# INTERACTIVE POVERTY AND AFFLUENCE OF A SOCIETY: CHANGE-POINT MODELS FOR STRUCTURAL CHANGES

by
Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1862

June 1989

# Interactive Poverty and Affluence of a Society : Change-Point Models for Structural Changes

Pranab Kumar Sen
*University of North Carolina at Chapel Hill*, N.C. 27599-7400, USA

**Summary**

For any society or community, there is ample room for a prismatic view of its *income distribution*. Conventional measures of income (*distributional*) *inequalities*, such as the *Gini coefficient*, may not depict a realistic picture accross the individual sectors of an income distributional spectrum. Stratification into *poor, middle class* and *affluent* sectors generally provides a better understanding of the socio-economic intricacies of such income distributional inequalities. Combining such component measures into a single index requires careful statistical considerations and entails a detailed analysis of the entire income distributional data. *Economic structural changes* may occur within each of these sectors and in plausibly rather diverse directions, so that the usual *linear models* may fail to be very appropriate for a composite analysis. A formulation of a *change-point model* in the setup of *constancy of regression surfaces* is therefore incorporated in the development of methodolgy for studying structural changes for such income distributions. Proper emphasis is placed on *nonparametric* as well as *robustness considerations* underlying such nonstandard analysis. Such considerations also play a vital role in *forecasting* of economic structural changes with respect to some income inequalities.

## 1. Introduction

In any society or community, the distribution of *real income* ( *or wealth*) is characterized by distinct *heterogeneity* (even within individual sectors of relatively homogeneous groups). The extent of this divergence generally depends on various *socio-economic factors* which may affect the individual sectors in a rather different manner. For example, a drought year may have serious impact on the income distribution of the agricultural households sector, quite perceptable for the lower income and the poor class, but not so much for the affluent people. There are literally thousands of such socio-economic factors influencing the *income pattern and distribution* of any society, and their impacts are generally perceived rather differentially accross the different strata. Thus, for a better understanding of the intricacies of various socio-economic factors affecting an income distribution and to relate them adequately tp plausible *structural changes*, it may be wiser to encompass a rational stratification of an income distribution. This stratification may be accomplished by identifying the relative levels of income ( viz., *poor, middle class and affluent people* ) resulting in non-overlapping strata, or by the conventional *agricultural, industrial and other profesional* sectors resulting in possibly overlapping income distributions. A combination of the two is also a possibility. With such a stratification, for drawing a neat picture for the entire spectrum, one needs to take into account both the *intra-strata* and *inter-strata variations*. In the current study, we shall mainly confine ourselves to the interactive features of **poverty** and **affluence** of a society, and incorporate them in the change-point models for a better understanding of some structural changes which may arise in this context.

In Section 2, we outline the general statistical considerations underlying the affluence and poverty indexes, and based on these findings, we procedd to consider a general breakdown of an overall index in terms of some component indexes. It turns out that in the case of non-overlapping strata, the Gini coefficient for the composite income distribution is explicitly expressible in terms of the component Gini coefficients , the relative proportion of the people in these strata and their relative total incomes. Some additional parameters enter into the picture when the strata income distributions are possibly overlapping. This gives us a strong motivation to incorporate these component Gini coefficients and related income inequality indexes in the formulation of a so called *response surface model* which may be more conveniently used to study suitable structural changes for such income distributions. This is considered in Section 3. The *change-point model* is then introduced in Section 4 with a view to studying plausible structual changes over a span of time. The model is generally more complicated than the classical change-point model as here progressive realignment of the strata may be necessary to cope with the chronological changes in living styles. These findings are finally incorporated in the last section in the forecasting of economic structural changes pertinent to the usual income distributions.

## 2. Income Inequality Indexes : A Prismatic View

The *Lorenz curve* for an income distribution provides a clear graphical picture of the overall concentration and dispersion of incomes of individuals or families in a society. However, as is generally the case, part of the income variation ( or inequality) may be explained better in the light of between sectors variation when a suitable system of sectors is brought in the picture. For example, a stratification into three strata: poor, middle class and affluent sectors, may explain some of the variations as due to inter-strata variations while the rest confined to the intra-strata ones. Hence, to study the income inequality picture more thoroughly (for the individual sectors as well), it may be wiser to look into the Lorenz curve a little bit more thonoughly with a view to depicting the picture for the component as well as the overall income inequality indexes. The sectors to be considered here are based on the relative levels of real income leading to a system of nonoverlappling strata. For simplicity (and practical relevance too), we confine ourselves to the case of three basic strata : *Poor, middle class and affluent people*. A brief treatment of the general case of $k(\geq 2)$ (possibly overlapping ) strata is also appended. In this context, there is a genuine need to pay attention to the following :

(i) For the correct labelling of the strata and real income, an assessment of real income of individuals or families in terms of a single *quantitative criterion* is needed. Once the real income is quantified , the drawing of the Lorenz curve and the associated indexes may not pose a serious problem. However, the issue of *robustness* remains as a pertinent one in such a quantification scheme.

(ii) Demarcation of the three strata rests on the proper fixation of the *line of poverty* and the *affluence line*. Poverty is usually defined as the extent to which individuals in a society or community fall below a *minimal acceptable standard of living*. Thus, quantified in terms of real income, the poverty line cuts off the lower tail of the income distribution: The truncated income distribution over this left hand tail is often called the *income distribution of the poor*. Affluence of a society or community is similarly quantified by the proportion of its rich (or affluent) people and by the concentration of their wealth or real income. Again, in terms of real income, this amounts to the right hand tail of the income distribution where the cut off point (the affluence line) is determined by various socio-economic factors. The truncated income distribution over this right hand tail is often called the *income distribution of the rich*. The income distribution truncated from below and above by the poverty and affluence line respectively is termed the *middle class income distribution*. It is quite clear that in the determination of poverty and affluence lines, various socio-economic and related monetary utility functions play a basic role. Statistical considerations are very important in this respect too.

(iii) For each of the three sectors, some measures of concentration or inequality of wealth (or real income) need to be developed. Statistical considerations are very pertinent in this context too.

Arbitration of affluence and poverty lines is generally a very delicate task. The criteria may differ considerably from a socialistic to a capitalistic society. Even, for the same society, they may vary progressively over time. These criteria appearing as deterministic in this context are also very relevant in the study of plausible structural changes of income indexes. In addition, there may be other important factors which should be taken into account in a proper formulation of a suitable response surface model for income inequality measures. Moreover, quantifications of all such basic factors are important for a proper formulation of real income on which everything is based.

With due considerations to this basic quantification of real income, we denote the income distribution (of a society or community in a given time period ) by $F = \{ F(x), x \in R^+ \}$ , $R^+ = [0,\infty)$. Also, with a proper arbitration of the poverty line ($\omega$ ) and the affluence line ($\rho$ ), we have two positive numbers $(\omega,\rho): 0 < \omega < \rho < \infty$ , such that the income in the ranges $[0,\omega]$, $(\omega,\rho)$ and $[\rho,\infty)$ characterizes the poor, middle class and affluent sectors, respectively. Thus, the income distribution of the poor is given by

$$F_P = \{ F_P(x) = \begin{cases} F(x)/F(\omega), & 0 \le x \le \omega; \\ 1, & x > \omega, \end{cases} \}, \tag{2.1}$$

and the proportion of the poor people is denoted by

$$\alpha_P = F(\omega) . \tag{2.2}$$

Similarly, the income distribution of the affluent people is given by

$$F_R = \{ F_R(x) = \begin{cases} 0 , & x < \rho ; \\ [F(x)-F(\rho)]/[1-F(\rho)], & x \ge \rho, \end{cases} \}, \tag{2.3}$$

and the proportion of affluent people is given by

$$\alpha_R = 1 - F(\rho). \tag{2.4}$$

Finally, the middle class income distribution is

$$F_M = \{ F_M(x) = \begin{cases} 0, & x \le \omega ; \\ [F(x)-F(\omega)]/[F(\rho) - F(\omega)], & \omega \le x \le \rho , \\ 1, & x > \rho , \end{cases} \} , \tag{2.5}$$

and the proportion of the middle class people is

$$\alpha_M = F(\rho) - F(\omega) = 1 - \alpha_R - \alpha_P . \tag{2.6}$$

Assume that for the entire income distribution F, the *mean real income* $\mu = \int_{R^+} ydF(y)$ is finite and positive. Also, let $F^{-1}(t) = \inf\{ x : F(x) \ge t\}, 0 \le t \le 1$. Define then $\xi = \{ \xi(t); t \in [0,1] \}$ by letting

$$\xi(t) = \mu^{-1}\{\int_0^{F^{-1}(t)} ydF(y)\}, \ t \in [0,1] . \tag{2.7}$$

Thus, the relative contributions of the poor, middle class and the affluent people to the total income are given by $\gamma_P$, $\gamma_M$ and $\gamma_R$, respectively, where

$$\gamma_P = \xi(\alpha_P), \quad \gamma_M = \xi(1- \alpha_R) - \xi(\alpha_P) \quad \text{and} \quad \gamma_R = 1 - \xi(1- \alpha_R). \tag{2.8}$$

Hence, the relative mean incomes of the poor, middle class and rich people are given by

$$\nu_P = \xi(\alpha_P)/\alpha_P \quad (\le 1) , \tag{2.9}$$

$$\nu_M = \gamma_M/\alpha_M = [ \xi(1-\alpha_R) - \xi(\alpha_P)]/\alpha_M , \tag{2.10}$$

$$\nu_R = \gamma_R/\alpha_R = [1- \xi(1-\alpha_R)]/\alpha_R \quad (\ge 1) , \tag{2.11}$$

where $\nu_M$ may be < , = or > 1 depending on $(\omega,\rho)$ and the income distribution F. Recall that

$$\nu_P = \mu^{-1}\{ \int_0^{\omega} ydF(y) \}/F(\omega)$$
$$= \{\text{average income of the poor}\}/\{\text{ mean income of all}\}, \tag{2.12}$$

$$\nu_M = \{\text{average income of the middle class}\}/\{\text{mean income of all}\} , \tag{2.13}$$

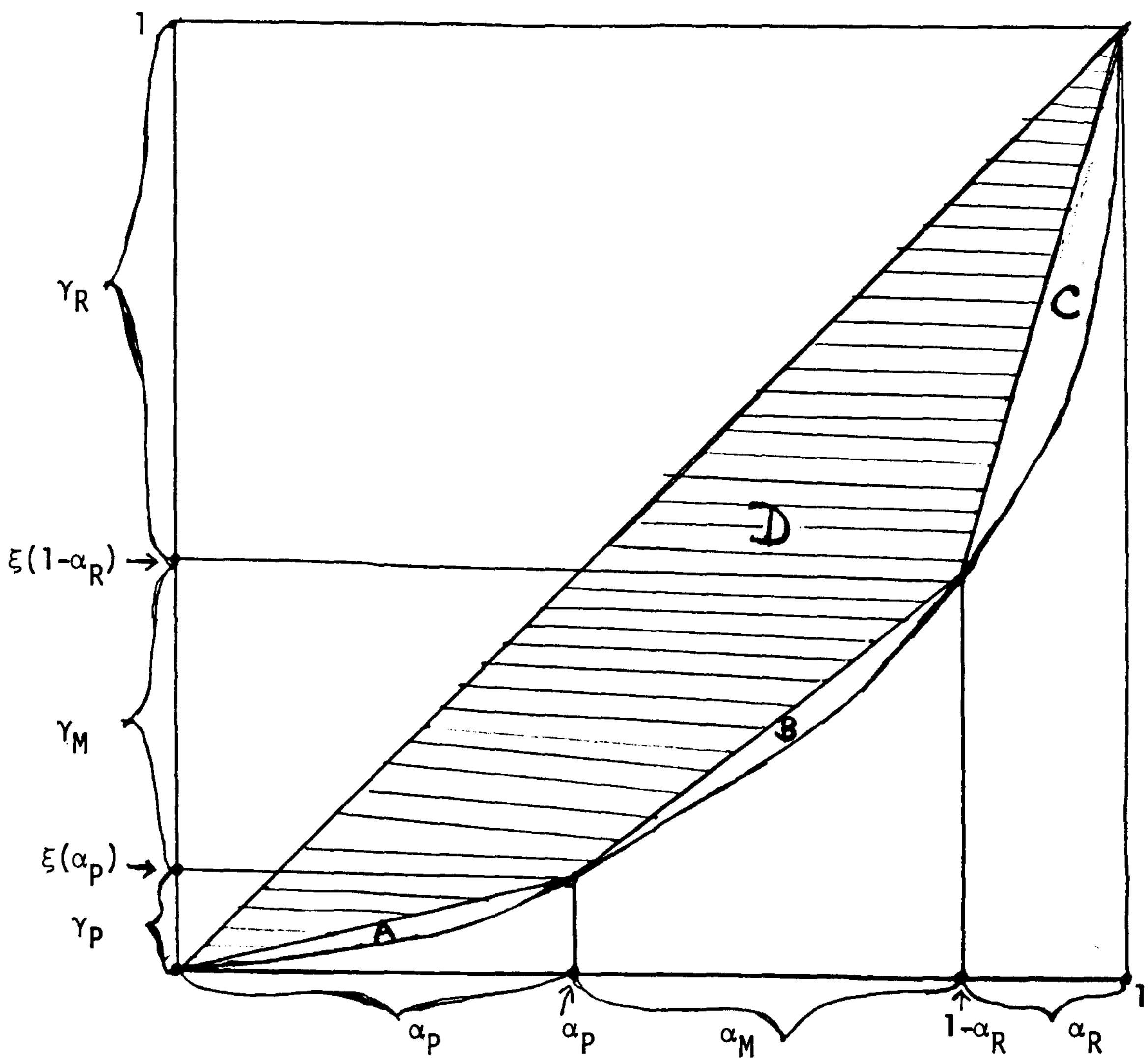
$$\nu_R = \{\text{average income of the rich }\}/\{\text{ mean income of all }\} , \tag{2.14}$$

and these reflect the between sector dispersion of the relative mean incomes ; we must have $\nu_P \leq \nu_M \leq \nu_R$ .

In a conventional setup, one plots $\xi(t)$ against t ($0 \leq t \leq 1$) and obtains the classical Lorenz curve for the income distribution F. In order to obtain the Lorenz curves for the individual sectors as well as for the entire distribution, we consider the following decomposition of the classical Lorenz curve.

*Figure 1*

*A Prismatic Decomposition of the Classical Lorenz Curve*




The entire picture is dictated by the proportions $\alpha_P$, $\alpha_M$, $\alpha_R$, $\nu_P$, $\nu_M$, $\nu_R$ and the individual sector Gini coefficients $G_P$, $G_M$ and $G_R$. In terms of the shaded areas A, B, C and D, the Gini coefficient (G ) for the entire income distribution is given by

$$G = 2(A + B + C + D) . \tag{2.15}$$

Similarly, the Gini coefficients for the poor, middle class and the rich people income distributions are given by

$$G_P = (2A)/(\alpha_P \gamma_P) , \tag{2.16}$$

$$G_M = (2B)/(\alpha_M \gamma_M), \tag{2.17}$$

$$G_R = (2C)/(\alpha_R \gamma_R) , \tag{2.18}$$

respectively. Moreover, D is a polygon whose area can easily be determined by some standard manipulations. Thus, we have

$$\begin{aligned} 2D &= \alpha_R\gamma_R - \alpha_R^2 + \alpha_P\gamma_P - \gamma_P^2 + (1- \alpha_R - \gamma_P)^2 - \alpha_M\gamma_M \\ &= ( \gamma_R - \alpha_R) + \gamma_M(\alpha_P - \gamma_P) + \gamma_P( \gamma_M - \alpha_M) \\ &= \alpha_R( \gamma_R/\alpha_R - 1) + \alpha_P\gamma_M(1- \gamma_P/\alpha_P) + \alpha_M\gamma_P( \gamma_M/\alpha_M - 1) \\ &= \alpha_R( \gamma_R/\alpha_R - 1) + \alpha_P\alpha_M( \gamma_M/\alpha_M - \gamma_P/\alpha_P ) \\ &= ( \nu_R - 1 )\alpha_R + ( \nu_M - \nu_P )\alpha_P\alpha_M . \end{aligned} \tag{2.19}$$

Also, note that by (2.9) through (2.14), we have

$$\nu_P \leq \nu_M \leq \nu_R \; ; \quad \nu_P \leq 1 \quad \text{and} \quad \nu_R \geq 1. \tag{2.20}$$

Thus, the first term on the right hand side of (2.19) represents the contributions of the affluent people [ through their excessive relative income ( $\nu_R - 1$ ) and their proportional representation ( $\alpha_R$ ), while the second term depicts the differential picture of relative mean incomes of the middle class and the poor people, adjusted by their relative proportions too. From (2.15) through (2.19), we have

$$\begin{aligned} G &= \alpha_P\gamma_P G_P + \alpha_M\gamma_M G_M + \alpha_R\gamma_R G_R + \alpha_R(\nu_R -1) + \alpha_M\alpha_P(\nu_M - \nu_P) \\ &= \alpha_P^2\nu_P G_P + \alpha_M^2\nu_M G_M + \alpha_R^2\nu_R G_R + \alpha_R(\nu_R - 1) + \alpha_M\alpha_P( \nu_M - \nu_P) \\ &= G( \alpha_P , \alpha_M, \alpha_R ; \nu_P , \nu_M, \nu_R ; G_P, G_M, G_R) , \text{ say.} \end{aligned} \tag{2.21}$$

This clearly shows the structural dependence of the overall Gini coefficient G on the individual sector Gini coefficients ($G_P$,$G_M$ and $G_R$), the relative proportions ($\alpha_P$,$\alpha_M$ and $\alpha_R$) and the relative mean incomes ( $\nu_P$, $\nu_M$ and $\nu_R$). Thus, for a better understanding of any plausible structural change in G ( due to a complex interplay of a (usually large) number of socio-economic factors), it may be better to look back into the vectors $\underset{\sim}{\alpha} = (\alpha_P, \alpha_M, \alpha_R)$, $\underset{\sim}{\nu} = ( \nu_P, \nu_M, \nu_R)$ and $\underset{\sim}{G} = (G_P,G_M,G_R)$, and to examine the extent of coherence of such a change accross the three sectors. We shall consider this model in greater details in the next section.

Let us next sketch the general case of $m(\geq 1)$ sectors with individual income distributions $F_1,\ldots,F_m$ and the relative proportions $\alpha_1,\ldots,\alpha_m$, respectively. Then

$$\Sigma_{i=1}^{m} \alpha_i = 1 \quad \text{and} \quad F(x) = \Sigma_{i=1}^{m} \alpha_i F_i(x) \text{ , } x \in R^+. \tag{2.22}$$

Note that these $F_j$ need not be all nonoverlapping. Then, we have

$$\mu = \int_0^\infty y dF(y) = \Sigma_{i=1}^{m} \alpha_i \int_0^\infty y dF_i(y) = \Sigma_{i=1}^{m} \alpha_i \mu_i \text{ ,} \tag{2.23}$$

where the $\mu_i$ are the individual sector mean incomes. Also let

$$\nu_i = \mu_i/\mu \text{ , } i=1,\ldots,m \text{ ( so that } \Sigma_{i=1}^{m} \alpha_i \nu_i = 1 \text{ ).} \tag{2.24}$$

Further, we let

$$\omega_{ij} = \mu^{-1}\{\int_0^\infty \int_0^\infty |x-y| dF_i(x)dF_j(y) \text{ , for } i,j = 1,\ldots,m. \tag{2.25}$$

Then, note that

$$\omega_{ii} = 2\mu^{-1}\mu_i G_i = 2\nu_i G_i \text{ , } i=1,\ldots,m, \tag{2.26}$$

where the $G_i$ is the usual Gini coefficient for the income distribution $F_i$ (i=1,..., m); for this definition of the $G_i$, we may refer to Sen(1986). For $i \neq j$, $\omega_{ij}$ stands for some average distance between $F_i$ and $F_j$ . For the income distribution F, we have

$$\begin{aligned} G &= (2\mu)^{-1}\{\int_0^\infty \int_0^\infty | x_1-x_2| dF(x_1)dF(x_2) \} \\ &= (2\mu)^{-1}\{ \Sigma_{i=1}^{m}\alpha_i^2 2\nu_i G_i + 2\mu \Sigma_{1\leq i<j\leq m} \alpha_i\alpha_j\omega_{ij} \} \\ &= \Sigma_{i=1}^{m} \alpha_i^2\nu_i G_i + \Sigma_{1\leq i<j\leq m} \alpha_i\alpha_j\omega_{ij} \text{ ,} \end{aligned} \tag{2.27}$$

so that the overall Gini coefficient can be expressed in terms of the individual $\alpha_i, \nu_i$, $G_i$ and the $\omega_{ij}$, $i\leq j = 1,\ldots,m$. In particular, if the strata income distributions are nonoverlapping (as in the case of the poor, middle class and affluent sectors), then for $F_1 \geq F_2 \geq \ldots \geq F_m$ , we have

$$\omega_{ij} = \nu_j - \nu_i \text{ , for } 1 \leq i < j \leq m, \tag{2.28}$$

so that in this special case (2.27) reduces to

$$G = \Sigma_{i=1}^{m} \alpha_i^2\nu_i G_i + \Sigma_{1\leq i<j\leq m} \alpha_i\alpha_j( \nu_j - \nu_i ), \tag{2.29}$$

and for the particular case of m = 3, (2.29) reduces to (2.21). Thus, for the general case of $m(\geq 2)$ and possibly overlapping income distributions, the only extra adjustment needed is to bring in the additional parameters $\omega_{ij}$ , $1\leq i<j\leq m$ which account fo

the between sectors distances. With this remark, and for the sake of simplicity of presentation, we shall only consider the case of m=3 and nonoverlapping income distributions; a similar picture holds for the general case.

Analysis of income pattern may be done in a parametric setup where F is assumed to be of a specified form [ viz., Pareto law ] and it involves a set $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_r)'$ of unknown parameters, or in a more general nonparametric setup where the functional form of F is not assumed to be given ( and it is assumed that F belongs to a general class, say, $\mathfrak{F}$ ). In a parametric model, the $\alpha_i$, $\nu_i$ and $G_i$ may all be expressed in terms of suitable functions of $\underset{\sim}{\theta}$ , so that the whole analysis may simply be done in terms of suitable parametric constraints on $\underset{\sim}{\theta}$ . There are, however, some general concerns with such parametric procedures :

(a) In practice, the actual functional form of the income distribution may never be known that precisely, and any simple form (such as the Pareto law) may not quite adequately fit the model for the entire range.

(b) As has been mentioned before, some of the socio-economic factors may affect the different sectors rather differently, and as such, in any structural model,with different roles of different input variables, it may be quite counter-intuitive to conceive of a common form of the income distribution for all the strata. That is, merely by varying the associated parameters ($\underset{\sim}{\theta}$), it may not be feasible to describe the component income distributions ( i.e., $F_P$, $F_M$ and $F_R$ ) in terms of a single parametric F. Inclusion of a large number of parameters may drastically reduce the sensitivity of the parametric models.

(c) With scope for plausible departures from an assumed parametric model ( rather differently in different sectors), the issue of robustness is quite an important one. The parametric models thus may not have good robustness properties (against possible departures from the assumed models).

(d) In a parametric framework, to describe a model adequately with a view to studying plausible structural changes, it may be necessary to bring in a large number of parameters ($\underset{\sim}{\theta}$). With the increase in the number of such parameters, the usual simplicity of a parametric approach may evaporate fast, and moreover, the efficiency of a parametric procedure may drastically go down.

There are other reasons too. As such, we would rather explore some alternative nonparametric approaches. Some nonparametric measures of income inequality (to be termed in the sequel as indexes) for the poor and affluent groups are available in the literature, and we shall find it convenient to extend such measures for the middle class sector as well.

For $F_P$, the income distribution of the poor, the average is $\mu_P = \mu\gamma_P/\alpha_P$ and it lies below the set poverty line $\omega$ . Thus, the income gap ratio ( $\beta_P$) for the poor people is defined by

$$\beta_P = 1 - \omega^{-1}\mu_P = 1 - \omega^{-1}\mu\,\alpha_P^{-1}\gamma_P . \tag{2.30}$$

A crude index of poverty is given by $\pi_P = \alpha_P\beta_P$, while a refined one is formulated in terms of the triplet ( $\alpha_P, \beta_P, G_P$). A popular form [ due to A.K.Sen(1976)], based on a set of axioms, is

$$\pi_{PS} = \alpha_P \{ \beta_P + ( 1 - \beta_P)G_P \} , \tag{2.31}$$

and a more robust version [ due to Sen(1986)] is given by

$$\pi^*_{PS} = \alpha_P \{ \beta_P^{1-G_P} \} . \tag{2.32}$$

An alternative poverty index, due to Takayama(1979), is based on the censored ( not truncated) income distribution of the poor, and it can be expressed as

$$\pi_{PT} = \alpha_P G_P + ( 1 - \alpha_P\beta_P )^{-1}(1 - \alpha_P)( \beta_P - G_P), \tag{2.33}$$

where $0 \le G_P \le \beta_P$ . It is known [viz., Sen(1986)] that

$$\pi_{PT} \le \alpha_P\beta_P = \pi_P \le \pi^*_{PS} \le \pi_{PS} \le \alpha_P\beta_P(2 - \beta_P) \le \alpha_P , \tag{2.34}$$

for all income distributions. There are some other forms [viz., Blackorby and Donaldson (1980)] which will not be considered here.

For the income distribution $F_R$ (of the affluent people), the average is $\mu_R = \mu\gamma_R/\alpha_R$, and , by definition, $\mu_R \ge \rho$ . Thus, a different definition of the income gap ratio is needed here. One way to define $\beta_R$, the income gap ratio of the rich is to take

$$\beta_R = 1 - \rho/\mu_R = 1 - \rho\alpha_R\mu^{-1}\gamma_R^{-1} = 1 - \rho\alpha_R\left( \int_\rho^\infty y\,dF(y)\right)^{-1} . \tag{2.35}$$

In this context, we may recall that wealth in a form other than income needs to be transferred into an income form, and there may be real difficulties for accurate assessment of wealth of excessively rich people. From this point of view, $\int_\rho^\infty ydF(y)$ may not be very robust, so that $\beta_R$ in (2.35) may not either be that robust against such measurement errors. The use of a harmonic income gap measure generally leads to a better robustness property [viz., Sen(1988)], and based on this consideration, we may set

$$\beta_R^* = 1 - (\rho/\alpha_R)\left( \int_\rho^\infty y^{-1} dF(y)\right) ; \tag{2.36}$$

it is known that $\beta_R^* \le \beta_R$ [ Sen(1988)]. For the Gini coefficient $G_R$ (for $F_R$), the same criticism (i.e., lack of robustness against measurement errors ) can be labelled to a greater extent, and for this reason, a harmonic Gini coefficient ($G_R^*$) has been advocated by Sen(1988). This is defined by

$$G_R^* = E[\ Y_1^{-1} - Y_2^{-1}\ ] \ / \ E[\ Y_1^{-1} + Y_2^{-1}] \ , \tag{2.37}$$

where $Y_1$ and $Y_2$ are independent random variables each having the distribution $F_R$ . Then parallel to (2.31) and (2.32), we may consider some indexes of affluence as :

$$\pi_{R1} = \alpha_R \{ \beta_R + (1-\beta_R)G_R \} \ , \quad \pi_{R2} = \alpha_R \{ \beta_R^* + (1-\beta_R^*)G_R^* \}, \tag{2.38}$$

$$\pi_{R3} = \alpha_R \{ \beta_R^{1-G_R} \} \quad \text{and} \quad \pi_{R4} = \alpha_R (\beta_R^*)^{1 - G_R^*} \ . \tag{2.39}$$

From what has been discussed in Sen (1988), we are biased towards the use of $G_R^*$, $\beta_R^*$ and $\pi_{R4}$ (= $\pi_R^*$ , say ).

The situation with the middle class income distribution is generally more manageable for two basic reasons :

(a) The real income of a person/family in the middle class group is bounded from below by $\omega$ ( $> 0$) and from above by $\rho$ ($< \infty$), and within these two bounds, the variation may be more smooth than in the affluent or the poor groups.

(b) There are generally far less difficulties in measuring the real income for the middle class people than for the rich or the poor ones. Thus, the robustness considerations relevant to the poor or rich sectors may not be that crucial for the middle class. Note that the average income for the middle class sector is $\mu_M = \mu\gamma_M/\alpha_M$ , and, by definition, $\omega \le \mu_M \le \rho$ . As such, by analogy with (2.30), we may define $\beta_M$ , the income gap ratio for the middle class people, as

$$\beta_M = (\rho - \mu_M)/(\rho - \omega) = 1 - (\mu_M - \omega)(\rho - \omega)^{-1}. \quad (2.40)$$

With this, an index of middle-classness ( $\beta_M$) may be defined as

$$\pi_{M1} = \alpha_M \{ \beta_M + (1-\beta_M)G_M \} \quad \text{or} \quad \pi_{M2} = \alpha_M \{\beta_M^{1-G_M}\} \quad . \quad (2.41)$$

For the general case of m($\geq$ 2) strata with possibly overlapping income distributions, similar indexes can be defined for the individual sectors, and these will usually provide with a more comprehensive picture than the alternative ones solely based on $\alpha_1$,...., $\alpha_m$ or the component Gini coefficients $G_1,\ldots,G_m$ . For this reason, we propose to replace the $\alpha$ by the $\pi$ , and to employ $(\pi_1,\ldots,\pi_m)$, $(\nu_1,\ldots,\nu_m)$ and $(G_1,\ldots,G_m)$ for a better understanding of plausible structural changes of various socio-economic factors affecting the original income distribution.

3. Income Inequality Structures and Response Surface Models

The results in Section 2 are incorporated in formulating a stochastic model for which some conventional response surface methodologies can be adopted. For simplicity of presentation, here we consider the case of three strata : Poor, middle class and affluent sectors; the general case can be handled in a similar manner. As has been mentioned before, instead of the overall income inequality measure G, we shall deal with the following vector :

$$\underset{\sim}{\theta} = ( \pi_P, \pi_M, \pi_R ; \nu_P, \nu_M, \nu_R ; G_P, G_M, G_R) , \quad (3.1)$$

where the $\pi$ stand for the individual sector income inequality indexes, the $\nu$ for their relative mean incomes, and the G for the corresponding Gini coefficients. We may also note that the $\nu$ are essentially measures of central tendencies (relative to the overall mean), while the $\pi$ and G are suitable within strata indexes. Although there may be some general concordance between the $\pi$ and the G , the multi-colinearity problem is not likely to arise here (as the $\pi$ depend on $\alpha$ and $\beta$, in addition to the G ). This dependent vector $\underset{\sim}{\theta}$ is conceivably related to a number of socio-economic factors (as independent variables or regressors ). It is quite clear that a formulation of a suitable response surface methodolgy may require that all the relevant input variables are capable of being quantified in, at least, on an ordinal categorical basis. Actually, for convenience, we assume that these input variables are more or less continuous. Among possible socio-economic variables , the arbitration of the poverty and affluence lines (i.e., $\omega$ and $\rho$ ) as well as the usual cost of living index (say, $\kappa$ ) are the most important ones. Again, it may be of advantage to decompose $\kappa$ into suitable components (viz., food and drinks, energy, education, housing, transportation, medical expenses, recreation etc.) which may have differential effects on the components of $\underset{\sim}{\theta}$ . In addition, distribution of social welfare or other forms of compensation may constitute relevant factors( especially, if they are likely to vary over time). Other economic factors, such as the prime interest rate for money lending, unemployment index, productivity index (for agricultural/industrial and other products) and GNP index may all be quite appropriate for explaining plausible variations in the components of $\underset{\sim}{\theta}$ (over time). With due considerations to all such plausible regressors, we conceive of a vector

$$\underset{\sim}{\xi} = ( \xi_1,\ldots,\xi_q)' \qquad ( \text{for some } q \geq 1) \tag{3.2}$$

of relevant input (factor) variables, and formulate a **regression model** :

$$\underset{\sim}{\theta} = \underset{\sim}{\psi}( \underset{\sim}{\xi}, F ) , \quad F \text{ being the income distribution,} \tag{3.3}$$

where $\underset{\sim}{\psi} = ( \psi_1,\ldots, \psi_9)'$ and each $\psi_k$ is a functional of the distribution F and a function of $\underset{\sim}{\xi}$ , k=1,...,9. The forms of these 9 functionals need not be all similar ( or linear in nature ).

The picture presented above relates to a given time period. In practice, we are interested in the composite picture over a span of successive time periods. For example, these time intervals may be the fiscal years, half-years or even the quarters, and we may like to study plausible changes over such intervals. In this setup, a change-point model becomes quite relevant . Keeping this in mind, we conceive of an index set

$$T = \{ t_1,\ldots, t_n : t_1 < \ldots < t_n \} , \tag{3.4}$$

where $t_j$ refers to the jth time intervals, quantified as its mid-point, j=1,...,n. For the time period indexed by $t_j$, we denote the income distribution by $F_j$ ( for the entire population), and let

$$\underset{\sim}{\theta}_j = \underset{\sim}{\theta}(t_j) \quad \text{and} \quad \underset{\sim}{\xi}_j = \underset{\sim}{\xi}(t_j) , \ j= 1,\ldots, n. \tag{3.5}$$

We may even allow the form of the functional in (3.3) to be time-dependent, and set

$$\underset{\sim}{\theta}_j = \underset{\sim}{\psi}_j( \underset{\sim}{\xi}_j , F_j) , \quad j = 1,\ldots,n . \tag{3.6}$$

The functional forms of the components of $\underset{\sim}{\psi}_j$ in (3.6) may remain stationary (over j) or may follow a change-point model. In this case, an abrupt change is likely to occur if the $\underset{\sim}{\xi}_j$ or $F_j$ itself udergo an abrupt change at some time point - a case that may arise due to some structural changes in the socio-economic factors affecting the income distributions $F_j$.It is in this general response surface model formulation , we intend to consider a change-point model for plausible structural changes over time. In this context, certain order relations and invariance properties of the income indexes [viz., Sen(1986)] need to be taken into account, and this results in a somewhat different formulation of the problem.

## 4. Change- Point Modelling for Income Distributions

Looking at the $\underset{\sim}{\theta}_j$ and $\underset{\sim}{\xi}_j$, we may gather that the $\underset{\sim}{\theta}_j$ are parameters (estimable functionals) of the distribution $F_j$ while the $\underset{\sim}{\xi}_j$ are also parametric quantities ( depending on some other distributions). The arbitration of the poverty and affluence lines or the adjustment for the cost of living index may change the form of the $F_j$ over time and also the $\underset{\sim}{\theta}_j$ may vary accordingly. Thus, we may have an implicit functional relationship between the $\underset{\sim}{\theta}_j$ and the $\underset{\sim}{\xi}_j$. To make this point clear, consider the simple situation where the poverty and affluence lines $\omega_j$ and $\rho_j$ are adjusted by the cost of living index, so that on denoting by $\kappa_j$, the cost of living index at time $t_j$ , we have

$$\omega_j = \omega.\kappa_j \quad \text{and} \quad \rho_j = \rho.\kappa_j \ , \text{ for } j = 1,\ldots,n \ , \tag{4.1}$$

where $\omega$ and $\rho$ are suitable positive numbers. If the income distributions $F_j$ ( at time point $t_j$) satisfy the condition that

$$F_j(y) = F(y/c_j) \ , \ j = 1,\ldots,n, \qquad [\ F \text{ arbitrary }\ ] \tag{4.2}$$

where the $c_j$ are proportional to the cost of living indexes $\kappa_j$ , then it is easy to show that the income indexes $\pi_P$, $\pi_M$ and $\pi_R$ , as well as $\nu_P$, $\nu_M$, $\nu_R$ and $G_P$, $G_M$ and $G_R$ remain the same over the entire time period. Thus, the $\underset{\sim}{\theta}_j$ remain stationary over the span of time . On the other hand, if the income distributions $F_j$ do not satisfy the scale-model in (4.2) and/or the poverty and affluence lines are not adjusted by the proper scale factors, the different components of the $\underset{\sim}{\theta}_j$ may be affected rather differently.In this context, it may also be noted that a change in $\alpha_P$ ( or $\alpha_M$ or $\alpha_R$)

may not necessarily lead to a change in $\nu_P$ , $G_p$ ( or $\nu_M$, $G_M$ or $\nu_R$,$G_R$ ) in the same direction, so that $\underset{\sim}{\theta}$ may not satisfy a partial ordering with respect to the $\alpha$ or the other parameters. However, the scale-model in (4.2) and the cost-of-living adjustment in (4.1) may work out quite well if the span of the time in T is not too large and there is no abrupt change in some socio-economic factors affecting the income distributions. As such, we may consider cost of living adjusted real income as well as poverty and affluence lines, so that in the ideal situation, $\underset{\sim}{\theta}$ based on this adjusted distribution would not depend on the cost of living index $\kappa$ , so that we have

$$\underset{\sim}{\theta}_j = \underset{\sim}{\theta} \text{ , for all } j = 1,\dots, n . \qquad (4.3)$$

On the other hand, when a cost of living adjustment may fail to induce a scale-equivariance of the income distributions at different time points, the $\pi$ , $\nu$ and G may not remain the same , and hence, the constancy of the elements of the $\underset{\sim}{\theta}_j$ over j may not hold. In the later case, the components of $\underset{\sim}{\xi}_j$ may provide suitable explanation for the variation in the components of the $\underset{\sim}{\theta}_j$ . This explains why in (3.6) for the $\underset{\sim}{\psi}_j$ being sufficiently smooth (i.e., locally linear), we may actually assume a linear response surface model for the $\underset{\sim}{\theta}_j$ in terms of the $\underset{\sim}{\xi}_j$ ( provided we use the cost of living adjusted income distributions). In the sequel, we therefore take the income distributions $F_j$ as the cost of living adjusted ones. In passing, we may remark that the elements of $\underset{\sim}{\theta}_j$ are regular functions of the distributioon $F_j$ , and hence, when the input variables $\underset{\sim}{\xi}_j$ affect the $F_j$ only locally , usual expansion of such functionals insures the linearity of the elements of $\underset{\sim}{\theta}_j$ in terms of the $\underset{\sim}{\xi}_j$.

In practice, the distributions $F_j$ as well as the $\underset{\sim}{\theta}_j$ and $\underset{\sim}{\xi}_j$ are unknown, and we need to estimate them. This does not pose any problem as usually large data sets are available for each of the time periods, and our estimates can be based on them. We denote such estimators by

$$\hat{F}_j \text{ , } \hat{\underset{\sim}{\theta}}_j = \underset{\sim}{Q}_j = (Q_{j1},\dots,Q_{j9})' \text{ and } \hat{\underset{\sim}{\xi}}_j = \underset{\sim}{U}_j = (U_{j1},\dots,U_{jq})' \text{ , } \qquad (4.4)$$

for j = 1,...,n. It may be quite reasonable to assume that $\hat{F}_j$ is a consistent estimator of $F_j$ (and other asymptotic optimality properties of these sample distributions may also be assumed). We may write

$$\underset{\sim}{Q}_j = \underset{\sim}{\theta}_j + \underset{\sim}{\varepsilon}_j \quad \text{and} \quad \underset{\sim}{U}_j = \underset{\sim}{\xi}_j + \underset{\sim}{\eta}_j , \ j = 1,\dots,n, \tag{4.5}$$

where, for each j, $(\underset{\sim}{\varepsilon}_j, \underset{\sim}{\eta}_j)$ has a (9+q)-variate joint distribution. As is usually the case, one may have large sample size pertaining to each time period, so that by an appeal to the classical large sanple theory, we may claim that

(i) the stochastic variability of $(\underset{\sim}{\varepsilon}_j, \underset{\sim}{\eta}_j)$ would be small, and

(ii) when suitably adjusted by these sample sizes, normalized version of these vectors $(\underset{\sim}{\varepsilon}_j, \underset{\sim}{\eta}_j)$ would have closely some multivariate normal distribution, although this asymptotic distribution may differ from one time period to another.

Based on the above observations, we may therefore assume that for each j (=1,...,n), the conditional distribution of $\underset{\sim}{Q}_j$ given $\underset{\sim}{U}_j$ is closely normal with the mean vector

$\underset{\sim}{\theta}_j + \underset{\sim}{\beta}_j(\underset{\sim}{U}_j - \underset{\sim}{\xi}_j)$ [where $\underset{\sim}{\beta}_j$ is a 9×q matrix of regression parameters] (4.6)

and a dispersion matrix $\underset{\sim}{\Sigma}_j$. In this setup, borrowing the(approximate) linearity of the $\underset{\sim}{\theta}_j$ [ as has been discussed after (4.3)], we may regard that the $\underset{\sim}{\theta}_j - \underset{\sim}{\beta}_j\underset{\sim}{\xi}_j$ behaves more or less as stationary, so that the main source of variability comes from the regression $\underset{\sim}{\beta}_j\underset{\sim}{U}_j$ ( for j=1,...,n). Note that in this setup, both $\underset{\sim}{\beta}_j$ and $\underset{\sim}{\Sigma}_j$ are unknown matrices and these can be estimated from the sample data as well. Generally, the $\underset{\sim}{\Sigma}_j$ depend on the sample sizes (say, $N_j$ ) on which the estimators $\underset{\sim}{Q}_j$ and $\underset{\sim}{U}_j$ are based, and they depend also on some underlying dispersion matrices. Since we assume that these $N_j$ are all large, we may use the classical *jackknifing* , *bootstrapping* or some other *resampling method* to estimate these $\underset{\sim}{\Sigma}_j$ consistently. On the other hand, the role of the $\underset{\sim}{\beta}_j$ is somewhat different, as they enter into the specification of the change-point model (and hence need to be estimated in a more refined manner). As such, for our statistical analysis, we may assume that the $\underset{\sim}{\Sigma}_j$ are all given although they may not be assumed to be homogeneous (i.e., $\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$ , for all j ). Usually the characteristic roots of the $\underset{\sim}{\Sigma}_j$ are all small ( when the $N_j$ are large), but the homogeneity of the $\underset{\sim}{\Sigma}_j$ may demand more restrictive conditions on the underlying dispersion matrices.

From what has been discussed above, we may conclude that in the formulation of our income pattern, structural changes, if any, should be attributed to the regression matrices $\underset{\sim}{\beta}_j$ , j =1,...,n. In this setup, we therefore confront a *constancy of regression surface model* where the *homoscedasticity* ( of the $\underset{\sim}{\Sigma}_j$ ) is not a part of the model (assumptions). As such, we frame the null hypothesis of no change-point as

$$H_0 : \quad \beta_1 = \ldots = \beta_n = \beta \quad (\text{unknown}), \tag{4.7}$$

and, side by side, we let

$$H_r : \beta_1 = \ldots = \beta_r \neq \beta_{r+1} = \ldots = \beta_n \quad , \text{ for } r = 1,\ldots,n-1. \tag{4.8}$$

Then, the usual change point alternative hypothesis is

$$H^* = H_1 \cup \ldots \cup H_{n-1} = \bigcup_{r=1}^{n-1} H_r . \tag{4.9}$$

In testing $H_0$ against $H^*$, we treat the $\Sigma_j$ as given (but not necessarily homogeneous). For this testing problem, we consider the following:

*I) Pseudo Two-Sample Approaches*. For each r ($1 \leq r \leq n-1$), consider a break up of the data set into two parts : $\{(Q_j, U_j, \Sigma_j), j \leq r\}$ and $\{(Q_j, U_j, \Sigma_j), j > r\}$. From the first set, by using the classical *wieghted least squares method*, we obtain the *weighted least squares estimator* (*WLSE*) $\hat{\beta}_{(r)}$ of $\beta$. Similarly, let ${}_{(n-r)}\hat{\beta}$ be the WLSE of $\beta$ from the second set. Let then

$$Z_r^o = \hat{\beta}_{(r)} - {}_{(n-r)}\hat{\beta} \text{ and } Z_{(r)} = \text{vec}(Z_r^o), \text{ for } r = 1,\ldots,n-1. \tag{4.10}$$

Thus, the $Z_{(r)}$ are 9q-vectors. Using the normality in (4.6) and the linearity of the WLSE, it is easily seen that under $H_0$ in (4.7), $Z_{(r)}$ has closely a multinormal distribution with null mean vector and dispersion matrix (say,) $\Gamma_{(r)}$. The $\Gamma_{(r)}$ can also be consistently estimated from the data set, and hence, we assume that these $\Gamma_{(r)}$ are all given. Let then

$$||Z_{(r)}||^2_{\Gamma_{(r)}} = Z'_{(r)} \Gamma^-_{(r)} Z_{(r)} \quad , \text{ for } r = 1,\ldots,n-1 , \tag{4.11}$$

and

$$T_{n1} = \max\{ \; ||Z_{(r)}||^2_{\Gamma_{(r)}} : 1 \leq r \leq n-1 \} . \tag{4.12}$$

In this context, recall that the $\Gamma^-_{(r)}$ are the generalized inverses of the $\Gamma_{(r)}$, and further, the characteristics roots of the $\Gamma_{(r)}$ are all small ( when the $N_j$ are all large), so that the characteristic roots of $\Gamma^-_{(r)}$ are all large. If the null hypothesis $H_0$ holds, for each r (=1,...,n-1), (4.11) is bounded in probability, so that $T_{n1}$ is also stochastically finite. On the other hand, if $H_0$ does not hold and $H_r$ holds for some r (=1,...,n-1), then at least some of these $Z_{(r)}$ would have a non-null mean vector, and hence, $T_{n1}$ will be large. Hence, the null hypothesis $H_0$ is to be rejected in favor of $H^*$ if $T_{n1}$ in (4.12) is significantly large. Thus, our main task is to find out a *critical value* $\tau_{n,\alpha}^{(1)}$ , such that

$$P\{ \ T_{n1} \geq \tau_{n,\alpha}^{(1)} \ | \ H_0 \ \} \quad \leq \quad \alpha \ , \tag{4.13}$$

where $\alpha$ ( $0 < \alpha < 1$) is the desired *level of significance* of the test.

Instead of (4.10), we may also consider

$$\underset{\sim}{Z}_r^{o*} = \hat{\underset{\sim}{\beta}}_{(r)} - \hat{\underset{\sim}{\beta}}_{(n)} \quad \text{and} \quad \underset{\sim}{Z}_{(r)}^{*} = \mathrm{vec}(\underset{\sim}{Z}_r^{o*}) \ , \ \text{for } r = 1,\ldots,n-1, \tag{4.14}$$

denote the dispersion matrix of $\underset{\sim}{Z}_{(r)}^*$ by $\underset{\sim}{\Gamma}_{(r)}^*$ , and let

$$T_{n2} \ = \ \max\{ \ \ ||\underset{\sim}{Z}_{(r)}^*||_{\underset{\sim}{\Gamma}_{(r)}^*}^2 \ : 1 \leq r \leq n-1 \ \}. \tag{4.15}$$

In this setup, parallel to (4.13), we need to find out a critical level $\tau_{n,\alpha}^{(2)}$ for $T_{n2}$ , such that $H_0$ would be rejected in favor of $H^*$ whenever $T_{n2} \geq \tau_{n,\alpha}^{(2)}$ . We shall consider suitable approximations for these critical levels later on.

*II) Recursive Residual CUSUM Procedures.* With the WLSE $\hat{\underset{\sim}{\beta}}_{(r)}$ defined as in before (4.10), let us define the recursive residuals as

$$\underset{\sim}{Y}_r^o = \underset{\sim}{Q}_r - \hat{\underset{\sim}{\beta}}_{(r-1)}\underset{\sim}{U}_r \ , \ \text{for } r = 2 \ ,\ldots,n \ ; \quad \underset{\sim}{Y}_1^o = \underset{\sim}{0} \ ; \tag{4.16}$$

$$\underset{\sim}{S}_r^o = \Sigma_{k\leq r} \underset{\sim}{Y}_k^o \ \text{and} \ \underset{\sim}{S}_{(r)} = \mathrm{vec}(\underset{\sim}{S}_r^o), \ \text{for } r = 2,\ldots,n. \tag{4.17}$$

Thus the $\underset{\sim}{S}_{(r)}$ are the CUSUM (vectors) of the recursive residuals (using the WLSE), and we denote the dispersion matrix of $\underset{\sim}{S}_{(r)}$ by $\underset{\sim}{\Lambda}_r$, $r = 2,\ldots,n$; again these $\underset{\sim}{\Lambda}_r$ can be consistently estimated from the data set, and hence, we assume that they are given. Let then

$$T_{n3} \ = \ \max \ \{ \ ||\underset{\sim}{S}_{(r)}||_{\underset{\sim}{\Lambda}_r}^2 \ : r = 2,\ldots,n \ \} \ . \tag{4.18}$$

The null hypothesis $H_0$ is to be rejected in favor of $H^*$ whenever $T_{n3}$ exceeds a critical level $\tau_{n,\alpha}^{(3)}$ , where $P\{ \ T_{n3} \geq \tau_{n,\alpha}^{(3)} \ | \ H_0 \ \} \leq \alpha$ .

These test procedures are formulated by analogy with the usual change-point tests for the classical regression model [ viz., Brown, Durbin and Evans (1975)]. However, to accommodate possible heterscedasticity, WLSE have been used instead of the classical least squares estimators. Consistency of these WLSE insures the consistency of the proposed tests against any fixed alternative covered by $H^*$ in (4.9). The crux of the problem is therefore to find out suitable approximations for the critical levels $\tau_{n,\alpha}^{(j)}$ for $j = 1,2,3$. Given that the sample sizes $N_j$, $j=1,\ldots,n$, are all large ( leading to the asymptotic normality in (4.6)), for these WLSE, we may as well assume that the asymptotic normality holds, and this is then transmitted on to the $\underset{\sim}{Z}_{(r)}$, $\underset{\sim}{Z}^{*}_{(r)}$ and $\underset{\sim}{S}_{(r)}$ . Thus, we are able to reduce the whole thing to a multi-normal setup. In this context, two basic things are to be paid due attention:

(i) The dispersion matrices of these vectors are in general of complicated structure, so that the results for multinormal distributions with identity dispersion matrix may not always be applicable;

(ii) The number of time points ( i.e., n) may or may not be small. For large values of n, the computational complexities may call for some further asymptotic approximations. In either situation, we shall see that suitable *simulation techniques* work out well. The basis for this simulation study is provided by the weak convergence results for the partial sequence $\underset{\sim}{Z}_{(r)}$ , $\underset{\sim}{Z}^{*}_{(r)}$ or $\underset{\sim}{S}_{(r)}$ . In either case, the asymptotic multinormality results for the finite dimensional dsitributions follow by an appeal to the classical Cramer-Wold theorem and the central limt theorem for WLSE (which are all linear estimators). In this context, one also needs to establish the *tightness* or *relative compactness* of the stochastic processes constructed from these vectors. When n is small, one does not need to construct such processes, so that the desired simulation results would directly fit with these asymptotic multinormal law, and we need to use consistent estimators of the associated dispersion matrices to generate these multinormal vectors. On the other hand, for large n, the process may turn out to be extremely tedious, and suitable Gaussian process approximations ( in law ) for the associated stochastic processes may provide simpler simulation prospects. In this context, the tightness condition may be verified by standard techniques ( applicable for linear processes), and hence, we omit the details here. In the context of survival analysis ( or in life testing models), for the weighted Kolmogorov-Smirnov statistics, a similar problem arises, and in Sinha and Sen (1982) some of these simulation studies have been reported in detail. In view of the similarity, we therefore avoid these details. In some specific (simple ) cases, one may use some standard results on the so called *Bessel processes* [ this is particularly the case where the homogeneity of the dispersion matrices (over j) can be assumed), and the detailed tables of De Long

(1981) can be used with some advantage. In this context, we may note that there are some simple relations between the critical levels of tied-down Bessel processes over a part of [0,1] and the usual Bessel procees on a part of $R^+$ [ viz., De Long (1981) and Sen(1981,Ch.2)], so that for $T_{n1}$ or $T_{n2}$ such relations can be used to adapt the critical levels from De Long's tables. If n is large enough then to apply these Bessel process approximations it is not necessary to assume the homogeneity of the dispersion matrices at various time points ( all we need that that the transformed points of the time arguments are dense on [a,b], for some $0 < a < b < 1$ , and this can be justified when these dispersion matrics are not too much different from each other).

5. Forecasting of Economic Structural Changes

The main focus of this study has been on a breakdown of the overall Gini coefficient in terms of a number of component Gini coefficients, relative mean incomes for these sectors and their relative proportions. It has also been shown in earlier sections that the usual income inequality indexes when suitably modified may provide some further insight in this probe. Thus, given the basic formulation of the (vector-) mod in (3.1) and the subsequent analysis in Section 4 [ viz., (4.1) through (4.9)] , it seems quite appropriate to concentrate our attention to a set of perninent socio-economic factors [ giving rise to the regressors $\underset{\sim}{\xi}_j$ ( or their estimators $\underset{\sim}{U}_j$ ) ] The basic issue is therefore to choose these $\underset{\sim}{\xi}_j$ in a most judicious manner. This choice is of course dependent on the particular society or community and the major socio-economic factors affecting the same. Although it may be intuitively appealing to have a large number of components in the $\underset{\sim}{\xi}_j$ ( to provide more explanation of the impact of various socio-econimic facrots ), from the statistical analysis point of view, there is a mixed reaction: The larger is the dimension of the regressor vectors, the greater should be the sample sizes so that the associated regression matrices can be estimated with comparable precision. The quadratic norms used in (4.12), (4.15) and (4.18) all relate to the dimension of the $\underset{\sim}{Z}_{(r)}$ ( or $\underset{\sim}{Z}^*_{(r)}$) and $\underset{\sim}{S}_{(r)}$ . The larger is the value of q, the greater will be the (stochastic) variablity of these norms, so that the criti levels [ $\tau_{n,\alpha}^{(j)}$ ] will be larger. This automatically points out that unless all the explanatory regressors are informative, the power or sensitivity of the tests in Section 4 may not increase with the increase in q . The situation is quite comparable to the usual chi squared (goodness of fit ) tests where an increase in the degrees of freedom may not necessarily lead to an increase in the power (unless the nocntrality increases at a commensurate rate ). Thus, in choosing q and the subset of the regressors, sufficient care must be taken to insure that unnecessary loss of efficiency does not arise due to redundant regressors or due to omission of some relevant ones.

For forecasting purposes, the above findings are quite important. First of all, if q and $\underset{\sim}{\xi}_j$ are chosen properly then a change in the structural form may be studied in terms of these explanatory regressors, and the model in Section 4 can be used with advantage. On the other hand, it may be such that some factor (or regressor) may be quite insignificant upto a certain time point, and then due to some extraneous factor, it may become quite an important one at some stage. Thus, if no attention is paid to this factor ( based on its past history), a forecasting formula may not work out that well. Thus, in a forcasting situation, one needs to pay attention to a possible change point model with respect to the regressors $\underset{\sim}{\xi}_j$ ; a progressive examination of these factors with a view to eliminating the redundant ones and accommodating new important ones should lead to a better forecasting. Secondly, in Section 4, to eliminate the redundant variation due to plausible scalar adjustments, the income distributions were taken as cost of living adjusted ones. It provided us with a reasonable way of achieving a linear model [c.f. Section 4 ] when the regressors are not too different from each other (i.e., there is no abrupt change in their realizations over the time period considered). In the conext of forecasting, this cost of living adjustment for the $F_j$ should also be examined carefully. If a cost of living adjustment fails to bring two distributions ( say, $F_j$ and $F_{j'}$ ) close to each other, the approximate linear expansion in Section 4 may not be that appropriate, and hence, a forecasting based on this model may not work out that well. This may particularly be very important when an existing important regressor phases out of the system and a new one enters into the scheme and the cost of living adjustment as adapted might not have paid due importance to this new factor. Nonlinear regression models may be necessary for the forecasting problem if no such cost of living adjustment is made to the income distributions and the regressors may vary considerably over time. Finally, in forcasting for a very near future time period, the methodology in Section 4 can be used with greater confidence. However, as the time gap between the time domain under study and the projected time increases, the effectiveness of forecasting models may decrease drastically, especially, in the context of structural changes. Nevertheless, the decomposition considered in Sections 2, 3 and 4 provide us with pertinent insight into the structure of such income patterns ( and inequalities), and a forecasting based on such a decomposition should be much more effective than the usual one based on the overall Gini coefficients or some other conventional measure of income inequality.

We conclude this section with a remark on the WLSE used in Section 5 (and in this one too). The justification for these WLSE is based on the asymptotic multi-normality result in (4.6). This is generally quite appropriate. If, however, (4.6) is not that appealing ( but still the linearity of regression may be tenable ), one may use some robust estimators instead of the WLSE. For the constancy of regression relationships over time, suitable M-tests based on recursive M-estimators have been considered by Sen(1984), Hušková (1988) and Hušková and Sen (1989), among others, and these may be considered here too. However, in all the other works, homoscedasticity has been a vital part of the basic model, while in the current setup, there may be good evidence that such an homoscedasticity assumption may not hold. Thus, it remains to study the effect of heteroscedasticity of the original model on the performance characteristics of such recursive M-tests for change-points or M-estimators in forecasting. We intend to take over this study in greater length and report the findings in a future communication. The basic reason for using the WLSE instead of the ordinary LSE is to take into account the possible heterscedasticity in the model, and this approach is likely to generate sufficient interest in other areas too.

## References

Blackorby, C., and Donaldson, D. (1980), Ethical Indices for the measurement of poverty. *Econometrica* 48, 1053-1060.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, B 37, 149-192.

DeLong, D. (1981), Crossing probabilities for a square root boundary by a Bessel process. *Communications in Statistics: Theory and Methods* A10, 2197-2213.

Hušková, M. (1988), Recursive M-test for detection of change. *Sequential Analysis* 7, 75- 90.

Hušková, M., and Sen, P.K. (1989),Nonparametric tests for shift and change in regression at an unknown time point. In *The Future of the World Economy : Economic Growth and Structural Changes* (Ed. W. Krille), Springer-Verlag, Berlin; pp. 73-87.

Sen, A.K. (1976), The measurement of poverty: An axiomatic approach. *Econometrica 44*, 219-232.

Sen, P.K. (1981), *Sequential Nonparametrics : Invariance Principles and Statistical Inference*, John Wiley, New York.

Sen, P.K. (1984), Recursive M-tests for the constancy of multivariate regression relationships over time. *Sequential Analysis,3,* 191-211.

Sen, P.K. (1986),The Gini coefficient and poverty indexes: Some reconciliations.*Journal of the American Statistical Association,81,* 1050-1057.

Sen, P.K. (1988), The Harmonic Gini coefficient and affluence indexes. *Mathematical Social Sciences, 16,* 65-76.

Sen, P.K. (1989a), Affluence and poverty indexes. In *Encyclopedia of Statistical Sciences* (Eds: S. Kotz and N.L. Johnson ), Supplement, pp.1-5.

Sen, P.K. (1989b), Change-point problem relating to the poverty structure. In *The Future of the World Economy: Economic Growth and Structural Changes* (Ed. W. Krille ), Springer-Verlag, Berlin; pp. - .

Sinha, A.N.,and Sen, P.K. (1982), Tests based on empirical processes for progressive censoring schemes with staggering entry and random withdrawal. *Sankhya, Series B*, 44, 1-18.

Takayama, N. (1979), Poverty, income inequality and their measures: Professor Sen's axiomatic approach reconsidered. *Econometrica 47,* 747-760.