No. 352

ON THE LOGNORMAL INCOME DETERMINATION HYPOTHESIS-LINEAR INCOME TAX, INCENTIVES AND ABILITIES

by

Sho-Ichiro Kusumoto

December 1987

Preliminary version Incomplete!

THE UNIVERSITY OF TSUKUBA

Institute of Socio-Economic Planning
Sakura, Ibaraki 305, Japan
Telephone : 0298-53-5182

Cable. UNTUKU TSUCHIURA
Telex : 3652580 UNTUKU J

ON THE LOGNORMAL INCOME DETERMINATION HYPOTHESIS — LINEAR INCOME TAX, INCENTIVES AND ABILITIES

By

Sho-Ichiro Kusumoto **

December 1987

ABSTRACT

The lognormal hypothesis and implications for the distributions of ability and income are investigated in a context of a simple general equilibrium model of redistributed workers with linear income taxation, in particular, from an incentive view point.

By isolating a hidden, vaguely recongnized effect in a more general formulation, from the income equalizing effect, we shall pay a special attention to an "adverse" effect which will possibly induce workers to manipulate costlessly their imperfectly observed present ability, in such a way that their future size will be in a skew way, as well as unevenly, distributed.

This manipulability features moral hazard the less able workers would fall into, as well as a hierarchical self-selection incentive the able workers would see. We formulate a time process of such costless ability manipulation as the process generating the size distributions of ability, depending upon which the authority and firms make decisions, hence the future sizes of income and consumption will be determined.

Under relevant conditions on the rates of manipulable change we show the process will lead the final distribution to lognormality if it is infinitely many times repeated, while it hardly converges to a stationary distribution, as long as such incentives motivates thus the workers.

Income Distributions

Neoclassical literature has abounded with alternative theories of distribution of income among factors of production. However there have been quite a few attempts to enunciate a theory of distribution among individuals from economic approach. Such an effort had begun more than three decades ago notably by Tinbergen[1957].

Although those distributions obtained by Roy[1951], Friedman [1953], Newman and Wolfe[1961] etc. were explained as the end of economic choice which could at most help the shape of distribution, they were each a drawing from a random process.

It was the attempt made by Stiglitz[1969] *1 that seems to have firstly and successfully completed an economic theory of wealth and income distribution among individuals, entirely based upon the basic principles that have proven in a neoclassical growth theory.

While elegant and enlightening, these, essentially stochastic, models, originated with Gibrat[1931], Kalecki[1945] and Champernoun [1953] and developed notably by Sargan[1957], Mandelbrot[1961] and most recently by Sharrocks[1975], hardly seem to have been integrated into a main stream of modern theory of income distribution. See Mincer[1970] and Blinder[1974] for survays.

Namely, an unsatisfactory state in the area of economic interpretation still has been continueing. This contrasts the fairly satisfactory state with respect to the statistical description of income distribution. *2

A purpose of the present work is to fill a part of this gap, by taking an approach from a theory of optimal taxation founded mainly by Mirrlees[1971,1974].

In a context of a simple general equilibrium of redistributive taxation, the lognormal hypothesis and implication for the size distributions of ability and income are fully investigated. Our final purpose is to lead the resulting distributions of ability and income asymptotically to lognormality whereas the process hardly converges to a stationary distribution.

Incentives of Income-Redistributed Workers

We begin by sketching our view of a mixed economy which will be our extensive concern, in particular from the point of a worker's view of incentives.

In the economic society, there are a large number of competitive workers*3 who are redistributed by means of income taxation policy. There are also a large number of competitive firms employing the workers in order to produce and sell outputs for consumption*4. The workers are essentially sharecropping in the sense that the output each hired individual adds can be distinguished perfectly from those which the others add. The authority offers alternative policy schemes of linear*5 income-taxes and transfers. Each scheme is budget-consistent and takes into account all potential employment contracts. These contracts each simply describe taxable labor supplies (wages) and will be offered by all firms and open to all potential workers.

The authority and each firm mutually inform and are informed of all knowledges they have obtained about the characteristics such as abilities and preferences of contracting workers, through the errors and trials process of contracting.*6 So they will be able to have a sufficient statistic on a unique distribution of the characteristics of all existing workers, upon which their final schemes and offers will depend.

Given a distribution of the recruited workers' characteristics, each firm, as a candidate employer, makes an offer to a candidate employee searching for the desired contracts; a set of labor contracts under alternative taxation schemes; a labor supply schedule. Each contract on offer maximizes a consumption-leasure utility of a potential worker, under each corresponding taxation scheme. The candidate worker would identify him/herself with the utility maximizer to find optimal each such contract offered if the firms can force workers to make market choice in such a way that they both reveal their characteristics and make the choice the firm would have wanted to make had their characteristics been publicly known.*7

The authority consists of officers called government (G) men, who are paid a constant portion of the tax revenues for supervising*8 the firms and workers in order to enforce the binding contracts.

The latent abilities of contracting workers are estimated and measured through each cropping to be realized. Each ability may be representable by a task he/she will perform in tilling the crop. The crop as a result of the work is used each as a useful proxy or indicator of the underlying ability of the individual in concern.

Thus, the stated crops are, each and together, regarded as serving a function of differentiating each worker from the others who will make different levels of performance in producing their outputs.

There is no technological uncertainty[*9] assumed to be involved between the expected and realized outputs.

There are pointed out two distinguished components to a worker's input: the work hours that he/she will put in and the (per hour) performance that he/she will make per hour. The performance reflects traits and attributes which cover effort and attitude toward the work or job of worker. While the first is easy to observe, the second can not without careful supervision which will cost relatively high.

In a private economy, the benefits of full and accurate screening information about the ability of a competitive worker would accrue to this individual, not accrueing to a competitive firm employing the worker. Hence the costs of obtaining such informations would be borne by the individual himself, if they are relatively small.[*10]

Under incomplete and asymmetric information, all able workers, who are perfectly informed of their own abilities,(perhaps except the least able) would each see a self-selection and hierachical screening incentive; an incentive to identify themselves with their own performances, in particular when they and their outputs could not be separated according to their different performances, so that, as in a wage-payment system, they would have been to receive an average wage of the entire or rest of work force. The less able workers would have an incentive to pretend to be more than they actually are so that they could be free-riding.

This is why and how the hierarchical screening/signaling device would be induced to be set and put into motion. Miyazaki [1977] and Riley[1979], among others, developed "imperfectly" competitive equilibrium[*11] concepts with a negative informational externality. See Riley[1979 Theorems 3,7 and final remark] for welfare implications of those non-Pareto optimal equilibria.

However, in the presence of its coordination with production, on the job screening[*12] with finite fines will likely to have a relevant implication here: The market equilibrium will be Pareto optimal and can be characterized by full screening without any resources on the screening itself, if it is perfectly accurate and if workers are perfectly informed about their abilities. For example, suppose, in the contract, the worker agrees to pay the firm a large fine if the contract turns out to be not carried out for such a reason as in case his/her ability has been misreported or overstated. Each firm has only to announce it will screen if the outputs differ from what they should be, given the abilities which the employed individuals have declared.

These workers are classified with sharecroppers and those who are costlessly and accurately observed. Unlike average wage earners, such workers are completely paid for the performances they have made, as well. They will be always identified with themselves.

However, if a taxation system is introduced in such a way that the outputs are each to be taxed in a direct proportion to them, even such contracting workers will see, more or less, an incentive to misreport both the expected and realized results of their performances, here, to for tax evasion understate the ex-ante outputs in their contracts. The incentives to engage in such tax-sheltering activities more likely to see for higher-ability workers.

Thus, as well as supervising through G men, that the juridical authority announce it will fine in case understating the taxable outputs (wages) would be, more or less, a disincentive to such manipulability by the workers.

Finally, given the contracted labor supply schedules of all existing workers, the authority will announce to implement an optimal tax rate which maximizes a social welfare (the aggregate of individual utilities[*13] over(a continuum of) workers of these different estimated abilities) function.

Summaries for Sections

In Section 1, we thus solve a simple problem of optimal taxation,[*14] with a linear income taxation for the income-redistributive purpose, a fixed (density function of) distribution of the present size of heterogeneous ability and with fixed homothetic preferences each representable by an identical Cobb-Douglas function which is linearly homogeneous and concave. Our first purpose is to work out the details of the solution properties which would be hidden in a more general formulation.

In such economy an optimal tax rate is determined so that a positive number of workers won't accept work for the tax rate. The possible largest ability such voluntarily unemployed possess is uniquely determined for the tax rate. We call this reservation ability for a clear reason. The reservation ability, associated with the reservation wage, will play an essential role throughout the whole sections. The income returned (for the policy purpose) to each worker will be distributed in the same way as the endowned ability is, but, now with a variance smaller by the square multiple of one minus the optimal tax rate. Consumptions would be confined to the range above the minimum (equal to the transfer) determined for the tax rate, as if the rest of distribution (for which they are less than the transfer) were removed. The consumption is thus trancatedly distributed and the point of trancation is also determined uniquely by the tax rate. The variance of consumption is much smaller (that is, by more than the square) than that of the income that would be earned by a worker him-/herself in case no such taxation policy was implemented at all. Theorem 1

There will be pointed out, however, a dynamic effect of the tax policy, in particular, on the future sizes of abilities.

Section 2 is thus for our second purpose. We shall deduce a different incentive that each worker will see depending upon a different position relatively to the reservation ability, in the distribution of abilities in the population.

In the mixed economy, the voluntarily unemployed are enjoying a full-time leasure by receiving the transfer, completely independent of their potentialized, now perfectly unobserved, abilities. Thus, on the one hand, the unemployed and those with lower abilities also would see an incentive to manipulate(costlessly)their abilities so as to pretend to be less,whenever it is advantageous. Such manipulability features the moral hazard the less able workers would fall into.

On the other hand, the hierarchical self-selection signaling incentive (which the more able would see in the case where the employers could not identify them with their different performances) may be extended to an incentive to achieve a hierarchical rise, since they always get identified with themselves. In a linear taxation economy like this, the more able workers will see an incentive to raise their present status quo (abilities), resulting in an increase in the taxable outputs.*15 The more able will see a larger utility gain from a unit increase in ability than the less able, which they would not see in case taxation was income-progressive.

Then, we formulate a time process,so that such incentive-consistent manipulable ability change will be a costless increase or decrease, depending, respectively, upon whether a present ability is higher or lower than the reservation ability for the present (optimal) taxation rate. In this sense, the process, determining the future size of ability, is not ad hoc nor stochastic.

We shall see a central limit porperty underlying the resulting distributions, with a set of relevant conditions on the (random) rates of manipulable change. We shall see the law of proportionate effect (pointed out by Gibrat[1931]) hold and extend the result of

Kalecki[1945], which was simply a drawing from an ad hoc assumption. Then, Liapounov's condition is met; see Loeve[1963 p.275], and we conclude the process lead the final distribution to lognormality, if it is infinitely repeated. We also see this determination process hardly converges to a stationary (equilibrium) distribution, as long as such self-destructing moral hazard and hierarchical self-improving incentives motivates a continuous increase in the inequality degree (measured by the variance of ability) in the future size distribution. Theorem 2

In Section 3, we shall establish a comparative statics theorem with a sequence of lognormal ability distributions which differ from each other in their means and variances. The variance of logarithm of ability monotonely increases as the time process evolves. In the case where the functional form of a lognormal density function is given and the standard deviation σ of log ability varies continuously with the mean constant, we prove the social welfare increases with the σ-parameter (which indicates the inequlaity degree of ability), in that, while both the optimal tax rate and the transfer (equal to the tax-revenues) increase, the inequality degree of redistributed income will decrease hence the (quasi-) concave social (aggregate, mean) welfare will increases. Theorem 3

Section 4 is our concluding remarks.

1. Redistributive Taxation Models Revisited with a Simple Utility Hypothesis

Assume that individuals of the society have identical preferences, but differ, to an interesting extent, in their ability producing output for consumption. While the preferences are concavifiable and representible by a linearly homogeneous Cobb-Douglas numerical function, the distribution of the dissimilar abilities is given in a general, unspecified form and fixed within a period though variable over the period.

Let y be an ability, endowned to and revealed by a worker, in the given period. It is intended to imply the maximum labor supply the individual would be able to choose. Thus, y is the maximum output he could add if he spent all his time working. Let L denote a supply that worker (characterized by) y will actually choose. It shows a quantity of output he will add when employed. We are assuming that y and L are nonnegative and measured in the unit of output (consumption good).

Designate by g(y) the density function of nonnegative variate y. g(y) expresses a relative number of workers whose abilities are identical to y. We normalize it so that $\int g(y)dy=1$, where $0<y<\infty$. The density function g is given, fixed and informed of the government who will make a public decision on the taxation policy.

The taxation system with its structure is specified simply by a linear income tax rate t such that $0\leq t\leq 1$. We follow a common premise underlying the redistributive taxation models, such as Mirrlees[1971], Sheshinski[1972], Rader[1983] etc..

That is, tax-revenues,net of a (λ) portion of which, are denoted by T, and, to be paid out to all individuals in equal quantities. The λ portion will be paid out to the government men. This modification, however, won't change any formal analysis and so we shall omit it except footnoting.

Parameterize the taxation system by tax rate t and lumpsum transfer T. Given (t,T), each worker y maximizes the identical utility of consumption c and the ratio r of leasure R=y-L over y,

$$c^{\alpha}r^{1-\alpha};\ 0<\alpha<1$$

subject to ($c\geqq 0$, $0\leqq r\leqq 1$) his full-income budget constraint,

$$c+(1-t)yr\leqq(1-t)y+T \qquad (1)$$

The solution gives the optimal supply L of worker y which is viewed as a function of (1-t)y and T, hence written as L=L{(1-t)y,T}. By the premise, the government's budget constraint holds as the equality*16

$$T=t\int L\{(1-t)y,T\}g(y)dy \qquad (2)$$

If leasure R is normal, which is satisfied by our utility hypothesis, there is a unique value of such T, satisfying (2) for each t. Call this T(t). Then, we can write the solution (c,r), each as a function of (y,t). Here we can specify them in terms of a new variable y(t), defined below. That is,

$$L(y,t)=\max[0,\alpha\{y-y(t)\}] \qquad (3)$$

where the new variable y(t) is associated with a boundary solution i.e. r(y(t),t)=1. From (2) and by taking T=T(t),

$$y(t)=(1-\alpha)T(t)/\alpha(1-t) \qquad (4)$$

Thus, $y(t)<y \leftrightarrow L(y,t)>0$, $y(t)\geqq y \leftrightarrow L(y,t)=0$.

Worker y will accept work at tax rate t if y>y(t) and won't otherwise.

Integration of (3) with respect to y over the continuum (0,1) of workers will bring out the aggregate (mean) labor supply in terms of (t,y(t)). Designate it L(t) for each t. Then,

$$L(t)=(1-t)\alpha\int_{y(t)} yg(y)\,dy/\{(1-t)+(1-\alpha)t\int_{y(t)} g(y)\,dy\} \qquad (5)$$

The lower bound of integration is y(t) defined in (4). Hence y(t) is obtained in an implicit form;

$$y(t)=(1-\alpha)t\int_{y(t)} yg(y)\,dy/\{(1-t)+(1-\alpha)\int_{y(t)} g(y)\,dy\} \qquad (6)$$

This ability,hereinafter I would say <u>reservation ability</u>, plays an essential role throughout the whole sections.

1.1 The Reservation Ability

We are able to see this function y(.) is strictly increasing over [0,1). Note y(0)=0 and y(1)=+∞.

<u>Lemma 1</u>: $y_t(t)>0$ for all t in [0,1).*17

<u>Proof</u>: Differentiate both hand-sides of (6) and rearrange them Then, we get the t-derivative of y(t) in terms of (t,y(t));

$$y_t(t)/y(t)=1/\{(1-t)+(1-\alpha)t\int_{y(t)} g(y)\,dy\}>0 \qquad (7)$$

The reservation ability y(t) defines <u>disincentive tax rate</u> s for each y. Define an inverse function s to be such that y(s) =y in the ability interval [0,∞). Then, s(y) is the least tax rate for which worker y won't work. s(y) strictly increases in y with its domain [0,∞) and range [0,1). Thus, positive correlation and one-to-one correspondence of ability y and disincentive tax rate s are established.

From (3)(4) and Lemma 1, it is immediate that there is no "backward bending" in the labor supply curve often explained in the literature. $L(y,t)$ with $L(y,0)>0$ decreases as t increases in $[0,s(y))$. $L(y,s(y))=L(y,t)=0$ for all $t\geq s(y)$. The mean labor supply $L(t)$ is always positive while it decreases in t in $[0,1)$. All these results are due to our utility hypothesis.

Remark 1: In case, for some numbers, $\underline{y}$ and $\bar{y}$, $0<\underline{y}\leq y\leq\bar{y}$, a corresponding pair of tax rates, $\underline{t}$ and $\bar{t}$, such that $\underline{t}=s(\underline{y})$ and $\bar{t}=s(\bar{y})$, are uniquely determined, respectively, as $\underline{y}/\{\alpha\underline{y}+(1-\alpha)E(y)\}$ and 1. All individuals will work whenever $t\leq\underline{t}$. But they decrease more or less supplies of their labor. Some choose unemloyment whenever $t>\underline{t}$. All will choose unemployment only at 100% tax rate.

1.2 Tax Revenues and Transfer

We easily see from (2)(4)(5) and from Lemma 1 that $T(0)=T(1)=0$, $T(t)>0$. By continuous differentiability, there exists a rate $t°$ such that $T(t°)\geq T(t)$ over $[0,1)$ and $T_t(t°)=0$.

Differentiate $T(t)$ to obtain in terms of $(t,y(t))$,

$$T_t(t)/T(t)=-(1-t)^{-1}+y_t(t)/y(t) \qquad (8)$$

Equate this to 0, then, from (7), we have

$$t°=1/\{1+\sqrt{(1-\alpha)\int_{y(t°)}g(y)dy}\}>1/\{1+\sqrt{(1-\alpha)}\}>1/2 \qquad (9)$$

again in an implicit form.

The uniqueness of such $t°$ is not obtained without a number of constraints. We can easily prove that $T(t)$ is strictly increasing in t in the local neighborhood of $t°$ if and only if $t<1/\{1+\sqrt{(1-\alpha)\int_{y(t)}g(y)dy}\}$. For a Pareto distribution with Pareto index β, $T(t)$ increases with t over $[0,t°)$, where $t°=1/\beta$.

1.3 The Break-Even Ability

Define and designate by τ a net tax rate assigned to a worker y to be such that $(1-\tau)L(y,t)=(1-t)L(y,t)+T(t)$. Then, from (4)-(6), where $\tilde{y}(t)=(1-\alpha t)y(t)/(1-\alpha)t$,

$$\tau(y,t)=t\{y-\tilde{y}(t)\}/\max\{0,y-y(t)\}$$

$$=i(y,t)\{c(y,t)-L(t)\}/c(y,t)\max\{0,y-y(t)\}.$$

A worker y is a tax payer if $\tau>0$, a recipient if $\tau<0$ and neither a payer nor a recipient if $\tau=0$. It is also immediate that $\tau\gtreqless 0$, according to whether $y\gtreqless\tilde{y}(t)$, or equivalently to whether $c(y,t)\gtreqless L(t)$, provided that $y>y(t)$. In case $y\leq y(t)$, $\tau=-\infty$.

We may call ability $\tilde{y}(t)$ the <u>break-even ability</u>, because it is associated with the break-even level income L(t) here, often referred to in the usual <u>negative income-tax</u> discussions; see J. Hirshleifer[1976,80 pp.454-9]. We can go further in this direction by making a good use of y(t).

1.4 The Optimal Tax (Taxation) Rates

To see how welfare is gained or lost by taxation for the policy purpose, we shall use a ratio of welfare from the redistributed income to that from the earned income in case no such policy is taken at all. We then take this as an objective function.

Define, for each (y,t), a <u>relative welfare</u> of worker y, as the ratio, designated by v(y,t), of his/her indirect utility u(y,t) over u(y,0), where $u(y,t)=c(y,t)^{\alpha}r(y,t)^{1-\alpha}$.

Then, we have

$$u(y,t)=(1-t)^{\alpha}\alpha^{\alpha}(1-\alpha)^{-\alpha}y^{\alpha-1}\{(1-\alpha)y+\alpha y(t)\} \quad \text{if } y>y(t) \quad (10)$$

$$=\{T(t)\}^{\alpha} \quad \text{if } y\leqq y(t) \quad (10')$$

Define likewise a relative welfare of the society, designated V(t), for each t, as the ratio of two aggregate (mean) welfares E(u(y,t)) and E(u(y,0)), where

$$E(u(y,t))=(1-t)^{\alpha}\alpha^{\alpha}(1-\alpha)^{-\alpha}[(1-\alpha)\int_{y(t)}y^{\alpha}g(y)dy$$
$$+\alpha y(t)\int_{y(t)}y^{\alpha-1}g(y)dy+\{y(t)\}^{\alpha}\int^{y(t)}g(y)dy] \quad (11)$$

Then, we have at last for each worker y, v(y,0)=1 and

$$v(y,t)=(1-t)^{\alpha}(1-\alpha)^{-1}\{(1-\alpha)y+\alpha y(t)\}/y \quad \text{if } y>y(t) \quad (12)$$

$$=\{(1-t)y(t)/y\}^{\alpha}(1-\alpha)^{-1} \quad \text{if } y\leqq y(t) \quad (12')$$

and, for the society (government), V(0)=1 and

$$V(t)=(1-t)^{\alpha}(1-\alpha)^{-1}\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\}/\tilde{\tilde{y}} \quad (13)$$

where

$$\tilde{\tilde{y}}(t)=\tilde{\tilde{y}}+\{(1-\alpha)\int^{y(t)}\{y(t)^{\alpha}-y^{\alpha}\}g(y)dy$$
$$+\alpha y(t)\int^{y(t)}\{y(t)^{\alpha-1}-y^{\alpha-1}\}g(y)dy\}/(1-\alpha)E(y^{\alpha-1}) \quad (14)$$

and

$$\tilde{\tilde{y}}=E(y^{\alpha})/E(y^{\alpha-1}),\ E(y^{\alpha})=\int y^{\alpha}g(y)dy\ \&\ E(y^{\alpha-1})=\int y^{\alpha-1}g(y)dy.$$

Note that, in the same way as y in (12), $\tilde{\tilde{y}}(t)$ may be interpretable in (13) as <u>ability</u> of an <u>aggregate</u> kind.

It is easy to check the existence of an (individual) optimum rate s* for worker y, which maximizes the ratio v(y,t) for each y, and the (social) optimum rate t* for the government, which maximizes V(t) over the tax interval.

1.4.1 Individual Optimum Rate s*

Differentiate (12) and (12') with respect to t and equate them to 0, then, we have a relation between t and y, characterized by

$$-1/(1-t)+y_t(t)/\{(1-\alpha)y+\alpha y(t)\}=0 \quad \text{if } y>y(t) \qquad (15)$$

$$-1/(1-t)+y_t(t)/y(t)=0 \quad \text{if } y\leqq y(t) \qquad (15')$$

Then, let s* denote such t given y, then, s* is viewed as a function of y in $(0,\infty)$. Alternatively, let y* denote such y given t, then, y* is viewed as a function of t in $[0,1)$.

Lemma 2: $y^*(t)$ is decreasing in t if and only if $y_{tt}(t)/y_t(t) <(1+\alpha)/(1-t)$ for t in $[0,t°)$.

By Lemma 2 we are able to prove $y^*>y(s^*)$ if and only if $s^*<t°$. At $t=t°$, y* is an image of a set-valued function, whereas y* is single-valued at each t in $(0,t°)$. For all $y\leqq y(s^*)$, $s^*=t°$.

s* is not unique, however, for each y, in general. Take a Pareto case for example. $y_{tt}(t)/y_t(t)=-\{(\beta-1)-2\alpha t\}/\beta(1-t)t \lesseqgtr (1+\alpha)/(1-t)$, according to whether $t\lesseqgtr(\beta-1)/\beta(1-\alpha)$. If $\beta\leqq 1+(1-\alpha)$, then, $s^*_y(y)>0$, $(\beta-1)/\beta(1-\alpha)<s^*\leqq 1/\beta$, where β is the Pareto index $\beta>1$.

1.4.2 Social Optimum Rate t*

Differentiate (13) and equate it to 0 to obtain for t*

$$\alpha/(1-t)=\{(1-\alpha)\tilde{\tilde{y}}_t(t)+\alpha y_t(t)\}/\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\} \qquad (16)$$

where

$$-\tilde{\tilde{y}}_t(t)=\{\alpha/(1-\alpha)E(y^{\alpha-1})\}y_t(t)\int^{y(t)}\{y^{\alpha-1}-y(t)^{\alpha-1}\}g(y)\,dy\geqq 0 \qquad (17)$$

An optimum tax rate must satisfy this. Recall this equation is only a necessary condition for t* to meet. From (17) $\tilde{\tilde{y}}_t(t)<0$ in (0,1), meaning that $\tilde{\tilde{y}}(t)<\tilde{\tilde{y}}$ there.

Taking y*(t), characterized by (15)and(15'), y(t), characterized by (14) and (17), y(t) characterized in Lemma 1, all into account of (16), we have, as what such t*, $0\leqq t^*$, must satisfy,

$$-\tilde{\tilde{y}}_t(t^*)/y_t(t^*)=\alpha\{y^*(t^*)-\tilde{\tilde{y}}(t^*)\}/\{(1-\alpha)y^*(t^*)+\alpha\tilde{\tilde{y}}(t^*)\}\geqq 0 \tag{18}$$

The equality holds only when t*=0. This will be eliminated by the assumption that $\tilde{\tilde{y}}<E(y)$.

We have here a main result.

Lemma 3: There exists an optimum tax rate t*, such that t*>0 and $y^*(t^*)>\tilde{\tilde{y}}(t^*)$, provided that $\tilde{\tilde{y}}<E(y)$.*18
The proof will be dealt with in Appendix A. The assumption is met in the cases including Pareto, lognormal distributions, etc..

The social optimum rate t* hardly seems to be unique. This, however, won't be a high hurdle for the following analysis.

1.5 The Policy Effect on Income and Consumption

We lastly work out a little detail of the consequeneces of the taxation policy especially on income and consumption. We also summarizes the main results so far obtained in a theorem.

To each worker y, the policy (t,T(t)) means after all a vector-valued function mapping each (y,t) into {c(y,t),r(y,t)}, where

$$c(y,t)=\alpha(1-t)R(y,t)/(1-\alpha) \qquad \text{if } y>y(t) \tag{19}$$

$$c(y,t)=\alpha(1-t)y(t)/(1-\alpha) \qquad \text{if } y\leqq y(t) \qquad (19')$$

$$r(y,t)=R(y,t)/y, \quad R(y,t)=\{(1-\alpha)y+\alpha y(t)\}$$

and

$$i(y,t)=(1-t)\{(1-\alpha)y+\alpha y(t)\}/(1-\alpha) \qquad (20)$$

Let δ_i^2 and δ_c^2 denote the variances of i(y,t) and c(y,t), respectively, for each t. Write the variance of y, δ^2. It is immediate that

$$\delta_i^2=(1-t)^2\delta^2 \qquad (21)$$

when the government implements the rate t.

The full incomes of all workers are distributed, in the same way as their abilities, but, with the variance smaller by a multiple $(1-t)^2$. The distribution of consumptions, c(y,t)'s, would be trancated with the point of trancation, c(y(t),t)=T(t). They would be confined to the range above the minimum T(t) as if the rest of the distribution, for which $c\leqq c(y(t),t)$, were removed. For all $y<y(t)$, c(y,t)=c(y(t),t).

For consumption, it is also immediate that

$$\delta_c^2=\alpha^2(1-t)^2\int\{y-E(y)-\int^{y(t)}\{y(t)-y\}g(y)dy\}^2g(y)dy$$
$$\leqq\alpha^2(1-t)^2\delta^2 \qquad (22)$$

where the equality holds only when t=0. $\alpha^2\delta^2$ is the variance of L(y,0).

The policy makes the variance of consumption much smaller That is, the variance becomes smaller than that of the incomes L(y,0) earned by the workers themselves in case of no taxation, by more than a multiple $(1-t)^2$.

Thus, we have established, for an often used class of distributions, which each satisfy the provision $E(y)>\tilde{\tilde{y}}$ of Lemma 3:

Theorem 1 (Optimal Tax Rate): There exists a strictly positive social optimum tax rate t* for each fixed (density function of) distribution of ability,

(i) for which a relative number of workers, given by

$$\int^{y(t^*)} g(y)dy > 0$$

where the reservation ability $y(t^*)$ is strictly positive and unique, will choose a voluntary unemployment (contract),

(ii) for which the returned full-incomes, $i(y,t^*)$, will be in the same way distributed as the endowned abilities of workers, y, but with a variance δ_i smaller by a scalar $(1-t^*)^2$ so that $\delta_i^2=(1-t^*)^2\delta^2$ for the variance δ^2 of ability, and

(iii) for which the consumption incomes, $c(y,t^*)$, will be trancatedly distributed with the point of trancation, $c(y(t^*),t^*)$, which is equal to the transfer $T(t^*)$, and with a much smaller variance than that of the earned incomes (equal to $L(y,0)$) in case of no policy at all.

Given a functional form g, those variables, such as $y(t)$, $y^*(t)$, $y(t)$ and t* etc. which have been designed under the transfer-taxation policy are obtained as the subjective and market equilibrium solutions.

2. Ability and Income Determination Process

Other than the policy effect on the incomes of all workers, one may point out an adverse effect which would tend to make their future abilities distributed more unevenly in a positively

We show first that every worker see, in a different way, an incentive to change his/her present ability relatively to the reservation ability for the present taxation rate. Then, if all workers attempt to engage in such incentive-consistent activities freely, there will be a process to be set up into motion, so that the resulting distributions would be unstable and entail with it no stationary distribution.

2.1 Ex-Post Effect of the Taxation Policy

Note, for each taxation rate t, utility level u(y,t) achieved by a working individual y, is precisely in one and the same order with ability. However what is achieved by each of not working individuals is identical to an α power of transfer, hence totally independent of the difference in their present abilities. There would be neither welfare gain from improving, nor loss from not sustaining, the present abilities, as long as they continue to choose unemployment.

Moreover, the abilities of the unemployed workers have been perfectly unobserved (private), as well as out of on-the-job training process.

However, this is not the case with a working and tax-paying worker, particularly, with such a higher-ability (as is located in the upper portion of the distribution). The welfare gain by ability improvement (on on-the-job training or screening) is more likely to be large and easy to obtain for each such individual.

Formally, the ability-elasticity, $\varepsilon(y,t)$, of welfare $u(y,t)$ is strictly positive whenever $y>y(t)$, while zero whenever $y\leqq y(t)$

$$\varepsilon(y,t) \begin{cases} =0 & \text{if } y\leqq y(t) \\ =(1-\alpha)\alpha\{y-y(t)\}>0 & \text{if } y>y(t) \end{cases} \tag{23}$$

Thus, $\varepsilon(y,t)$ increases with y and has a range $(0,\infty)$.
This shows that, as y increases, the utl ty level of a worker y increases increasingly with y, while the levels of the unemployed stay constant.

The ability-elasticity, $\upsilon(y,t)$, of the relative utility gain $v(y,t)>1$, or loss $v(y,t)<1$, is negative for every y:

$$\begin{aligned} \upsilon(y,t) &= -\alpha & \text{if } y\leqq y(t) \\ \upsilon(y,t) &= -\alpha/\{\alpha+(1-\alpha)y/y(t)\} & \text{if } y>y(t) \end{aligned} \tag{23'}$$

Thus, $\upsilon(y,t)$ increases with y and has a range $[-\alpha, 0)$ if $y>y(t)$, while constant if $y\leqq y(t)$. The gain decreases at a constant elasticity for every unemployed and at a less elasiticity for each $y<E(y)$. The loss increases with y but at a much less elasticity for every $y\geqq E(y)$.

Note this would'nt be so if taxation was income-progressive. Or, this is the more so because of the linearity of income-tax.

Thus, considering, in particular in (iii), the results in Section 1, we conclude thereof as follows:

(i) For each of the unemployed who won't accept work for the present tax rate t, there is no incentive to improve or sustain the present, now observed only privately and out of the process of on-the-job training or screening, abilities. This incentive is more likely to see whenever he/she expects a continual or continuous increase (non-decrease) in the tax rate in the future.

(ii) Every worker, accepting work, sees an incentive, more or less, to improve upon his/her ability on the work. The strongness of such an incentive is to a different extent and increases with the difference of present ability larger than the reservation ability.

(iii) For each increase in taxation rate, there will be the newly unemployed and the more workers who will stay unemployed. For a larger number of workers, each with ability low or close relatively to the rising reservation ability, the higher taxation and benefit will create substitution effects in favor of greater density (population) of lower abilities and longer duration periods of unemployment.

2.2 A Formulation of Dynamic Process

A relevant formulation, which takes the above three points into account, may be as follows:

$$\log y_1/y_0=\lambda_1 \log y_0/y_0(t)+\nu_1, \quad \lambda_1 \geqq 0 \qquad (24)$$

where $y_0(t)$ is the reservation ability determined at the 0 th (present) period for the social optimum tax rate $t=t_0$, and ν_1 is independent of y_0 (unexplained term).*19

Note the asymmetry between point (i) and point (ii) has been taken into account by formulating the process in a logarithmic form.

(24) is equivalent to

$$y_1 = y_0^{1+\lambda_1} y_0(t)^{-\lambda_1} \exp \nu_1. \qquad (25)$$

If $\lambda_1 > 0$ and $\nu_1 = 0$, then y_1 is determined, respectively, to be larger equal or smaller than y_0, depending upon whether y_0 is larger, equal or smaller than $y_0(t)$, by a portion $|\{y_0/y(t)\}^{\lambda_1} - 1|$.

2.2.1 Central Limit Theorems

A remarkable result from this process is an application of the well-known central limit theorem, here implying the generating distribution approaching approximately to the Gibrat law under a number of limitations on λ_i and ν_i.

In (24), take $y_i = y_0$, $y_{i+1} = y_1$, $i=1,2,..,n$

$$\log y_i/y_{i-1} = \lambda_i \log y_{i-1}/y_{i-1}(t_{i-1}) + \nu_i \qquad (26)$$

If λ_i and ν_i, $i=1,...,n$ all are independent of each other and of y_{i-1}, and so small that $\lambda_i |\log y_{i-1}/y_{i-1}(t)| + |\nu_i| < 1$. Then we can assume the standard deviation of $\Sigma_{i=1}^{n} \log y_i/y_{i-1}^{*}$, $i=1,..,n$, are equal to or larger than 1, as far as the standard deviation of $\log y_n/y_{n-1}$ does not fall below a certain level, because, by (26) $|\log y_i/y_{i-1}|$, $i=1,..,n$ are small as compared with 1. Thus, this assumption allows us to take $\log y_i/y_{i-1}$ small as compared with the standard deviation of the sum. Suppose further if, for n large enough, the standard deviation of the sum is large as compared with that of $\log y_0$, then, Liapounov's condition will be met and his theorem will be applied to obtain that $\log y_n$ is approximately lognormal as $n\to\infty$

Let us denote the third moment about mean by ρ_i^3, for each $i=0,1,..,n$, where $\mu_i=E|\log y_i|$, $\sigma_i^2=E(\log y_i-\mu_i)^2$ and

$$\rho_i^3=E|\log y_i-\mu_i|^3 \qquad (27)$$

Then, the above assumptions imply that, as $n\to\infty$,

$\rho_n^3/\sigma_n^3\to 0$ Liapounov's condition

Apply Liapounov's theorem (Loeve[1963 pp.275-277] for example) to obtain, as $n\to\infty$, $y_n\to G(\mu_n,\sigma_n^2)$

In stead of the previous assumption for λ_i and ν_i in (26) suppose

$\lambda_i<0$, $i=1,..,n$.

Then the process will lead the resuting distribution to the lognormality, although this supposition is inconsistent with our observation (i)(ii) and (iii) in subsection 2.1. It is, in fact, an <u>ad hoc</u> assumption[*20] made by Kalecki[1945] for the variate y indicating a personal income determined by economic forces.

Then, there are Λ_i, $i=1,..,n$, such that

$\Lambda_i<0$, and $\Pi_{h=i+1}^{n}(1+\lambda_h)=\Pi_{h=i+1}^{n}(1+\Lambda_h)\sigma_n/\sigma_i$.

Hence by (26), we have

$$\log y_n=\{(\log y_0)/\sigma_0\}\Pi_{i=1}^{n}(1+\Lambda_i)\sigma_n$$
$$+\Sigma_{i=1}^{n}(\nu_i/\sigma_i)\Pi_{h=i+1}^{n}(1+\Lambda_h)\}\sigma_n$$
$$-\Sigma_{i=1}^{n}[\{(\lambda_i/\sigma_i)\log y_{i-1}(t_{i-1})\}\Pi_{h=i+1}^{n}(1+\Lambda_h)]\sigma_n \quad (28)$$

If $0<1+\Lambda_i<1$ $(0<1+\lambda_i<1)$ for each i, and if $|\nu_i|$ and $|\lambda_i|\log y_{i-1}(t_{i-1})$, i=1,..,n, are small as compared with σ_i, then, it follows that the distribution of $\log y_n$ is approximately normal.

Note here that the proof may contain a special case in which the standard deviation is constant, i.e. $\sigma_i=\sigma, i=1,2,..,n$.

2.2.2 Stationary Distribution

Suppose y_0 is $G(\mu,\sigma^2)$, then, by the simple reporductive property, it is immediate that y_1 is $G(\kappa_{1,0}+(1+\lambda_1)\mu,(1+\lambda_1)^2\sigma^2)$ where $\kappa_{1,0}(t)=\nu_1-\lambda_1\log y_0(t)$ and t is t_0 determined and implemented at 0 the period as an optimum rate t*. The new distribution function is given by a lognormal formula of the Chapman-Kolmogorov equation (see Cox and Miller[1965 p.205]):

$$G(y_1;\kappa_{1,0}+(1+\lambda_1)\mu,(1+\lambda_1)^2\sigma^2)$$
$$=\int G(y_1/y_0;\kappa_{1,0}+\lambda_1\mu,\lambda_1^2\sigma^2)dG(y_0;\mu,\sigma^2) \quad (29)$$

This reproductive property extends to any finite set of independent lognormal variates and also to an infinite sequence, provided that that some conditions of convergence are fulfilled. In general it follows that

$$\Pi_{i=1}^{n}(y_{i-1})^{\lambda_i}$$

$$G(\Sigma_{i=1}^{n}\lambda_i\mu_i, \Sigma_{i=1}^{n}\lambda_i^2\sigma_i^2)$$

if y_i is $G(\mu_i,\sigma_i)$ for each i.

(See, for example, Aitchison and Brown[1975 p.11] for this.)

In (24) take $y_i = y_0$, $y_{i+1} = y_1$, and take $\mu_i = \mu$ and $\sigma_i = \sigma$ for each i. Then, provided that both of

$$\Pi_{i=1}^{n}(1+\lambda_i) \text{ and } \Pi_{i=1}^{n}(1+\lambda_i)^2 \tag{30}$$

converge as $n\to\infty$, it follows that

$$y_0^{\Pi_{i=1}^{n}(1+\lambda_i)} \tag{31}$$

is $G(\mu_n^{\dagger}, \sigma_n^2)$,

where $\mu_n^{\dagger} = \Pi_{i=1}^{n}(1+\lambda_i)\mu$ and $\sigma_n^2 = \Pi_{i=1}^{n}(1+\lambda_i)^2\sigma^2$.

Let $y_n^{\dagger}$ denote (31). Then, as both of (30) converge as $n\to\infty$,

$$y_n = y_n^{\dagger}\ \Pi_{i=1}^{n}\{y_{i-1}(t_{i-1})\}^{-\lambda_i\Pi_{h=i+1}^{n}(1+\lambda_h)}\exp\{\Sigma_{i=1}^{n}\nu_i\Pi_{h=i+1}^{n}(1+\lambda_h)\}$$

is also $G(\mu_n,\sigma_n)$, where

$$\mu_n = \Sigma_{i=1}^{n}\kappa_{i,i-1}(t_{i-1}) + \mu_n^{\dagger}$$

provided that, in addition,

$$\Sigma_{i=1}^{n}\ \nu_i \tag{32}$$

converges, as $n\to\infty$.

To see this, it sufices to show that $y_n/y_n^{\dagger}$ converges, under the conditions that all of (30) and (32) converge.
By Theorem 1 (lemma 3), there exists an optimal tax rate t* determined for the i th period within which (μ_i,σ_i) is given.

If the density function $g_i(y)$ is lognormal so that

$$g_i(y_i)=(\sqrt{2\pi}\sigma_i y_i)^{-1}\exp\{\log y_i-\mu_i\}^2/2\sigma_i^2$$

then, clearly it is continuous in (μ,σ) in any compact local neighborhood of (μ_i,σ_i). Since $y(.)$ is continuous in t in $[0,1)$ for each (μ_i,σ_i) and so is in (μ,σ) for each t, it follows that $y(t^*,\mu,\sigma)$ is continuous in (μ,σ). It is positive and bounded, since $0<t^*<t°<1$ by Lemma 3 for each (μ_i,σ_i) and $0<y(t^*)<y(t°)<\infty$ by Lemma 1. Hence,

$$\Sigma_{i=1}^{n}\{y_{i-1}(t_{i-1})\}^{-\lambda_i\Pi_{h=i+1}^{n}(1+\lambda_h)}$$

is bounded, if $\Sigma_{i=1}^{n}(1+\lambda_i)$ converges, because λ_i must become 0 so that each term $y_{i-1}(t_{i-1})^{-\lambda_i\overline{\Sigma_{h=i+1}^{n}(1+\lambda_h)}}=1$, $i>N$ for some finite number N. Therefore, we will complete if we show

$$\Sigma_{i=1}^{n}\nu_h\Pi_{h=i+1}^{n}(1+\lambda_h)$$

converges. In fact it does by almost the same reason under the convergence conditions.

Thus, y_n asymptotically becomes y^*, where
y^* is $G(\mu^*,\sigma^{*2})$, $y^*=\lim_{n\to\infty}y_n$, $\mu^*=\lim_{n\to\infty}\mu_n$ and $\sigma^*=\lim_{n\to\infty}\sigma_n$.

Remark 2: Assume that λ_i continuously depends upon $y_{i-1}(t)$. Then, since $y_{i-1}(.)$ varies continuously in the tax interval $[0,t°]$, it follows that λ_i also varies in the interval. Hence $\Pi_{i=1}^{n}\{1+\lambda_i(y_{i-1}(t)\}$ uniformly converges, if $\Sigma_{i=1}^{n}|\lambda_i|$ converges uniformly, and continuous in the interval. This gives the continuous variation of y in the process which evolves.

Theorem 2: (Dynamic Process) A dynamic process is set up to generate the size distributions of ability and income for the next period, such as follows:

(i) Under a set of fairly general conditions on the manipulability, the resulting size distribution will be approximately lognormal, if the process is infinitely many times repeated

(Central Limit Theorem)

(ii) Under some convergence conditions, the size distributions will entail with them a stationary (equilibrium) distribution to which they converge, if they are initially lognormal.

(Stationary Distribution)

The convergence conditions hardly seem to be fulfilled, if all workers continue to engage in such costless self-destructing moral hazard and self-improving activities.

(iii) The inequality degree of workers' ability, measured by the variance, monotonely increases as the process evolves.

3. Comparative Statics

In order to work out a much detailed investigation concerning how each key variable varies over periods of time in which structural changes in g is possible to arise, we shall exploit the density function of a lognormal distribution that has just been justified in Section 2 for our analysis.

3.1. Lognormal Distribution Function

We consider an essentially positive random variable y, such that $0<y<\infty$, $\log y$ is normally distributed with mean μ and variance σ^2. Let us write the lognormal distribution function $N(\log y;\mu,\sigma^2)$ of $\log y$ as $G(y,\mu,\sigma^2)$. Then, the density function $g(y;\mu;\sigma^2)$ of y is given so that

$$g(y;\mu,\sigma^2)dy=dG(y;\mu,\sigma^2)=dN(\log y;\mu,\sigma^2)$$

$$=(\sqrt{2\pi}\sigma y)^{-1}\exp-\{(\log y-\mu)/\sigma\}^2/2\ dy$$

We write often that y is $G(\mu,\sigma^2)$ when y is distributed with a distribution function $G(y;\mu,\sigma^2)$, in other words, when y is lognormally distributed with mean μ and variance σ^2.

A (continuous) change in the structure of the lognormal distribution is given and represented by a continuous change in the σ parameter. We shall show how this change will produce effects on the key variables, $y(t)$, $\tilde{\tilde{y}}(t)$, $y^*(t)$ and t^*, and the others.

For the density function $g(y;\mu,\sigma^2)$, our key variables; $y(t)$, $\tilde{\tilde{y}}(t)$, $y^*(t)$, t^* and the others, all are continuously differentiable functions of the σ parameter. Write $y(t)$ as $y(t,\sigma)$, etc.. Then, we write their first derivatives w.r.t. σ as $y_\sigma(t)$ etc.. Likewise we write the σ-derivative of g as g_σ, where $g=g(y;\mu,\sigma^2)$, etc..

3.2 Comparative Statics Results and Theorem

The results will be formally summarized in Lemma 4 through 8 and rigorous proofs for them will be provided in Appendix B.

Then we have established:

Lemma 4: $y_\sigma(t,\sigma)/y(t,\sigma) \geqq \sigma$ for each $t \varepsilon (0,1)$.

The reservation ability $y(t)$ shifts continuously upward as σ increases in the t,y plane. Similarly, $T_\sigma(t)/T(t) \geqq \sigma$.

Lemma 5: Suppose $\alpha \leqq 1/2$. Then, $\tilde{\tilde{y}}_\sigma(t,\sigma) \leqq 0$ with its range $(0,-\infty)$.

$\tilde{\tilde{y}}(t)$ shifts downward as σ increases. In case $\alpha > 1/2$, there is an interval $(0,t)$ for some $t>0$, $\tilde{\tilde{y}}_\sigma(t)>0$ while $\tilde{\tilde{y}}_\sigma(t)<0$ otherwise. This, however, won't be any obstacle to our following analysis.

Lemma 6: $y^*_\sigma(t,\sigma)>0$ for each t in $(0,t°)$.

The inverse $y^*(t)$ of individual optimum tax rate $s^*(y)$, that is, the ability of a worker who regards t as his optimum tax rate, also shifts upwards in the t,y plane, as σ increases.

Lemma 7: $V_\sigma(t)/V(t)>0$ for each t in $(0,t°)$

This is somewhat surprising. We consider this later.

By the well-known envelope property, we immediately have,

Lemma 8: $\partial V(t^*)/\partial\sigma = V_\sigma(t^*)>0$.

Similarly, we also have $\partial T(t^*)/\partial\sigma = T_\sigma(t^*)>0$.

Lastly, we prove our main result here; the positivity of the σ-derivative of a solution to (17) and (18):

Lemma 9: $t^*_\sigma > 0$ if $(1+\alpha)/(1-\alpha) > \partial\{y^*(t^*)/y(t^*)\}/\partial t$.

The condition of Lemma 2 is sufficient for the condition of Lemma 9, if it holds only at t*. This excludes the possibility of a local minimum.

We have thus established a comparative statics theorem under a mild condition:

Theorem 3: (Comparative Statics)

(i) For a class of lognormal ability distributions, each having a density function $g(y;\mu,\sigma^2)$ and differing from other in mean μ and variance σ^2; any tax rate t*, which is determined as a local optimum rate for each (μ,σ^2), will, with μ constant, increase continuously with the σ-parameter.

(ii) The social (aggregate, mean) welfare V(t*) increases continuously with the σ-parameter, in that the inequality degree of the redistributed income measured by $(1-t^*)^2\delta^2$ decrease, hence the (quasi-)concave social welfare V(t*) increases.

4. For Concluding Remarks

In a mixed economy of individuals of different abilities and identical (homothetis) preferences, where their earnings are redistributed by a linear income-tax and transfer policy, we have seen two economic forces working contradictorily in determining present and future distribution structures. The one is the income-equalizing effect which is appropriate to take this policy. The other an adverse effect which will determine the future sizes of their abilities more unevenly in a distorted way.

The income-insurance purpose will motivate a self-destructing moral hazard to the lower-ability workers. The linear income-tax will allow the higher-ability workers to see the more self-improving incentives to increase the taxable earnings. Thus, a dynamic process is set up and put into motion.

However, the social welfare increases as the time process evolves. While the inequality degree of ability, the optimal tax rate and the tax revenues hence transfer, all increase, the inequality degrees of both income and consumption decrease hence the quasi-concave social welfare increases, as the time process is repeated infinitely many times.

** Although this work is still preliminary and incomplete, I should acknowledge the discussions I had about this subject with Professors Hirofumi Uzawa, Yoshihiko Otani, Takatoshi and Shigeru Matuskawa, among the others.

REFERENCES

Aitchison,J.and J.A.C.Brown: The Lognormal Distribution with special reference to its uses in economics Cambridge University Press Cambridge, New York London Melbourne 1976

Blinder,Alan S.: Toward an Economic Theory of Income Distribution M.I.T. Press, Cambridge,Massachusetts and London, England 1977

Champernowne,D.G.."A Model of Income Distributions" Economic Journal 63(1953) 318-351

Diamond,P.,Hems,L.and J.Mirrlees:"Optimal Taxation in a Stochastic Economy" Journal of Public Economics 14(1980) 1-29

Dvoretzky,Aryeh: "Asymptotic Normality For Sums of Dependent Random Variables" Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability Vol.II Probability Theory ed.by L.M.Lecam,J.Neyman and E.L.Scott. 1972 University of California Press. Berkeley and Los Angeles. 513-535

Easley,D.,Kiefer,N.M. and U.Possin:"An Equilibrium Analysis of Unemployment Insurance and Taxation" Quarterly Journal of Economics 100(1986) 989-1010

Friedman,M.:"Choice, Chance and the Personal Distirbution of Income" Journal of Political Economy 61(1953) 273-290.

Gibrat,Robert:"On Economic Inequalities" International Economic Papers 7(1931) 53-70

Hirshleifer,Jack:"Price Theory and Applications Prentice-Hall Inc. Englewood Cliffs. New Jersey 1976,1980.

Kalecki,Michael:"On the Gibrat Distribution" Econometrica 13(1945) 161-

Loeve,M.: Probability Theory Chapter VI. Van Nostrand Reinhold Company, New York, Cincinnati Tront, Rondon, Melbourne, 1963

Lillard,L.A. and R.J.Willis:"Dynamic Aspects of Earning Mobility" Econometrica 46(1978) 985-1012

Mandelbrot,Benoit:"Stable Paretian Random Functions and the Multiplicative Variation of Income" Econometrica 29(1961) 517-543

Mincer,J.:"The Distribution of Labor Incomes:A Survay with Special Reference to the Human Capital Approach" Journal of Economic Literature 8(1970) 1-26

Mirrlees,James A.:"An Exploitation in the Theory of Optimal Income Taxation" Review of Econimic Studies 38(1971) 175-208

________:"Notes on Welfare Economics,Information and Uncertainty" in Essays on Economic Behavior under Uncertainty ed.by M.S. Balch,D.L.McFadden and S.Y.Wu. North-Holland Publishing Company Amsterdam Oxford, American Elsevier Publishing Comaany, Inc. New York 1974

________:"The Optimal Structure of Incentives and Authority within an Organization" Bell Journal of Economics 7(1976) 105-131

Miyazaki,Hajime:"The Rat Race and Internal Labor Market" Bell Journal of Economics 8(1977) 394-418

Newnan,Peter and J.N.Wolfe:"A Model for the Long-Run Theory of Value" Review of Economic Studies 29(1961) 51-61

Rader,John Trout:"The Cost of Redistributive Taxation" Mimeographed University of Washington 1983

Riley,John G.:"Informational Equilibrium" Econometrica 47(1979) 331-359

Rothschild,Mickael and Joseph Stiglitz: "Equilibrium in Comepetitive Insurance Markets: An Essay on the Economics of Imperfect Information" Quarterly Journal of Economics 90(1976) 629-649

Roy,A.D.:"Some Thoughts on the Distribution of Earnings" Oxford Economic Papers 3(1951) 135-146

Sargan,S.D:"The Distribution of Wealth" Econometrica 25(1961) 568-590

Sharrocks,A.F.:"On Stochastic Models of Size Distributions" Review of Economic Studies 42(1975)

Sheshinski,Eytan:"The Optimal Linear Income Tax" Review of Economic Studies 39(1972) 297-302

Stiglitz,Joseph E.:"Distributions of Income and Wealth among Individuals" Econometrica 39(1969)

_________: "Incentives and Risk Sharing in Sharecropping" Review of Economic Studies 41(1974) 219-255

_________: "The Theory of"Screening", Education and Distribution of Income" American Economic Review 75(1975) 283-300

Tinbergen,Jan:"Welfare Economics and Income Distribution" American Economic Review Papers and Proceedings 47(1957) 490-503

_________:"A Positive an. a Normative Theory of Income Distribution" Review of Income and Wealth 16(1971) 221-234

The Wall Street Journal:"The Redistributionalist Tax Reduction" June 24 1984

"Saoking the Rich Through Tax Cuts" March 21 1985

FOOTNOTES

*1 As a result, he has shown that the size distribution of wealth will asymptotically approach to the Pareto Law..

*2 See Lillard and Willis[1978] for a most recent contribution relevant here. They deal with life cycle earnings and mobility among earnings classes, by proposing a new econometric approach and comparing it to Markovian models.

*3 They are perfectly and privately informed about their own ability.

*4 The markets for output (consumption goods)is assumed to be clear at the equilibrium price 1.

*5 Mirrlees pointed out the cases where such a linear tax system is approximately optimal[1971]. However, we shall take this assumption for our own reason. We shall see a linear income taxation will give more incentives for the more able workers to increase the taxable incomes.

*6 This process is implicit. The knowledge about the characteristics about ability and preferneces of each worker will be obtained indirectly only through the quantity of contracted labor supply (wage) (and consumption demand).

*7 Quoted from Rothschild and Stiglitz[1976 p.632].

*8 Supervision contains a general education, as well as the juridical aspect of supervising. The cost of educational training or screening may thus be borne by the authority, at least in part.

*9 Formulations that will take account of technological uncertainty are mathematically very complicated, even if they are the simplest possible ones. The difficulty lies with the problems themselves.

See Mirrlees[1974] for such a general and fundamental formulation. Also see Stiglitz[1974] for a sharecropper who takes risk to bear crop failure, Diamond,Hems and Mirrlees[1980] and most recently see Easelay,Kiefer and Possin[1986].

*10 See Stiglitz[1975] for a seminal but general discussion on the theory of srceening.

*11 in which performance-equivalent price or wage searching agents take account of possible reactions by other agenes. The per-hour performance an agent makes is formulated as an increasing function of unobservable characteristics and the observable sales-activity (effort) intensity which operates as a signal to potential buyers (employers). Hence, the more able, being on the higher intensity assembly line, will be able to separate themselves from the less able.

*12 The process of learning-by-doing is the process of screening the workers' quality on the job at the same time.

*13 The utilities are assumed to be identical, homothetic and representable by a Cobb-Douglas function.

*14 within the contracting period.

*15 Also larger than what would be gained from an income-progressive taxation. For example, see two interesting articles of the Wall Street Journal[1984,5], for a leveraged effect on the incentives of the high-income tax-payers to increase the taxable incomes hence upper-income earners, paying a greater share of the tax burden afetr the tax rate cut. I owe this refernec to Yoshihiko Otani.

*16 If the λ portion of tax revenues are paid to G men, then, the eqaution (2) should read ;

$$T=(1-\lambda)t\int L\{(1-t)y,T\}g(y)dy \qquad (2')$$

Consequently, the equations (5) for $L(t)$, (6) for $y(t)$, (7) for $y_t(t)$ and (9) for $t°$ will be subject to a minor change, that is, $(1-\lambda)(1-\alpha)$ instead of $(1-\alpha)$, whi e the others won't be subject to any.

*17 By L'Hospital's rule, $\lim_{t\to 1} y_t(t)/y(t)=+\infty$.

*18 In view of Remark 1, if there is a lower bound $\underline{y}$ for y, the condition should be replaced by a general one, that is,

$$\tilde{\bar{y}} < E(y)-[\{\alpha\underline{y}/(1-\alpha)\}+1]\underline{y}$$

Taking $\underline{y}\to 0$, this will become $\tilde{\bar{y}}<E(y)$. That $\underline{y}\to 0$ holds will be shown in Section 2 but implicitly.

*19 This exogenous terms may include the returns of educational training or screening which are randomized by technological innovations or discoveries.

*20 Another applicable ad hoc assumption, for which Central Limit theorem will hold, is that the mean of the sums of log y_i depends linearly upon the sums of the previous terms. See for this Dvoretzky [1972, p.532].

APPENDIX A

Let us denote by $\zeta^1(t)$ the R.H.S.of (18), by $\zeta^2(t)$ the R.H.S., devided by $y_t(t)$, of (17) and by $\zeta^3(t)$ the following; $-\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\}/\{(1-\alpha)y^*(t)+\alpha y(t)\}$. Then, (18), which takes (17) into account, may become $\zeta^1(t)=\zeta^2(t)$ or $(1-\alpha)\zeta^2(t)=\alpha\{1+\zeta^3(t)\}$. The optimal tax rate t* is one of the solutions to this.

Proof of Lemma 3: In view of Remark 1, we show $t^*>\underline{t}$, if $\tilde{\tilde{y}}<E(y)-\{\alpha\underline{y}/(1-\alpha)E(y)+1\}\underline{y}$. Note $\zeta^1(\underline{t})>0$ by assumption, while $\zeta^2(\underline{t})=0$ $\zeta^2(t^\circ)>0$. This implies $t^*>\underline{t}$, whatever $\zeta^1(t^\circ)$ may take. Taking $\underline{y}\to 0$ will lead to Lemma 3.

APPENDIX B (Lognormal Distribution)

Proof of Lemma 4: Differentiate both hand sides of (6) w.r.t. σ we have for each t, and for $g(y)=(\sqrt{2\pi}\sigma y)^{-1}\exp-\{(\log y-\mu)/\sigma\}^2/2$,

$$y_\sigma(t)/y(t)=[\sigma E(y)\int^{y(t)}\{y(t)-y\}g_\sigma dy]/\int_{y(t)}yg(y)dy \qquad \text{(B1)}$$

Change variables so that $x=\log y-\mu$, and, $x(t)=\log y(t)-\mu$. Then,

$$y_\sigma(t)/y(t)=\sigma+\exp[-\{x(t)-\sigma\}^2/2]/\int_{x(t)}\exp\{-(x-\sigma)^2/2\}dx>0 \qquad \text{(B2)}$$

The 2nd term of the R.H.S. of (B2) is clearly positive and increases with t with its range $[0,\infty)$, because, if we denote it by $\Phi(t)$,

$$\text{sign }\Phi_t(t)=\text{sign}[\int_{z(t)}\{z-z(t)\}\exp-z^2/2\,dz\,/\int_{z(t)}\exp-z^2/2\,dz]>0$$

where $z=x-\sigma$, and, $z(t)=x(t)-\sigma$. Thus, $y_{\sigma t}(t)>0$.

Proof of Lemma 5: Let

$$\eta^0(t)=\tilde{\tilde{y}}_\sigma+\{\tilde{\tilde{y}}-\tilde{\tilde{y}}(t)\}(1-\alpha)^2\sigma,\ \eta^1(t)=[(1-\alpha)\int^{y(t)}\{y(t)^\alpha-y^\alpha\}g_\sigma dy$$
$$+\alpha y(t)\int^{y(t)}\{y(t)^{\alpha-1}-y^{\alpha-1}\}g_\sigma dy]/(1-\alpha)E(y^{\alpha-1}),\ \eta^2(t)=\alpha y_\sigma(t)\int^{y(t)}\{y(t)^{\alpha-1}$$
$$-y^{\alpha-1}\}g(y)/(1-\alpha)E(y^{\alpha-1}).$$ Then, $\tilde{\tilde{y}}_\sigma(t)=\eta^0(t)+\eta^1(t)+\eta^2(t)$.

Note first $\tilde{\tilde{y}}_\sigma(0)=\eta^0(0)=-(1-2\alpha)\sigma\leqq 0$ by assumption. $\tilde{\tilde{y}}_\sigma(1)=-\infty$, and $\eta_t^0(t)+\eta_t^1(t)+\eta_t^2(t)=0$ for $t=0,1$. We have the t-derivative of $\tilde{\tilde{y}}_\sigma(t)$ in what follows: $(1-\alpha)E(y)\tilde{\tilde{y}}_{\sigma t}(t)=\eta_t^1(t)+$

$$\alpha\{y_{\sigma t}(t)-y_t(t)(1-\alpha)^2\sigma\}\int^{y(t)}\{y(t)^{\alpha-1}-y^{\alpha-1}\}g\,dy$$

$$-y_t(t)y_\sigma(t)\int^{y(t)}(1-\alpha)y(t)^{\alpha-2}g(y)\,dy.$$

$\eta_t^1(t)=\alpha y_t(t)\int^{y(t)}\{y(t)^{\alpha-1}-y^{\alpha-1}\}g_\sigma(y)\,dy<0$ and recall $y_\sigma(t)>0$ and $y_{\sigma t}(t)>0$ in the previuos results. Then, it suffices to show $y_{\sigma t}(t)-(1-\alpha)^2y_t(t)>0$. We may have this as follows;

$$y_{\sigma t}(t)-(1-\alpha)^2\sigma y_t(t)=y_t(t)\{(2-\alpha)\sigma+\Phi(t)\}+\Phi_t(t)y(t)>0 \quad //$$

By definition

$$\zeta^1(t)=\alpha\{y^*(t)/y(t)-\tilde{y}(t)/y(t)\}/\{(1-\alpha)y^*(t)/y(t)+\alpha\} \qquad (B3)$$

and from (15)(15'),

$$y^*(t)/y(t)=(1-t)\{y_t(t)/y(t)-\alpha/(1-t)\}/(1-\alpha) \qquad (B4)$$

Hence by differentiation

$$-\zeta_\sigma^1(t)/\alpha=[\partial\{\tilde{\tilde{y}}(t)/y(t)\}/\partial++\zeta^3(t)\partial\{y^*(t)/y(t)\}/\partial\sigma]/\{(1-t)y_t(t)/y(t)\}<0$$

because, by Lemma 1,4 $\partial\{\tilde{\tilde{y}}(t)/y(t)\}/\partial\sigma<0$, by definition $\zeta^3(t)<0$ and

$$\partial\{y^*(t)/y(t)\}/\partial\sigma=(1-t)t^2\{y_t(t)/y(t)\}^2\phi(t)>0 \qquad (B5)$$

where

$$\phi(t)=\int^{y(t)}g_\sigma(y)\,dy+g\{y(t)\}y_\sigma(t)=\{(\sqrt{2\pi}\sigma)^{-1}\exp-x(t)^2/2\}\int_{z(t)}\{z-z(t)\}\exp-z^2/2\,dz\,/\int_{z(t)}\exp-z^2/2\,dz>0.$$

(B5) in view of Lemma 4, provides a <u>proof to Lemma 6</u>.

For $\zeta^2(t)$, we have

$$(1-\alpha)E(y^{\alpha-1})\zeta_\sigma^2(t)/\alpha=\int^{y(t)}y(t)^{\alpha-1}g(y)\,dy\{(1-\alpha)^2\sigma+(1-\alpha)y_\sigma(t)/y(t)\}$$

$$+\int^{y(t)}\{y^{\alpha-1}-y(t)^{\alpha-1}\}g_{\sigma}(y)dy-(1-\alpha)^2\int^{y(t)}y^{\alpha-1}g(y)dy \qquad \text{(B6)}$$

by changing variables,

$$=(1-\alpha)\{(1-\alpha)\sigma+\sigma+\exp-z(t)^2/2/\int_{x(t)}\exp-z^2/2\ dz\}\{y(t)^{\alpha-1}\int^{x(t)}\exp$$
$$-x^2/2\ dx\}\ +E(y^{\alpha-1})(\sqrt{2\pi})^{-1}\exp-w^2(t)\ >0$$

where $w=x-(\alpha-1)\sigma$. As σ increases, the graphs of $\zeta^1(t)$ and $\zeta^2(t)$ shift upwards in the t,y plane.

In order to see how a solution to $\zeta^1(t)=\zeta^2(t)$ varies, it suffices to show how far these two graphs shift in the t,y plane, with respect to t as well as σ. That is, what signs do the differences $\zeta^1_\sigma(t^*)-\zeta^2_\sigma(t^*)$ and $\zeta^1_t(t^*)-\zeta^2_t(t^*)$ take, respectively, where for any solution t*,

$$\zeta^1_\sigma(t^*)-\zeta^2_\sigma(t^*)=-\{\zeta^1_t(t^*)-\zeta^2_t(t^*)\}t^*_\sigma \qquad \text{(B7)}$$

First we show $\zeta^1_t(t)-\zeta^2_t(t)< 0$ for any density function, if if $y^*(t)$ is decreasing in t. See Lemma 2. If, at $t=t^*$, the condition holds in Lemma 9, then $\zeta^1_t(t^*)-\zeta^2_t(t^*)<0$.

Differentiation gives

$$\zeta^2_t(t)/\zeta^2(t)=\{\tilde{\tilde{y}}_{tt}(t)/\tilde{\tilde{y}}_t(t)-y_{tt}(t)/y_t(t)\}$$
$$=\alpha y(t)^{\alpha-2}\{\int^{y(t)}g(y)dy\}y_t(y)/E(y^{\alpha-1})>0 \qquad \text{(B8)}$$

and

$$(1-t)\zeta^1_t(t)/\alpha=\{-\zeta^2(t)-\tilde{\tilde{y}}(t)/y(t)\}-\zeta^3(t)\partial\{y^*(t)/y(t)\}/\partial t/y_t(t)/y(t).$$

(B9)

(B9) may reduce

$$(1-t^*)\zeta^1_t(t^*)=-\alpha\zeta^3(t^*)(1-\alpha)[(1+\alpha)/(1-\alpha)-\partial\{y^*(t^*)/y(t^*)\}/\partial t]$$

which is negative if

$$(1+\alpha)/(1-\alpha)>\partial\{y^*(t^*)/y(t^*)\}/\partial t \qquad \text{(B10)}$$

(Condition of Lemma 9).

Then, we have only to show $\zeta^1_\sigma(t)-\zeta^2_\sigma(t)>0$. at t^*.
After a tedious calculation, we can have

$$(1-\alpha)E(y^{\alpha-1})/\alpha\}\{\zeta^1_\sigma(t)-\zeta^2_\sigma(t)\}=\{\alpha/(1-t)-y_t(t)/y(t)\}\Delta^1(t)$$

$$+\{\alpha/(1-t)y_t(t)\}\Delta^2(t)$$

$$+\{(1-\alpha)E(y^{\alpha-1})y^*(t)/(1-t)y_t(t)\}\{1-\int^{y(t)}y^{\alpha-1}g(y)dy/E(y^{\alpha-1})\}$$

$$+(1-\alpha)[y^*_\sigma(t)/(1-\alpha)y^*(t)+\alpha y(t)\}-y_\sigma(t)/y(t)\}]y(t)^{\alpha-1}\int^{y(t)}g(y)dy$$

Here, in the 3rd term, $1-\int y^{\alpha-1}gdy$ takes 1 at t=0, 0 at 1 and it is easy to check it decreases as t increases. Hence the term is positive in the interval [0,1). In the fourth term, the value inside the square blacket is always larger than $(1-\alpha)\{y^*_\sigma(t)/y^*(t)-y_\sigma(t)/y(t)\}$ which is positive in the interval from (B5).

Since $\alpha/(1-t)-y_t(t)/y(t)>0$ if $y^*(t)>0$, it remained to prove

$$\Delta^1(t)=(1-\alpha)^2\sigma\int^{y(t)}\{y(t)^{\alpha-1}-y^{\alpha-1}\}gdy-\int^{y(t)}\{y(t)^{\alpha-1}-y^{\alpha-1}\}g\ dy<0 \text{ and}$$

$$\Delta^2(t)=(1-\alpha)^2\sigma\int^{y(t)}\{y(t)^{\alpha}-y^{\alpha}\}gdy-\int^{y(t)}\{y(t)^{\alpha}-y^{\alpha}\}g_\sigma dy>0.$$

Change the variables y to w, defined before, then,

$$\Delta^1(t)=(1-\alpha)\sigma E(y^{\alpha-1})(1-\alpha)\int^{w(t)}\exp\{-w^2/2+(\alpha-1)\sigma(w(t)-w)\}dw$$

$$-\exp-w(t)^2/2\ ,\ \Delta^1(0)=\Delta^2(t)=0$$

$$\Delta^1_t(t)=\{w_t(t)\exp-w(t)^2/2\}\{x(t)-2(\alpha-1)\}>0 \leftrightarrow x(t)>-2(1-\alpha).$$

Hence, $\Delta^1(t)$, starting with the value 0, decreases and increases, ending with 0. That is, $\Delta^1(t)<0$.

For $\Delta^2(t)$, we can have for $w=x-(\alpha-1)\sigma$,

$$\Delta^2(t)/E(y\ )=(1-\alpha)^2\sigma\int^{w(t)-\sigma}\exp\{-(w-\sigma)^2/2+\alpha\sigma\{w(t)-w\}\ dw$$

$$+\alpha\exp-\{w(t)-\sigma\}^2/2-(1-2\alpha)\sigma\int^{w(t)-\sigma}\exp-(w-\sigma)^2/2\ dw$$

$$\Delta^2(0)=\Delta^2(1)=0$$

and

$$-\Delta^2_t(t)=[\alpha\exp-\{w(t)-\sigma\}^2/2]w_t(t)\{x(t)-2\alpha\sigma\}\gtreqless 0 \quad \text{as } x(t)\gtreqless 2\alpha\sigma$$

Hence $\Delta^2(t)>0$ in the tax interval.

This completes the proof of Lemma 9.

Dofferentiation of (13) w.r.t. σ yields

$$\{V_\sigma(t)/V(t)\}\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\}$$
$$=\{(1-\alpha)\tilde{\tilde{y}}_\sigma(t)+\alpha y_\sigma(t)\}-\tilde{\tilde{y}}_\sigma\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\}/\tilde{\tilde{y}} \qquad \text{(B11)}$$

Let Γ denote the R.H.S.of (B11), then, we show the t-derivative of $\Gamma(t)$ positive in the interval. Note $\Gamma(0)=0$. In

$$\alpha^{-1}E(y^{\alpha-1})\Gamma_t(t)=-y_t(t)\Delta^1(t)+\alpha\{y_{\sigma t}(t)-(2\alpha-1)\sigma y_t(t)\}\{E(y^{\alpha-1})$$
$$-\int^{y(t)}y^{\alpha-1}g\,dy\}$$
$$+\,y_\sigma(t)[y_{\sigma t}(t)/y_\sigma(t)-y_t(t)/y(t)$$
$$+y_t(t)\{1-\sigma\{1-(1-\alpha)/\alpha\}/y_\sigma(t)/y(t)\}]\int^{y(t)}y(t)^{\alpha-2}g\,dy$$

we already know that $\Delta^1(t)<0$, $y_{\sigma t}(t)-(2\alpha-1)\alpha y_t(t)>y_{\sigma t}(t)-\sigma y_t(t)>0$ from the previous results. Since

$1-\sigma\{1-(1-\alpha)/\alpha\}/y_\sigma(t)/y(t)=1-\sigma\{1-(1-\alpha)/\alpha\}/\{\sigma+\Phi(t)\}>1-\{1-(1-\alpha)/\alpha\}>0$

and $\Phi(t)>0$ in the proof of Lemma 4, it follows that $\Gamma_t(t)>0$ hence $\Gamma(t)>0$. Note $V_\sigma(0)/V(0)=0$ and $(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)>0$ in the closed interval.

Thus the proof is complete for Lemma 7 hence Lemma 8.