ECONOMIC GROWTH CENTER

YALE UNIVERSITY

Box 1987, Yale Station
New Haven, Connecticut

CENTER DISCUSSION PAPER NO. 223

MARCH 1975

GROWTH AND THE FAMILY DISTRIBUTION OF INCOME BY FACTOR COMPONENTS: THE CASE OF TAIWAN

John C. H. Fei, Gustav Ranis, and Shirley W. Kuo

Note: Center Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Discussion Papers should be cleared with the authors to protect the tentative character of these papers.

The research reflected in this paper was financed by the World Bank under RPO/284. However, the views expressed do not necessarily reflect those of the Bank.

The authors gratefully acknowledge the assistance of Julia Collins and Yu Yuan Bien.

Introduction

Economic Development and Family Income Distribution

The inequality of Family Income Distribution (FID) has long been recognized as an important social problem in both the rich and the poor countries. Any respectable theory of FID must give prominent recognition to the fact that the total income pattern $(Y_1, Y_2, \ldots, Y_n)$ of n-families has many additive components $W^i = (W^i_1, W^i_2, \ldots, W^i_n)$ $(i = 1, 2, \ldots, r)$.[1] For example, the traditional theory of functional income distribution focuses on the distinction between wage and property income. In case the labor force provided by families is heterogeneous, the wage income components are again the additive sum of homogeneous subcomponents, characterized by differences in age, sex and educational level. Thus, if an index of income inequality such as the Gini Coefficient, G_y, is adopted, the same index can also be used to describe the equity of any factor component, $G(W^i)$. The basic purpose of FID analysis must be to "explain" the total Gini, G_y, which is obviously affected by the component factor Ginis $G(W^i)$.

Any positive theory of FID must thus examine the forces determining the pattern of the income components, W^i. The two main components of total income customarily are wage and property income which are, in turn, traceable mainly to the family ownership of human and physical capital assets, respectively.[2] The heart of economic development is a concern with the accumulation of these human and physical assets over time; hence, any useful investigation of changes in FID must be inextricably linked up with the theory of development.

Two facets of this link are essential: the functional distribution of income and the equity of the distribution of family assets (or incomes).The

---

[1] i.e., $Y = W^1 + W^2 + \ldots + W^r$.

[2] Additional decomposition, by age, sex and education level, would clearly be desirable and is contemplated but not attempted in this paper.

traditional functional income distribution theory which can be linked directly to growth theory tends to determine factor prices and factor shares (ϕ_i). The equity of the existing family factor incomes, $G(W^i)$, is affected by the differentiated patterns of family acquisition of human and physical assets in the past.

In this paper we attempt to design an abstract quantitative framework for FID analysis and then to apply it empirically to Taiwan, a specific case of contemporary development. The quantitative framework we have developed aims, in the first instance, at deriving a decomposition equation linking G_y, the overall Gini, to the $G(W^i)$, the component factor Ginis, and ϕ_i, the component factor shares. A full development of this decomposition technique will be presented in the Appendix. In the text we shall only present the major results with the aid of a concrete example.

Our major purpose is the application of this quantitative analytical framework to an investigation of the impact on FID of growth in a labor surplus developing economy. Guided by the growth theoretic notions relevant to such an economy, we identify the existence of dualism between rural and urban activities as a pronounced feature of the landscape. The above framework can therefore be applied to the economy as a whole and separately to the rural and urban sectors. In the urban sector wage and property income are the main factor income components; in the rural sector the main components are agricultural income as well as the wage and property income arising in rural industries. Our empirical work will focus on the case of Taiwan.

The relationship between growth and the distribution of income in the course of development is gaining increasing attention in recent years. Some claim there exists a necessary inverse U-shaped relationship between equity and growth so that "things have to get substantially worse before they can get better."[1] While such conclusions are based on inductive evidence,[2] there remains the question of the logical necessity of the asserted trade-off between growth and distribution. By linking the inquiry of income distribution and growth we hope to advance our theoretical understanding of this important issue.

---

[1] e.g., Adelman, I., Economic Growth & Social Equity in Developing Countries (Stanford: Stanford University Press, 1973); Kuznets, S., "Economic Growth and Income Inequality," American Economic Review, March 1955, 45(1), pp. 1-28.

[2] The empirical evidence has been mainly of a cross-country type, which, in our view, lends itself less well to the task at hand than longitudinal studies.

Section I: Total Income and Factor Component Patterns

Let the pattern of the income of n-families be denoted by the row vector $Y = (Y_1, Y_2, \ldots, Y_n)$.[1] At any point in time Y may be the additive sum of r-factor income components:

1.1) a) $Y \equiv (Y_1, Y_2, \ldots, Y_n) = W^1 + W^2 + \ldots + W^r$

b) $W^i = (W^i_1, W^i_2, \ldots, W^i_n) \qquad i = 1, 2, \ldots, r$

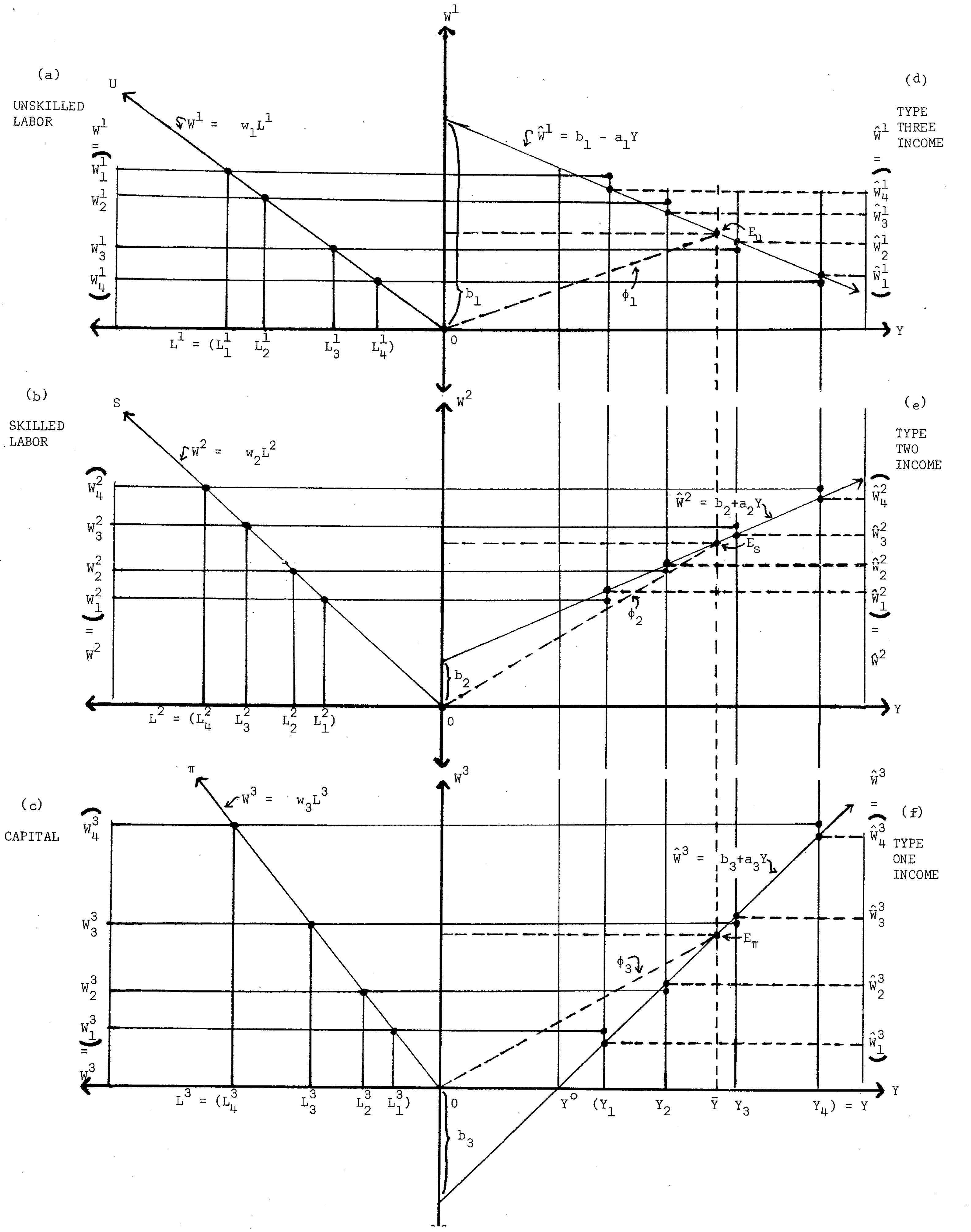
To illustrate the ideas involved, let us assume that there are three factor components, i.e., physical capital (i=1), skilled labor (i=2), and unskilled labor (i=3).

At any point in time the difference between a wealthy and a poor family is due basically to their possession of different amounts of these assets. Suppose there are four families (n=4); the family ownership patterns of assets are indicated by the points on the horizontal axis of Diagrams labc (pointing to the left). When these family ownership patterns are given "by history," we know the total primary factor endowment patterns (i.e., total capital stock and total skilled and unskilled labor force) for the whole economy. Given the factor endowment, the factor prices, (i.e., the rate of return to capital as well as the skilled and unskilled wage rates) are determined by the theory of the functional distribution of income.[1]

---

[1]Such a theory is usually an aggregative theory relating the level of factor prices to some combination of the neo-classical marginal productivity theory as adjusted by considerations of market imperfections, e.g., it may seek to differentiate between the factor returns of a particular homogeneous asset depending on whether it is owned by a "rich" or a "poor" family. For purposes of this paper we intend to accept, rather than improve upon, whatever theory of the functional distribution of income seems appropriate.

DIAGRAM 1


(a)
UNSKILLED LABOR
U
W¹ = w₁L¹
Ŵ¹ = b₁ - a₁Y
b₁
φ₁
E_u
(d)
TYPE THREE INCOME
L¹ = (L¹₁ L¹₂ L¹₃ L¹₄)
(b)
SKILLED LABOR
S
W² = w₂L²
Ŵ² = b₂+a₂Y
b₂
φ₂
E_s
(e)
TYPE TWO INCOME
L² = (L²₄ L²₃ L²₂ L²₁)
(c)
CAPITAL
π
W³ = w₃L³
Ŵ³ = b₃+a₃Y
b₃
φ₃
E_π
(f)
TYPE ONE INCOME
L³ = (L³₄ L³₃ L³₂ L³₁)
Y° (Y₁ Y₂ Ȳ Y₃ Y₄) = Y

In Diagram 1abc, suppose the slopes of the straight radial lines $O\pi$, OS, and OU represent the factor prices, w_1, w_2, w_3. Then the indicated assets ownership patterns are transformed linearly through these radial lines to the factor income patterns, W^i (1.1b) on the vertical axis. The total family income pattern, Y, (1.1a) shown on the horizontal axis of Diagram 1def, is thus determined at the closed end of the model. We can also determine per family income $\bar{Y}$, per family factor income $\bar{W}^i$ and the distributive shares of national income ϕ_i. For the n-factor case, we have:

1.2) a) $\bar{Y} = (Y_1 + Y_2 + \dots + Y_n)/n$

b) $\bar{W}^i = (W_1^i + W_2^i + \dots + W_n^i)/n \qquad i = 1, 2, \dots, r$

c) $\phi_i = \bar{W}^i/\bar{Y} \qquad i = 1, 2, \dots, r$

d) $\phi_1 + \phi_2 + \dots + \phi_r = 1$

Since the asset ownership patterns are the result of cumulative family investments in physical and/or human capital in the past, we may expect the higher income families to own more capital assets as well as more of the highly educated labor force (doctors, lawyers, engineers, etc.). On the other hand, the low income families own, and derive income from, more of the society's unskilled labor force. Thus when the linear regressions of factor income on total income of (1.3a)

1.3) a) $\hat{W}^i = b_i + a_i Y$ $(i = 1, 2, 3, \ldots, r)$

b) $b_1 + b_2 + \ldots + b_r = 0$

c) $a_1 + a_2 + \ldots + a_r = 1$

are fitted to the data in (1.1), in the general case, we expect the regression coefficients, a_i, to be positive for such components as property and skilled wage income and to be negative for such components as unskilled wage income and transfer (i.e., welfare) income. In the example of Diagram 1def, regression lines with these different characteristics are shown for these three types of incomes.[1]

Since the total income pattern Y is the sum of all its factor components (1.1), the linear regression equations of (1.3) are not independent of each other. Thus the sum of the regression coefficients, a_i, must equal one and that of the regression constants, b_i, add up to zero, as stated in (1.3bc).[2] Furthermore, in a growing economy, all family factor incomes, W_j^i (and hence total family incomes, Y_i), are non-negative (i.e., the scatter diagrams in 1def consist only of points lying in the first quadrant).[3] Thus

1.4) $a_i < 0$ implies $b_i > 0$

---

[1]It should be emphasized that these regression relationships are determined at the closed end of the model, i.e., they are used for purely descriptive purposes and are devoid of the behavioristic connotations usually associated with regression lines (e.g., the consumption function) in economics.

[2]This follows readily from the fact that a_i and b_i are estimated by the method of least squares.

[3]In other words, we assume there is no dissaving.

i.e., for any regression line in the system, the regression constant, b_i and the regression coefficient, a_i, cannot both be negative. It follows that the 2r parameters, a_i and b_i in system (1.3) can be classified into three types of cases:

Type One: $T_1 = (a_1, a_2, \ldots, a_{r_1}; b_1, b_2, \ldots, b_{r_1}); a_i \geq 0, b_i < 0$

Type Two: $T_2 = (a_{r_1+1}, a_{r_1+2}, \ldots, a_{r_2}; b_{r_1+1}, b_{r_1+2}, \ldots, b_{r_2}); a_i \geq 0, b_i > 0$

Type Three: $T_3 = (a_{r_2+1}, a_{r_2+2}, \ldots, a_r; b_{r_2+1}, b_{r_2+2}, \ldots, b_r); a_i < 0, b_i > 0$

These cases have a natural economic interpretation. With reference to our examples of Diagram 1def, suppose ($a_1 = -.2$; $a_2 = + .5$; $a_3 = + .7$). Thus whenever the income of two families differs by one dollar, the wealthier family has 50 (70) cents more skilled wage income (property income), and 20 cents less unskilled wage income than the poorer family. Thus the a_i's indicate the income sensitivity of the ith factor. The distinguishing characteristic of a Type Three case is that the income sensitivity is negative. Government transfer payments calculated to promote income distribution equity are typical of the Type Three case. Type Three income thus serves as an FID "equalizer."

To help us distinguish T_1 from T_2, let us divide (1.3a) by Y

1.5) a) $S \equiv \hat{W}^i/Y = \frac{b_i}{Y} + a_i \qquad i = 1, 2, \ldots, r$

b) $S_1 + S_2 + \ldots S_r = 1 \qquad$ (by 1.3bc)

and refer to S^i as the family's income share of the ith factor. For example, if Y = \$100 and the family's property income is \$40, then the family property income share is .4. The S^i's permit us to distinguish between T_1 and T_2.

Equation (1.5) shows that the distinguishing characteristic of T_1 income is that the income share increases for the richer family, while for T_2 income the income share decreases.

As shown in Diagram 1f, property income is typically a T_1 income, i.e., the percentage of property income is higher for the wealthier family. On the other hand, the income from skilled wages in Diagram 1e illustrates a T_2 case, i.e., while this income component increases absolutely, its percentage contribution declines for the wealthier family. The T_3 or unskilled wage income component falls absolutely, as well as relatively. Thus, T_1, T_2 and T_3 cases, in this order, represent decreasing contributions to overall FID inequality.[1]

In the normal case in which there exist several factor income components, (1.3b) implies

1.6) T_1 not empty; $T_2 \cup T_3$ not empty;

namely, there must be at least one T_1 asset and at least one T_2 or T_3 asset.

From Y and the ith regression equation, we can calculate an estimated pattern of factor income $\hat{W}^i$:

1.7) a) $\hat{W}^i = (\hat{W}^i_1, \hat{W}^i_2, \ldots, \hat{W}^i_n)$ $i = 1, 2, \ldots, r,$ where

b) $\hat{W}^i_j = b_i + a_i Y_j$ $j = 1, 2, \ldots, n; i = 1, 2, \ldots, r$

---

[1]At any moment of time, the fact that the wealthier family owns more property and more educated manpower (and thus derives more income from these factors) is a piece of common knowledge which provides the major incentive for family investment in human and/or capital resources. Most families want to move up the ladder and emulate the wealthier families by acquiring more assets of the T_1 and T_2 type, subject, of course, to the "family budget constraint" (Y^o). Thus while the classification of factor income and assets (i.e., into T_1, T_2, T_3) results at the closed end of a static model (see above), the information thus revealed may not be irrelevant to an understanding of family saving and investment behavior ultimately essential to the construction of a dynamic theory of income distribution. Such a theory must take into consideration not only an individual family's utility in isolation but also the impact of other families' income and expenditure patterns on its own behavior.

The estimated patterns of the various incomes (1.7) can be read off on the right hand vertical axes of Diagrams 1def. The following are elementary properties for the estimated factor income patterns which follow directly from (1.3bc):

1.8) a) $Y = \hat{W}^1 + \hat{W}^2 + \ldots \hat{W}^r$

b) $\bar{W}^i \equiv (W^i_1 + W^i_2 + \ldots + W^i_n)/n = (\hat{W}^i_1 + \hat{W}^i_2 + \ldots + \hat{W}^i_n)/n \equiv \bar{\hat{W}}^i$

c) $\bar{\hat{W}}^i = b_i + a_i\bar{Y}$

d) $\phi_i = \bar{\hat{W}}^i/\bar{Y} = (b_i + a_i\bar{Y})/\bar{Y}$ (by (1.2c) and (1.8b))

(1.8a) states that Y is the additive sum of the estimated pattern. Thus the estimated wage patterns, $\hat{W}^i$ ($i = 1, 2, \ldots, r$), may be viewed as a system of linear approximations of the original factor income components, W^i ($i = 1, 2, \ldots, r$), as defined in (1.1b).

Equation (1.8b) states that the average factor income per family as defined in (1.2b) can be calculated from the estimated pattern. The average factor income of the estimated pattern is a linear function of average total income (1.8c). It follows that the distributive shares of (1.2c) can be calculated from the estimated pattern as shown in (1.8d). In our example, in Diagram 1def, the average values of total and factor incomes are represented by the points E_π, E_s, E_u on the regression lines. Then the slopes of the radial lines OE_π, OE_s, and OE_u represent the three distributive shares. In the next section, the decomposition of Y into factor components (1.1) can be studied in two steps. The first step is to study the decomposition of Y into the approximated system (1.8a), i.e., under linearity assumptions. The second step is to study the deviation of the original system (W^i) from the linear approximation.

Section II: Inequality[1]

Let us adopt the familiar Gini Coefficient $G(X)$ as a measurement of the degree of inequality of the non-negative vector X. When total family income Y and its factor components W^i (1.1) are given, we can define the total Gini, $G_y = G(Y)$, as well as the factor Ginis, $G_i = G(W^i)$ $(i=1,2,\ldots,r)$. Furthermore, from the linear approximations $\hat{W}^i$ cf(1.7a), we can define the "estimated" factor Ginis $G(\hat{W}^i)$. We then have the following theorem (proved in the Appendix):

Theorem One

2.1) a) $G(\hat{W}^i) = (a_i/\phi_i) G_y$ for $a_i \geqq 0$ (Type One, Type Two)

b) $G(\hat{W}^i) = (-a_i/\phi_i) G_y$ for $a_i < 0$ (Type Three)

Notice that $a_i/\phi_i = a_i/(\bar{\hat{W}}^i/\bar{Y})$ (see 1.8d) is the elasticity of the regression line at the "mean point" (in Diagram 1def, the mean points are E_π, E_u, E_s). In the case of a Type One or Type Two factor, the elasticity is positive (since $a_i \geqq 0$), and the "estimated" factor Gini $G(\hat{W}^i)$ is the product of the total Gini G_y and the elasticity (2.1a). In the case of a Type Three factor (as $a_i < 0$), the elasticity is negative and thus a negative sign must be attached in (2.1b), since the Gini coefficient is always a non-negative fraction.

---

[1]For a fuller treatment of the materials of this section, see the Appendix.

When (1.8d) is substituted in (2.1a), we have (with $\bar{Y} \geqq 0$)

$$2.2) \quad a) \quad G(\hat{W}^i) = \frac{a_i \bar{Y}}{b_i + a_i \bar{Y}} G_y. \qquad \text{Hence}$$

$$b) \quad G(\hat{W}^i) > G_y \quad \text{if} \quad a_i \geqq 0 \quad \text{and} \quad b_i < 0 \quad \text{(Type One income)}$$

$$c) \quad G(\hat{W}^i) < G_y \quad \text{if} \quad a_i \geqq 0 \quad \text{and} \quad b_i > 0 \quad \text{(Type Two income)}$$

The theorem states that in the case of a Type One income, the estimated factor Gini is greater than the total Gini (2.2b) and the opposite is true for Type Two income (2.2c).[1] When both sides of equation (2.1) for each factor component are multiplied by the relevant ϕ_i, we have, after adding:[2]

$$2.3) \quad a) \quad G_y = F^+ - F^- \qquad \text{where}$$

$$b) \quad F^+ = \phi_1 G(\hat{W}^1) + \phi_2 G(\hat{W}^2) + \ldots + \phi_{r_2} (G(\hat{W}^{r_2}) \quad \text{(Summed over } T_1 \text{ and } T_2 \text{ factors by (2.1a))}$$

$$c) \quad F^- = \phi_{r_2+1} G(\hat{W}^{r_2+1}) + \ldots + \phi_r G(\hat{W}^r) \quad \text{(Summed over } T_3 \text{ factor (by 2.1b))}$$

Thus G_y is the difference between two non-negative terms, F^+ and F^-, where F^+ is the weighted sum of Type One and Type Two terms (i.e., for non-negative regression coefficients, $a_i \geq 0$) and where F^- is the weighted sum of the Type Three terms (i.e., for $a_i < 0$).

The economic interpretation of the "decomposition equation" (2.3) is that the inequality of total family income (G_y) may be seen as contributed to

---

[1]This is what we would expect since Type One income is an increasing relative share of total income as a family is wealthier and Type Two income is a decreasing relative share.

[2]From (2.1) we have $G(\hat{W}^1)=(a_1/\phi_1)G_y$; $G(\hat{W}^2)=(a_2/\phi_2)G_y$, ..., $G(\hat{W}^{r_2})=(a_{r_2}/\phi_{r_2})G_y$ for Type One and Type Two incomes, and $G(\hat{W}^{r_2+1}) = -(a_{r_2+1}/\phi_{r_2+1})\, G_y$, ..., $-(a_r/\phi_r)G_y$ for Type Three incomes. After multiplying by ϕ_i and adding we have $G_y(a_1 + a_2 + \ldots + a_r) = H^+ - H^-$ which reduces to (2.3) by (1.3c).

by the inequality of factor distributions (i.e., $G(\hat{W}^i)$) depending upon the relative size of the distributive shares (ϕ_i) and depending upon the type of income. Type Three income, as a total income equalizer (see last section), contributes to equality rather than inequality as seen from the negative Type Three effects, $-F^-$. The Type One and Type Two factors enter into equation (2.3a) through the positive H^+ effect. This term can be further divided into the sum of the Type One terms and the sum of the Type Two terms. In the Type One term, every $G(\hat{W}^i)$ is greater than G_y (see (2.2b)) and in the Type Two term, every $G(\hat{W}^i)$ is less than G_y (see (2.2c)). In this sense a Type One term contributes more heavily to inequality than a Type Two term.

In the real world, the original data W^i (1.1) does not, of course, show perfect linear correlation with Y . In that case, we can define a non-linearity error term for every factor income component:

2.4) a) $\theta_i = (G(\hat{W}^i) - G_i)/G_i$ or

b) $G(\hat{W}^i) = G_i + \theta_i G_i$ $(i = 1, 2, \ldots, r)$

i.e., θ_i is the deviation of the estimated factor Gini from the "true factor Gini"--expressed as a fraction of the latter. When W^i is correlated nearly perfectly with Y, $\hat{W}^i$ is approximately the same as W^i and hence θ_i tends to be zero. For this reason, θ_i is called a non-linearity error. Substituting (2.4b) into (2.3), we have:

2.5) a) $G_y = \hat{\hat{G}}_y - \hat{\theta}$ where

b) $\hat{\hat{G}}_y = H^+ - H^-$ and

c) $H^+ = \phi_1 G_1 + \ldots + \phi_{r_2} G_{r_2}$

d) $H^{-} = \phi_{r_2+1}G_{r_2+1} + \ldots + \phi_r G_r$

e) $\hat{\theta} = -[\theta^{+} - \theta^{-}]$

f) $\theta^{+} = \phi_1\theta_1 G_1 + \phi_2\theta_2 G_2 + \ldots + \phi_{r_2}\theta_{r_2}G_{r_2}$

g) $\theta^{-} = \phi_{r_2+1}\theta_{r_2+1}G_{r_2+1} + \ldots + \phi_r\theta_r G_r$

In (2.5a) $\hat{\hat{G}}_y$ may be interpreted as an estimator of G_y while $\hat{\theta}$ is the non-linearity error of estimation. The estimator $\hat{\hat{G}}_y$ is the difference of two positive terms, H^{+} and H^{-} (corresponding to F^{+} and F^{-} in (2.3a), except that the true factor Ginis now replace the estimated factor Ginis of the estimated income pattern). The non-linearity error term $\hat{\theta}$ tends toward zero, with increasing high correlation between Y and every W^{i}. Equation (2.5a) also permits us to assign "blame" for overall inequality to the various factor components. Dividing both sides by G_y (and assuming $\hat{\theta}$ is small), we have

2.6) a) $1 = (F_1 + F_2 + \ldots + F_{r_2}) - (F_{r_2+1} + F_{r_2+2} + \ldots + F_r)$ where

b) $F_i = \phi_i G_i / G_y \qquad i = 1, 2, \ldots, r$

where the factor inequality weights (F_i) indicate that factor's contribution to overall inequality or equality.

Notice that in the special case when there is no Type Three income. $\hat{\hat{G}}_y$ of (2.5) reduces to

2.7) a) $G_y = \hat{\hat{G}}_y - \hat{\theta}$

b) $\hat{\hat{G}}_y = F^{+} = \phi_1 G_1 + \phi_2 G_2 + \ldots + \phi_r G_r$ (no Type Three income)

in which the estimator $\hat{G}_y$ is simply the weighted sum of the factor Ginis. In this special case, it can be proved (see Appendix) that the non-linearity term of (2.7a) is always non-negative, i.e.

2.8) $\hat{\theta} \geqq 0$ (no Type Three income)

so that $\hat{\hat{G}}_y$ always tends to overestimate the true Gini, G_y. Thus in the special case when there are no Type Three assets, we can use the approximation equation of (2.7b) when the non-linearity error term is small. That these conditions are satisfied in our particular empirical application to the case of Taiwan will be shown below.

Section III: Income Distribution and Growth

We are, of course, not merely interested in the forces determining G_y at any point in time but even more in what may occasion changes in equity over time. The approximation equation which we have just derived can be used to analyze the two types of forces that affect the value of G_y over time--on the assumption of no Type Three asset and small non-linearity error. To illustrate this, let us assume r = 2 and that the two factor components are capital (K) and labor (L). Then

3.1) a) $G_y = \phi_\pi G_\pi + \phi_w G_w$

b) $\phi_w + \phi_\pi = 1$

K and L represent the economy's primary factor endowment and ϕ_w and ϕ_π the distributive shares. Differentiating (3.1) with respect to time "t", we have

3.2) a) $\frac{dG_y}{dt} = A + B$ where

b) $A = \phi_w(G_w - G_\pi)\frac{d\phi_w}{dt}/\phi_w$ (Functional Distribution Effect)

c) $B = \phi_\pi \frac{dG_\pi}{dt} + \phi_w \frac{dG_w}{dt}$ (Factor Gini Effect)

The term A, the functional distribution effect, describes the change of G_y due to changes in the relative shares; and B, the factor Gini effect, describes the change of G_y due to changes in the factor Ginis. Thus the change in overall G_y can be traced partly to changes in the functional distribution of income and partly to changes in family asset ownership patterns.

We may assume that capital is a T_1 asset and labor a T_2 asset (see (1.6)). Thus we have, by (2.2):

3.3) a) $G_w < G_\pi$ which implies that

b) The A-effect of (3.2b) is negative if and only if the rate of growth of ϕ_w is positive (i.e., $(d\phi_w/dt/\phi_w) > 0$).

Thus, looking only at the A-effect, we see that FID equity improves through time when the labor share increases. People are inclined to believe that any change in the functional distribution of income favoring the laboring class would automatically benefit FID. Our analysis establishes the necessary conditions for this to be true, i.e., wage income must be distributed more equally than property income. Moreover (see below) it is not unambiguously true without regard to the (much larger) B-effect.

This analysis of the direction of change of the distributive shares can in turn be tied up with the theory of development. In the case of a labor surplus economy, as in Taiwan (see below), we can further distinguish between two cases, i.e., before the "turning point," (or before the labor surplus has been fully absorbed), and after the turning point. Before the turning point wages may be assumed to be approximately constant and after the turning point to be rising rapidly. The growth equations relevant to the analysis of the direction of change of ϕ_w after the turning point is as follows:

3.4) $$\eta_{\phi_w} = (1 - \phi_w)\, \eta_{K/L} \left(\frac{1}{\varepsilon} - 1\right) + B_L$$

where ε is the elasticity of substitution; B_L the degree of labor-using bias of innovations; (and η_x denotes the time rate of change of any "x").

Substituting (3.4) in (3.2b), the A-effect becomes:[1]

3.5) $$A = \phi_w (G_w - G_\pi) [(1 - \phi_w) (\frac{1}{\varepsilon} - 1) \eta_{K/L} + B_L] .$$

This permits us to see that FID improves via a negative A-effect (with $G_w < G_\pi$) when technology change is biased in a labor-using direction ($B_L > 0$) and when there is overall capital deepening ($\eta_{K/L} > 0$). For these are the conditions that improve labor's share (3.4).[2]

Before the turning point, on the assumption of an unlimited supply of labor and near constancy in the real wage, (3.4) reduces to the following special form:[3]

3.6) a) $\eta_{\phi_w} = B_L\varepsilon - J(1 - \varepsilon)$ implying that

b) $\eta_{\phi_w} > 0$ if and only if $B_L > J(\frac{1}{\varepsilon} - 1)$ (with $\frac{1}{\varepsilon} - 1 > 0$).

Combining (3.5a) with (3.2b), we thus have, before the turning point:

3.7) $$A = \phi_w (G_w - G_\pi) [B_L\varepsilon - J(1 - \varepsilon)]$$

---

[1]For a fuller exposition and derivation of both these equations, see Chapter 3 (especially Table 1 and the appendix to Chapter 3) in the authors' Development of the Labor Surplus Economy: Theory and Policy, Irwin, 1964. It may be noted that the ε used in this paper coincides with the more conventional definition but is the reciprocal of the definition used in the aforementioned volume.

[2]This is true in the "normal" case, i.e., that of production complementarity, i.e., when $\varepsilon < 1$. In the case of production substitutability ($\varepsilon > 1$), it is true, i.e., when there is capital shallowing ($\eta_{K/L} < 0$) instead of capital deepening.

[3]When $\eta_w=0$, the rate of labor absorption is $\eta_L=\eta_K+\frac{B_L+J}{\varepsilon_{LL}}$ (see Fei and Ranis, Development of the Labor Surplus Economy, Chapter 3). When $\eta_{K/L} = -(B_L+J)/\varepsilon_{LL}$ is substituted in (3.4) we have $\eta_{\phi_w} = -(\phi_\pi/\varepsilon_{LL})(\frac{1-\varepsilon}{\varepsilon})(B_L+J) + B_L$

cont'd/

Thus, during the unlimited supply of labor phase, the FID becomes better (lower G_y) through a negative A-effect when (3.6b) holds, i.e., when technology change is sufficiently biased in a labor-using direction to overcome the innovation intensity effect.[1]

Thus for a labor surplus economy, a high J and a high B_L always contribute to elimination of unemployment and the arrival of the turning point. However, income distribution may get worse, through the A-effect, when the intensity effect $J(\frac{1}{\varepsilon} - 1)$ overwhelms B_L. (Statistically this is seen via a decrease of ϕ_w and an increase of G_y.) After the turning point, a higher value of B_L, combined with capital deepening, will contribute to the improvement of FID. (Statistically, this is seen via a decrease of G_y and an increase of B_L and $\eta_{K/L}$.) Whether or not these conditions are met lies behind the Kuznets hypothesis concerning the inverse U-shaped time path of G_y.

As far as the B term, i.e., the factor Gini-effect of equation (3.2c), is concerned, we see that a negative B, which would contribute to growing FID equity, may be caused by a change in the asset ownership patterns of either capital or labor or both. On the one hand, the fact that the capital ownership pattern has become more equal through time, i.e., $dG_\pi/dt < 0$, can occur when the lower income families acquire capital assets faster than the

---

which can be reduced to (3.5a) in the text with the help of $\varepsilon = \phi_\pi/\varepsilon_{LL}$ (see op. cit., p. 85). Comparing (3.5) and (3.7) we can see that the behavior of ϕ_w is caused by quite different types of forces before and after the turning point. This difference is traced basically to the fact that, before the turning point, employment is causally determined by capital accumulation via labor absorption. After the turning point capital and labor are symmetrical and the real wage is determined endogenously.

[1]In the normal case, $\varepsilon < 1$, a high innovation intensity leads to more labor absorption and thus a lower K/L ratio, decreasing labor's share. Thus a high B_L contributes to both employment and FID objectives while a high J contributes to the first but not the second objective. In the abnormal case, when the technology is substitutable, $\varepsilon > 1$, a high J also contributes to a better FID. In that case, a high J and a high B_L would contribute to both elimination of unemployment and the improvement of FID.

rich families--through a combination of higher savings rates or favorable inheritance laws--or if there is a land or capital reform. On the other hand, the pattern of ownership of skilled labor can become more equal over time, i.e., $dG_w/dt < 0$, when lower income families manage to acquire more skilled labor through their own education expenditures and/or governments' providing free education to the lower income groups.

Thus, changes in G_y through time can be traced to two types of forces. First, there is the impact on the functional distribution of income arising from such growth-relevant forces as capital accumulation, population growth and technology change; here the link between FID and growth theory is quite direct. Second, there is the impact on G_y caused by changes in the equity of factor asset (and income) distribution patterns as traced to the differentiated patterns of family acquisition of human or physical assets, as impacted by public policy. Here the relationship to growth theory is less direct and requires additional research effort.

The above analysis can now be generalized when a third factor component is added, i.e., $r = 3$. If that third component is agricultural income A, (3.1) becomes[1]

3.8) a) $G_y = \phi_w G_w + \phi_\pi G_\pi + \phi_A G_A$

b) $\phi_w + \phi_\pi + \phi_A = 1$

---

[1] Assuming, as before, that there is no Type Three income and that the non-linearity error is small.

We can easily imagine a dualistic developing country--such as Taiwan (see below)--in which families receive income from two production sectors: the agricultural sector from which they derive undifferentiated agricultural income with distributive shares ϕ_A, and a non-agricultural sector (industry and services) with differentiated distributive shares ϕ_W and ϕ_π for wage and property income, respectively. To emphasize this dualistic feature of the economy, equation (3.8) can be rewritten as:

3.9) a) $G_y = \phi_x G_x + \phi_A G_A$ where

b) $\phi_x + \phi_A = 1$

c) $G_x = \phi'_W G_W + \phi'_\pi G_\pi$ where

d) $\phi'_W = \phi_W/(\phi_W + \phi_\pi)$; $\phi'_\pi = \phi_\pi/(\phi_W + \phi_\pi)$

e) $\phi'_W + \phi'_\pi = 1$

In (3.9a), ϕ_x is the distributive share of national income generated by the non-agricultural sector and hence G_y is the weighted average of the two sectoral Ginis G_A and G_x. From (3.9d), we see that ϕ'_W and ϕ'_π are the two functional distributive shares generated by the non-agricultural sector, with (3.9c) indicating that the Gini Coefficient of non-agricultural income (G_x) is the weighted average of the Ginis of non-agricultural wage and property income G_W and G_π. Notice that (3.9ac) are in the form of (3.1ab); hence, we can apply (3.2) to obtain:[1]

---

[1]Notice that (3.10) is really a generalization of (3.2), namely, when $\phi_A = 0$, (3.10) becomes (3.2). As will be clear below, (3.10) is useful for our empirical application to Taiwan; however, it can be generalized to additional factor components.

3.10) a) $\frac{dG_y}{dt} = A' + A'' + B$

b) $A' = (G_A - G_x) \frac{d\phi_A}{dt}$ (Reallocation Effect)

c) $A'' = (G_w - G_\pi) \phi_x \frac{d\phi'_w}{dt}$ (Functional Distribution Effect)

d) $B = \phi_A \frac{dG_A}{dt} + \phi_w \frac{dG_w}{dt} + \phi_\pi \frac{dG_\pi}{dt}$ (Factor Gini Effect)

The term B represents the factor Gini effect comparable to (3.2c) above. On the other hand, A' and A" represent two types of "A effects" due to the change of factor shares (corresponding to (3.2b)). One prominent feature of the development of a dualistic economy is that of continuous reallocation of labor from the agricultural to the non-agricultural sectors over time. This gradual shift of the economy's center of gravity can be proxied by a decline of ϕ_A, i.e., the share of total national income in the form of agricultural income. Thus the term A' in (3.10b) may be referred to as the reallocation effect (i.e., related to the reallocation of resources from agriculture to non-agriculture). Notice that when such reallocation takes place, $d\phi_A/dt$ is negative. Thus the impact of the reallocation effect on overall income distribution equity G_y in (3.10b) depends upon the sign of $G_A - G_x$. When agricultural income is a Type Two income (i.e., $G_A < G_y$), (3.9a) shows G_x must be larger than G_y. Hence, $G_A - G_x < 0$. In this case, A' is positive which means that the reallocation effect worsens the overall equity of income distribution through the A' effect. Conversely, when agricultural income is a Type One income, implying $G_A - G_x > 0$, the reallocation effect A' is negative and hence reallocation will improve overall income distribution equity. In summary, 3.11) When the agricultural income share ϕ_A decreases, the impact of the reallocation effect on G_y is negative ($A' < 0$) when agriculture is a Type Two income and positive ($A' > 0$) when it is a Type One income.

The term A'' in (3.10c) describes the impact on G_y as traced to the functional distribution of income forces within the non-agricultural sector of a dualistic economy. All the results obtained earlier for linking growth and distribution in the two factor incomes model described by (3.1) can be directly applied here.

Thus we see that in a dualistic economy the equity of total family income G_y is determined by three types of forces: A', the reallocation effect; A'', the functional distribution effect (in non-agriculture); and B, the factor Gini effect. Equation (3.10a) therefore provides a quantitative theoretical framework for linking the theory of FID determination with growth.

Section IV: Application to Taiwan

The above ideas on distribution and growth will now be applied to the case of Taiwan, which fits the above dualistic economy typology. The n-households[1] are divided into c urban and n-c rural households. The total income of the n-families is $(Y^u, Y^r) = (Y_1^u, Y_2^u, \ldots, Y_c^u, Y_1^r, Y_2^r, \ldots, Y_{n-c}^r)$. The urban households receive wage income W^u $(= W_1^u, W_2^u, \ldots, W_c^u)$ and property income π^u $(= \pi_1^u, \pi_2^u, \ldots, \pi_c^u)$ generated in urban production activities, such as large and small-scale urban industries and services. The rural households receive wage income $W^r = (W_1^r, W_2^r, \ldots, W_{n-c}^r)$ and property income $\pi^r = (\pi_1^r, \pi_2^r, \ldots, \pi_{n-c}^r)$ from rural industrial and service activities as well as income from agricultural activities A $(= A_1, A_2, \ldots, A_{n-c})$. Thus we may employ three models of decomposition to analyze income distribution trends in Taiwan, i.e., for urban families, for rural families and for all families:

4.1) a) Urban family decomposition: $Y^u = W^u + \pi^u$

b) Rural " " $Y^r = W^r + \pi^r + A$

c) All " " $(Y^u, Y^r) = (W^u, W^a) + (\pi^u, \pi^a) + (0, A)$

The urban model of (4.1a), with two income components, can make use of the framework of analysis provided by (3.1) in the last section, while the rural and all family models (4.1bc) containing three factor components, can make use of (3.8).

---

[1]Ideally, one should be working with individual families. But, for practical purposes, since it is difficult to work with millions of individual observations, income distribution data is usually grouped by income intervals. In the application to Taiwan (see below), the household surveys used actually grouped the n-families into 23 to 32 such class intervals.

The data for Taiwan are based on sample surveys starting in 1964 and have been processed in accordance with this understanding.[1] The results are summarized in Tables 1 a, b, c.

For "all households," Table 1a [rows (1), (2), (3), (4)] presents the factor Ginis for each of the three component factor incomes (G_W, G_π, G_A) and the total Ginis (G_y). The factor distributive shares (ϕ_W,ϕ_π,ϕ_A) are found in rows (5),(6),(7). The regression coefficients a_i and the regression constants

---

[1] The first estimates of income distribution for the whole Taiwanese economy were undertaken by Professor Kowie Chang of National Taiwan University for 1953 and again for 1959 and 1961. These surveys were based on samples of only 301 and 812 observations, respectively. The first official surveys were made for 1964 by the Directorate General of Budget, Accounting and Statistics (DGBAS). Surveys have been carried on for every alternate year between 1964 and 1970 and for every year since then, with sample size ranging between 1.3% and 1.9%. These data are subject to a number of deficiencies, including the underestimation of high incomes, the inability to separate property from wage income in the (lumped) category of agricultural income, and the separation of Taipei City from the rest of Taiwan since 1970. (For a fuller statement of the data availabilities, as well as deficiencies, see Wanyong Kuo, "Income Distribution by Size in Taiwan Area--Changes and Causes," paper presented to joint JERC-CAMS Seminar on "Income Distribution, Employment and Economic Development in Southeast and East Asia," Tokyo, December 1974.) However, the overall quality of the DGBAS data--which we have concentrated on to date--compares favorably with that of most countries, certainly other LDC's. The information is available by type of income and is broken down by farm and non-farm families--as well as by education, sex and age of head of household, breakdowns which have not yet been utilized at this stage of our analysis. The farm, non-farm breakdown unfortunately only begins in 1966.

In transforming the raw data into the simplifying framework of (4.1) the following assumptions should be noted: 1) a category of unallocatable "miscellaneous income" for all families is neglected, as is the agricultural income of urban families since both are quantitatively small (see below); 2) the absence of inter-sectoral payments (e.g., W^r may contain some "farmers' daughters'" income in the city) due to the impossibility of isolating such payments in the data; 3) agricultural income A is not functionally disaggregated into a wage and property share; in a family farm type of agriculture, such a separation would have required a rather arbitrary imputation procedure.

[2] The relative shares do not add up to 1 since we neglected the merged category of "mixed incomes" which adds up to less than 10% of the total.

Table 1a

All Households

			1964	1966	1968	1970	1971	1972
1	Factor Ginis	G_W	0.2434	0.2760	0.3037	0.2763	0.2558	0.2549
2		G_π	0.4664	0.4233	0.4721	0.4451	0.4319	0.4035
3		G_A	0.3669	0.3564	0.2286	0.1437	0.1771	0.1736
4	Total Gini	G_y	0.3282	0.3300	0.3346	0.2951	0.2837	0.2823
5	Factor Distributive Shares	ϕ_W	0.4324	0.4760	0.5066	0.5405	0.5847	0.5884
6		ϕ_π	0.2401	0.2557	0.2777	0.2370	0.2353	0.2343
7		ϕ_A	0.2754	0.2118	0.1523	0.1614	0.1233	0.1277
8	Weighted Linear Regression: Slopes (a)	a_W	0.2457	0.3658	0.4774	0.4347	0.4471	0.4859
9		a_π	0.4849	0.3653	0.3547	0.4055	0.4014	0.3470
10		a_A	0.2288	0.1944	0.0555	0.0584	0.0852	0.0816
11	Weighted Linear Regression: Constants (b)	b_W	5.5821	3.6950	1.1849	4.6705	6.6717	5.8077
12		b_π	-7.3210	-3.6783	-3.1270	-7.4438	-8.0907	-6.3891
13		b_A	1.3952	0.5840	3.9331	4.5506	1.8545	2.6108
14	Weighted Linear Regression: Correlations (r)	r_W	0.8849	0.9743	0.9545	0.9561	0.9286	0.9789
15		r_π	0.8896	0.9696	0.9272	0.9445	0.9411	0.9794
16		r_A	0.8125	0.8871	0.5972	0.7923	0.5372	0.8680
17	Weighted Gini	$\hat{G}_y$	0.3341	0.3347	0.3454	0.3004	0.2893	0.2854
18	Non-Linearity Error	$\hat{\theta}=\hat{G}_y-G_y$	0.0059	0.0047	0.0108	0.0053	0.0056	0.0031
19	Degree of Overestimation	$D_y=\hat{\theta}/G_y$	0.0178	0.0142	0.0324	0.0178	0.0198	0.0109
20	Factor Inequality Weights	$F_W=\phi_W G_W/G_y$	0.3205	0.3980	0.4598	0.5061	0.5272	0.5313
21		$F_\pi=\phi_\pi G_\pi/G_y$	0.3411	0.3279	0.3918	0.3575	0.3582	0.3349
22		$F_A=\phi_A G_A/G_y$	0.3078	0.2287	0.1041	0.0786	0.0770	0.0785
23	Relative Wage-Profits: Gini	G_W/G_π	0.5219	0.6520	0.6433	0.6208	0.5923	0.6317
24	Relative Wage-Profits: Share	ϕ_W/ϕ_π	1.8009	1.8616	1.8243	2.2806	2.4849	2.5113
25	Relative Factor Ginis	G_W/G_y	0.7416	0.8364	0.9077	0.9363	0.9017	0.9029
26		G_π/G_y	1.4211	1.2827	1.4109	1.5083	1.5224	1.4293
27		G_A/G_y	1.1179	1.0800	0.6832	0.4870	0.6243	0.6149

Table 1b

Nonfarm Households

			1966	1968	1970	1971	1972
1	Factor Ginis	G_w	0.2870	0.2845	0.2347	0.2232	0.2305
2		G_π	0.4362	0.4394	0.3897	0.4160	0.3700
3		G_A	0.3259	0.4079	0.2720	0.2946	0.2689
4	Total Gini	G_y	0.3315	0.3386	0.2839	0.2710	0.2724
5	Factor Distributive Shares	ϕ_w	0.5925	0.5627	0.6137	0.6519	0.6468
6		ϕ_π	0.3218	0.3366	0.2914	0.2686	0.2792
7		ϕ_A	0.0217	0.0288	0.0292	0.0262	0.0235
8	Weighted Linear Regression: Slopes (a)	a_w	0.4539	0.4981	0.4292	0.4650	0.5200
9		a_π	0.4539	0.3620	0.4445	0.4632	0.3871
10		a_A	0.0132	0.0201	0.0271	0.0219	0.0135
11	Weighted Linear Regression: Constants (b)	b_w	4.7243	3.0842	9.0566	9.6453	7.6094
12		b_π	-4.5008	-1.1337	-7.5145	-10.0475	-6.4728
13		b_A	0.2873	0.3877	0.1056	0.2200	0.5992
14	Weighted Linear Regression: Correlations (r)	r_w	0.9721	0.9585	0.9412	0.9245	0.9733
15		r_π	0.9805	0.9056	0.9433	0.9472	0.9648
16		r_A	0.5835	0.5453	0.6989	0.7871	0.6455
17	Weighted Gini	$\hat{\hat{G}}_y$	0.3367	0.3472	0.2892	0.2768	0.2768
18	Non-Linearity Error	$\hat{\theta}=\hat{\hat{G}}_y-G_y$	0.0052	0.0086	0.0053	0.0058	0.0044
19	Degree of Overestimation	$D_y=\hat{\theta}/G_y$	0.0159	0.0254	0.0187	0.0214	0.0167
20	Factor Inequality Weights	$F_w=\phi_w G_w/G_y$	0.5130	0.4767	0.5073	0.5369	0.5475
21		$F_\pi=\phi_\pi G_\pi/G_y$	0.4200	0.4368	0.4000	0.4070	0.3794
22		$F_A=\phi_A G_A/G_y$	0.0213	0.0347	0.0280	0.0285	0.0232
23	Relative Wage-Profits: Gini	G_w/G_π	0.6580	0.6475	0.6023	0.5365	0.6230
24	Relative Wage-Profits: Share	ϕ_w/ϕ_π	1.8412	1.6717	2.1060	2.4270	2.3166
25	Relative Factor Ginis	G_w/G_y	0.8658	0.8402	0.8267	0.8236	0.8462
26		G_π/G_y	1.3158	1.2977	1.3727	1.5351	1.3583
27		G_A/G_y	0.9831	1.2047	0.9581	1.0871	0.9872

Table 1c

Farm Households

				1966	1968	1970	1971	1972
1		Factor Ginis	G_w	0.2245	0.3121	0.2138	0.2332	0.2439
2		Factor Ginis	G_π	0.3497	0.3084	0.3766	0.3562	0.3594
3		Factor Ginis	G_A	0.3612	0.3453	0.3217	0.3263	0.3085
4		Total Ginis	G_y	0.3263	0.2902	0.2831	0.2969	0.2908
5		Factor Distributive Shares	ϕ_w	0.2016	0.3228	0.3602	0.3572	0.4230
6		Factor Distributive Shares	ϕ_π	0.0998	0.0994	0.1029	0.1224	0.1072
7		Factor Distributive Shares	ϕ_A	0.6595	0.5263	0.4869	0.4523	0.4226
8	Weighted Linear Regression	Slopes (a)	a_w	0.1078	0.1375	0.2162	0.2224	0.2810
9	Weighted Linear Regression	Slopes (a)	a_π	0.1113	0.1392	0.1540	0.1605	0.1743
10	Weighted Linear Regression	Slopes (a)	a_A	0.7205	0.6553	0.5244	0.4932	0.4396
11	Weighted Linear Regression	Constants (b)	b_w	3.0317	5.9228	5.1025	5.5068	6.9641
12	Weighted Linear Regression	Constants (b)	b_π	-0.3721	-1.2715	-1.8091	-1.5552	-3.2934
13	Weighted Linear Regression	Constants (b)	b_A	-1.9727	-4.1258	-1.3265	-1.6693	-0.8328
14	Weighted Linear Regression	Correlations (r)	r_w	0.7535	0.6524	0.8486	0.8924	0.8959
15	Weighted Linear Regression	Correlations (r)	r_π	0.8659	0.7665	0.8328	0.7869	0.6832
16	Weighted Linear Regression	Correlations (r)	r_A	0.9906	0.9865	0.9601	0.9260	0.9076
17		Weighted Gini	$\hat{\hat{G}}_y$	0.3389	0.3031	0.2901	0.3054	0.2947
18		Non-Linearity Error	$\hat{\theta}=\hat{\hat{G}}_y-G_y$	0.0125	0.0129	0.0070	0.0085	0.0135
19		Degree of Overestimation	$D_y=\hat{\theta}/G_y$	0.0385	0.0448	0.0247	0.0285	0.0135
20		Factor Inequality Weights	$F_w=\phi_w G_w/G_y$	0.1387	0.2371	0.2720	0.2806	0.3548
21		Factor Inequality Weights	$F_\pi=\phi_\pi G_\pi/G_y$	0.1069	0.1057	0.1369	0.1468	0.1325
22		Factor Inequality Weights	$F_A=\phi_A G_A/G_y$	0.7301	0.6264	0.5533	0.4971	0.4483
23		Relative Wage-Profits: Gini	G_w/G_π	0.6420	0.6910	0.5677	0.6547	0.6786
24		Relative Wage-Profits: Share	ϕ_w/ϕ_π	2.0200	3.2475	3.5005	2.9183	3.9459
25		Relative Factor Ginis	G_w/G_y	0.6880	0.7343	0.7555	0.7852	0.8381
26		Relative Factor Ginis	G_π/G_y	1.0717	1.0627	0.3303	1.1997	1.2359
27		Relative Factor Ginis	G_A/G_y	1.1070	1.1899	1.1363	1.0990	1.0609

b_i are shown in rows (8) to (13).[1] The correlation coefficients are shown in rows (14),(15),(16).

In verifying (2.7a), $\hat{\hat{G}}_y$ of (2.7b) (i.e., the weighted average of the true factor Ginis) is shown in row (17). The value of $\hat{\theta}$ (i.e., the non-linearity error defined as $\hat{\hat{G}}_y - G_y$ in (2.7a)) is shown in row (18). Finally, we have computed D_y $(=\hat{\theta}/G_y)$, i.e., the linearity error as a fraction of the true Gini in row (19). The factor inequality weights of (2.6) are shown in rows (20)-(22) and selected relative distributive shares and relative Ginis are presented in rows (23)-(27). Tables 1c and 1b are identical for farm households and non-farm households, respectively.[2]

To provide better visual overview of the time series the data in Tables 1abc are also summarized in Diagrams 2, 3, and 4, respectively, for the three models; the six panels (ABCDEF) of these diagrams correspond to the classification by rows of Table 1abc. Our empirical analysis will concentrate on the pattern of these indices, central to the analytical framework we have evolved, in the course of Taiwan's development. Each of the major relevant findings (F) will be stated and its significance in the context of our analytical framework explored.

F.1: For all three models, property income is a Type One income and wage income is a Type Two income. Hence, $G_w \leq G_y \leq G_\pi$.

This is seen from the fact that the regression coefficients a_w and a_π are both positive (panel E, Diagrams 2, 3, 4) while b_w is positive and b_π is

[1]For the same reason cited in footnote 1 above, the a's don't quite add up to 1 and the b's don't quite add up to 0.

[2]The reader should note from Table 1b that agricultural income accruing to non-farm households (ϕ_A) is uniformly small, i.e., less than three percent, which conforms to our model framework in (4.1a) which neglects this type of income.

DIAGRAM 2
INEQUALITY TRENDS FOR ALL HOUSEHOLDS

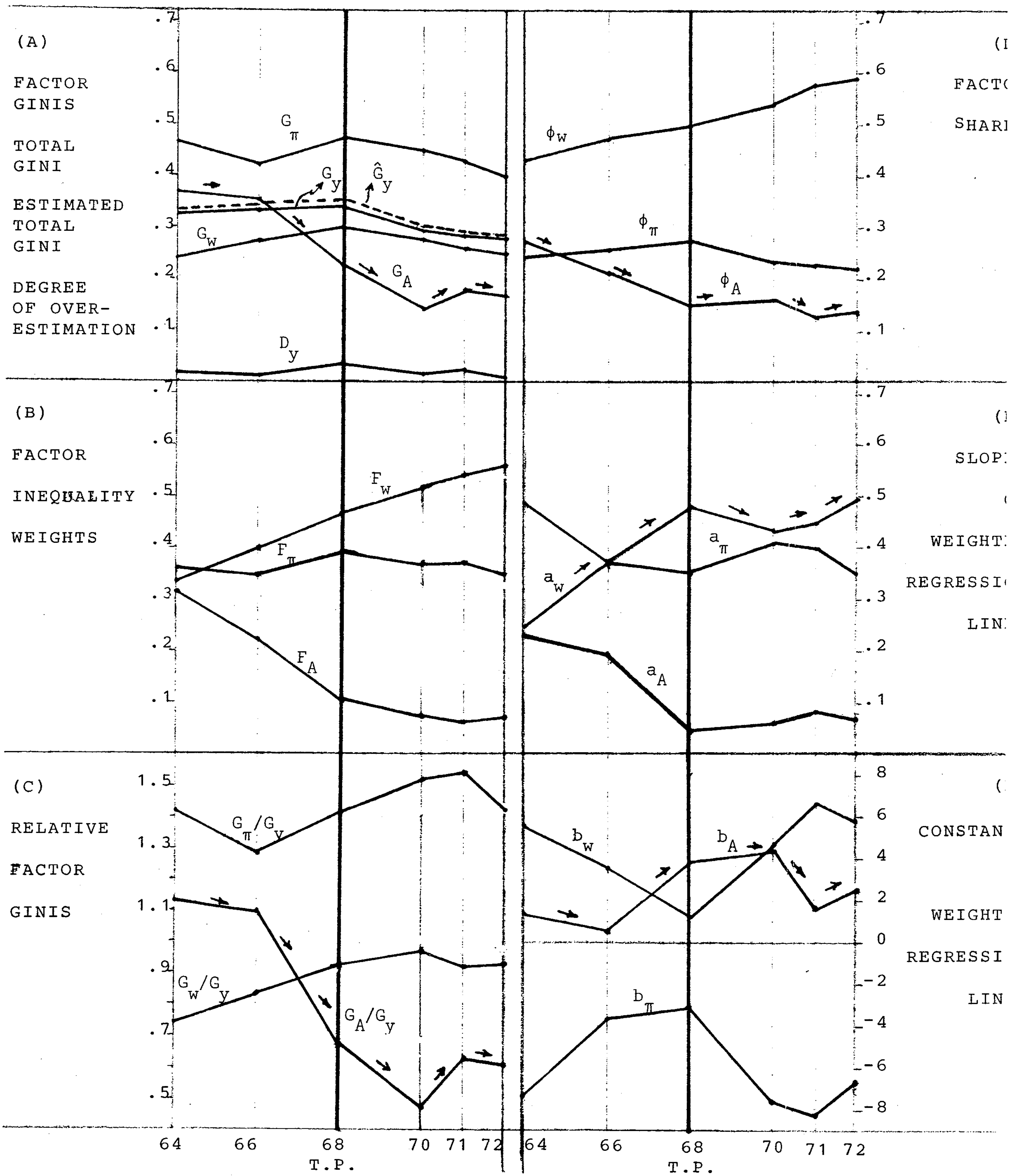


DIAGRAM 3

INEQUALITY TRENDS FOR NONFARM HOUSEHOLDS

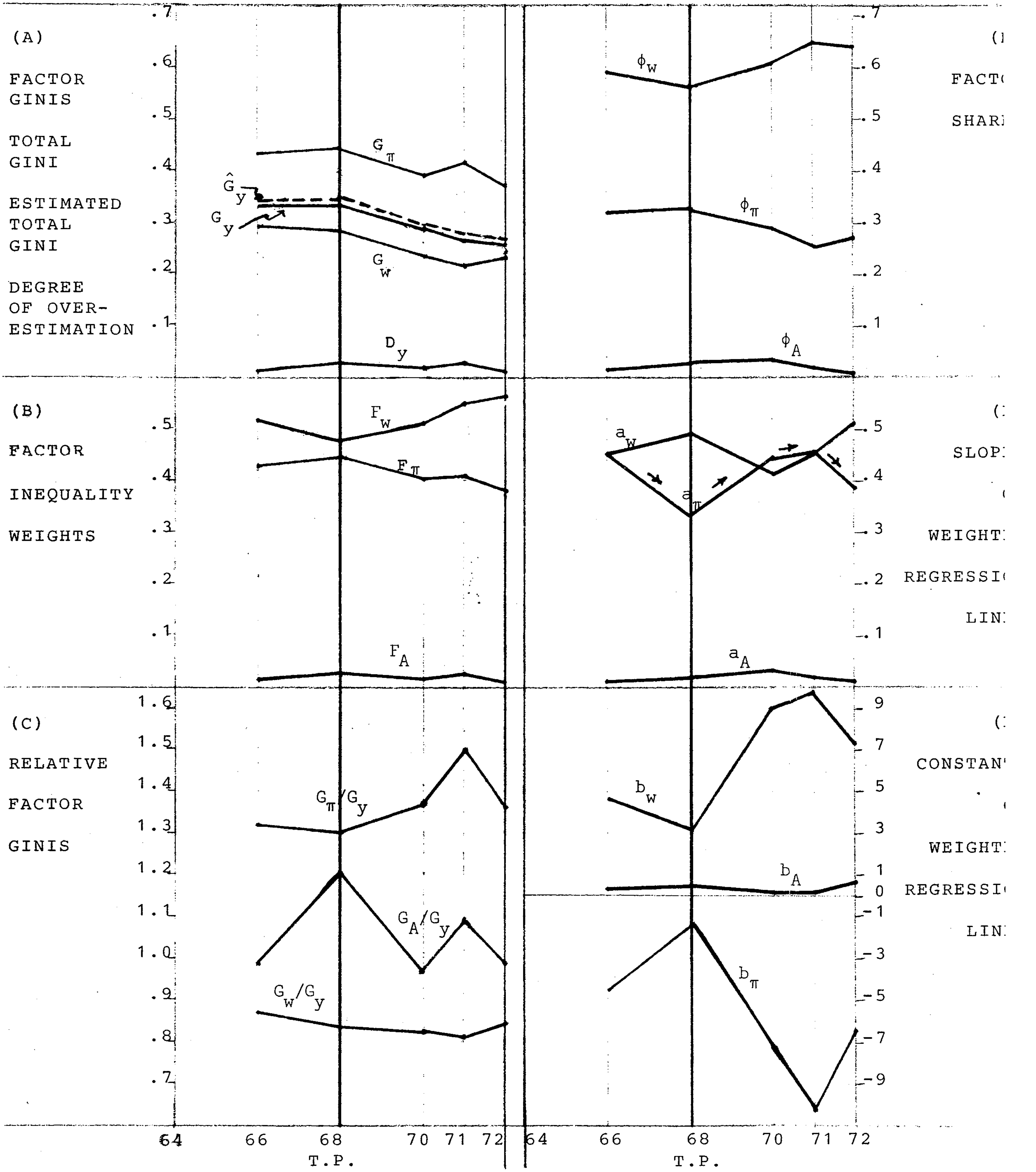
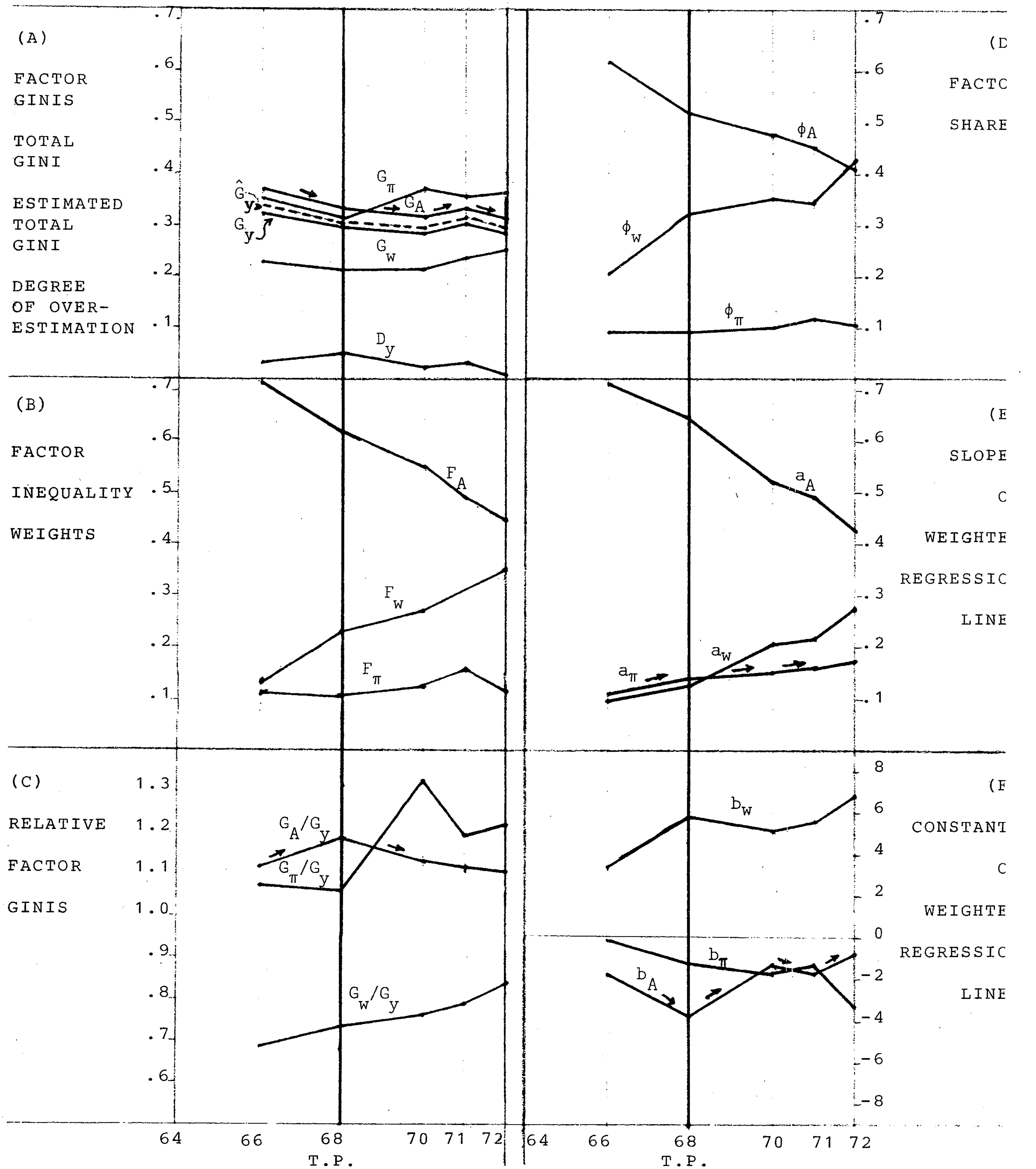


DIAGRAM 4

INEQUALITY TRENDS FOR FARM HOUSEHOLDS



negative (panel F, Diagrams 2,3,4). The economic interpretation of this is that the share of wage income decreases and the share of property income increases as the family gets wealthier. This leads to the phenomenon shown in panel A (Diagrams 2,3,4), i.e., a consistent straddling of the total Gini curve G_y by the two factor Ginis G_w and G_π, as wage income is distributed more equally than total income which is, in turn, distributed more equally than property income.[1]

F.2: Agricultural Income is a Type One income for Rural Households and a Type Two income for All Households.

This finding can be readily detected in Panels E and F of Diagrams 2 and 4.[2] It suggests that the G_A curve lies above the G_y curve for farm households (panel A, Diagram 4) while the opposite is generally true for the all households model (Panel A, Diagram 2). The economic interpretation of this is that, for rural families, the income share traced to agricultural income increases with total family income, while, by implication, the share of non-agricultural income (from rural industries) declines. Thus, while non-agricultural income sources are unusually important in Taiwan for all rural families (see below), they are especially important for the relatively poor rural families. Rural industries thus serve as an important income "equalizer" as far as the rural families are concerned. From the point of view of the whole economy, i.e., the All Families model, on the other hand, the share of income from agriculture decreases with total family income--

[1]See (2.1) and (2.2). The reader might recall that this theorem was established for the "expected" rather than the actual factor income pattern. However, in view of the high correlation coefficients (rows (14)-(16)) indicating a high degree of linearity, the $G_w \lesssim G_y \leq G_\pi$ relation is seen to be valid for the actual factor Ginis as well. An examination of row (23) in Tables 1b and 1c shows no significant difference in the G_w/G_π ratio for the two sectors.

[2]In other words, a_A is always positive while b_A is positive in Diagram 2 but negative in Diagram 3.

for the obvious reason that the wealthier families derive an increasing portion of their incomes from non-agricultural sources, especially property income (also see F.1).

<u>F.3: It follows from (F.1) and (F.2) that there is no Type Three income in the Taiwan case.</u>[1]

This means that the Gini approximation equation in (2.7) above can be used in which the overall Gini is simply the weighted average of the factor Ginis. In the absence of any Type Three income (or error) the remaining non-linearity error is always non-negative (see (2.8)) and $\hat{\hat{G}}_y$ overestimates G_y. Moreover, the size of the non-linearity error $\hat{\theta}$ and of the Gini Error Fraction D_y (see row (19), Tables 1a, b, c and curve D_y in Panel A of Diagrams 2, 3, 4) is uniformly small. In fact, the non-linearity error is so small[2] that D_y never exceeds 3% for all three models and that the approximation equation therefore overestimates the true Gini by less than 3%.[3] For this reason we can safely neglect the non-linearity error term and use equation (3.1) for the non-farm family model and equation (3.8) for the farm family and all families models.

---

[1]A quantitatively small category of "transfers" which we have neglected does show up in the data but it does not decrease absolutely with total family income. At a more disaggregated level some Type Three income would undoubtedly appear, but, it does not appear at any reasonable level of aggregation. However, since the analysis of Type Three income is theoretically interesting and may be of practical policy importance, e.g., in a fiscal redistribution context, we have pursued it fully in the Appendix.

[2]The very low non-linearity error is also demonstrated by the consistently high values of the correlation coefficients in rows (14),(15), & (16) of Tables 1a,b.

[3]In Panel A of Diagrams 2, 3, 4, the estimated Gini $\hat{\hat{G}}_y$ is shown by the dotted curve. The gap between it and the true Gini, the G_y curve, indicates the size of the (non-negative) Gini error term.

In summary, then, the following decomposition equation may be employed in our further empirical analysis of the Taiwanese case:

4.2) a) All Households:

$$G_y = \phi_W G_W + \phi_\pi G_\pi + \phi_A G_A \quad (\text{with } G_W \leq G_y \leq G_\pi \quad \text{and} \quad G_A \leq G_y)$$

b) Farm Households:

$$G_y = \phi_W G_W + \phi_\pi G_\pi + \phi_A G_A \quad (\text{with } G_W \leq G_y \leq G_\pi \quad \text{and} \quad G_A \geq G_y)$$

c) Non-Farm Households:

$$G_y = \phi_W G_W + \phi_\pi G_\pi \quad (\text{with } G_W \leq G_y \leq G_\pi)$$

The primary objective of this analysis is, of course, to attempt to identify the causes of the change in G_y through time. The decomposition equations of (4.2) provide a framework for this effort which is partly deductive and partly inductive. We begin by identifying the three effects on the Gini, i.e., the reallocation effect, the functional distribution effect and the factor Gini effect as derived from (3.2) and (3.10). Since these three causative effects are defined in terms of the various factor Ginis G_i and factor shares ϕ_i, our analysis proceeds by first inductively examining the time pattern of these G_i's and ϕ_i's and, by appealing to growth theory, explaining their pattern of behavior and their relation to the overall Ginis. The impact of these effects on the three overall Ginis (for all families, farm families and non-farm families) are then handled both qualitatively--in terms of the direction of impact--and then quantitatively--in terms of their precise empirical contribution.

Our findings are recorded in Table 2 below; its organization is self-explanatory. Our findings, first qualitative, i.e., the F's, then quantitative, i.e., the numerical (or percentage) contributions of the various effects to changes in the relevant Ginis, are recorded in each cell.

We begin with an examination of the overall index of FID equity in Taiwan and observe that

F.4: For the whole economy (all families) G_y shows a slightly increasing trend between 1964 and 1968 and declines consistently and markedly thereafter (Diagram 2, Panel A).

The recognition of the existence of a "turning point" around 1968 indicates that the so-called Kuznets thesis concerning an inverse U-shaped pattern of distribution over time applies, although it is much less pronounced than in other LDCs.[1] The fact that 1968 appears as something of a turning point from the viewpoint of FID analysis is highly significant also because it confirms conclusions reached independently with respect to the overall development process in post-war Taiwan. Some of the authors' earlier work indicates that it was around this time that the labor supply condition came to an end as real wages began to rise in a sustained fashion.[2] This suggests the existence of a close relationship between growth and income distribution patterns, in particular that once a labor surplus economy has succeeded in making full use of its labor force and turned labor into a scarce factor for the first time, the resulting marked increases in the real wage are likely to lead to a marked improvement in FID.[3]

Since the whole economy--or all families--represents merely an aggregation of the rural and urban families, an examination of G_y for the two sectors separately permits a number of additional observations:

---

[1]This "mildness" of the Kuznets effect when coupled with the (by comparative international standards) overall low level of G_y (see, e.g., Redistribution with Growth, Chenery, Ahluwahlia, Bell, Duloy and Jolly, Oxford University Press, 1974) provided, in fact, a major motivation for the present study.

[2]See Fei and Ranis, "A Model of Growth and Employment in the Open Dualistic Economy: The Cases of Korea and Taiwan," Journal of Development Studies, forthcoming. A vertical line is drawn in Diagrams 2, 3, 4 to indicate the role of 1968 as a turning point (or part of a "turning range"); similarly, Table 2 is broken down into the period before and after 1968.

[3]More work on the historical relationship between growth phases and FID behavior is indicated, especially for the earlier (pre-1964) period. Kowie Chang's data indicate a much higher overall Gini (in the .5 range) for the 50's than the 60's, indicating the possibility of another substantial turning point (and marked improvement in FID) as the economy moved from import-substitution to export substitution in the early 60's. (On a discussion of this export substitution point, see Fei and Ranis, ibid.) But the present paper restricts itself to an analysis of the post-1964 period.

F.5: For farm households, the total Gini G_y declines slightly before the turning point (1968) and remains relatively constant thereafter. For non-farm households, on the other hand, the slight increase of G_y before the turning point gives way to a consistent and significant decline thereafter (Diagrams 3, 4, Panel A).

The differentiated sectoral G_y pattern observed here gets at the heart of the problem of FID which we seek to explain. We see that the slightly inverse U-shaped pattern of G_y for the whole economy (see (F.4)) is basically a reflection of the G_y for the urban families. This is due primarily to the fact that the income of the urban families accounts for a larger and increasing share of the national income relative to that of the rural families. The role of the rural and urban sectors of Taiwan in the course of development is by no means symmetrical. Quite to the contrary, the exhaustion of the economy's labor surplus around 1968 was the result of a process of rapid labor reallocation from the agricultural to the industrial sectors during the 60's.[1] The empirical evidence suggests that, before the turning point, as the agricultural sector gradually loses its surplus labor, the income distribution improves for the rural households while worsening for the urban households. It is reasonable to expect such a result as labor reallocation can be expected to reduce the weight of the poorest families in lower average income agricultural activities and at the same time increase the weight of the poor families in the higher average income non-agricultural activities. This pattern can no longer be observed after the turning

---

[1] As we have found earlier (see Fei, Ranis, *op. cit.*), this rate of labor reallocation proceeded at a rate in excess of 6% annually during the 60's.

point when there is a reduction in the average income gap between the rural and urban families.[1]

The pattern of the income distribution equity over time can now be analyzed more precisely with the help of equation (3.10)--for the all families and rural families models--and equation (3.2)--for the urban families model. These equations, it will be recalled, permit us to decompose changes in the overall Gini into reallocation effects, functional distribution effects and factor Gini effects.[2] Concentrating first on the reallocation effect, i.e., the decline of the agricultural income (share ϕ_A), this effect is relevant to both the all households and the rural households models. Our empirical evidence indicates the following:

F.6: For the whole economy, the distributive share of agricultural income (ϕ_A) drops consistently for the whole period ('64-'72)--with the drop more pronounced before the turning point (Diagram 2, Panel D).

This decline of ϕ_A is associated with the rapid reallocation of labor from agricultural to non-agricultural activities in the course of a rapid industrialization process.[3] The most rapid reallocation actually took place after 1960 during the so-called "export substitution phase" of Taiwanese economic development characterized by the rapid absorption of rural manpower

---

[1]For more discussion on the inter- vs. intra-sectoral explanation of the overall FID pattern, see below. It is true that labor reallocation continues after the turning point; however, labor is now a scarce commodity, the wage gap between the sectors has narrowed and further reallocation no longer improves G_y for farm families or worsens it for non-farm families.

[2]They are derived by differentiating the approximation equations summarized in (4.2) with respect to time.

[3]The proportion of the labor force in non-agriculture increased from 42% of the total in 1952-54 to 58% of (a much larger) total in 1967-69.

in labor intensive export-oriented industries.[1] It was this particular pattern of rapid growth which produced the "reallocation effect" on the Gini for the whole economy. Its impact on FID can be traced as follows:

F.7: For the whole economy, the reallocation of labor from agricultural to non-agricultural production (i.e., the decline of ϕ_A) contributes to the worsening of the overall equity of income distribution, i.e., an increase of G_y.

The fact that the reallocation effect works against the equity of income distribution follows directly from (F.6) (i.e., ϕ_A declines) and (F.2) (i.e., $G_A < G_y$). Since agricultural income acts as an "income equalizer" for the whole economy, the decline of its importance or relative share thus worsens the overall distribution of income. (Technically, (F.7) can be seen from the fact that the term A' in (3.10b) is positive.) The labor allocated out of agriculture will move partly into larger-scale urban non-agricultural activities and partly into rural industry and services. This means that rural households receive an ever-increasing share of their incomes from non-agricultural employment. That this has happened in a quantitatively very significant way is confirmed by the following observable characteristic:

F.8: For the rural households, the distributive shares of non-agricultural income ($=\phi_W+\phi_\pi$) represent a quantitatively important income component which increases consistently throughout the whole period (Diagram 4, Panel D).

---

[1]Industrial exports as a percentage of (a rapidly growing) total export rose spectacularly from 40% of the total in 1958 to 89% in 1967-69. (Industrial exports rose from $16 per capita in 1952-54 to $223 in 1967-69.) (See Fei, Ranis, op. cit.)

The fact that non-agricultural income ($1-\phi_A=\phi_\pi+\phi_W$) generated by rural industry and services has always been an important source of rural family incomes is seen from the fact that $\phi_W + \phi_\pi$ was about 30% from the very beginning. It is even more noteworthy that this share has climbed to 60% in the short span of six years (68-72) so that it has become the dominant source of income at the end of our period of observation. The pattern which has emerged gives evidence of a rather unique spatially dispersed industrialization pattern. The relative importance of this non-agricultural rural activity can also be seen from the fact that value added in rural industry constituted about 20% of value added in urban industry, on the average, during this period. Hence, a study of the FID of Taiwan, especially its rural households, must pay special attention to this unusually large rural industry component.[1] This substantial shift of the source of rural family income helped produce a reallocation effect beneficial to the income distribution pattern of these families, as follows:

F.9: For the rural households, the increased share of non-agricultural income ($\phi_\pi + \phi_W$) contributes to an improvement in the equity of their distribution of income, G_y.

That the above reallocation effect works to improve equity (lower G_y) follows directly from (F.5) and (F.6). As we pointed out earlier, for rural households non-agricultural income is an "income equalizer" and hence the increase of its weight or distributive share will improve the rural families' equity of income distribution. (Technically, this is seen from the fact that the term A' in (3.10b) is now negative.)

---

[1]Certainly by any international LDC standards (see Sharing in Development, ILO, 1974, Special Paper #9 on "Medium-Scale and Small-Scale Industry."

Let us turn now to the functional distribution effect, i.e., the impact (on changes in FID) caused by variations in the relative wage-profit share ratio ϕ_w/ϕ_π prevailing in non-agricultural activities. This effect may be traced to the production conditions and the functional distribution forces within the non-agricultural sector.[1] The time path of ϕ_w/ϕ_π for all three models is shown in Diagram 5 for purposes of comparison. From this diagram we can see that

F.10: The relative share ratio ϕ_w/ϕ_π is higher for rural non-agricultural activities than for urban non-agricultural activities. Moreover ϕ_w exceeds ϕ_π. (i.e., $1 < (\phi_w/\phi_\pi)_{urban} < (\phi_w/\phi_\pi)_{rural}$.)

The economic interpretation of this finding is that the wage share is larger than the property share in both sectors[2] and that rural industries (and services) are relatively more (in fact, almost twice as) labor intensive than their urban counterpart.[3] Thus one unit of capital provides more employment in rural than in urban non-agricultural activities. Such a characterization of the comparative advantage between the two types of industries is indeed what we would expect to find.[4]

---

[1] It is obviously not relevant in the case of lumped agricultural income.

[2] Notice that the ϕ_w/ϕ_π is merely a weighted average of the sectoral values. Thus in Diagram 5, the ϕ_w/ϕ_π curve for the whole economy lies in the middle and close to that of the urban ratio because of its relatively heavier weight. Since the overall ϕ_w/ϕ_π ratio is merely an aggregative index for the two sectoral ratios, we shall only emphasize the two sectoral values in our analysis of this section.

[3] If p is the wage-profit or factor-price ratio applying to both sectors (i.e., there are no factor-price distortions), and K_r^* and K_u^* represent the capital-labor ratio for rural and urban non-agriculture, respectively, the inequality in (F.10) implies that $p/K_r^* > p/K_u^*$ or $K_u^* > K_r^*$, i.e., for the same p, the capital per head in urban industries is larger, which is the conventional definition of "relatively high capital intensity."

[4] It supports our earlier finding that the development of rural industries played a crucial role in solving the unemployment or underemployment problem in Taiwan.

The time pattern of the relative shares ϕ_W/ϕ_π in Diagram 5 shows a generally increasing trend over the time span (66-68) although there are periods of decline--indicated by the "broken line" portions of these curves--in both sectors. These observed patterns may be described as follows:

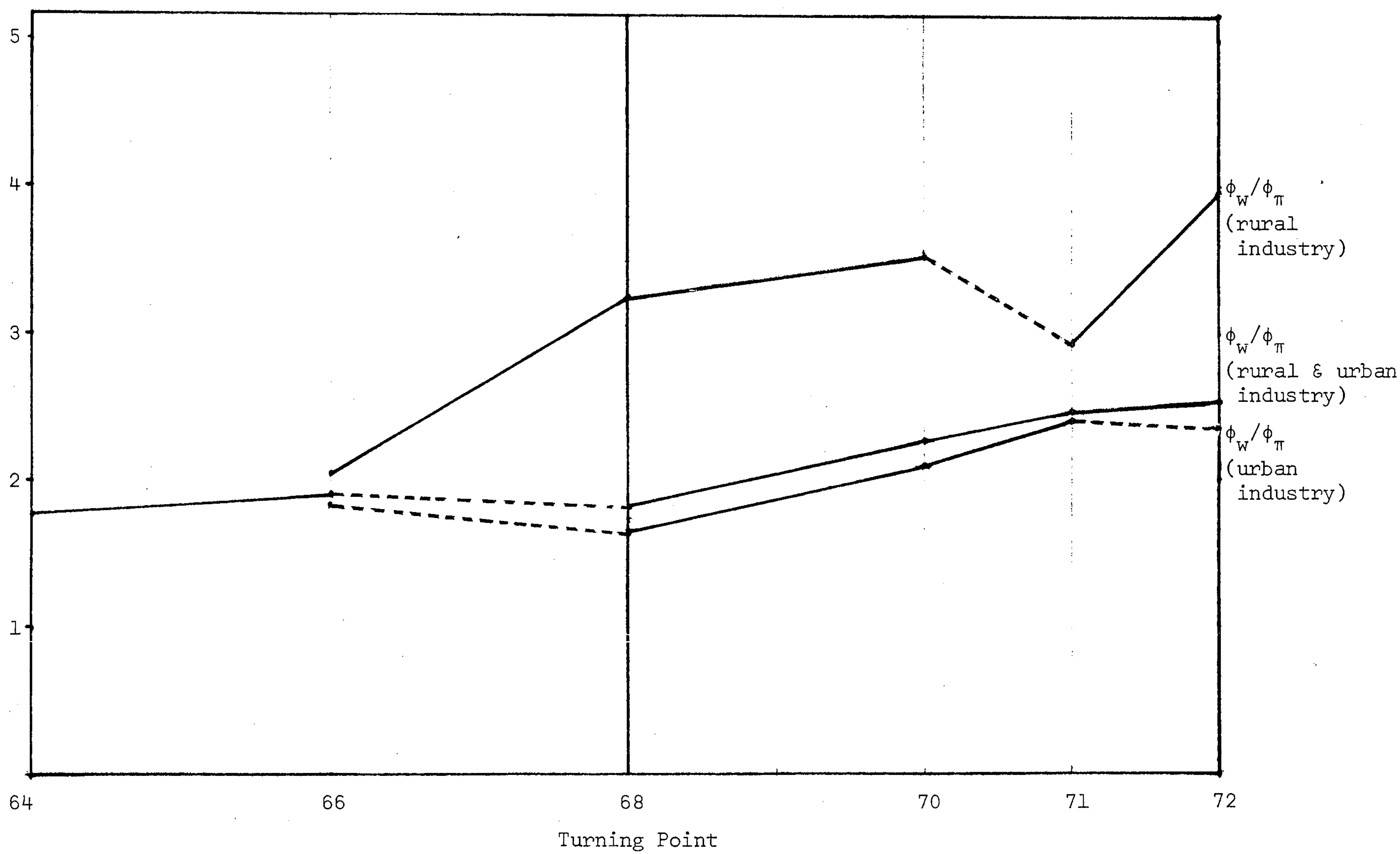
F.11: For urban industries, the time path of ϕ_W/ϕ_π is U-shaped, i.e., the relative labor share decreases before the turning point (1968) and increases after the turning point. For the rural industries, the relative labor share (ϕ_W/ϕ_π) increases consistently over the whole period (with the exception of one year).

These time patterns may be explained with the help of growth theory introduced earlier. In the period before the turning point when the real wage is relatively stable (i.e., the surplus labor condition exists), the necessary and sufficient condition for the upturn of ϕ_W/ϕ_π is that the labor using bias effect of technology change, B_L, overwhelms the intensity of innovation effect, J (see (3.6)). Finding (F.11) immediately suggests that this holds in rural industry but fails to hold in large scale urban industry. The significance of this is that when the surplus labor condition obtains, i.e., before the turning point, rural non-agricultural activities are more technologically responsive to the ample supply of cheap labor than large scale urban industries.

In the period after the turning point, when the labor surplus condition is exhausted and real wages begin to increase significantly, the necessary and sufficient condition for ϕ_W/ϕ_π to increase through time is given by (3.4), namely, a combination of industrial capital deepening and labor using technology bias. Since capital deepening does occur in both industrial sectors after the turning point, all we need is that technology

Diagram 5

Comparison of Relative Wage-Profit Shares of Urban and Rural Industries



bias is not so labor-saving as to overwhelm the increased capital intensity to ensure the observed result.[1] It should also be noted (see Diagram 5) that the difference in relative factor shares (or capital intensity) in the two sectors narrows somewhat after the turning point.

The gain in Taiwan's competitive position in world export markets at this time may be traced mainly to increases in the productive efficiency of her labor intensive industrial exports (i.e., due to the combination of entrepreneurial capacities with abundant labor, utilizing labor-using technological innovations). Assuming that the origin of these exports is initially in rural based industry but, with increasing skill and technology content, shifts gradually toward urban industry permits us to explain why the urban and rural ϕ_W/ϕ_π ratios diverge before the turning point but move closer together thereafter when wage gaps narrow and industrial dualism disappears.

Having identified the growth related causation of the change in the relative distributive shares ϕ_W/ϕ_π, we can easily summarize the impact of these growth related forces on FID with the aid of term A" corresponding to the "functional distribution effect" in (3.10c). Generally,

F.12: An increase of the relative distributive share ϕ_W/ϕ_π in favor of the laboring class always improves the equity of overall FID.[2]

Hence, before the turning point, the strong labor using bias of innovations contributes to FID equity among rural families while a relatively lower labor using bias serves to slightly worsen FID among urban families. After the turning point, capital deepening and the labor using bias of innovations affect FID favorably for both rural and urban households and for the whole economy.

---

[1]This, again, is under "normal" assumptions about factor non-substitutability ($\varepsilon < 1$).

[2]We should recall that this follows directly from the fact (see (F.1)) that wage income is always distributed more equally than property income.

Turning, finally, to the <u>factor Gini effect</u>, let us recall that the reallocation effect and the functional distribution effect refer to the impact on FID traced to changes in the functional distribution of income, ϕ_i, holding the factor Gini constant. The factor Gini effect, on the other hand, refers to the impact on FID traced to variations of the factor Gini through time. It should be noted that we have tried to understand the underlying causation of the variation of the income shares (ϕ_i) in terms of certain familiar growth relevant forces--reallocation between agriculture and non-agriculture in the dualistic economy and changes in the factor shares. The causation of the variation of factor Ginis through time, on the other hand, represents a more difficult area which has to rely on as yet less well understood theoretical issues. While we shall proceed to identify certain important observable characteristics of the time patterns of the factor Ginis, attempts to "explain" these patterns can be little more than intelligent guesses at this stage of our understanding.

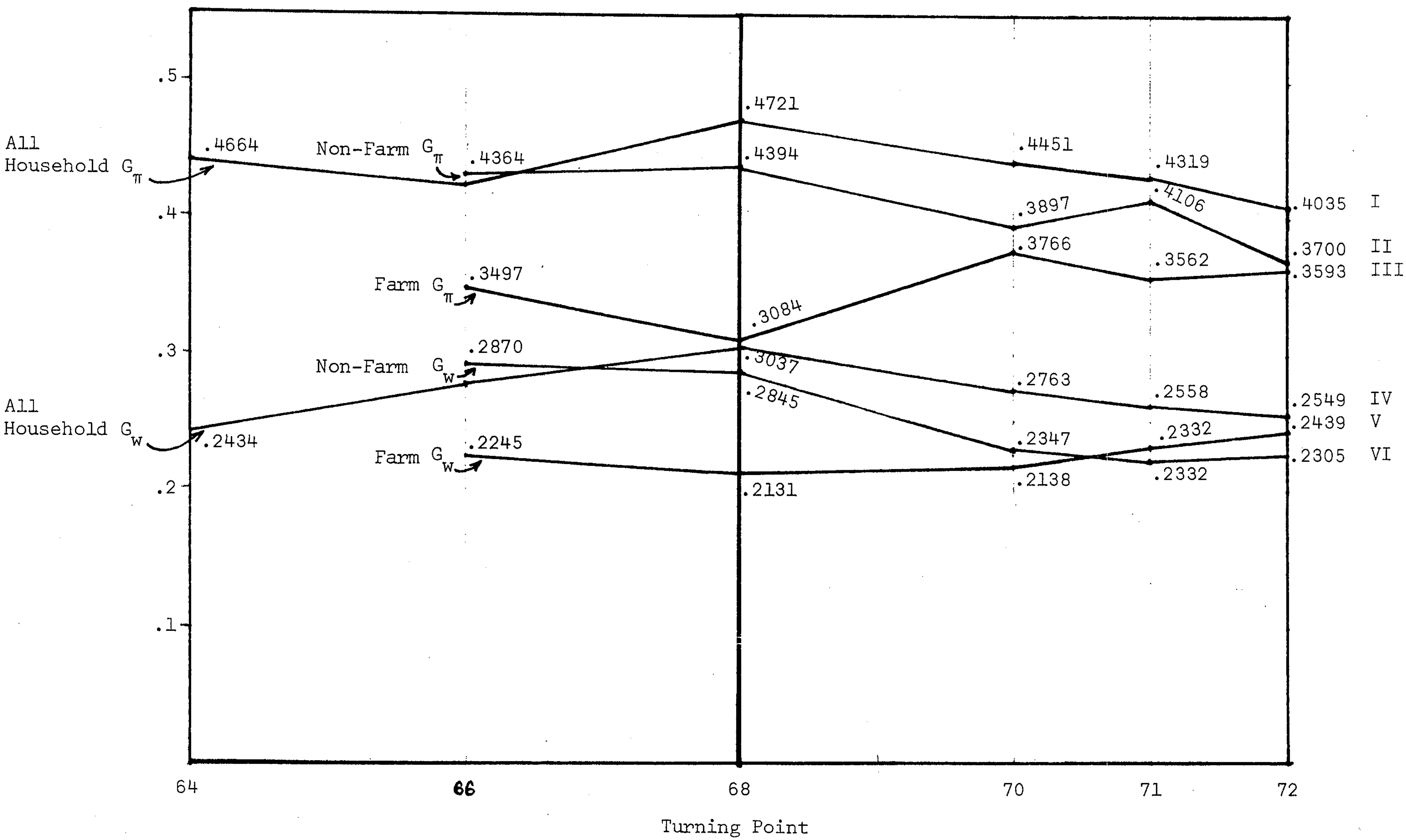
Since family incomes are generated by two types of productive activities (agricultural and non-agricultural) in the dualistic economy, let us first concentrate on non-agricultural income--i.e., property and wage income arising in industry and services. The time pattern of the factor Ginis for these incomes are shown in Diagram 6. From this diagram, a number of observable characteristics can be identified:

<u>F.13</u>: <u>The total wage income Gini is lower than the total property income Gini, i.e., total wage income is distributed more equally than total property income (Curves I, II, III lie above Curves IV, V and VI in Diagram 6)</u>

This observable characteristic is merely a further strengthening of (F.1). While (F.1) noted that $G_W < G_\pi$ <u>within</u> any sector, the statement

Diagram 6

Cross-Sectional Comparison of Time Pattern of Wage Gini and Property Gini



here is stronger in that it refers to cross-sectoral comparisons, e.g., the G_w for all households is less than the G_π for farm households. This stronger conclusion enables us to make the intuitively obvious and unconditional statement that it is the existing inequality in the distribution of the ownership of property which makes a substantial contribution to overall income distribution inequity.

F.14: <u>For each functional type of income (i.e., wage or property), both sectoral factor Ginis (i.e., the factor Gini for the farm and for the non-farm households) lie below the factor Ginis for the economy as a whole.</u>

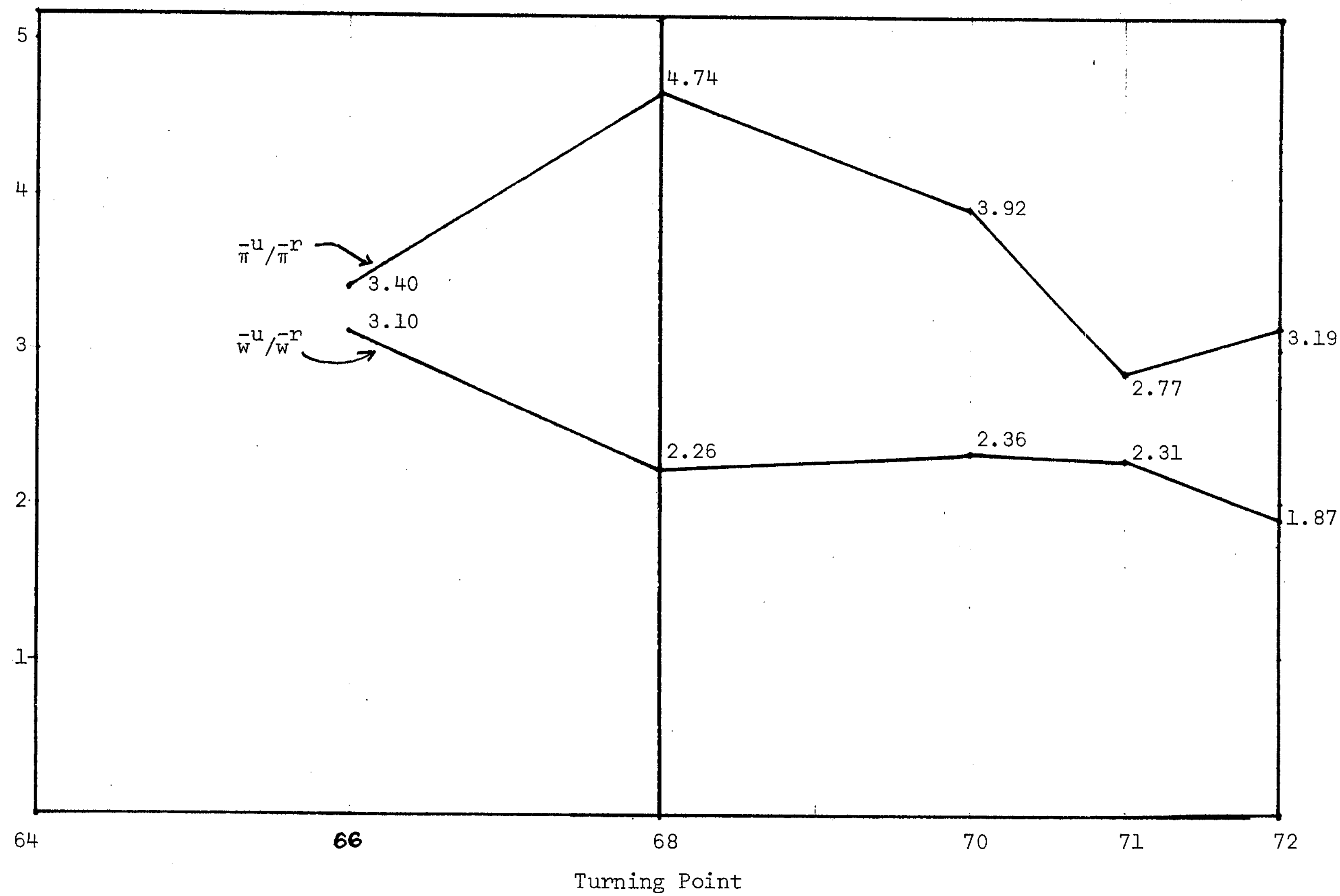
With respect to property income, Curve II and III lie below Curve I (Diagram 6). Similarly, for wage income, Curves V and VI lie below Curve IV. The common sense of this phenomenon is that if you segment a population into a relatively rich and relatively poor component, i.e., there is a significant gap between the average incomes of the two components, the Ginis for the two (now more homogeneous) components tend to be lower than for the population as a whole.[1] Diagram 7 indicates a substantial gap between the average wage income in the urban and the rural sectors ($\bar{w}^u$ and $\bar{w}^r$). The same thing holds

---

[1]For example, taking the extreme case of seven families segmented into three relatively low income rural families and four relatively high income urban families, $Y = (Y^u, Y^r) = [(1,1,1),(100,100,100)]$ so that within each sector (i.e., (1,1,1) and (100,100,100)) the degree of equity of income distribution is perfect, while for the whole economy the Gini is quite high. In Theil's terminology, T(Y) is the sum of the weighted average of inequality within each sector and the inequalities between the sectors.

Diagram 7

Time Pattern of Relative Urban-Rural Average Wage and Profit Income



narrowing of this gap follows directly from the argument above (F.15) explaining the existence of the gap. Thus, after the turning point in 1968, the forces making for a gap between rural and urban industries tend to weaken. As the labor surplus is exhausted and the entire economy becomes increasingly subject to competitive market forces, we can expect the trend for a narrower gap with respect to both the ownership of capital and the heterogeneity in quality and bargaining power of the labor force, as between rural and urban industries, to set in.

F.18: For each functional income component, the gap between the total Gini and the sectoral Gini tends to narrow.

This is due to the narrowing of the gap between the average income from wages--and profits--in the urban and rural sectors after the turning point, as clearly shown in Diagram 7. This gap ($\bar{w}^u/\bar{w}^r$) for wages declined from 2.3 in 1968 to 1.9 in 1972, and ($\bar{\pi}^u/\bar{\pi}^r$) for profits declined from 4.7 in 1968 to 3.2 in 1972.

Thus, in the brief post-turning point era under observation, there is a tendency for a convergence in the rural and urban property and wage Ginis. The gap between the equity of wage and property incomes as a whole, on the other hand, is not reduced and must be viewed as a longer term phenomenon.

Up to this point we have been focusing mainly on the comparative magnitudes of the factor Ginis for industrial activity as identified in Diagram 6. Concentrating more fully on the time pattern of these curves, a look at Curves I and V permits us to observe that

F.19: Total industrial property and wage income show an inverse U-shaped pattern, with 1968 as a turning point.

for the gap between average urban and rural property incomes ($\bar{\pi}^u$ and $\bar{\pi}^r$).

F.15: For each functional factor income component, the factor Gini for the farm households is smaller than that of the non-farm households.

In Diagram 6, Curve III lies below Curve II--for property income--and Curve V lies above Curve VI--for wage income. This means that the ownership of capital assets in the larger-scale urban industries is more unequally distributed than that of rural industries. Such a high concentration of ownership supported by a highly imperfect capital market is a pronounced feature of the industrial structure in LDCs. Furthermore, the wage income generated by urban industries is also more unequally distributed than that by rural industries, since urban industries are characterized by a greater skill heterogeneity of the labor force, more unionized labor, greater impact of minimum wage legislation, etc..

F.16: Notice that the above three properties, (F.13), (F.14), and (F.15), together establish the complete ordering of all six components shown in Diagram 6.

The net result is that property income in the urban industries tends to be more unequally distributed than the wage income of rural industries. It is for this reason, everything else being equal, that overall inequality is more related to the concentration of capital in the urban sector and less to the inequality of wage income in the growing rural industries.

F.17: For each functional income component, the gap between the two sectoral factor Ginis tends to narrow after the turning point.

In Diagram 6, we may note that, for property income, the gap between Curves II and III tends to narrow over time. Similarly, for wage income, the gap between Curves V and VI tends to narrow. The explanation for the

We will recall (from (F.4)) that the overall distribution of income, G_y, exhibits the same inverse U-shaped characteristic, which is, in fact, what we are basically trying to explain. Thus, there is a prima facie case for the factor Gini effect due to industrial income providing a large part of the explanation for the pattern of the overall Gini. We may also note that (Diagram 6, Curves II and V, and III and VI)

F.20: Within each of the (urban and rural) industrial production sectors, the two factor Ginis (G_W and G_π) behave consistently, i.e.,

a) for the urban industries they are nearly constant before the turning point and decline consistently thereafter (i.e., inverse U-shaped)

b) for the rural industries they fall somewhat before the turning point and rise consistently thereafter (i.e., U-shaped).

c) for both industries, taken together, they increase before and fall after the turning point.

These results merit a number of additional observations:

1) With respect to the consistency of the movement of G_W and G_π, and taking urban industries as an example, we know there exists a dualistic structure within such industries--with the large-scale organized subsector hiring higher quality and/or unionized workers and the smaller-scale subsector hiring lower quality and/or unorganized labor. The existence of such dualism usually connotes the prevalence of a substantial wage gap as well as a gap in the rate of return to capital favoring the larger scale.[1]

[1] See Sharing in Development: A Programme of Employment, Equity and Growth for the Philippines, ILO, Geneva, 1974; also, Little, Scitovsky and Scott, Industry and Trade in Some Developing Countries, Oxford University Press, 1970.

The stronger this dualism, the wider these gaps and therefore the more unequal the distribution of wage and property income. The inverse U-shaped pattern shown by urban industry indicates the increasing orientation toward dualism before the turning point and the decreasing orientation thereafter.[1]

2) Given the time pattern of the six industrial factor Ginis pictured in Diagram 6, we can thus unambiguously discuss the direction of the factor Gini effects for each of the relevant total G_y. Since wage and property Ginis for each model move consistently

F.21: The factor Gini effect of total non-agricultural income (i.e., wage and property income) on the G_y of

a) all households is somewhat unfavorable before the turning point and favorable after the turning point (Curves I and IV)

b) urban households is neutral before the turning point and favorable after the turning point (Curves II and V)

c) rural households is neutral before the turning point and unfavorable after the turning point (Curves III and VI).

Returning to agricultural income:

F.22: The G_A curves of Diagrams 2 and 4 (Panel A) permit the conclusion that the factor Gini effect of agricultural income is, on the whole, favorable to the Gini of both the farm and total family models.

This results from the fact that lower income agricultural families are seen to be consistently improving their position relative to higher income agricultural families.

---

[1]The explanation for the U-shaped pattern for rural industries is less obvious.

We are now in a position to supplement the above "qualitative analysis" with a more precise "quantitative analysis." The actual annual average changes of G_y (i.e., dG_y/dt), before and after the turning point, for the three models, are indicated in column (1) of Table 2.[1] Notice that in the All Households model, G_y increases moderately before the turning point while declining significantly after the turning point. This contrast of a "moderate change" with a "significant change," visually apparent from Diagram 2 (Panel A), can now be seen precisely in terms of the fact that the absolute magnitude of decline after the turning point (-.0146) is nearly 10 times as large as the moderate annual increase before the turning point (.0013). Similarly, for the Rural Households model, the contrast between a significant decline (-.0168) before, with a moderate decline (-.0014) after the turning point should be noted. In the Non-Farm Households model, the contrast between a moderate increase (.0020) before the turning point with a substantial decline (-.0167) after the turning point should also be noted.

In the case of a significant change of the G_y for any of the three models, in either direction, we want to determine the quantitatively dominant causative factors. On the other hand, in the case of moderate or no changes in G_y, we want to determine whether this is due to the stability of all the causative factors or, alternatively, to the cancellation of "positive" causative factors with "negative" causative factors.

---

[1]The numbers in brackets in each cell of column (1) represent observed annual average changes of G_y, while the numbers without brackets represent the sum of the estimated changes of the components (i.e., A' + A" + B) according to equation (3.10). The small differences between the estimated and exact changes is due to three sources, the first two of which have been referred to earlier: (i) the existence of a small non-linearity error; (ii) the fact that the distributive shares do not exactly add up to one, since some "mixed income" components were excluded; and (iii) the need to make discrete approximations to continuous changes.

Table 2

Decomposition of the Changes in FID

		Average Annual Change of G_y (1)		(Net) Factor Gini Effect (2)	Reallocation Effect (3)	Functional Distribution Effect (4)	Decomposition of (Net) Factor Gini Effect: Agriculture and Non-Agriculture: Non-Agriculture (5)	Agriculture and Non-Agriculture: Agriculture (6)	Property Income and Wage Income: Property (7)	Property Income and Wage Income: Wage (8)
ALL HOUSEHOLDS	Before 1968 (1)	+.0013 [.0016] (F4)	=	+.0012 (92%)	+.00002 (15%) (F7)	-.00001 (-8%) (F12)	.0077 (592%) (F21a)	-.0065 (-500%) (F22)	.0006 (8%) (F20c)	.0071 (92%) (F20c)
	After 1968 (2)	-.0146 [-.0131] (F4)	=	-.0131 (-90%)	-.0010 (7%) (F7)	-.0025 (-17%) (F12)	-.0108 (-74%) (F21a)	-.0023 (-16%) (F22)	-.0042 (39%) (F20c)	-.0066 (61%) (F20c)
FARM HOUSEHOLDS	Before 1968 (3)	-.0168 [-.0181] (F5)	=	-.0083 (-49%)	-.0068 (-40%) (F9)	-.0017 (-11%) (F12)	-.0036 (-21%) (F21c)	-.0047 (-28%) (F22)	-.0021 (-58%) (F20b)	-.0015 (-42%) (F20b)
	After 1968 (4)	-.0014 [~ 0] (F5)	=	+.0009 (+64%)	-.0019 (-136%) (F9)	-.0004 (-29%) (F12)	.0053 (379%) (F21c)	-.0044 (-314%) (F22)	+.0012 (23%) (F20b)	.0041 (77%) (F20b)
NON-FARM HOUSEHOLDS	Before 1968 (5)	+.0020 [.0036] (F5)	=	-.0003 (-15%)		+.0023 (115%) (F12)	-.0003 (-15%) (F21b)		+.005 (167%) (F20a)	-.0008 (-267%) (F20a)
	After 1968 (6)	-.0167 [-.0166] (F5)	=	-.0132 (-79%)		-.0035 (-21%) (F12)	-.0132 (-79%) (F21b)		-.0051 (39%) (F20a)	-.0080 (61%) (F20a)

Concentrating first on the All Households model, our major quantitative finding is that

(1) For the All Households model, the dominant causative factor of change is the Factor Gini Effect. More precisely:

(a) Before the turning point, the modest gain in G_y is due to the dominance of the unfavorable contribution of the non-agricultural Gini effect over the favorable agricultural Gini effect.

(b) After the turning point, the significant decline of G_y is contributed to by the favorable contribution of both the agricultural and non-agricultural Gini effects, especially the former.

From Table 2, Column (2),(Rows 1 and 2), we see that for both before and after the turning point, the factor Gini effect accounts for over 90% of the total causation of change of G_y. The reallocation effect is 15% before and 7% after the turning point, i.e., losing importance; the functional distribution effect (8% before and 17% after) is gaining weight. Together they account for only 8-10% of the total change of G_y, in either direction. Thus, for the All Households model, the factor Gini effect clearly dominates and warrants more careful examination. It should be noted that we have labeled Column (2) of Table 2 the (Net) Factor Gini Effect. This is because the factor Gini effect can be decomposed further, by applying equation (3.10). The results are shown in columns (5) to (8) of Table 2. In columns (5) and (6), the decomposition is performed with respect to agricultural and non-agricultural income.[1] The non-agricultural income

[1]In other words, the values in columns (5) and (6) add up to the value in column (6). Likewise, the percentages, in (5) and (6), add up to those of column (2).

Gini effect of column (5) can be further decomposed into its wage and property Gini effects as shown in columns (7) and (8).[1]

There is a basic difference between the factor Gini effect before and after the turning point. Before the turning point, the factor Gini effect for the two production sectors (agricultural [-500%] and non-agricultural [-592%]) work in opposite directions (see columns (5) and (6), Table 2). After the turning point, they work in the same favorable direction with non-agriculture making the larger contribution. This means that the industrial factor Gini coefficient, G_x, (in equation (3.10)) itself has a very pronounced inverted U-shaped characteristic,[2] reminding us of the Kuznets hypothesis. However, the existence of a continuously falling agricultural Gini, G_A, somewhat overcomes the inverse-U-shaped characteristic of G_x-- producing the observable pattern of G_y.

The implications of this conclusion for distribution and growth in the labor surplus economy are three-fold. First, the "Kuznets effect" is not very pronounced, i.e., the "conflict" between development and FID before the turning point is relatively mild. Second, the "switch" from a slightly worsening to an improving FID after the turning point may be said to be largely due to a significant shift in the contribution of the non-agricultural Gini effect (from +592% to -74%). In other words, before the turning point, it is the non-agricultural sector which is almost entirely responsible for the (slight) worsening of the overall Gini, since rapid industrialization apparently requires an increasingly heterogeneous labor force and a differentiated industrial structure, with

---

[1]The values of (7) and (8) add up to the value in column (5). (But the two percentages in (7) and (8) add up to 100%.)

[2]This can be easily shown when the time series G_x is plotted.

wealthier families owning the large scale and poorer families the smaller scale.[1] Third, we see that, from the very beginning, the agricultural Gini effect is powerful and favorable and thus represents an important instrument for reducing the initial (upward) steepness of the inverted U-shaped pattern, and possibly eliminating it altogether.

(2) For the farm households, the favorable reallocation effect has a quantitative significance which is equal to or greater than that of the factor Gini effect.

(a) Before the turning point, the significant decline of G_y was contributed to by a favorable factor Gini effect and a favorable reallocation effect of approximately the same weight.

(b) After the turning point, the moderate decline of G_y is due to the fact that the unfavorable factor Gini effect is overwhelmed by a favorable reallocation effect.

Thus, before the turning point, both the factor Gini effect (-49%) and the reallocation effect (-40%) are favorable and quantitatively important. The functional distribution effect accounts for only 11% of the total change of the Gini. After the turning point, the favorable reallocation effect (-136%) is nearly twice as large and thus overwhelms the unfavorable factor Gini effect (+64%). As compared with the All Households model, we see that the reallocation effect is a much more important factor here in explaining the consistent improvement in the equity of FID for rural households. This is due to the

[1] Notice that the wage income Gini effect (column (8)) contributes the lion's share (92%) during the pre-turning point (unfavorable) phase but gradually gives way to the property income Gini effect in the post-turning point (favorable) phase. This is undoubtedly related to the reduction of market imperfections favoring the larger capitalists and the unionized workers after 1968.

obvious fact that, for rural households, agriculture provides a dominant source of income. If we recall that G_A exceeds G_X (the Gini of income from rural industries), we can readily see why the relative decline of the weight of agricultural income over time and the relative gain in the weight of non-agricultural income (favoring the relatively poorer families) represents a dominant factor in the improvement of rural household FID.

Notice that the decomposed factor Gini effects behave differently before and after the turning point. While the agricultural factor Gini effect is consistently favorable (i.e., -28% before and -314% after the turning point), the non-agricultural Gini effect changes from making a mildly favorable (-21%) to a strongly unfavorable (379%) impact after the turning point. The non-agricultural factor Gini effect (i.e., income from rural industries) thus has a distinct U-shaped characteristic while the agricultural income factor Gini effect has a consistent impact on improving the overall FID of rural households.

The practical policy significance of our findings is that, before the turning point, rural industries grow in a context of abundant supplies of surplus labor while, after the turning point, the environment becomes much more commercialized, with labor becoming a scarce and more expensive factor. Thus, before the turning point we expect rural industries to benefit the relatively low income families in the mopping up of unemployed rural labor. After the turning point, on the other hand, the changed environment leads us to expect the income distribution generated by rural industries to begin to contribute to an unfavorable factor Gini effect as labor becomes more heterogeneous and rural industry more differentiated by scale. Our analysis has shown that this unfavorable development can be offset by two favorable trends (i.e., the allocation effect and the agricultural income Gini effect) leading to a possible overall improvement in FID equity. The new rural industries and services can thus produce a slight overall improvement throughout in the FID for rural families, even though in this process forces are generated (e.g., the factor Gini effect associated with the creation of relatively modern rural industries) which work against improving FID.

(3) For non-farm households, the functional distribution effect represents a quantitatively significant causative factor in determining changes in overall FID.

(a) Before the turning point, the modest rise in G_y is due to the fact that an unfavorable functional distribution effect overwhelmed the favorable factor Gini effect.

(b) After the turning point, the significant decline of G_y is mainly contributed to by a favorable factor Gini effect which is reinforced by a favorable functional distribution effect.

Thus, before the turning point, the unfavorable functional distribution effect (115%) is almost eight times as large as the favorable factor Gini effect (-15%) and hence overwhelms it. After the turning point, the decline of G_y is accounted for jointly by the favorable factor Gini effect (-80%) and a favorable functional distribution effect (-20%). Notice that, in contrast with the two early models, we now cannot neglect the functional distribution effect on overall FID--for the obvious reason that it is in the relatively large scale urban industries that capital deepening and technology change are prominent features of industrialization. In particular, our earlier conclusion pointed to the fact that while high innovation intensity contributes unfavorably to the FID of urban families before the turning point, capital deepening induced by real wage increases plus the force of innovations made a favorable contribution to FID after the turning point.

Notice that while the factor Gini effect had a consistent favorable impact over the whole period, its negative contribution after the turning point (-.0132 per year) is more than 40 times as large as before the turning point (-.0003 per year). It is thus clear that the coming of the turning point signals a substantial improvement in the contribution of both wage (-.0091) and property (-.0080) income in the urban centers--as the strength of dualism within urban industry is reduced. The overall pattern of non-farm family FID thus fits the Kuznets pattern quite well--as in the All Families case.

Summary and Conclusions

In this paper we have attempted to analyze the distribution of family income at the aggregate level, and have shown that FID, as measured by G_y, is much affected by the particular forces of growth that the country is experiencing. In the case of Taiwan, for example, the arrival of the turning point (signifying the exhaustion of surplus labor) can bring about a marked difference in the behavior of G_y. In analyzing the causation of the change of G_y and relating it to growth, it is crucial to give prominent recognition to the fact that family income has several factor components, differing in type and impact, on overall equity (in our empirical case, agriculture, wage and property incomes). The change of G_y through time, as affected by growth, can then be analyzed in terms of reallocation effects, factor Gini effects, and functional distribution effects, defined in terms of these factor components, in quantitative as well as qualitative terms.

Along with the importance given to historical phases in growth theory goes the recognition of the relevance of sectoral dualism. Such dualism is a root cause of some of the complexities inherent in FID analysis because the incomes generated by different production sectors tend to behave differently in their respective impact on G_y. The relevance of this structure is reflected in the identification of the three separate models--all households, farm households, and non-farm households--according to which our theoretical as well as empirical analysis of the Taiwan case was organized.

A number of our findings with basic implications for both policy and the future strategy of research deserve a succinct summary. With respect to the implications for policy, our findings tend to support the notion that the most effective method of tackling the maldistribution of income is via a change in the nature of the growth path itself. Income distribution can, of course, be affected directly, e.g., via fiscal policy and transfers (e.g., Type Three incomes);[1] but the experience of Taiwan--with unusually low levels of Gini and an unusually mild Kuznets effect over time, and without any significant Type Three income--bears out our conviction that significant and sustained changes in FID equity are achieved mainly through the modification of the forces underlying the pattern of growth.

Our first conclusion relates to the possibilities demonstrated here for a substantial reduction (and possibly elimination) of the so-called conflict between growth and income distribution. Our work clearly demonstrates how and why the arrival of the turning point and the exhaustion of labor surplus benefit the overall distribution of income along with the objective of growth. But even more interesting is the finding that, even before the turning point, the agricultural income Gini can make a substantial contribution to the moderation of any worsening of the distribution of income.

The non-farm household distribution pattern follows the overall pattern most closely (i.e., it is also inverse U-shaped) but it is substantially more pronounced thanks to the ameliorating effect of the farm household patterns. Turning to these farm households, we may note that the initially favorable pattern of the distribution of income before the turning point derives essentially from two sources: one is the unusually dispersed location pattern of Taiwanese industries providing employment and profit opportunities to the relatively poor rural families. (This is reflected in the reallocation effect showing the shift of the poorest members of the rural population into small-scale labor-intensive rural industry as indexed by a decline of the

---

[1]But keep in mind the very limited fiscal capacity of most LDC governments.

agricultural share and an increase of the non-agricultural share of rural family incomes). The second beneficial contribution is made by the fact that agricultural income itself (i.e., the agricultural Gini effect) shows a pattern of consistent and sustained distributional equity, i.e., a low G_A both before and after the turning point. It is this effect more than any other that improves the overall equity of farm families and softens the Kuznets tendency toward a worsening of the FID for all families before the turning point. Rapid agricultural mobilization based on the improvement of the efficiency of relatively poor, small farmers, both as farmers and as participants in the rapidly growing rural industries and services, thus played a decisive role.

Our analysis, and its inadequacies, also point out directions future research might usefully take. What we have tried to do, essentially, is to decompose the overall pattern of income distribution by type of income, and, via a time-phased examination of changes in different growth regimes, to tie the analysis of FID to the analysis of the underlying factors determining growth. The decomposition of total FID change into its component effects also gives us an indication as to where future effort might have the largest explanatory pay-off for each of the models under consideration. For example, when the factor Gini effect accounts for 80% of the total explanation of overall FID changes, it clearly pays to expend more research effort on the behavior underlying it; similarly, the functional distribution effect warrants more attention in some cases than in others.

Secondly, we can already see the need to further integrate two types of analytical procedures: ours, which decomposes family income by additive factor income components, and Theil's which segments family income by homogeneous groups.[1] Our division of total households into farm and non-farm goes some distance in this direction but does not permit as clear-cut a division into inter- and intra-sectoral inequity.[2] Theil's segmentation as employed by Fishlow,[3] on the other hand, does not lend itself to a search for real causation via linkage to growth-relevant phenomena. The attempt to marry these two procedures should prove useful and productive.

Thirdly, we may note that, of the three effects isolated in our paper, the reallocation effect (related to the relative weight of the agricultural sector) and the functional distribution effect (related to the change in distributive shares) can be linked with certain familiar notions in growth theory. On the other hand, we do not as yet have the same level of formal theory for tackling an explanation of the powerful factor Gini effect. We know that the wage Gini and property Gini effects, for example, must be tackled at more of a micro level. The wage Gini effect through time is evidently related to the changing heterogeneity of labor as differentiated by age, sex, educational attainment, etc., which is, in turn, related to private and public decisions on education for different income levels in the course of growth. Similarly, the property Gini effect is related to the changing structure of the family ownership of capital stock in terms of size, location and monopoly power which is, in turn, related to gradual changes in differential saving

---

[1] Henry Theil, Economics and Information Theory, Chicago, Rand McNally, 1967.

[2] We can and do deal with the subject on a more intuitive level (see our discussion relating to the consistency and convergence phenomena pictured in Diagram 6 above).

[3] See A. Fishlow, "Brazilian Size Distribution of Income," AER, May 1972.

behavior, and the more radical effects of inheritance and tax laws. The agricultural factor Gini effect, in much the same way, can be seen to relate to the changing family ownership of land and heterogeneous labor, which can be modified over time by such gradual changes as saving and educational investments and such radical changes as land reform. Our analysis of FID can thus provide us, at this stage, with a framework for determining where the major problems are and where future efforts must be made if we are to achieve a fully deterministic theory of the functional distribution of income and be in a position to deploy those elements of policy which reduce or eliminate the conflict between FID and growth.

Appendix

The abstract framework for the quantitative analysis utilized in the paper will be rigorously deduced in this Appendix and illustrated with the help of a numerical example to guide our discussion. In the example we shall assume there are n=5 families with r=3 factor income components: wage income, property income, and transfer income.

Table A1

	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

Table A1 contains all the primary data needed for our analysis. Notice that property income (transfer income) is concentrated heavily among the wealthy families (poor families). The total family income indicated at the bottom of Table A1 is arranged in a monotonically non-decreasing order (i.e., $Y_1 \leq Y_2 \leq \ldots \leq Y_5$). (The rank index of each factor income component is indicated in parentheses.) Thus there exists an assumed perfect positive (negative) rank correlation between total family income (Y_i) and property (transfer) income--describing the difference in characteristics of these two types of incomes which will be essential to our analysis.

To help our discussion, all the computation results based on Table A1 are shown in Table A2. The last column, (4), of this table always indicates the sum of all entries in the first three columns. For example, the three distributive shares are indicated in row (1); their sum is 1, as shown in column (4).

Gini Coefficient, Pseudo Gini Coefficient and Gini Error

We begin with a rigorous definition of a Gini coefficient, G_y, (and analogously, a Pseudo Gini coefficient, $\bar{G}_y$, which can be defined for any non-negative vector, Y

A1) a) $Y = (Y_1, Y_2, \ldots, Y_n) \geq 0$

b) $S_y = Y_1 + Y_2 + \ldots + Y_n > 0$

c) $y_i = Y_i/S_y \qquad i = 1, 2, \ldots, n$

d) $y_1 + y_2 + \ldots + y_n = 1$

Notice that y_i $(i=1,\ldots,n)$ are fractions of total income which add up to 1 in (A1d). Since the elements in Y may or may not satisfy the following "monotonic" condition

A2) $0 \leq Y_1 \leq Y_2 \leq \ldots \leq Y_n$

there is a permutation $(i_1, i_2, \ldots, i_n)$ of the first "n" integers $(1, 2, \ldots, n)$ such that the monotonic condition

A3) $0 \leq y_{i_1} \leq y_{i_2} \leq \ldots \leq y_{i_n}$

is satisfied. The Lorenz curve is a real valued function defined on $(1/n, 2/n, \ldots, n/n)$ such that

Table A2

				Wage (1)	Property (2)	Transfer (3)	Total (4)
ϕ_i	(1)		Factor Share	.4500	.3500	.2000	1.0
G_i	(2)	Gini Coefficient	Factor Gini	.3912	.6628	.6400	
$\phi_i G_i$	(3)	Gini Coefficient	Weighted Gini	.1760	.2320	.1280	$\hat{G}_y$ = .5360
$\bar{G}_i$	(4)	Pseudo-Gini Coefficient	Pseudo-Gini	.2308	.6628	-.5600	
$\phi_i \bar{G}_i$	(5)	Pseudo-Gini Coefficient	Weighted Pseudo-Gini	.1039	.2320	-.1120	G_y = .2239
ε_i	(6)	Error	Gini Error	.1604	.0000	1.2000	
$\phi_i \varepsilon_i$	(7)	Error	Weighted Error	.0722	.0000	.2400	E = .3121
b_i	(8)	Regression Analysis	**Regression Equation**	1.857	-15.143	13.286	$\sum_{i=1}^{n} b_i = 0$
a_i	(9)	Regression Analysis	$\hat{W}^i = b_i + a_i Y$	.3571	1.1071	-.4643	$\sum_{i=1}^{n} a_i = 1$
	(10)	Regression Analysis	Income Type	Type 2	Type 1	Type 3	
$\hat{W}_1^i$	(11)	Regression Analysis	**Estimated Factor**	5.786	-2.964	8.179	Y_1 = 11
$\hat{W}_2^i$	(12)	Regression Analysis	Income Patterns	6.500	-0.750	7.250	Y_2 = 13
$\hat{W}_3^i$	(13)	Regression Analysis		8.643	5.893	4.464	Y_3 = 19
$\hat{W}_4^i$	(14)	Regression Analysis		10.071	10.321	2.607	Y_4 = 23
$\hat{W}_5^i$	(15)	Regression Analysis		<u>14.000</u>	<u>22.500</u>	<u>-2.500</u>	Y_5 = 34
Total S^i	(16)	Regression Analysis		45.000	35.000	20.000	Y = 1000
$G(\hat{W}^i)$	(17)	Approximation Error	Gini for	0.1777	0.7083	0.5202	$\hat{\hat{G}}_y = F^+ - F^-$
$\phi_i G(\hat{W}^i)$	(18)	Approximation Error	Estimated	F^+ = 0.0800	F^+ = 0.2479	F^- = .1040	(.2239 = .3279 - .1040)
	(19)	Approximation Error	Pattern				$\hat{G}_y$ = .4319; $\hat{E} = \hat{G}_y - G_y$ = .2080
$\phi_i G_i$	(20)	Approximation Error	Non-Linearity and	H^+ = .1760	H^+ = .2320	H^- = .1280	$\hat{\hat{G}}_y = H^+ - H^-$; (.2800 = .4080 - .1280)
$\theta_i = \frac{G(\hat{W}^i)-G_i}{G_i}$	(21)	Approximation Error	Type Three	θ_1 = -0.5458	θ_2 = 0.0686	θ_3 = -.1876	$\hat{\theta} = -[\theta^+ - \theta^-]$; .0561 = -[-.0801-(-.0240)]
$\phi_i \theta_i G_i$	(22)	Approximation Error	Error	θ^+ =-.0960	θ^+ = .0159	θ^- = -.0240	$G_y = \hat{\hat{G}}_y - \hat{\theta}$; (.2239 = .2800 - .0561)

A4) $L_Y(j/n) = y_{i_1} + y_{i_2} + \dots + y_{i_j}$ (j = 1,2, .., n)

For example, if $(y_1, y_2, y_3, y_4) = (.2, .4, .1, .3)$ then $(i_1, i_2, i_3, i_4) = (3,1,4,2$ and the Lorenz curve is shown in diagram A1(a). Let the area under the Lorenz curve of the unit square of this diagram be denoted by B. Then the Gini coefficient is

A5) $$G_y = (\frac{1}{2} - B)/(1/2) = 1 - 2B$$

and we have the following theorem:

Theorem A1 The Gini Coefficient of Y (A1a) is

A6) a) $G_y = \alpha u_y - \beta$, where

b) $\alpha = 2/n;\ \beta = (n + 1)/n$, and

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

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

Proof: The area above the Lorenz Curve is

$$1 - B = (y_{i_1}/2 + y_{i_2} + \dots + y_{i_n})/n + (y_{i_2}/2 + y_{i_3} + \dots + y_{i_n})/n + \dots (y_{i_n}/2)/n$$

$$= (y_{i_1} + y_{i_2} + \dots + y_{i_n})/n + (y_{i_2} + y_{i_3} + \dots + y_{i_n})/n + \dots + y_{i_n}/n -$$

$$(y_{i_1} + y_{i_2} + \dots y_{i_n})/2n$$

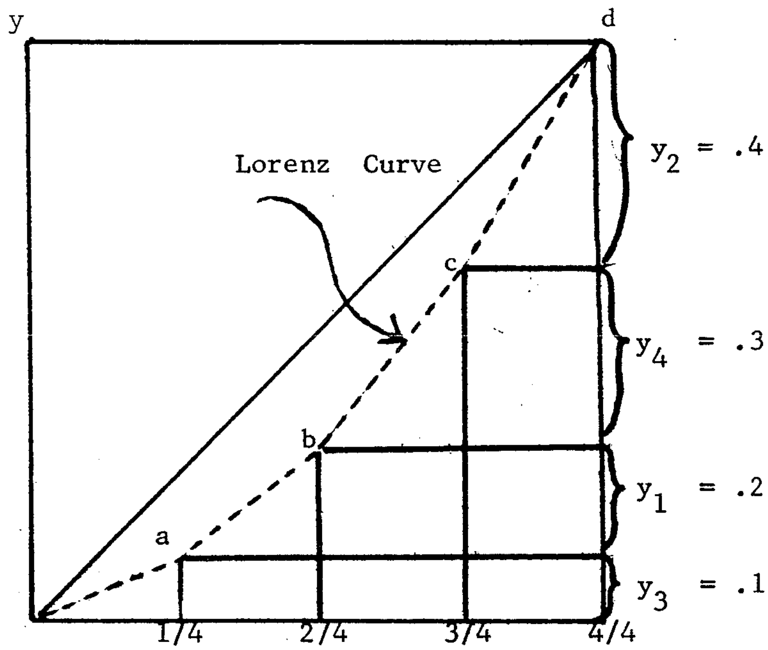
$$= (1y_{i_1} + 2y_{i_2} + \dots + ny_{i_n})/n - 1/2n \text{ by (A1d)}$$

$$G_y = 1 - 2B = 1 + 2\,[\sum_{i=1}^{n} \lambda_i y_{i_j})/n - \frac{1}{2n} - 1] \qquad \text{Q.E.D.}$$

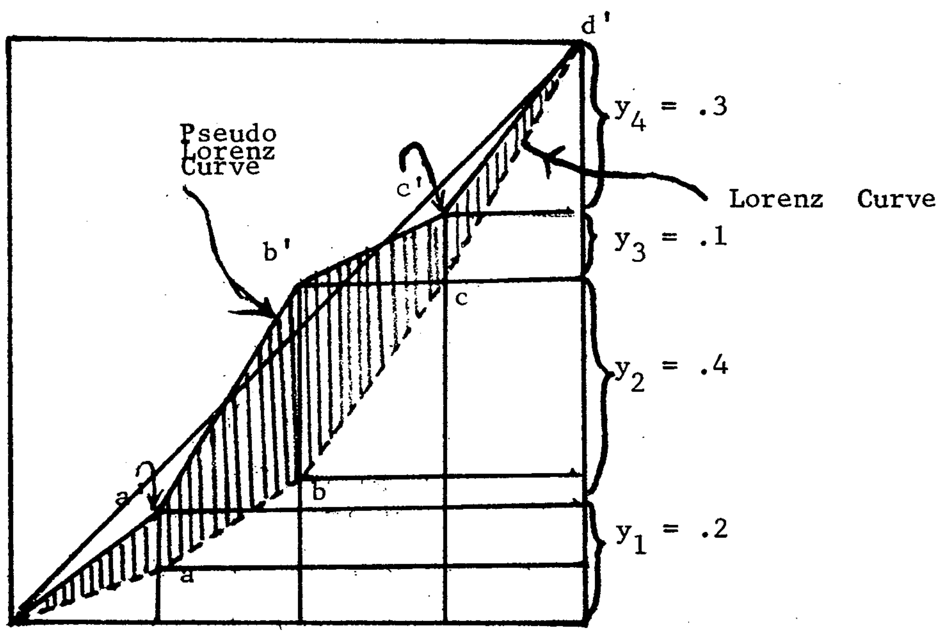
The theorem states that the Gini coefficient, G_y, is a linear transformation of the "u-index" of Y defined in (A6c).

When Y (A1a) is given, we may also define a Pseudo-Lorenz Curve which is similar to the formal definition of a Lorenz Curve except that the income fractions y_i in (A1c) are not arranged in a monotonically non-decreasing

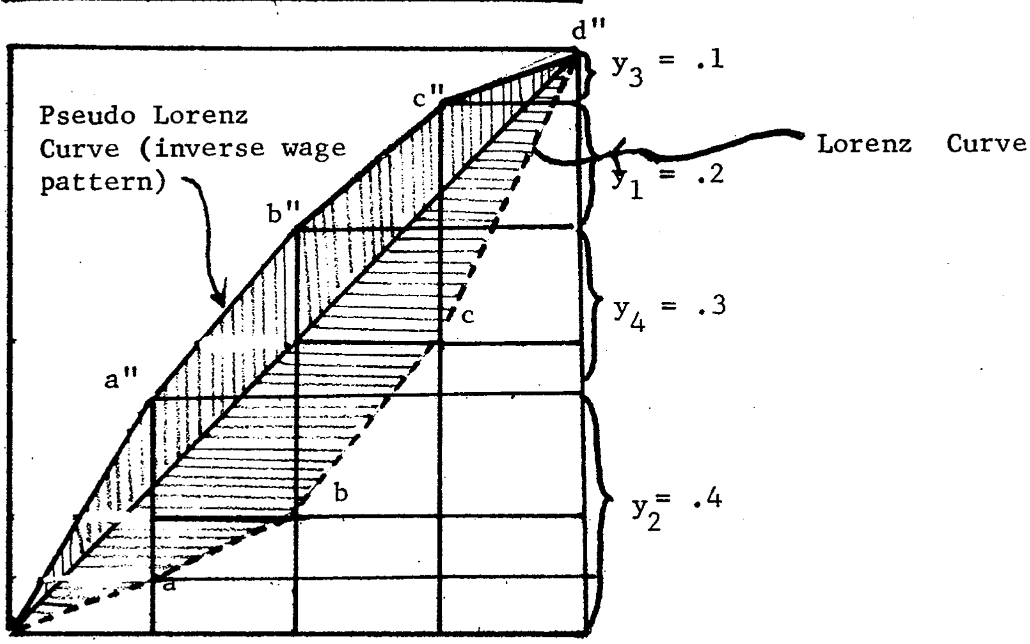
A1a


y
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

A1b


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

A1c


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

order. Thus, formally, the Pseudo-Lorenz Curve is defined as:

A7) $L_Y^S(j/n) = y_1 + y_2 + \ldots + y_j \quad j = 1, 2, \ldots, n$

For the same raw data of Diagram (A1a), the Pseudo-Lorenz Curve is indicated by the dotted curve a'b'c'd' of Diagram (A1b). ...

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 (A5), namely:

A8) $\bar{G}_y = 1 - 2\bar{B}$

We can then deduce:

Theorem A2, i.e., the Pseudo-Gini Coefficient of Y (A1a) is

A9) a) $\bar{G}_y = \alpha\bar{u}_y - \beta$, where

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

and where α, β and λ_i are defined in (A6).

The Gini error of Y in (A1a) can be defined in terms of G_y and $\bar{G}_y$:

A10) $E = G_y - \bar{G}_y$

In other words, the Gini error E is the difference between the Gini and the Pseudo-Gini coefficient of Y. We then have the following theorem:

Theorem A3 For any non-negative vector Y (A1a), the Gini error is non-negative. The Gini error is zero if the monotonic condition (A2) is satisfied.

Proof: From (A10), (A9) and (A6), we have

$$E = G_y - \bar{G}_y = (\alpha u_y - \beta) - (\alpha \bar{u}_y - \beta) = (2/n)V \quad \text{where}$$

$$V = u_y - \bar{u}_y = [\lambda_1 y_{i_1} + \lambda_2 y_{i_2} + \ldots + \lambda_n y_{i_n}] - [\lambda_1 y_1 + \lambda_2 y_2 + \ldots + \lambda_n y_n]$$

It is sufficient to investigate the sign of V. We can use the numerical example of Diagram (A1) to illustrate the logic of the proof. In this example:

$$V = [1\,(.1) + 2\,(.2) + 3\,(.3) + 4\,(.4)] - [1\,(.2) + 2\,(.4) + 3\,(.1) + 4\,(.3)]$$

There is a finite number of steps through which (y_1, y_2, y_3, y_4) can be rearranged into a monotonically increasing order:

Step One:	$(y_1, y_2, y_3, y_4) = (.2, .4, .1, .3)$	$J_1 = 1(.2)+2(.4)+3(.1)+4(.3) = 2.5$
Step Two:	$(y_1, y_3, y_2, y_4) = (.2, .1, .4, .3)$	$J_2 = 1(.2)+2(.1)+3(.4)+4(.3) = 2.8$
Step Three:	$(y_1, y_3, y_4, y_2) = (.2, .1, .3, .4)$	$J_3 = 1(.2)+2(.1)+3(.3)+4(.4) = 2.9$
	$(y_3, y_1, y_4, y_2) = (.1, .2, .3, .4)$	$J_4 = 1(.1)+2(.2)+3(.3)+4(.4) = 3.0$

In each step a larger y_i is moved to the right by an adjacent interchange. The largest y_i (in our case $y_2 = .4$) is first moved to its proper place. Then the next largest y_i (in our case $y_4 = .3$) is moved to its proper place, etc. Thus J_1, J_2, J_3 and J_4 increase monotonically to $J_4 = u_y$. Hence $V = u_y - J_1 = .5$, $u_y - J_2 = .2$, $u_y - J_3 = .1$, $u_y - J_4 = 0$ monotonically decreases to zero. Thus V must be non-negative. Q.E.D.

A geometric interpretation of the Gini error is seen from the fact that

A11) $$E = (1 - 2B) - (1 - 2\bar{B}) \ \ldots \text{ by (A5) and (A8)}$$
$$= 2(\bar{B} - B)$$

and hence the area between the Pseudo-Lorenz Curve and the Lorenz Curve is one-half the value of the Gini error (see Diagram (A1b)).

If Y of (A1a) satisfies the following "reverse" monotonic order:

(A12) $Y_1 \geqq Y_2 \geqq \ldots \geqq Y_n \geqq 0$

Then we have the following theorem:

Theorem A4: If $Y = (Y_1, Y_2, \ldots, Y_n)$ satisfies the conditions of reverse monotonic order (A12), then

$$E_y = 2\,G_y \quad \text{or} \quad \bar{G}_y = -\,G_y$$

i.e., the Pseudo-Gini is the negative of the true Gini.

Proof: The proof can be seen from Diagram (A1c) in which the dotted curve a"b"c"d" is the Pseudo-Lorenz Curve for ($y_1 = .4$, $y_2 = .3$, $y_3 = .2$, $y_4 = .1$) which satisfies the reverse monotonic order. Because the Pseudo-Lorenz Curve and the Lorenz Curve abcd are "symmetrical" with respect to the 45 degree line, the area between them is twice A, the area between the 45 degree line and the Lorenz Curve. Thus, by (A11),

$$E_y = 2(2A) = 2 \cdot A/.5 = 2A(B+A) = 2G_y$$

The fact $\bar{G}_y = G_y$ follows directly from (A10). Q.E.D.

Total Income and Factor Components

Suppose the total income pattern Y (A1a) is the vector sum of r-factor components:

A13) a) $Y = W^1 + W^2 + \ldots + W^r$

b) $W^i = (W^i_1, W^i_2, \ldots, W^i_n) \geqq 0$, $i = 1, 2, \ldots, r$

c) $\phi_i = S^i/S_y$ where

d) $S^i = W^i_1 + W^i_2 + \ldots + W^i_n$ $i = 1, 2, \ldots, r$

e) $S_y = Y_1 + Y_2 + \ldots + Y_n$

f) $1 = \phi_1 + \phi_2 + \ldots + \phi_r$

Notice that the ϕ_i (i=1,2,...,r) of (A13f) are the distributive shares. We then have the following Lemma:

Lemma A1. The fraction of the total income of the jth family is the weighted average of the factor shares:

$$y_j = \phi_1(W_j^1/S^1) + \phi_2(W_j^2/S^2) + \ldots + \phi_r(W_j^r/S^r) \qquad \text{A14)}$$

Proof:

$$\begin{aligned} y_j &= (S^1/S_y)(W_j^1)/S^1 + (S^2/S_y)(W_j^2)/S^2 + \ldots + (S^r/S_y)(W_j^r)/S_y \\ &= (W_j^1 + W_j^2 + \ldots + W_j^r)/S_y \\ &= Y_j/S_y \qquad \text{Q.E.D.} \end{aligned}$$

Let us assume that the ordering of the families has been arranged so that the total income patterns Y of (A13a) satisfies the monotonically non-decreasing condition (A2). Under this convention the factor income pattern of (A13b) may or may not satisfy the monotonic condition. Hence we can define a factor Pseudo-Gini coefficient for each factor, $\bar{G}_i$, in the same way as in Theorem A2 for total Gini, and prove the following theorem:

Theorem A5: The Gini coefficient of total family income (G_y) is the weighted sum of the factor Pseudo Gini coefficients when the distributive shares ϕ_i are the weights:

$$G_y = \phi_1\bar{G}_1 + \phi_2\bar{G}_2 + \ldots + \phi_r\bar{G}_r \qquad \text{15)}$$

Proof: By (A6), the u-index of Y is

$$\begin{aligned} u_y &= \lambda_1 y_{i_1} + \lambda_2 y_{i_2} + \ldots + \lambda_n y_{i_n} \qquad (i_1 = 1,\ i_2 = 2,\ \ldots\ i_n = n) \\ &= \lambda_1[\phi_1(W_1^1/S^1) + \phi_2(W_1^2/S^2) + \ldots + \phi_r(W_1^r/S^r)] \qquad \text{by Lemma (A1)} \\ &+ \lambda_2[\phi_1(W_2^1/S^1) + \phi_2(W_2^2/S^2) + \ldots + \phi_r(W_2^r/S^r)] + \ldots + \\ &+ \lambda_n[\phi_1(W_n^1/S^1) + \phi_2(W_n^2/S^2) + \ldots + \phi_r(W_n^r/S^r)] \\ &= \phi_1\bar{u}_1 + \phi_2\bar{u}_2 + \ldots + \phi_r\bar{u}_r \end{aligned}$$

where $\bar{u}_i$ is the pseudo $\bar{u}$-index of the ith factor component (A9b).
Thus the Gini coefficient of Y, by (A6a) is:

$$\begin{aligned} G_y &= (2/n)u_y - (n+1)/n \\ &= (2/n)[\phi_1\bar{u}_1 + \phi_2\bar{u}_2 + \dots + \phi_r\bar{u}_r] - (n+1)/n \\ &= \phi_1[(2/n)\bar{u}_1 - (n+1)/n] + \phi_2[(2/n)\bar{u}_2 - \frac{(n+1)}{n}] + \dots + \phi_r[(2/n)\bar{u}_r - \frac{n+1}{n}] \\ &+ \phi_1(n+1)/n + \phi_2(n+1)/n + \dots + \phi_r(n+1)/n - (n+1)/n \\ &= \phi_1\bar{G}_1 + \phi_2\bar{G}_2 + \dots + \phi_r\bar{G}_r \quad \dots \text{ by (A13f) and (A9)} \qquad \text{Q.E.D.} \end{aligned}$$

For each factor income component (A13b) we can also define a factor Gini coefficient, G_i, and a factor Gini error, ε_i:

A16) $\varepsilon_i = G_i - \bar{G}_i \geqq 0 \quad (i = 1, 2, \dots, r)$ ... by (A10)

which are non-negative, by Theorem (A3). Substituting $\bar{G}_i$ of (A16) into (A15), we have

A17) $G_y = \phi_1(G_1 - \varepsilon_1) + \phi_2 (G_2 - \varepsilon_2) + \dots + \phi_r (G_r - \varepsilon_r)$

From this result we immediately have the following theorem:

<u>Theorem A6</u>: <u>The Gini-coefficient of total income G_y (A13a) is the difference between an approximate Gini $\hat{G}_y$ and a weighted Gini-error term, E:</u>

A18) a) $G_y = \hat{G}_y - E$ where

b) $\hat{G}_y = \phi_1G_1 + \phi_2G_2 + \dots + \phi_rG_r$

c) $E = \phi_1\varepsilon_1 + \phi_2\varepsilon_2 + \dots + \phi_r\varepsilon_r \geqq 0$

where E is non-negative.

Proof: by (A17) and (A16).

The term $\hat{G}_y$ in (A18c) is the weighted sum of the factor Gini coefficients and E (A18c) is the weighted sum of the factor Gini errors. Notice that when the errors are small, when E is negligible, $\hat{G}_y$ can be taken as an approximation of G_y. When this is the case, we can use this approximation and say that a factor income contributes to inequality of total family income distribution when the factor itself is very unequally distributed, i.e., G_i is large, and/or when a factor's distributive share is large, i.e., ϕ_i is large. Such an approximation tends to slightly overestimate the true G_y. However, when the Gini errors are large, i.e., when E is not negligible, the overestimation of the true Gini is too large and the approximation cannot be taken for granted. In the following sections, we shall analyze the forces that determine the magnitude of the Gini error.

Returning to our example, in Table (A2), the factor Gini coefficients are indicated in row (2). The weighted overall Gini coefficient, $\hat{G}_y$, is shown in row (3), column (4), showing a value of .5360. The Pseudo-Ginis for each factor income are indicated in row (4) and the computation of the true Gini, G_y, as the weighted sum of these factor Pseudo-Ginis is shown in row (5) leading to G_y = .2239. Thus when $\hat{G}_y$ is used as an estimator of the true Gini, G_y, we have an overestimation of the true Gini by $E = \hat{G}_y - G_y$ = .3121, or more than 100%. The Gini errors for each factor in row (6) are the differences between rows (4) and (2) and the weighted Gini error term is computed in row (7) leading to E = .3121 which is the same value as $\hat{G}_y - G_y$, above.

In the next sections the Gini error will be fully decomposed into two types of errors: the error due to the existence of Type Three income, and the error due to non-linearity.

Linear Approximation

Given the total income pattern, Y, and the factor components, W^i (A13b), we can find a linear regression of W^i on Y in the form:

A19) $$W^i = b_i + a_i Y \qquad i = 1, 2, \ldots, r .$$

When the parameters a_i and b_i are estimated by the method of least squares, it is easy to prove:

A20) a) $b_1 + b_2 + \ldots + b_r = 0$

b) $a_1 + a_2 + \ldots + a_r = 1$

From the regression lines of (A19) we can compute the estimated factor income patterns for each factor component $\hat{W}^i = (\hat{W}^i_1, \hat{W}^i_2, \ldots, \hat{W}^i_n)$ $(i=1,2,\ldots,r)$ satisfying:

A21) $$\hat{W}^i_j = b_i + a_i Y_j \qquad i = 1, 2, \ldots, r; \quad j = 1, 2, \ldots, n$$

We then have the following lemma:

Lemma A2.

a) $Y \equiv (Y_1, Y_2, \ldots, Y_n) = \hat{W}^1 + \hat{W}^2 + \ldots \hat{W}^r$

b) $\phi_i \equiv (W^i_1 + W^i_2 + \ldots + W^i_n)/(Y_1 + Y_2 + \ldots + Y_n)$
$= (\hat{W}^i_1 + \hat{W}^i_2 + \ldots + \hat{W}^i_n)/(Y_1 + Y_2 + \ldots + Y_n)$.

Proof: The jth component of the vector on the right hand side of (a) is

$$\hat{W}^1_j + \hat{W}^2_j + \ldots + \hat{W}^r_j = (b_1 + a_1 Y_j) + (b_2 + a_2 Y_j) + \ldots (b_r + a_r Y_j)$$
$$= a_1 Y_j + a_2 Y_j + \ldots a_r Y_j \quad \text{(by A20a)}$$
$$= Y_j \quad \text{(by A20b)}$$

This proves Lemma A2a. The validity of A2b follows as a property of linear regression. Q.E.D.

Lemma A2a states that total income is a vector sum of the estimated pattern of factor incomes. Lemma A2b states that the distributive shares defined in (A13c) can also be computed from the estimated pattern of factor income.

Returning to Table A2, the regression coefficients b_i and a_i are shown in rows (8) and (9). Notice that conditions (A20) are satisfied. Property income is type one; wage income is type two; and transfer income is type three, as indicated in row (10). (See Section II in text.) With the aid of the linear regression equations the estimated factor income patterns are indicated in rows (11) - (15). Lemma A2 is satisfied as can be seen by row (16) and column (4) (rows (11)-(15)) which are identical to the original income pattern. This "expected pattern" of factor components may be interpreted as a linear approximation of the original data. In this section we shall concern ourselves only with this system of linear approximations to reality.

We will now trace the component parts of the Gini error term E (as defined in (A10)) in two steps. First, we shall see that, if we accept the linear approximation, the Gini error is traced entirely to the existence of Type Three income and thus a "Type Three error" can be defined. Second, when we examine the deviation of the original data from linearity, we can identify another component of the Gini error, namely the so-called non-linearity error.

Type Three Error

In this section, we shall only be concerned with the Type Three error using the linear approximation data.[1] In the next section, we shall be concerned with the deviation of the original data from the linear approximation. We can apply (A17) to the linear approximation to obtain:

$$\text{A22)} \qquad G_y = \phi_1 [G(\hat{W}^1) - \hat{\varepsilon}_1] + \phi_2 [G(\hat{W}^2) - \hat{\varepsilon}_2] + \ldots + \phi_r [G(\hat{W}^r) - \hat{\varepsilon}_r]$$

where $G(\hat{W}^i)$ is the Gini coefficient of the pattern of estimated factor income components $\hat{W}^i$ and where $\hat{\varepsilon}_i$ is defined by

$$\text{A23)} \qquad \hat{\varepsilon}_i = G(\hat{W}^i) - \bar{G}(\hat{W}^i)$$

In other words, $\hat{\varepsilon}_i$ is the difference between the Gini of the estimated pattern and the pseudo-Gini of the estimated pattern. Since we have assumed that Y satisfies the monotonic conditions (A2), there are two types of cases, corresponding to income types 1 and 2, on the one hand, and income type 3, on the other (see text).

A24) a) Income Types One and Two:

If $a_i \geqq 0$, the estimated pattern satisfies the monotonic condition: $\hat{W}^i_1 \leq \hat{W}^i_2 \ldots \leq \hat{W}^i_n$.

Hence: $\hat{\varepsilon}_i = 0$, $\bar{G}(\hat{W}^i) = G(\hat{W}^i)$ ... by Theorem (A3)

b) Income Type Three:

If $a_i \leq 0$, the estimated pattern satisfies the inverse monotonic condition: $\hat{W}^i_1 \geq \hat{W}^i_2 \geq \ldots \geq \hat{W}^i_n$.

Hence: $\hat{\varepsilon}_i = 2G(\hat{W}^i)$; $\bar{G}(\hat{W}^i) = - G(\hat{W}^i)$ ... by Theorem (A4).

---

[1] A simple interpretation of the work in the rest of this section is therefore as follows: Suppose the linearity condition is satisfied by the original data; then the Gini error has no "non-linearity error" and E consists only of "Type Three error."

Thus, if the first r_2 terms of (A22) correspond to non-negative a_i and the last $r - r_2$ terms correspond to negative a_i, we have the following theorem by (A22) and (A24):

Theorem A7: In the linear approximation (A21), when the last $r-r_2$ regression coefficients, a_i, are strictly negative:

A25)

a) $G_y = F^+ - F^-$ where

b) $F^+ = \phi_1 G(\hat{W}^1) + \phi_2 G(\hat{W}^2) + \ldots + \phi_{r_2} G(\hat{W}^{r_2})$

c) $F^- = \phi_{r_2+1} G(\hat{W}^{r_2+1}) + \phi_{r_2+2} G(\hat{W}^{r_2+2}) + \ldots + \phi_r G(\hat{W}^r)$

Notice that the correlation coefficient between each estimated pattern $\hat{W}^i$ and total income is, of course, perfect, i.e., $r = +1$ for Type One and Type Two incomes and $r = -1$ for Type Three income (since perfect linearity is implied by definition). Thus, when the original data W^i has an almost perfect linear correlation with Y, the above theorem suggests that a correct approximation equation should be

$$\text{A26)} \quad \hat{\hat{G}}_y = (\phi_1 G_1 + \phi_2 G_2 + \ldots + \phi G_{r_2}) - (\phi_{r_2+1} G_{r_2+1} + \phi_{r_2+2} G_{r_2+2} + \ldots + \phi_r G_r)$$

instead of G_y in (A18b). In other words, the correct approximation formula includes positive weights for Type One and Type Two income components and negative weights for Type Three income components. Intuitively, this is due to the fact that Type Three income contributes to equality rather than inequality.

The factor Gini coefficients of these patterns are shown in row (17) of Table A2. The computation of the F^+ and F^- terms of (A25) is shown in row (18), i.e., $F^+ = .3279$ and $F^- = .1040$. The true Gini coefficient is their difference which is exactly equal to .2239, the G_y in row (5), column (4). Thus, in the case of perfect linear approximation, the estimator $\hat{\hat{G}}_y$ is a perfect estimator because no error whatsoever is committed.

If we were to apply the estimator, $\hat{G}_y$, (as defined in (A18) which can be applied to any set of data) to the linear approximation data of Table A1, rows (11)-(15), the approximation of G_y is indicated in row (19) leading to $\hat{G}_y = .4319$ and a Gini error of $\hat{E} = .2080$. Thus when the linear approximation holds, the use of the estimator $\hat{\hat{G}}_y$ eliminates Type Three error.

Another theorem relevant to the linear approximation which we use in the text deals with the relationship between the overall and factor Gini.

<u>Theorem A8:</u> $\bar{G}(\hat{W}^i) = (a_i/\phi_i)G_y$

Proof: $\bar{u}_{\hat{W}^i} = (\lambda_1\hat{W}_1^i + \lambda_2\hat{W}_2^i + \dots + \lambda_n\hat{W}_n^i)/n\bar{W}^i$ by (A6c)

$= [\lambda_1(b_i+a_iY_1) + \lambda_2(b_i+a_iY_2) + \dots + \lambda_n(b_i+a_iY_n)]/n\bar{W}^i$ by (A19)

$= [b_i(\lambda_1+\lambda_2 + \dots + \lambda_n) + a_i(\lambda_1Y_1 + \lambda_2Y_2 + \dots + \lambda_nY_n)]/n\bar{W}^i$

$= [b_i(n+1)n/2 + a_in\bar{Y}\ (\lambda_1Y_1 + \lambda_2Y_2 + \dots + \lambda_nY_n/n\bar{Y})]/n\bar{W}^i$

by $1 + 2 + \dots + n = (n+1)n/2$ by (A9b)

$= b_i(n+1)/2\bar{W}^i + a_i\bar{Y}\ u_y/\bar{W}^i$ by $\lambda_1+\lambda_2+ \dots + \lambda_n = (n+1)n/2$, (A9b)

$\bar{G}_{\hat{W}^i} = \frac{2}{n}[(b_i[n+1]/2\bar{W}^i + a_i\bar{Y}\ u_y/\bar{W}^i] - (n+1)/n$ by (A6a)

$= b_i(n+1)/n\bar{W}^i + (a_i\bar{Y}/\bar{W}^i)[2u_y/n - (n+1)/n]$

$+ (a_i\bar{Y}/\bar{W}^i)(n+1/n) - n+1/n$

$= n+1/n\ [(b_i+a_i\bar{Y})/\bar{W}^i - 1] + (a_i\bar{Y}/\bar{W}^i)\ G_y$ by (A6a)

$= n+1/n\ [(b_i+a_i\bar{Y})/(b_i+a_i\bar{Y}) - 1] + a_i(\bar{W}^i/\bar{Y})\ G_y$

$= (a_i/\phi_i)G_y$ by (1.2c)

Q.E.D.

Substituting (A24) in the above we have the following:

A27) a) Type One and Two: $G(\hat{W}^i) = a_i/\phi_i\ G_y$

b) Type Three: $G(\hat{W}^i) = (-a_i/\phi_i)G_y$

which is (2.1) in the text.

Non-linearity error

In the more normal case when the correlation coefficient between W^i and Y is less than perfect, we can define a non-linearity error for each factor component.

A28) $$\theta_i = \frac{G(\hat{W}^i) - G_i}{G_i} \qquad i = 1, 2, \ldots, r$$

Substituting $G(\hat{W}^i)$ of (A28) into (A25) we obtain the following theorem:

Theorem A9: The Gini coefficient, G_y, can be estimated by the estimator $\hat{\hat{G}}_y$:

A29) a) $G_y = \hat{\hat{G}}_y - \hat{\theta}$ where

b) $\hat{\hat{G}}_y = H^+ - H^-$ and

$H^+ = \phi_1 G_1 + \phi_2 G_2 + \ldots + \phi_{r_2} G_{r_2} \qquad (a_i \geqq 0)$

$H^- = \phi_{r_2+1} G_{r_2+1} + \phi_{r_2+2} G_{r_2+2} + \ldots + \phi_r G_r \qquad (a_i \leqq 0)$

c) $\hat{\theta} = -[\theta^+ - \theta^-]$

$\theta^+ = \phi_1 G_1 \theta_1 + \phi_2 G_2 \theta_2 + \ldots + \phi_{r_2} G_{r_2} \theta_{r_2}$

$\theta^- = \phi_{r_2+1} G_{r_2+1} \theta_{r_2+1} + \ldots + \phi_r G_r \theta_r$

When $\hat{\hat{G}}_y$ is used as an estimator of G_y, the error term, $\hat{\theta}$, may be referred to as the non-linearity error. Since the estimator G_y of (A18) can be written as $H^+ + H^-$, comparing (A18) and (A29), we have

A30) a) $\hat{\theta} = E - 2H^-$ or

b) $E = \hat{\theta} + 2H^-$

which shows the relation between the Gini error, E, and the non-linearity error $\hat{\theta}$.

We see from (A30b) that the non-negative Gini error E has two components: the linearity error $\hat{\theta}$ and the Type Three error $2H^-$. Notice that the type three error term is non-negative, i.e.,

A31) $2H^- \geq 0$

and is zero in the linearity case when there is no Type Three income. The fact that $\hat{G}_y$ overestimates G_y (because the Gini error E is non-negative) is due (at least partly) to the fact that there is at least one non-negative component when there is type three income.

The comparative merits of the two estimators $\hat{G}_y$ and $\hat{\hat{G}}_y$ can be seen from two extreme cases. First, suppose there is no type three income (i.e., $2H^- = 0$). Then the two estimators are the same, $\hat{G}_y = \hat{\hat{G}}_y$ (i.e., $\hat{\theta} = E$ in (A30)).[1] Hence the two estimators have the same merit. Next, suppose there is Type Three income and there is no non-linearity error (i.e., suppose $\hat{\theta} = 0$). Then $\hat{\hat{G}}_y$ is a perfect estimator and $E = 2H^-$ from (A30). This means that the Gini error E consists only of type three error and $\hat{G}_y$ is thus definitely inferior to $\hat{\hat{G}}_y$ as an estimator.

(A30) indicates that if $\hat{\theta}$ is a non-negative number, it is always smaller than the non-negative Gini error E. Thus, when $\hat{\hat{G}}_y$ overestimates G_y, the degree of overestimation is smaller than the overestimation by $\hat{G}_y$.

In row (20) of Table A2, we show the estimated value of G_y using estimator $\hat{\hat{G}}_y$ of (A29b). The result is $\hat{\hat{G}}_y = .2800$ and smaller than $\hat{G}_y = .4319$. To compute the non-linearity error $\hat{\theta}$ of (A29c) we first calculate θ_i of (A28) as shown in row (21). Then we obtain θ^+ and θ^- as shown in row (22), leading to a value of $\hat{\theta} = .0561$. Thus the difference $\hat{\hat{G}}_y - \hat{\theta} = .2800 - .0561 = .2239$ which is the value of the true Gini, illustrating (A29a).

[1]This leads to (2.8) in the text, i.e., $\hat{\theta} \geq 0$ when there is no Type Three income.

Thus, in our example, the total Gini Error E = .3121 (row (7)) can be decomposed into its non-linearity and Type Three error components, according to (A30b). This yields E (.3121) = $\hat{\theta}$ (.0561) + $2H^-$ (.2560) indicating that, in our example, 80% of the Gini Error is accounted for by the presence of Type Three income. In such cases, it is not safe to use $\hat{G}_y$ as an estimator of G_y. On the other hand, $\hat{\hat{G}}_y$ as the estimator results in a non-linearity error of only .0561 and thus the degree of overestimation is substantially reduced.