1, H_i + 2, \ldots H_i + n_i \tag{AII.12}$$

while the total income rank of these families are:

$$S_i + 1, S_i + 2, \ldots, S_i + n_i \tag{AII.13}$$

where $S_i$ is defined in AII.5c. The sum of the squares of the differences in rank (d) for the $n_i$ families in this class is then:

$$D_i = n_i(S_i - H_i)^2 \tag{AII.14}$$

The computational procedure for the Rank Correlation Coefficient (R') is illustrated for the numerical example within the brackets of Table AII.1.

The values for $S_i$ are shown in Column 12. The values of $H_i$ are shown in Column 13. (Notice that for each income class, $H_i$ is the sum of the frequencies of classes with income less than that of the i-th class). The values of $(S_i - H_i)^2$ and $n_i(S_i - H_i)^2$ are shown in Columns 13 and 14 leading to the sum of the differences squared:

$$\sum_{i=1}^{N} d_i^2 = \sum_{i=1}^{N} D_i = n_1(S_1 - H_1)^2 + n_2(S_2 - H_2)^2 + \ldots + n_m(S_m - H_m)^2 \qquad \text{(AII.15)}$$

indicated at the bottom of Column 15 ( $\Sigma d_i^2 = 90{,}000$). Thus the Rank Correlation Coefficient is:

$$R' = 1 - \frac{6 \Sigma d_i^2}{n^3 - n} = \frac{6 \times 90{,}000}{100^3 - 100} = .46 \qquad \text{(AII.16)}$$

## Appendix III

The derivation of (6.2) from (6.1) in the text:

The five row vectors L, K, v, p and m in 6.1 bcd can be normalized by the following magnitudes

A3.1) a) $S_L = L_1 + L_2 + \ldots + L_n$

b) $S_k = K_1 + K_2 + \ldots + K_n$

c) $m_a = (m_1L_1 + m_2L_2 + \ldots + m_nL_n)/S_L$

so that the wage and property income vector (6.1a) can be written as

A3.2) a) $W = (W_1,W_2,\ldots,W_n) = v_1 m_a S_L\ (v_1^* m_1^* L_1^*, v_2^* m_2^* L_2^*, \ldots, v_n^* m_n^* L_n^*)$

b) $K = (K_1,K_2,\ldots,K_n) = p_1 S_k\ (p_1^* k_1^*, p_2^* k_2^*, \ldots, p_n^* k_n^*)$ where

c) $(v_1^*,v_2^*,\ldots,v_n^*) = (v_1/v_1, v_2/v_1, \ldots, v_n/v_1)$

d) $(p_1^*,p_2^*,\ldots,p_n^*) = (p_1/p_1, p_2/p_2, \ldots, p_n/p_1)$

e) $(L_1^*,L_2^*,\ldots,L_n^*) = (L_1/S_L, L_2/S_L, \ldots, L_n/S_L)$

f) $(k_1^*,k_2^*,\ldots,k_n^*) = (K_1/S_k, K_2/S_k, \ldots, K_n/S_k)$

g) $(m_1^*,m_2^*,\ldots,m_n^*) = (m_1/m_a, m_2/m_a, \ldots, m_n/m_a)$

W and K are seen to be determined by certain aggregate economic variables, i.e., the total labor force ($S_L$) and capital stock ($S_k$), the absolute factor price level ($v_1$) and ($p_1$) and the average employment rate ($m_a$) as well as certain "structural vectors" representing relative factor prices ($v_i^*$, $p_i^*$ in A3.2cd), relative factor ownership ($L_i^*$, $k_i^*$ in A3.2ef) and relative employment rate ($m_i^*$ in A3.2g). Thus the vectors $(g_1,g_2,\ldots,g_n)$ and $(h_1,h_2,\ldots,h_n)$ in (6.2) in the text are seen to be determined completely by the structural vectors.

In the aggregate model described by (6.2) in the text, the underlying assumption is that all the structural vectors are held constant through time. Under this assumption, we see that the Pseudo-Gini Coefficient as well as the Gini Coefficient are invariant through time (by theorem 1.4 in the text). Thus, these coefficients are determined by the structural vectors (A3.2c-g). Therefore, we see that theories on the Gini and Pseudo-Gini Coefficients belong to the domain of general equilibrium theory rather than macro or micro economics.