No. 374

ON THE LOGNORMAL HYPOTHESIS-LINEAR TAXATION, INCENTIVES AND ABILITIES IN SCREENING

by
Sho-Ichiro Kusumoto

July 1988

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 HYPOTHESIS — LINEAR TAXATION, INCENTIVES AND ABILITIES IN SCREENING

By

Sho-Ichiro Kusumoto *1

July 1988

ABSTRACT

The lognormal hypothesis and implications for the distributions of ability and income are investigated in a simple general equilibrium labor-contract model of heterogeneous workers redistributed by means of linear income-tax and transfer policy. In particular, we shall pay an attension to the various incentives and decisions of these workers who are screened by firms, affecting upon the (shape of) distributions.

By isolating from the income-equalizing effect, we shall consider of the possible incentive effects that would differently render the workers to manipulate their imperfectly observed (present) ability, in such a way that their future size would be distributed in a positively skewed way.

This manipulability, which could be assumed relatively costless, may feature the difficulties with moral hazard which the less able would fall into on one hand, and, on the other hand, the self-selection activities in which the more able workers more intensely engage, given the ability-dependent distribution of screening error and the costless productivity structure of an educational system of various intensities.

Under relevant conditions on the manipulable rates of change in ability, we show a time process of such ability manipulation formulated as generating the size distribution of true ability, will lead the final distribution to a skewed one such as lognormal.

A complete treatment of such a model allows us to reason the whole paths of movement which would take place in the economy.

To be presented at the Annual Meetings of the Japan Association of Economics and Econometrics, September 23 1988.

A. Income Distribution

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] though. Although the size distributions (of income and wealth), such as obtained by Roy[1951], Friedman[1953], Newman and Wolfe[1961], among others, were each explained as the end of economic choice, it could actually have done no more than help the shape of distribution drawn from a random process ad hoc assumed.

It was the attempt made by Stiglitz[1969] that seems to have firstly and completely argued an economic theory of wealth and income distributions among individuals, entirely based upon the basic principles that have proven in a neoclassical (growth) theory.*2

While elegant and enlightening, those, essentially stochastic, models originated with Gibrat[1931], Kalecki[1945], 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 economic theory of income distribution. See for survays Mincer [1970] and Blinder[1974].

Namely, an unsatisfactory state in the area of economic interpretation still has been continuing. This sharply contrasts the fairly satisfactory state with respect to the statistical description of income distribution.*3

A purpose of the present work is to fill a part of this gap by taking a unified approach from theory of optimal taxation founded by

Mirrees[1971,1974] and theory of screening founded by Stiglitz[1975]. In a simple general equilbrium of repeated employment-contracts and redistributive taxation, implications for the size distributions of ability rather than income will be attempted to be fully investigated. Our final purpose is to lead the limiting distributions asymptotically to lognormal ones in a rather wide variety of formulations.

B. Incentives, Abilities and Employment Contracts in a Mixed Economy with Taxation-Transfer Policy: Introductory Remarks

We begin by sketching our view of a mixed economy, which will be in our extensive concern.

1. Many Workers, Many Firms, the Authority (Government) and Employment Contracts

In the economy, there are a large number of workers[*4] who are perfectly informed about their own abilities and are redistributed by means of income-tax and transfer policy. There are also a large number of firms[*5] employing, through contracting, the heterogeneous workers in order to produce and sell outputs for consumption.[*6] The workers are almost sharecropping in the sense that the output each hired individual or grouped[*7] individuals add could be distinguished almost perfectly by expensive monitoring and detection, from those which the other hired add.

The authority consists of officers who are fully paid a constant[*8] portion, as a whole, of the tax revenues for public, including educational, legal and informational, services supervising all the firms and workers in order to enforce binding and obligatory contracts described below. Among alternative policy schemes[*9] of linear income-tax and transfer, which are each budget-consistent and take into a full account all screened contracts under the schemes, the authority will finally choose and offer the best (soical optimum) policy scheme.

The contracts each simply describe taxable labor supplies (wage-incomes[*10] and tax returned under the policy schemes) of an individual or grouped individuals. They will be offered competitively by all firms and open to all potenstial employees outside or inside each firm.

They are binding and obligatory once contracted.

The authority and each firm mutually inform and are informed of all knowledges about workers' characteristics, such as abilities and preferences, they have obtained through the errors and trials process of contractual arrangement. Thus, they will, after all, have a, more or less, satisfactory labor statistics on a (unique) distribution of the population, upon which their final choices will depend effectively within a contracting period.

Given a distribution of the potential workers in screening tests, each firm competitively makes an offer as a candidate employer to a candidate employee. A potential employee is competitively and internally searching for the desired firm offering a set of optimal contracts under alternative taxation-transfer schemes; a labor supply schedule.[*11]

We can thus presume each recruiting ("identical") firm, both competitively (externallly) and internally offering efficient labor-contracts, will be faced with an identical distribution of potential employees' heterogeneous abilities screened.*12

Given the ability distribution determined based upon the contracted taxable labor schedules of all workers sorted in a certain screening system, the authority will determine and implement an optimal tax rate maximizing a social welfare function (the aggregate of individual utilities) defined over a continuum of the individuals of those estimated (maximum) abilities.

2. Screening Information and Cost:

The latent abilities of contracting workers are pointwise estimated and measured in the unit of output, through the process of sorting the workers based upon the results, such as credentials, credit, other qualifications, of both educational and on-the-job trainings at a given intensity level. The crop returned by a worker as a result of his/her work may be used each as ascertaining the ability estimated or as a proxy indicating the underlying ability of his/her own.

There is no technological uncertainty assumed to be involved between the expected (at the begining of a contracting period) and realized (at the end) outputs.*13

Thus, the approved crops or labor supplies returned by workers may be, more or less, purposely misreported but respectively revealing different levels of performance in producing their outputs, and one another viewed as serving a function of sorting the workers one another. The performance that he/she will make per hour reflects traits and attributes which cover effort and attitude toward his/her work. While the work-hour put in is usually easy to observe perfectly, the (per-hour) true performance is not. Even in a careful, additional screening which will cost relatively high, it can not be observed perfectly.

In a private economy, the benefits of full and accurate screening information about the ability of a competitive worker would accrue to the individual, not accrueing to a competitive firm employing the worker. Hence the cost of obtaining such information would be borne by the worker him/herself, it is relatively small.

In the mixed economy characterized by offering public facilities and services, such as yielding legal and informational externality, the running cost should be borne in large part by paying from the revenues collected from individuals through taxes.

3. Self-Identity and Free-Riding Incentives in Imperfect Screening System:

Under incomplete and asymmetric information (in the screening system of a certain intensity), the more able, who are perfectly informed of their own ability on an assembly line while the firms are not, would each see an incentive to identify themselves with their own performances (the self-identity incentive). When the more able could not be separated in the work force on the assembly line according to their performances, they all would have been to receive an average wage of the entire or rest of the work force. The less able, thus, would see an incentive to pretend to be better than they actually are, because, by so doing, they could be freely given by the wage subsidization within each work cohort.

Thus, a device for such screening/signaling would be induced to be set for private uses. As such arrangement, we take an educational system of the variable intensities which can be individually chosen with relatively higher costs, as well as on-the-job screening with relatively low costs..

Further, we consider both the productivity and the screening effects of the educational system of an intensity, one-period lagged behind the educational investment within a period for that intensity, while the on-the-job pure screening effects immediate. Indeed, the result of each training may screen for some different characteristics than other, but we regard the return to on-the-job screening as dependent more largely upon the amount of education.*15

In a private economy, if it is perfectly accurate, full screening would imply a Pareto-optimal market equilibrium without expending any resources on screening on the job. For example, suppose a worker agree in the contract to pay 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 overstated. Each employer has only to announce it will screen individuals on each assembly line if the output of that line differs from what it should be, given the ability which each of the employed has returned. The workers who are accurately observed would be completely paid for their performances they have made.

Even if screening need not be full nor accurate at a high intensity level, it is possible to creat work incentives. For example, a worker on an assembly line might be compared with the contracted standard on the line, resulting in one payment awarded if output falls anywhere below the standard, and another (higher) payment awarded if output falls anywhere above the standard. Creating in such a way the incentives for the contracted workers to perform their own productivity within each work cohort, may convert the continuous distribution of ability into a discrete (binominal) distribution of income. See Lazear and Rosen[1981] for a rank-order payment scheme compared with a piece rate.

4. Tax-Evasion and Moral Hazard: Incentives with Taxation-Transfer Policy:

Since taxation system is institutionally set in a way that the crop realized is to be taxed in a direct proportion to itself, a tax paying, able worker, to whom a (net) tax rate becomes very high, would see the incentives to misreport the result of his/her performance (ability) in order to evade paying large taxes. Such individuals might understate the contractual outputs hence have their ability underrated. This would be much more so if taxation were income-progressive.*16

Under imperfect screening, there would be some incentives to engage in such tax-avoiding activities more or less intensely for working individuals.

Among the difficuties, some of which are already referred to, with moral hazard, perhaps the worst would arise when the difference in the ability of members can become large so that the utilities-sum social welfare criterion will determine an optimum transfer as an amount large or larger than enough to feed a number of individuals and prevent them from working.

Workers, once unemployed, will be perfectly"unobserved" and in fact they are out of any skill acquisition or on the job training process. Thus, they would freely manipulate, in particular by lowering, their present abilities for any purposes. The (minimum) income insurance policy would, thus, help the lower-endowned individuals to be further lower-rated.

We shall refer to this as a self-destructing (lowering) moral hazard and consider a possible distorting effect which will affect the shape of distribution in particular in the lower portion.

5. <u>The Incentives Compared</u>:

The four main types of incentives workers would seem to see are listed below in a table.

Under perfect screening, the incentives to engage in free-riding or tax-sheltering activities would vanish while the incentives to engage in self-identity or self-destructing moral hazard activities need not do so.

We shall argue, in Section 1 and 2, that tax-paying workers situated in the upper portion of distribution would see more or less the self-selection incentives and engage more or less intensely in such screening/signalling activities, whereas those in the lower portion would be more likely to engage more or less intensely in self-lowering activities. The self-selection incentives are hierarchically encouraged by the productivity (growth) effects of the educational system of various intensities, but do not seem to be discouraged by taxing because of the <u>linear</u>[*17] proportionality and decreasing marginal tax rate. On the other hand, The moral hazards are encouraged by the income transfer policy rather than discouraged.

We shall discuss these in more details in Section 2 in Part C and mathematically also in Section 2 in Part D.

Incentives / Workers With	Within Each Work Force on an Assembly Line Incentives	By Tax and Transfer Policy Incentives
Higher-Abilities	Self-Identity (Self-Selection)	Tax-Evasion (Underrated)
Lower-Abilities	Free-Riding (Overrated)	Self-Destruction Moral Hazard

C. Summaries and Interpretation of the Results in Formal Analysis:

In Section 1, we thus solve a simple problem of optimal taxation, with a linear income taxation for transfer, a fixed (density function of) distribution of the present size of heterogeneous ability, and with fixed homothetic (identical and truely known) preferences each assumed representable by a Cobb-Douglas function which is linearly homogeneous and concave. Our first aim is to work out the details of the designed solution properties under full screening at a given intensity.

1. Perfect Screening Information and Policy Effect:

For the given distribution, an optimal tax rate t* is determined so that a positive number of workers won't work for the rate t*. The possible largest ability the voluntarily unemployed possess is uniquely determined for the tax rate t*. It is an increasing function in tax rate t. We call this the reservation ability for a clear reason. It will play an essential role throughout the whole sections. See Subsection 1.1 in Part D.

The full-income of each worker will be distributed in the same way as the "endowned" ability y, but, now, with a variance δ^2 smaller (small by 100t*(2-t*)%). Consumption, which is equal to the sum of transfer and wage income net of tax returned, would be confined to the range above the minimum (equal to the transfer which is also uniquely determined for the rate t*), as if the rest of distribution (for which they are less than the transfer) were removed. The consumption is thus trancatedly distributed where the point of trancation is associated with the reservation ability y(t*). The variance of consumption is much smaller (small by more than 100t*(2-t*)%) than of the wage income which would be earned by a worker him/herself, in case of no such policy at all.

If screening is perfect, then, (y,g(y)) are the true data screened. The designed solutions based on the data are a market equilibrium under the optimal tax rate t* determined above. Theorem 1

2. Imperfect Screening Information, Policy Effects and Incentives:

2.1 A Larger Policy Effect:

However, if screening is not accurate at intensity Λ, then, ability, labelled y(Λ), can not be estimated without screening error. This results in a, more or less, "incorrect" distribution (density). Let us designate the density function g(.,Λ) at Λ. Then, applying the above results (in Section 1) to the data (y(Λ),g(y(Λ),Λ)) with a larger variance,[*18] we shall be likely to have a larger "optimal" rate t*(Λ), larger revenues and transfer T(t*(Λ)) and a larger reservation ability y(t*(Λ),Λ) hence more unemployed, so that the income equalizing effect of the public policy will be realized more largely. In fact, we shall see this more precisely in Section 3 with a lognormal distribution.

This will lead an intuition to a somewhat surprising statement: In imperfect screening, a market equilibrium is established with a larger social welfare than in perfect screening. This is because of not only the wage (private) subsidization within each cohort, but also the larger income-equalizing effects, both being due to screening error.

2.2 Wage-Subsidization and the Distribution Density:

A market equilibrium is obtained under the taxation rate $t^*(\Lambda)$ and the feasible (budget consistent) revenues $T(t^*(\Lambda))$ at intensity Λ, for the observed distribution density. However, within each cohort $g(y(\Lambda),\Lambda)$ wage subsidization (from the more able to the less able within the work cohort) will go in a different way, depending upon where a worker in concern is labelled in the true distribution on the whole. A worker of the work cohort $g(y(\Lambda),\Lambda)$, labelled $y(\Lambda)$ at which the true density is decreasing, i.e. $g'(y(\Lambda))<0$, will be paid less than the contractual wage $L(y(\Lambda), t^*(\Lambda))$, while he/she, labelled where $g'(y(\Lambda))>0$, will be paid more. This is because, if he/she is labelled where the true density is decreasing (resp. increasing), he/she will be averaged with less (resp. more) individuals who are better than he/she but underrated, and with more (resp. less) who are worse but overrated. Subsection 1.7 in D.

2.3 Differentiating An Observed Density from the True Density will Strengthen or Weaken the Incentives:

Furthermore, wherever more individuals are thus underrated on an assembly line y (than overrated, within the range of error given at a screening intensity Λ), the observed density $g(y,\Lambda)$ will be larger (resp. smaller) than the true density $g(y)$, if the excess of the underrated over the overrated is increasing (resp. decreasing) there.

This means, not only that there will be a larger number of workers screened wrongly and underrated into a work-cohort on an assembly line where the density is increasing, but also that, due also to the screening error, the excess number will be large (resp. small) by a certain amount over the number just enough to be offset by the overrated so that the observed density can be equated with the true one.

Conversely for the portion where more overrated workers exist and the excess of the overrated over the underrated is increasing (resp. decreasing). The observed density will be smaller (resp. larger) than the true density. By screening error, a smaller (resp. larger) number of workers will be overrated, small (resp. large) by an amount over just enough to be so offset by the underrated.

Thus, wherever the observed population density may be above (resp. below) the true one, both the incentives to be "overrated" and "self-identified" will be strengthened (resp. weakened). For this see (23) in Subsection 1.7 in Part D, in a formal detail.

2.4 The Incentive Effects Due to Screening Error in Unimodal Distribution:

In a unimodal distribution, relevant here, such as lognormal, observation error will thus stimulate to see the hierarchical sceening incentives more for the higher-ability workers whose expected abilities are in the high portion of the distribution. For the lower-ability workers in the low portion and employed, on the other hand, screening error will stimulate to see more incentives to be overrated.

Thus, incentive-motivate activities would possibly characterize the future size of ability hence the shape of distribution. They would each result in an increase or decrease in private expenditure, relatively to an expenditure in case of no screening error. The relative increase would be larger for those located in the high portion relatively to the mode, while the relative decrease larger for those in the low portion, provided that the costs of screening are borne privately.

Thus, screening intensities would be ability-dependent and different from those determined under full screening without error.

3. Intrinsic Dynamic Factors:

Other than the income-equalizing effect, which may possibly be enlarged by observation error, we shall see some dynamic effects of the linear income-tax and transfer policy,affecting upon the future size of ability. They would not vanish even under perfect screening.

Thus, we shall point out some intrinsic dynamic effects induced by the economic policy, other than those effects dependent upon the various incentives induced by screening error. See Section 2 in D.

A worker will see quite different incentives, depending upon where he/she expects to be rated in the overall distribution relatively to the current reservation ability. We shall classify them two ways:

3.1 Moral Hazard; Self-Destructing Activities:

In this mixed economy, the voluntarily unemployed are enjoying a maximum utility he can attain by unemployment and transfer income. Whether or not they earn a utility gain is completely independent of any changes in their ability, in so far as they continue to be unemployed. Thus, each of the unemployed could manipulate their, now potentialized and unobserved, abilities, in particular, by lowering them without any cost. In fact, they are out of any skill acquisition or ability improvement or sustaining process on the job. Possibly and it should be emphasized that those whose true abilities are larger but their expected abilities are less than present reservation ability would fall into this self-destructing moral hazard if the tax rate be determined high enough. See Subsections 2.1 and 2.2 and also Theorem 3 in D.

Such costless manipulability, combined with the overrated incentives, may, thus, feature the moral hazard which relatively lower-ability workers would fall into. The income insurance policy would support their engagement in such moral hazard.

3.2 Self-Selection Encouraged by Productivity Effect:

Taking the productivity effect of education into full account, the self-selection will go beyond the scope of pure screening within each work cohort. An incentive to acquire an increase in the true ability for a worker is motivated by an attempt to earn a larger utility gain. In fact, by the linear income taxation, the more able will see a larger gain from a unit increase in ability than the less able will see, if they are sorted perfectly. It is large by the amount which would not have been expected if taxation were income-progressive. This amount is an increasing function of the ability of a working individual, while it is constant for all the unemployed. See for this, Subsections 2.1 and 2.2 in Part D.

Such utility-motivate ability acquisition incentives of the relatively higher-ability workers could be flustrated only by a possible high(er) cost of education and skill acquisition that they will have to bear privately.

4. Central Limit Property in the Ability Determination Process:

Finally, incorporating (concave) production and (convex) cost structure into the integrated incentive structure, we formulate a time process of the incentive-consistent ability manipulation. It aims to describe the manipulable change in ability that will be a relatively costless increase or decrease, asymmetrically depending upon whether the expected (present) ability is larger or smaller than the reservation ability for the present tax rate $t^*(\Lambda)$ at screening intensity Λ. In this sense the process is not ad hoc nor stochastic in determining the future size of ability hence of income and consumption. For this see Subsection 2.2.

In a set of relevant conditions on the rates of ability change which are asymmetrically different, we see a central limit property underlying the resulting distributions of true ability from the process. While the rate, taken by an individual whose expected ability is not larger than the reservation ability, is randomized in a way of somehow converging, the growth rate, which is taken by an individual whose expected ability is larger, consists of two terms; a random term with zero mean, hence unexplained, and the explained term maximizing, as an economic choice, the expected value of his/her true ability with respect to educational and screening intensity that he/she is able to choose for the future contract. See for this Subsection 2.3.

We shall see the law of proportionate effect (pointed out by Gibrat[1931])holding. We also extend the result of Kalecki[1945], which was simply a direct drawing from an ad hoc assumption, from our approach. We argue the process will lead the final distributions to the lognormal type in infinitely many periods of contracting. It increases the variances of true ability in the meanwhile, but does not necessarily converge to a stationary distribution. Theorem 2

5. Lognormal Distribution and Implication:

In Section 3, where, in a sequence of lognormal distributions of ability, their means and variances differ and increase as the time process evolves, we shall prove the social welfare will increase.

Given a lognormal density function with two parameters; the variance σ^2 and the mean μ of logarithm of ability y, then, we prove the optimum social welfare will increase, as the σ-parameter increases. Theorem 3.

Section 4 is our concluding discussions and remarks. We shall interpret the comparative statics results obtained in Section 3 as implying the whole dynamic movement of the economy.

D. Formal Analysis

1. Redistributive Taxation Models; Revisited Under A Simple Utility Hypothesis:

Let y be an ability, endowned to and returned by a worker, in a given period. We intend it to imply the maximum homogeneous labor or output the worker would be able to choose or produce, if he/she spend all time working when employed. Let L denote a non-negative quantity of labor supply this worker, characterized by y, will choose. We thus presume it to show a quantity of output he/she will add when employed. Ability y, labor L, and output are measured in the unit of output (for consumption).

Designate by g(y) the population density function of variate y. We normalize it so that $\int g(y)dy=1$, where we assume throughout the whole sections $g(y)>0$, $0<y<\infty$, $g(0)=g(\infty)=0$. Then, the value of g(y) expresses a relative number of workers whose ability is y. Within each period, the density function g is fixed and informed at a certain screening intensity level to the government authority which will make a public decision on taxation-transfer policy to implement.

The taxation system with its structure is specified by a linear income tax rate t such that $0 \leq t \leq 1$. We revise a common premise underlying the income-redistributive models, such as of Mirrlees[1971], Sheshinski[1972], Rader[1983] etc.. That is, the tax revenues, denoted by T, are net of a constant (Θ) portion of them and are to be paid out to all individuals in eqaul quantities. The Θ portion will be paid to the government officers for their public services. This modification, however, won't change any formal analysis and therefore 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[*19] of leasure R=y-L over y,

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

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

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

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 [*20]

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

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 differentiable function of (y,t). We can specify them in terms a new variable, y(t), defined below and give them in (19)-(20), in particular for L(y,t),

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

Note that the new variable y(t) is associated with the boundary solutions c(y(t),t) given in (19) and r(y(t),t)=1 in (19'). From (2) and by taking T=T(t),

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

Thus, $y(t)<y \leftrightarrow L(y,t)>0$ and $y(t)\geq y \leftrightarrow L(y,t)=0$

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,

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

The lower bound of integration is y(t) defined in (4). Hence y(t) is obtained in an implicit form;[*21]

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

This ability,hereinafter I would say reservation ability, 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 $y(1)=+\infty$.[*17]

Lemma 1: $y_t(t)>0$ for all t in [0,1).

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

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

The reservation ability y(t) defines a critical tax rate s for each y. Define an inverse function s to be such that $y(s)=y$, $y\epsilon[0,\infty)$. Then, s(y) is the least tax rate for which worker y won't work. It increases in y with its range [0,1).Thus, the positive correlation and one-to-one correspondence of ability y and tax rate s are established.

From (3)(4) and Lemma 1, it follows immediately there is no "backward bending" in the labor supply curve often discussed and 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:[*22] 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; Revenues Maximizing Tax Rate $t°$:

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 $\bar{y}(t)=(1-\alpha t)y(t)/(1-\alpha)t$,

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

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

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 $y \gtreqless \bar{y}(t)$, 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 $\bar{y}(t)$ 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 Rates:

To see how welfare is gained or lost by the taxation-transfer policy, we shall define and use the relative utility of worker y as the ratio, designated $v(y,t)$, of $u(y,t)$ over $u(y,0)$, where

$$u(y,t)=c(y,t)^{\alpha}r(y,t)^{1-\alpha} , \quad u(y,0)=c(y,0)^{\alpha}r(y,0)^{1-\alpha}$$

$$c(y,t)= (1-t)yr(y,t)/(1-\alpha), \text{ and } r(y,t)=\{(1-\alpha)y+\alpha y(t)\}/y.$$

We shall call this ratio v a <u>relative individual welfare</u> and view it as an objective function of worker y.

Then, we may have

$$u(y,t)=\{T(t)\}^{\alpha}\{y/y(t)\}^{\alpha}r(y,t) \qquad \text{if } y>y(t) \qquad (10)$$

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

where, from (4),

$$\{T(t)/y(t)\}^{\alpha}=(1-t)^{\alpha}\alpha^{\alpha}(1-\alpha)^{-\alpha}$$

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

$$E(u(y,t))=\{T(t)/y(t)\}^{\alpha}[(1-\alpha)\int_{y(t)} y\, g(y)dy$$

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

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

$$v(y,t)= \alpha^{-\alpha}(1-\alpha)^{1-\alpha}\{T(t)/y(t)\}^{\alpha}r(y,t)$$

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

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

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

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

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

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 $\tilde{\tilde{y}}=E(y^{\alpha})/E(y^{\alpha-1})$, $E(y^{\alpha})=\int y^{\alpha}g(y)dy$ and $E(y^{\alpha-1})=\int y^{\alpha-1}g(y)dy$. Note that, in the same way as y in (12), $\tilde{\tilde{y}}(t)$ and $\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\}/\tilde{\tilde{y}}$ may be interpretable in (13) as <u>ability</u>, and the <u>leasure ratio</u>, of an <u>aggregate</u> kind, respectively.

It is easy to check the existence of an (individual) optimum rate s^* maximizing the ratio $v(y,t)$ for each y, and the social optimum rate t^* maximizing $V(t)$ in the tax interval $[0,t°]$.

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 $y_{tt}(t)/y_t(t)$ [*23] $<(1+\alpha)/(1-t)$ for t in $[0,t°)$.

Further 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 $y\leqq y(s^*)$, $s^*=t°$.

However, for each y, s^* is not always unique. Let us consider 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 rate t^* 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,

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

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 $\tilde{\tilde{y}}<E(y)$.

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, either.

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

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

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

$$r(y,t)=R(y,t)/y, \quad R(y,t)=\{(1-\alpha)y+\alpha y(t)\} \qquad (19'')$$

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.

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, and $\alpha^2\delta^2$ is the variance of wage income $L(y,0)$ when no taxation nor transfer are implemented.

Thus, we have established a theorem for an often used class of distributions satisfying the condition of Lemma 3.

<u>Theorem 1</u>:(<u>Social Optimum Policy</u>): For each fixed (density function of) distribution which satisfies the condition $E(y)>\tilde{\tilde{y}}$, there exists a strictly positive social optimum tax rate t^*,

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

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

where the reservation ability $y(t^*)$ is unique and strictly positive, will choose unemployment (full time leasure),

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

(iii) for which the consumption income, $c(y,t^*)$, will be trancatedly distributed with the point of trancation, $c(y(t^*),t^*)=T(t^*)$ and, thus, with a much smaller variance, $\delta_c^2<(1-t^*)^2\alpha^2\delta^2$.

1.6 Perfect Screening Information:

If screening is perfect, the data $(y,g(y))$ are the true value of ability and the true density function. The designed contractual solutions are a market equilibrium under the optimal tax rate t^* determined by the authority. Theorem 1 applies.

1.7 Imperfect Screening Information:

However if screening is not accurate, then, the data $(y,g(y))$ won't be true anymore. Let us label ability and density function obtained at screening intensity Λ, $y(\Lambda)$ and $g(y(\Lambda),\Lambda)$, respectively.

Specify and assume the probability, that worker y be labelled $y(\Lambda)$, in a screening system of intensity Λ, is given by a density function $\Omega(\varepsilon)$ of random error ε, where $\varepsilon=y-y(\Lambda)$, $E(\varepsilon)=\int\varepsilon\Omega(\varepsilon,\Lambda)d\varepsilon=0$, $E(\varepsilon^2)=\int\varepsilon^2\Omega(\varepsilon,\Lambda)d\varepsilon=\delta_\varepsilon^2(\Lambda)\ \varepsilon\ C^2$. Then, it follows that

$$g(y(\Lambda),\Lambda)=\int\Omega(y-y(\Lambda),\Lambda)g(y)dy$$

where $(y,g(y))$ are the true data. This approximates in a familiar way to

$$g(y(\Lambda),\Lambda)=g(y(\Lambda))+g''(y(\Lambda))\delta_\varepsilon^2(\Lambda)/2 \qquad (23)$$

Hence, the variance of $y(\Lambda)$, designated $\delta^2(\Lambda)$, is obtained as the sum of variance of y and the variance of error ε, that is,

$$\delta^2(\Lambda)=\delta^2+\delta_\varepsilon^2(\Lambda) \qquad (24)$$

The average ability, designated $A(y(\Lambda))$, of the work force $g(y(\Lambda),\Lambda)$, is obtained as

$$A(y(\Lambda))=y(\Lambda)+g'(y(\Lambda))\ \delta^2{}_\varepsilon(\Lambda)/g(y(\Lambda)) \qquad (25)$$

Thus, a worker on an assembly line $g(y(\Lambda),\Lambda)$ is paid the average wage associated with the average ability $A(y(\Lambda))$ of the work force on the line. From (3) (for a working individual) it follows that, for the taxation rate $t^*(\Lambda)$ (obtained by applying Theorem 1 to the data $(y(\Lambda),g(y(\Lambda),\Lambda))$),

$$\{\int L(y,t^*(\Lambda))\Omega(y-y(\Lambda),\Lambda)g(y)dy\}/g(y(\Lambda),\Lambda)$$

$$=\alpha\{A(y(\Lambda))-y(t^*(\Lambda))\}=L(A(y(\Lambda)),t^*(\Lambda)).$$

Hence, the expected ability of the true ability y of an individual, is obtained as at screening intensity Λ,

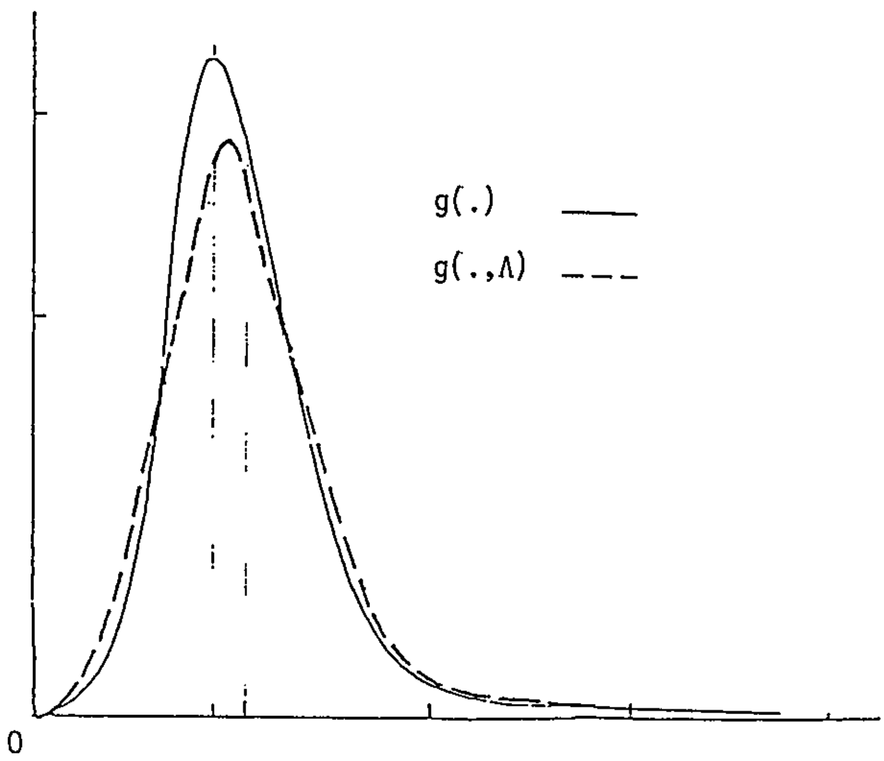
$$\int A(y(\Lambda))\Omega(y-y(\Lambda),\Lambda)dy(\Lambda)$$

This approximates a simple form, similar to (25)

$$y+g'(y)\delta_{\varepsilon}^{2}(\Lambda)/g(y) \qquad (26)$$

In fact, this is the expected value of true ability y if the distribution of error is normal.

Interpretation and implication for (23)-(26) have been made in Section 2 in Part C. Here we add a diagram below that shows a difference of an observed from the true density.



1.8 Screening Error and Policy Effects:

Next we wish to see how the aggregate variables, such as reservation ability, its tax elasticity, revenues and transfer, the aggregate labor supply (taxable income base), etc. will differ under imperfect screening from those under perfect screening.

Based upon the data $(y(\Lambda), g(y(\Lambda),\Lambda)$, we may have, for each (t,Λ),

$$y(t,\Lambda)/y(t) = \frac{1+\delta_\varepsilon^2(\Lambda)\{g(y(t,\Lambda))-y(t,\Lambda)g'(y(t,\Lambda))\}/\int_{y(t,\Lambda)} y(\Lambda)g(y(\Lambda))dy(\Lambda)}{1+\delta_\varepsilon^2(\Lambda)\{-y(t,\Lambda)g'(y(t,\Lambda))\}/\int_{y(t,\Lambda)} y(\Lambda)g(y(\Lambda))dy(\Lambda)} > 1$$

provided that, in addition to that $g(0)=g(\infty)=0$,

$g'(0)=g'(\infty)=0$.

We have made an approximation here. Let $y(t,\Lambda)=y(t)$ when $\delta_\varepsilon^2(\Lambda)=0$. Otherwise we approximate, for each t, and for $\delta_\varepsilon^2(\Lambda)$ small enough, $\int_{y(t,\Lambda)}^\infty y(\Lambda)g(y(\Lambda))dy(\Lambda) \doteq \int_{y(t)}^\infty yg(y)dy$, $\int_{y(t,\Lambda)}^\infty g(y(\Lambda))dy(\Lambda)=\int_{y(t)}^\infty g(y)dy$.

Likewise for $ty_t(t,\Lambda)/y(t,\Lambda)$, we may have, for each (t,Λ),

$$y_t(t,\Lambda)/y(t,\Lambda)/y_t(t)/y(t) = 1/\{1+\delta_\varepsilon^2(\Lambda)\{-y(t,\Lambda)g'(y(t,\Lambda))/\int_{y(t,\Lambda)} y(\Lambda)g(y(\Lambda))dy(\Lambda)\}\} \gtreqless 1 \quad (27)$$

according to $g'(y(t,\Lambda)) \gtreqless 0$.

How $L(t)$, $T(t)$, $y^*(t)$ etc. differ follows directly from (4)(5)(8)(15) etc..

In order to see how the aggregate variable $\tilde{y}(t)$, and its derivative $\tilde{y}_t(t)$ change, and finally to see how the social optimal tax rate t^* changes by screening error, we shall have to investigate how $\tilde{y}(t,\Lambda)$, $\tilde{y}_t(t,\Lambda)$ differ from $\tilde{y}(t)$, $\tilde{y}_t(t)$. In the same way of approximation, we may have, for each (t,Λ),

$$\{-\frac{\tilde{\tilde{y}}_t(t,\Lambda)}{y_t(t,\Lambda)}\}-\{-\frac{\tilde{\tilde{y}}_t(t)}{y_t(t)}\} =$$

$$\delta_\varepsilon^2(\Lambda)\{\alpha/(1-\alpha)E(y^{\alpha-1},\Lambda)\}(1-\alpha)\{(2-\alpha)\int^{y(t,\Lambda)}y(\Lambda)^{\alpha-3}g(y(\Lambda))dy(\Lambda)$$

$$+y(t,\Lambda)^{\alpha-2}g(y((t,\Lambda))\}>0$$

The approximation proceeded is; $-\tilde{\tilde{y}}_t(t)$ in (17) equalized with

$$\alpha/(1-\alpha)E(y(\Lambda)^{\alpha-1})\int^{y(t,\Lambda)}\{y(\Lambda)^{\alpha-1}-y(t,\Lambda)^{\alpha-1}\}g(y(\Lambda))dy(\Lambda)$$

multiplied by $y(t)$, where

$$E(y(\Lambda)^{\alpha-1})\overset{\Delta}{=}E(y^{\alpha-1},\Lambda)=\int y(\Lambda)^{\alpha-1}g(y(\Lambda),\Lambda)dy(\Lambda).$$

Note here $E(y)=E(y(\Lambda))$ but $E(y(\Lambda)^{\beta})\neq E(y^{\beta})$, in general.

Thus, we have for each (t,Λ) such that $\delta_\varepsilon^2(\Lambda)>0$,

$$\frac{\tilde{\tilde{y}}_t(t,\Lambda)}{y_t(t,\Lambda)} < \frac{\tilde{\tilde{y}}_t(t)}{y_t(t)} < 0. \qquad (28)$$

Similarly, we may have for $y(t,\Lambda)$ and $y(t)$, from (14),

$$\{\tilde{\tilde{y}}(t,\Lambda)-\tilde{\tilde{y}}(\Lambda)\}-\{\tilde{\tilde{y}}(t)-y\}$$

$$=\delta_\varepsilon^2(\Lambda)\{\alpha/E(y^{\alpha-1},\Lambda)\}(2-\alpha)\int^{y(t,\Lambda)}y(\Lambda)^{\alpha-3}\{y(\Lambda)-y(t,\Lambda)\}g(y(\Lambda))dy(\Lambda)$$

$$< 0,$$

where $\tilde{\tilde{y}}(\Lambda)=E(y^{\alpha},\Lambda)/E(y^{\alpha-1},\Lambda)$ and $\tilde{\tilde{y}}(\Lambda)<\tilde{\tilde{y}}$ in general.

Hence,

$$\tilde{\tilde{y}}(t,\Lambda)-\tilde{\tilde{y}}(t)<\tilde{\tilde{y}}(\Lambda)-\tilde{\tilde{y}}<0.$$

It immediately follows that

$$0 < \tilde{\tilde{y}}(t,\Lambda)/y(t,\Lambda)<\tilde{\tilde{y}}(t)/y(t) \qquad (29)$$

and also from (16)(28) and (29) that

$$\{1/(1-t^*(\Lambda))\}-1/(1-t^*)<\{y_t(t,\Lambda)/y(t,\Lambda)\}/\{(1-\alpha)\tilde{\tilde{y}}(t,\Lambda)/y(t,\Lambda)+\alpha\}$$
$$-y_t(t)/y(t)/\{(1-\alpha)\tilde{\tilde{y}}(t)/y(t)+\alpha\} \qquad (30)$$

If the right hand-side of (30) is negative, then, we may have $t^*>t^*(\Lambda)$.

However we easily see that, if $y(t,\Lambda)$ is such that $g'(y(t,\Lambda))\geqq 0$,

$$y_t(t,\Lambda)/y(t,\Lambda)/y_t(t)/y(t)\geqq 1> \tilde{\tilde{y}}(t,\Lambda)/y(t,\Lambda)/\tilde{\tilde{y}}(t)/y(t)$$

Therefore it is necessary for the right hand-side to be negative that $g'(y(t,\Lambda))<0$.

Taking variance $\delta^2_\varepsilon(\Lambda)$ so small that the inequality in (30) can be regarded almost as the equality, we may view the optimal tax rate $t^*(\Lambda)$ as being determined to be larger than the optimal tax rate t^*, except in case $g'(y(t,\Lambda))<0$ and

$$1>\{(1-\alpha)\tilde{\tilde{y}}(t,\Lambda)/y(t,\Lambda)+\alpha\}/\{(1-\alpha)\tilde{\tilde{y}}(t)/y(t)+\alpha\} > y_t(t,\Lambda)/y(t,\Lambda)/y_t(t)/y(t) \quad (31)$$

We shall reinforce the results obtained in this section including the above conjecture in Section 3, by analysis with a lognormal density function for g.

Considering also the (negative) substitution effects due to the change in tax rate, we may have, in case $t^*(\Lambda)>t^*$, larger revenues and transfer and government expenditures, though we can't be sure to have the larger taxable income base (aggregate labor). See Lemma 8 in Section 3.

2. Ability Determination Process:

Other than the equalizing effect on the incomes of all workers, one may point out an adverse effect of the public policy which would tend to render their future abilities to be more unevenly distributed, even if screening is perfect.

We show first that every worker will see, in a different way, an incentive to change his/her present ability, relatively to the reservation ability determined for the present tax rate. Then, if all workers attempt to engage in such incentive-consistent activities freely, there would 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 How Much Utility Gain A Worker Will See From Skill Acquisition:

Note, for each tax rate t, utility level u(y,t) achieved by a working individual y is precisely in one and the same order with ability y, if screening is perfect, whereas, what is achieved by each of not working individuals is identical to an α power of transfer, that is, totally independent of the difference in their abilities.

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

$$\varepsilon(y,t)\begin{cases} = 0 & \text{if } y\leq y(t) \\ =(1-\alpha)\alpha\{y-y(t)\}>0 & \text{if } y>y(t) \end{cases} \tag{32}$$

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;

$$\upsilon(y,t)=-\alpha \qquad \text{if } y\leq y(t)$$
$$\upsilon(y,t)=-\alpha/\{\alpha+(1-\alpha)y/y(t)\} \quad \text{if } y>y(t) \qquad (33)$$

It is immediate that $\varepsilon(y,t)$ increases with y and has a range $(0,\infty)$. As y increases, the utility of worker y increases increasingly with y, while, the utilities of the unemployed remain constant. Likewise, $\upsilon(y,t)$ increases with y and has a range $(-\alpha,0)$ if $y>y(t)$, while, constant if $y\leq y(t)$. The utility gain decreases at a constant rate for every unemployed and at a less rate for each worker $y<E(y)$. The (relative) utility loss of taxation increases with y but at a much less rate for every worker $y\geq E(y)$.

Note this would not be so if taxation was income-progressive.

Thus, considering in particular the results under imperfect screening in Section 1, we can summarize thereof as follows:

(i) For each of the workers whose expected ability is not larger than the present reservation ability, there is no incentive to improve or sustain the present true ability of his/her own, though he/she would accept a work contract offer by screening error.

(ii) Every worker, whose expected ability is larger and who might not be offered a work contract, sees an incentive, more or less, to improve upon his/her present true ability. The strongness of such an incentive is to a different extent and it increases with the difference of the expected ability from the reservation ability.

In case of an increase in taxation rate which will entail an increase in transfer, there will be created "substitution effects" for a larger number of workers, each having the expected ability less or close relatively to the rising reservation ability, in favor of greater density of lower abilities and longer periods of unemployment. A relevant formulation, which takes the above two points into account, will imply this condition in Section 3.

2.2 A Formulation of the Dynamic Process:

A mathematical formulation, which is able to take into account all the points reflecting upon the various incentives, including the above two points (i) and (ii), may be as follows:

$$\log y_1/y_0=\lambda_1 \log A(y_0,\Lambda_0)/y_0(t(\Lambda_0))+\nu_1, \quad \lambda_1>0 \qquad (34)$$

where the suffix 0, and 1 denotes present and future, respectively,

$A(y_0,\Lambda_0)=y_0+g'(y_0)\delta^2_\varepsilon(\Lambda_0)/g(y_0)$

$y_0(t(\Lambda_0))$, reservation ability; $t(\Lambda_0)$, optimal tax rate;

λ and ν are independent, ν_1 independent of y_0, and λ_1 is independent of y_1.

Note that, while ν_1 is unexplained, λ is explicable but here left open. See Section 2.3. for an extension in this direction.

(34) is equivalent to

$$y_1=y_0\{A(y_0,\Lambda_0)/y_0(t(\Lambda_0))\}^{\lambda_1} \exp \nu_1 \qquad (34')$$

If $\nu_1=0$, then y_1 is determined to be larger, equal, or less than y_0, respectively, depending upon whether $A(y_0,\Lambda_0)$ is larger, equal or less than $y_0(t(\Lambda_0))$. How much larger or smaller depends also upon the open parameter λ_1, as well as the difference of the expected (present) ability from the reservation ability. Note also the asymmetry between (i) and (ii) has been partly considered by formulating the process in a logarithmic form.

2.2.1 Central Limit Theorem:

A remarkable result from this formulation is an application of the well known central limit theorem, here implying the generated distributions approaching approximately to the Gibrat law under a number of limitations on λ and ν.

In (34) or (34'), take

$$y_i=y_0,\ y_{i+1}=y_1,\ i=1,2,..,n$$
$$\log y_i/y_{i-1}=\lambda_i \log A(y_{i-1},\Lambda_{i-1})/y_{i-1}(t(\Lambda_{i-1})) + \nu_i \qquad (35)$$

Suppose λ_i, i=1,..,n, ν_i, i=1,..,n all are independent of each other, λ_i independent of y_i, ν_i independent of y_{i-1}, and they are so small that

$$|\lambda_i||\log A(y_{i-1},\Lambda_{i-1})/y_{i-1}(t(\Lambda_{i-1}))| + |\nu_i| <1 \qquad (36)$$

Then, we shall see how Liapounov's condition (Loeve[1963 pp.275 -277] for example) will be satisfied hence the central limit property will be implied, irrespective of if λ_i is positive or negative.

Under the condition (36), we can assume the standard deviation of $\Sigma_{i=1}^{n} \log y_i/y_{i-1}$, n=1,2,.., is 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. This is because $\log y_i/y_{i-1}$, i=1,..,n, are small as compared with 1 by (36), for large n. Thus, for n large enough, we can take $\log y_i/y_{i-1}$ small as compared with the standard deviation of the sum. Further for this large n, the standard deviation of the sum is as large as compared with that of $\log y_0$. Then, the condition of Liapounov is satisfied. 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

$$\rho_i^3=E|\log y_i-\mu_i|^3 \qquad (37)$$

Then, the condition (36) implies that, as $n\to\infty$,

$$\rho_n^3/\sigma_n^2\to 0 \qquad \text{(Liapounov's condition)}$$

and apply his theorem to obtain

$$y_n \to G(\mu_n, \sigma_n^2) \quad \text{as } n \to \infty$$

where $G(\mu_n, \sigma_n^2)$ is a lognormal distribution with the normal density having mean μ_n and variance σ_n^2 of variate $\log y_n$.

In stead of the assumption (36), suppose

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

Then, the process (34) will lead the resulting distributions also to the lognormal type, although this supposition is inconsistent with the above two points (i) and (ii) in Subsection 2.1. It is, in fact, an ad hoc assumption made by Kalecki[1945] for the variate indicating a personal income determined by economic forces. Along the line of proof taken by Kalecki[1945, pp.162-166], we can extend his result from our approach with a condition weaker than (36).

2.2.2. Stationary Distribution:

First assume screening is perfect here, i.e. $\delta_\varepsilon^2(\Lambda) = 0$.

Suppose y_0 is $G(\mu, \sigma^2)$, then, by the simple reproductive property it is immediate that y_1 is also $G(\kappa_{1,0} + (1+\lambda_1)\mu, (1+\lambda_1)^2\sigma^2)$, where $\kappa_{1,0} = \kappa_{1,0}(t) = \nu_1 - \lambda_1 \log y_0(t_0)$.

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

In general, if y_i, i=1,..,n, are $G(\mu_i, \sigma_i^2)$, i=1,..,n, then,

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

is

$$G(\Sigma_{i=1}^{n} \lambda_i \mu_i, \Sigma_{i=1}^{n} \lambda_i^2 \sigma_i^2), \quad {}^{*24}$$

Taking $\mu_i = \mu$, $\sigma_i = \sigma$ for each i and from (35), it follows that

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

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

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

provided that both of

$$\Pi_{i=1}^{n}(1+\lambda_i) \text{ and } \Pi_{i=1}^{n}(1+\lambda_i)^2 \qquad (38)$$

converge as $n\to\infty$.

Let $y_n^{\dagger}$ denote this product. Then, if both of (38) converge,

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

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,

$$\Sigma_{i=1}^{n}\ \nu_i \qquad (39)$$

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 (35) and (39) converge. By Theorem 1 (lemma 3), there exists an optimal tax rate t_i determined for the i th period within which (μ_i,σ_i) is given.

Given the density function g_i such that, for each i=1,..,

$$g_i(y_i)=(\sqrt{2\pi}\sigma_i y_i)^{-1}\exp-\{\log y_i-\mu_i\}^2/2\sigma_i^2 \qquad (40)$$

Then, it is continuous in (μ,σ). Since $y(t)$ is continuous in t in [0,1) for each (μ_i,σ_i) and so is in (μ,σ) for each t,it follows

$y(t_i(\mu_i,\sigma_i))$ is continuous in (μ_i,σ_i). It is also bounded, since by Lemma 3 $0<t_i^*<t_i^\circ$ for each (μ_i,σ_i) and $0<y(t_i^*)<y(t_i^\circ)<\infty$ 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 $y_{i-1}(t_{i-1})^{-\lambda_i\Sigma_{h=i+1}^{n}(1+\lambda_h)}=1$, 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, which in fact does under the convergence conditions for (38) and (39).

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 $[0,t^\circ]$, 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.

2.3 Self-Selection Activities; Screening and Productivity Effects; Determination of Educational Intensity Levels:

Publicly or privately, how the intensity levels of education be determined has not been made as a choice yet. Here we shall deal with it.

The public services here we are considering include such activities as general education and trainings, as well as legal and informational services for smoothing contractual arrangement and transaction. The government authority arranges and equips the educational, legal and informational externalities.

We have been assuming a constant portion of the tax revenues collected and paid in part for such government expenditures. The portion determines an intensity Λ^1 which is homogeneous and identical to every individual. Therefore this is open to any rationality for each individual.

A worker will be able to choose an optimal intensity Λ^* by choosing an intensity Λ^2, given the external intensity Λ^1, by bearing privately the cost of the additional intensity.

Consider a simple case,*25 in which the productivity-cost structure of these two-dimensional intensities, is given as follows:

Assume that the (gross) growth rate $\xi(\Lambda)$ at Λ, taken by worker y whose true ability is y, increases with his/her own choice intensity (an element of Λ) at a decreasing (non-increasing) rate, whereas the cost denoted by $c(\Lambda)$ of acquiring the growth in ability increases at an increasing rate. We may presume that $\xi(\Lambda)-c(\Lambda)<1$ for each (y,Λ). Formally, by letting one prime and two primes denote respectively the first and second partial derivatives (with respect to the choice element), we assume that when we write, for $(\Lambda^1,0)$, $\xi(\Lambda^1;0)=\xi(0)$, $c(\Lambda^1;0)=c(0)$ etc.,

$$\begin{cases} c(0)=0, \quad \xi(\Lambda)-c(\Lambda)<1, \ \xi'(\Lambda)>0, \ \xi''(\Lambda)\leqq 0 \\ c'(\Lambda)>0, \ c''(\Lambda)>0 \text{ and } \xi'(0)>c'(0) \end{cases} \tag{41}$$

To worker y_0 (for the present), an optimal intensity Λ^* for the future will thus be obtained by maximizing an <u>expected value</u>;

$$\{y_0+g'(y_0)\delta_\varepsilon^2(\Lambda)/g(y_0)\}\{1+\xi(\Lambda)-c(\Lambda)\} \qquad (42)$$

with respect to his/her choice element of Λ_1 where $\Lambda_1=(\Lambda_1^1,\Lambda_1^2)$

Note that Λ_1^1 is determined by the authority. The expected value takes into account the (net) growth factor in the present true ability y_0. We may assume both growth rate $\xi(\Lambda)$ and cost $c(\Lambda)$ are independent of the intensities which others will choose, but we would rather assume the variance $\delta_\varepsilon^2(\Lambda)$ be dependent.

The first condition for an interior solution Λ^* may be arranged in a way to imply that (from (43') below)

$$g'(y_0)\gtreqless 0 \leftrightarrow \xi'(\Lambda^*)\gtreqless c'(\Lambda^*) \leftrightarrow \Lambda_1(y_0)\lesseqgtr\Lambda^{**}$$

where $\Lambda^*=(\Lambda_1^1,\Lambda_1^2(y_0))=\Lambda_1(y_0)$. Λ^{**} is an intensity maximizing the aggregate (mean) of abilities (over the whole distribution);

$$\int\{1+\xi(\Lambda)-c(\Lambda)\}y_0\{1+g'(y_0)\delta_\varepsilon^2(\Lambda)/g(y_0)y_0\}g(y_0)dy_0$$

which is equal to $\int\{1+\xi(\Lambda)-c(\Lambda)\}y_0g(y_0)dy_0$, hence, $\xi'(\Lambda^{**})=c'(\Lambda^{**})$,

$$\Lambda^{**}=\Lambda_1(y_0) \text{ for } y_0;\ g'(y_0)=0.$$

The <u>first-order</u> condition is given in a neat form;[*26]

$$\{-g'(y_0)\delta_\varepsilon^{2\prime}(\Lambda_1(y_0))/g(y_0)\}\{1+\xi(\Lambda_1(y_0))-c(\Lambda_1(y_0))\}=\{y_0+g'(y_0)\delta_\varepsilon^2(\Lambda_1(y_0))/g(y_0)\}\{\xi'(\Lambda_1(y_0))-c'(\Lambda_1(y_0))\} \qquad (43')$$

Thus,[*27] a sequence $\{\Lambda_1(y_0)\}$ is monotone in y_0, that is, for any $y_0'>y_0$

$$\Lambda_1^2(y_0)<\Lambda_1^2(y_0') \qquad (44)$$

The maximand (42), evaluated at $\Lambda_1(y_0)$, is the expected value of the true ability of worker y_0 for the future contract. Hence the true ability $y_1=y_0\{1+\xi(\Lambda_1(y_0))-c(\Lambda_1(y_0))\}$ gives an account of (34) in terms of the λ_1-parameter, by taking the expected value of λ_1 in a way that it will hold ;

$$E(\lambda_1) \doteq \log\{1+\xi(\Lambda_1(y_0))-c(\Lambda_1(y_0))\}/[\log\{y_0+g'(y_0)\delta_\varepsilon^2(\Lambda_0)/g(y_0)\}/y(t(\Lambda_0))]$$

(36) may apply so that for each i=1,2,..,n

$$|\log\{1+\xi(\Lambda_i(y_{i-1}))-c(\Lambda_i(y_{i-1}))\}|+|\upsilon_i| < 1 \qquad (45)$$

Thus, the central limit property will obtain if (45) holds for each i=1,..,n and $n\to\infty$. If we neglect the unexplained term υ_i, i=1,..,n, then, that $\{\xi(\Lambda_i(y_{i-1}))-c(\Lambda_i(y_{i-1}))\}$ is small enough for each i, is sufficient for the central limit theorem to obtain.

Thus, the variance of ability $y(\Lambda_1)$, where $y(\Lambda_1)$ is observed at intensity $\Lambda_1=(\Lambda_1^1,\Lambda_1^2(y_0))$, is dependent upon each of all the variances $\delta_\varepsilon^2(\Lambda_1(y_0))$, $y_0\varepsilon[0,\infty)$. It is easy to see from (44) that

$$\delta_\varepsilon^2(\Lambda_1(y_0))>\delta_\varepsilon^2(\Lambda_1(y_0')) \qquad (46)$$

for every pair (y_0,y_0') such that $T_0+g'(T_0)\delta_\varepsilon^2(\Lambda_0)/g(T_0)>y(t(\Lambda_0))$ where $T_0=y_0,y_0'$, and $y_0>y_0'$.

This shows how intensely the relatively able, working individuals would engage in self-selection activities. Each such activity strictly decreases the errors within a certain range arround y_0, in a cooperative way with the authority's systematic screening. All such activities together will contribute to diminishing the variance δ^2 of observed ability $y(\Lambda_1)$, in a way of decreasing errors more and more along the higher and higher portion of the size distribution.

Thus, self-selection activities are not compatible with a stationary distribution, if screening is imperfect.

2.4 Theorem

We summarize our results in Section 2 in a theorem:

Theorem 2: (Dynamic Processes)

A set of dynamic processes is made up to generate the size distributions of ability and income, such as follows;

(i) Central Limit Theorem: Under a set of fairly general conditions on the manipulable (growth) factors, the size distributions will approximately become lognormal, if the processes are infinitely many times repeated.

(ii) Stationary Distribution: Under some convergence conditions on the manipulable rates, the size distributions will entail with them a stationary (equilibrium) distribution, to which they will converge, if they are initially lognormal and if screening will in the meanwhile become perfect.

Remark 3: The variance of true ability monotonely increases as the processes evolve. This reflects upon the incentive effects of the policy enlarged by screening error.

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, under perfect screening.

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, $y(t)$, $\tilde{\bar{y}}(t)$, $y^*(t)$ and t^*, and the others.

For the density function $g(y;\mu,\sigma^2)$, our key variables; $y(t)$, $\tilde{\bar{y}}(t)$, $y^*(t)$, 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: $y_\sigma(t,\sigma)/y(t,\sigma)\gtreqless\sigma$ for each $t\varepsilon(0,1)$.

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

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

$$y_\sigma(t^*)/y(t^*)=T_\sigma(t^*)/T(t^*)=L_\sigma(t^*)/L(t^*)>0$$

and

$$\partial y(t^*)/\partial\sigma/y(t^*)=\{y_t(t^*)/y(t^*)\}t^*_\sigma+y_\sigma(t^*)/y(t^*)$$

$$\partial L(t^*)/\partial\sigma/L(t^*)=\{y_t(t^*)/y(t^*)-1/t^*(1-t^*)\}t^*_\sigma+L_\sigma(t^*)/L(t^*)$$

$$\partial T(t^*)/\partial\sigma/T(t^*)=\{y_t(t^*)/y(t^*)-1/(1-t^*)\}t^*_\sigma+L_\sigma(t^*)/L(t^*)$$

Hence,

$$t^*_\sigma \gtreqless 0 \leftrightarrow \partial y(t^*)/\partial\sigma/y(t^*) \gtreqless \partial T(t^*)/\partial\sigma/T(t^*)\} \gtreqless \partial L(t^*)/\partial\sigma/L(t^*)$$

The overall changes, due to a change in the σ-parameter, in the aggregate variables $y(t^*)$, $L(t^*)$ and $T(t^*)$ can be decomposed into two terms; the direct effects and the indirect effects (via a change in tax rate). Suppose tax rate t^* increases with the σ-parameter. The direct effect on the aggregate supply or taxable income base $L(t^*)$ is offset at least in part by the negative substitution effect. Possibly it might be fully offset hence the taxable base would decrease. However, since the partial elasticity of $L(t^*)$ with respect to tax rate is inelastic (less than 1), tax revenues will increase owing to both an expansion in the tax base due to the direct effect and revenue gains due to the indirect effect by the higher tax rate.

Lastly we prove the positivity of the σ-derivative of a solution satisfying both (17) and (18):

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

The condition of Lemma 2, which has only to hold at t^*, is sufficient for this condition to hold. It excludes the possibility of a local minimum.

We have thus obtained a comparative statics theorem.

Theorem 3: (Comparative Statics Under Perfect Screening)

Given a lognormal density function $g(y;\mu,\sigma^2)$ of ability variate y, where μ is the mean of log y and σ^2 the variance:

(i) A tax rate t^*, which is determined as a local optimum for each pair (μ,σ^2), will continuously increase with the parameter σ, provided that the condition of Lemma 9 holds.

(ii) The social (aggregate or mean) welfare $V(t^*)$ continuously increase with the parameter σ.

4. Concluding Discussions and Remarks:

The Gibrat Law

In the early discussion of income analysis, the law of proportionate effect was considered as almost equivalent to the asymptotic lognormality. If earnings are dependent upon ability and if ability is a product of a large number of independent characteristics such as skills, then, the earnings will finally become lognormal. This central limit property was directly related to the old (and perhaps pressent) idea; income follows a skewed distribution, while skills are apparently normally distributed. See Pigou[1924] for this.

Taking a new and economic approach, we have shown this conjecture correct in a general equilibrium contract setting of different-ability workers who are screened by firms, where their earnings are linearly taxable and redistributed from a social welfare view point. The skewedness and the asymptotic lognormality of their true ability was explained in terms of their asymmetric incentives and activities which are encouraged by the policy effects, as well as private wage subsidization, both being enlarged by screening error.

Long-Run Effects of Policy:

We have seen two policy effects, enlarged by screening error and working in opposite directions in determining the present income size and the future size of true ability. We have then emphasized a dynamic aspect of the overall incentive effects which differ across work cohorts, affecting upon the shape of distribution.

Individually rational self-selecting activities through an educational system of more or less intensities will characterize the future sizes of those abilities realatively high to current reservation ability, while self-destructing moral hazard activities here randomized, will characterize those relatively low abilities. The variance of true ability will thus become larger in the process. The resulting distributions will become asymptotically lognormal. However, in the presence of costly but imperfect detection, such incentive-motivate activities hardly seem to be uniform and compatible with a stationary distribution.

A remarkable result in the limiting continuity case is that the social welfare optimized in each period will continue to be also larger. This is because the direct effect of an increment in the ability variance, will give rise to an increment in reservation ability, while the indirect due to an entailed change in optimal tax rate will vanish, hence bringing upon an increment in the leasure ratio of an aggregate kind.

With the variance enlarged, there will be, for each tax rate, an expansion in the taxable income base, i.e. the aggregate labor, in proportion to the increment in reservation ability.

Dynamics:

The complete treatment of such a model thus allows us to suggest a dynamic interpretation of our results in terms of comparative statics.

We have seen the economy will have two different phases.

In a phase where there is brought about a rise in the tax rate, there will be a negative substitution effect. It will shrink the inelastic tax base, but the loss will be fully offset by revenue gains due to the higher tax rate. Thus, there will be an additional increase in the tax revenues collected. Correspondingly, larger transfer and public expenditures will be paid out. A larger number of individuals will be voluntarily unemployed, and less labor quantities will be as a whole supplied by working individuals. Thus, the mixed economy, entering upon this phase, will functionate aggravatingly in encouraging the moral hazard of relatively lower-ability workers.

Conversely for the case in which there is a reduction in the tax rate, which would not happen under the condition of Lemma 9. The substituion effect will bring on an expansion in the tax base but this gain will be fully offset by revenue losses due to the lower rate. Thus, there will be a relative decrease in the revenues, bringing about smaller transfer and expenditures. A smaller number will be unemployed and larger labor quantities will be supplied by working individuals. Thus, it will be functioning well in discouraging the moral hazard.

Perhaps there would be a critical tax rate, which is stationary with respect to the σ-parameter and to which optimal tax rate, determined in one phase or the other, will tend to approach. There, reservation ability hence the newly unemployed, tax revenues hence transfer and expenditures, and the aggregate (mean) labor or tax base, all are proportionately increasing as the σ-parameter increases.

We also note this model seems to suggest: Even if the existence of multitple equilibria and optimal tax rates takes place, an intervention by the planning authority, such as choosing an optimal tax rate in such a way as of regulating the path of movement of the economy, can be justified.

A Basic Framework Offered:

In the above analysis, an emphasis has been put on growth factors larger than cost. In fact this regulates the whole path of movement of the economy, making it possible to derive the above conclusion. However, for obtaining the asymptotic lognormality only, relatively small growth factors may possibly be sufficient as well (a proof of which has been contained and noted in Subsection 2.2.1). Thus, the framework investigated here includes, as a special case, the case in which there will be no productivity effects but screening effects only.

A rather wide variety of interpretations may be possible. Our interpretation made above may be appropriate for a one-generation model of workers, where the taxation-income subsidization is made only within the same generation but not intergenrationally. There, first thought of a rather homogeneous group of wage earners and each claiming an equal share, a generation of workers will become heterogeneous in the above described repeated process, so that each worker will be able to possess attributes and talents which influence the magnitude of his/her claim, to a different extent reflecting upon the various incentives and activities, good or bud lucks, and so on. The second, third,.., generations, follow likewise. Thus an audacious, rough reasoning is as follows in this case: Provided that a generation will suddenly vanish after a finite number of periods, a finite number of (lognormal) distributions, each belonging to a generation and each having a different set of mean and variance making a monotone sequence, will sum up to a whole distribution of the workers sumed up intergenerationally, both mean and variance of which are the averages and may be stable. If such being the case, the authority, independent of each of the existing generations, would determine an intergenerationally optimal taxation-income subsidization policy with those policy dependent variables having medium values, given the whole distribution with stable mean and variance.

Alternatively, by viewing a period as indicating a life-period and the precommited investment for education for his/her child, we would be allowed to interpret it as an intergenerational model. As already pointed out elsewhere, we could regard this as an analytical framework for social insurance. For this, read the ability as a risky random variable, the density as the probability density having that random ability, the social welfare as the expected utility etc. An individual may be characterized by a probability distribution. The model thus interpreted can be extended to many individuals to whom different probability distributions will be given, hence different optimal premium rates and minimum income insurances will be determined if they are perfectly screened by insurance companies (firms). See Easley, Kiefer and Possin[1986] for a special case for this.

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-

Lazear,P.Edward and Sherwin Rosen:"Rank-Order Tournaments as Optimum

Contracts" Journal of Political Economy 89(1980) 841-864

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

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

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

Pigou,Arthor: The Economics of Welfare London 1924.

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

"Soaking the Rich Through Tax Cuts" March 21 1985

FOOTNOTES

*1 I thank in particular Professors S.Fujishige, K.Kishimoto, S.Matsukawa, Y.Otani, T.Tabuchi, H.Uzawa and Y.Yamamoto for discussions that were very suggestive concerning this research.

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

*3 See Lillard and Willis[1978], in which they deals with life cycle earnings and mobility among earnings classes, by proposing a new econometric approach and comparing it to Markovian models.

*4 We assume each worker does not communicate or collude with others and does not know who his/her colleagues will be on the same assembly line at the time all decisions about investments(for educational training or screening his/her ability)are made.

*5 We assume each firm possesses the (identical) constant returns to scale technology with additively separable inputs of the different workers, where human capital is only necessary for production.

*6 The markets for outputs (consumption goods) are always clear.

*7 Workers in the same cohort, such as individuals on an assembly line. We take the continuum approach, rather than the discrete one.

*8 The constancy, in general, has nothing to do with optimality of any kind.

*9 For practical calculation, only local schemes which can be obtained *a priori* in a neighborhood of the optimum policy scheme, may be of interest to the central authority.

*10 The contractual wage is provisional in the sense that a worker on an assembly line will be paid the average of the workers on the same line.

*11 See footnote 9.

*12 if screening is perfect.

*13 We would like to concentrate on incentive effects due to screening error. Formulations that will take account of technological uncertainty are mathematically very complicated, even if they are the simplest. See Mirrlees[1974] for such a formulation. Se also Easelay Kiefer and Possin [1986] for a most recent work along our line. For a risk taking worker paid prize on the rank-order basis in a contest, see Lazear and Rosen[1981].

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

*14 See Miyazaki[1977], Riley[1979], among the others, where performance-equivalent price or wage searching agents take into account possible reactions by other agents. The per-hour performance an agent makes is formulated as an increasing function of unobservable characteristics and the observable sales-activities (effort) intensity which operates as a signal to potential buyers (employees). Hence the more able, by being on the higher intensity assembly line, may be able to separate themselves from the less able.

*16 See two interesting articles on the Wall Street Journal[1984,5] on the incentive effects of a tax reduction across income cohorts. I owe this references to Professor Yoshihiko Otani.

*17 We shall argue that the linearity of income taxation will encourage self-selecting activities combined with the productivity effect of educational investment.

*18 It is large by the error variance, if the distribution of error is independent of ability. See (24) in Subsection 1.6. See Subsection 1.8 for implications from this distribution.

*19 We call this ratio a leasure ratio.

*20 If the Θ portion of tax revenues is paid to G men, then, equation (2) should read;

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

Consequently, equations (5), (6), (7) and (9), will be subject to minor changes, that is, put $(1-\Theta)(1-\alpha)$ instead of $(1-\alpha)$, while the others won't.

*21 That is, $L(t)=\alpha(1-t)y(t)/(1-\alpha)t$

*22 Added, thanks to the comment made by Professor H.Uzawa.

*23 $y_{tt}(t)$ is the second derivative of $y(t)$ with respect to t.

*24 See, for example, Aitchison and Brown[1976], Theorem 2.3, p.11.

*25 We can extend this to a more general case in which the growth-cost factor depends also upon ability in a way of incorporating into this model the ability meant by a greater ability to learn more easily.

*26 For the second-order condition, see Appendix B, where the boundary solution case is also discussed.

*27 For the proof this, see the latter part of Appendix B.

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

Determination of Intensity Λ:

In more general, that $\xi'(0)-c'(0)>0$ is sufficient for an interior solution Λ^* for such y_0 that $g'(y_0)\leqq 0$ and necessary for each y_0 where $g'(y_0)>0$, provided that $\Xi'(\Lambda,y_0)<0$ for Λ large enough, where $\Xi'(\Lambda,y_0)$ is the partial derivative with respect to Λ^2,

$$\Xi(\Lambda,y_0)=\{y_0+g'(y_0)\delta_\varepsilon^2(\Lambda)/g(y_0)\}\{1+\xi(\Lambda)-c(\Lambda)\} \qquad \text{(B1)}$$

Proof: Let us write $\Lambda=0$, where $\Lambda=(\Lambda^1;0)$.

$$\Xi'(\Lambda,y_0)=\{g'(y_0)\delta_\varepsilon^{2\prime}(\Lambda)/g(y_0)\}\{1+\xi(\Lambda)-c(\Lambda)\}+\{y_0+g'(y_0)\delta_\varepsilon^2(\Lambda)/g(y_0)\}\{\xi'(\Lambda)-c'(\Lambda)\}, \text{ and hence}$$

$$\Xi'(0,y_0)=\{g'(y_0)\delta_\varepsilon^{2\prime}(0)/g(y_0)\}+\{y_0+g'(y_0)\delta_\varepsilon^2(0)/g(y_0)\}\{\xi'(0)-c'(0)\} >0 \text{ for each } y_0 \text{ such that } g'(y_0)<0.$$

Clearly, $0<\xi'(0)-c'(0)$ is necessary for each y_0 such that $g'(y_0)>0$. For the boundary solution $\Lambda^2=0$, that $\xi'(0)-c'(0)<0$ is sufficient for such y_0 that $g'(y_0)\geqq 0$ and necessary for each y_0 such that $g'(y_0)<0$.

Also we can show if δ_ε^2 has the second partial derivative;

$$\Xi''(\Lambda^*,y_0)<0 \leftrightarrow -\delta_\varepsilon^{2}{}''(\Lambda^*)\{\xi'(\Lambda^*)-c'(\Lambda^*)\}+\delta_\varepsilon^{2}{}'(\Lambda^*)\{\xi''(\Lambda^*)-c''(\Lambda^*)\}>0$$

Thus, for a local maximum Λ^* to obtain, it is sufficient that

$\delta_\varepsilon^{2}{}''(\Lambda^*)>0$ for each y_0 that $g'(y_0)<0$,

$\delta_\varepsilon^{2}{}''(\Lambda^*)<0$ for each y_0 that $g'(y_0)>0$.

When $g'(y_0)=0$, it gives the sufficiency for a global maximum. In general that $|\{\xi''(\Lambda^*)-c''(\Lambda^*)\}/\{\xi'(\Lambda^*)-c'(\Lambda^*)\}|>|\delta_\varepsilon^{2}{}''(\Lambda^*)/\delta_\varepsilon^{2}{}'(\Lambda^*)\}|$ is sufficient.

To see the ordering in (39), it sufices to show that,

$$\partial\Xi'(\Lambda,y_0)/\partial y_0>0 \qquad \text{(B2)}$$

where $d\Lambda^2(y_0)/\ y_0=\{\partial\Xi'(\Lambda,y_0)/\partial y_0\}/\{-\Xi''(\Lambda,y_0)\}$.

In fact, we have

$$\{\partial\Xi'(\Lambda,y_0)/\partial y_0\}/y_0[\{1+\xi(\Lambda)-c(\Lambda)\}g'(y_0)\delta_\varepsilon^{2}{}'(\Lambda)/g(y_0)\{y_0+g'(y_0)\delta_\varepsilon^2(\Lambda)/g(y_0)\}]$$

$$=y_0\{g''(y_0)/g'(y_0)-g'(y_0)/g(y_0)\}-1/y_0\}>0$$

provided that

$$g''(y_0)/g'(y_0)\ -\ g'(y_0)/g(y_0)\ -1/y_0>0$$

This means that

$$\partial\Xi'(\Lambda,y_0)/\partial y_0 \gtreqless 0 \quad \text{according to} \quad g'(y_0)\lesseqgtr 0$$

Take a lognormal density function for g, then we have

$$g''(y_0)/g'(y_0)-g'(y_0)/g(y_0)-y_0^{-1}=(\log y_0-\mu+\sigma^2)^{-1}$$

and therefore it holds everywhere that

$$\partial\Xi'(\Lambda,y_0)/\partial y_0>0,$$

hence, if $\Xi''(\Lambda,y_0)<0$, then, for every y_0,

$$d\Lambda^2(y_0)/dy_0>0.$$

APPENDIX C

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

$$y_\sigma(t)/y(t)=[\sigma E(y)+\int^{y(t)}\{y(t)-y\}g_\sigma dy]/\int_{y(t)}yg(y)dy \qquad (C1)$$

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

$$y_\sigma(t)/y(t)=\sigma+\exp[-\{x(t)-\sigma\}^2/2]/\int_{x(t)}\exp\{-(x-\sigma)^2/2\}dx>0 \qquad (C2)$$

The 2nd term of the R.H.S. of (C2) 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 $\tilde{\tilde{y}}_\sigma(0)=\eta^0(0)=-(1-2\alpha)\sigma\tilde{\tilde{y}}\leqq 0$ by assumption. $\tilde{\tilde{y}}_\sigma(1)=-\infty$, and $\eta^0_t(t)+\eta^1_t(t)+\eta^2_t(t)=0$ for $t=0,1$. We have the t-derivative of $\tilde{\tilde{y}}_\sigma(t)$ in what follows: $(1-\alpha)E(y)\tilde{\tilde{y}}_{\sigma t}(t)=\eta^1_t(t)+$

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

$$-y_t(t)y_\sigma(t)\int^{y(t)}(1-\alpha)y(t)^{\alpha-2}g(y)dy\ <0,\ \text{where}$$

$\eta^1_t(t)=\alpha y_t(t)\int^{y(t)}\{y(t)^{\alpha-1}-y^{\alpha-1}\}g_\sigma(y)dy<0$ 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;

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

By definition

$$\zeta^1(t)=\alpha\{y^*(t)/y(t)-\tilde{\tilde{y}}(t)/y(t)\}/\{(1-\alpha)y^*(t)/y(t)+\alpha\} \qquad (C3)$$

and from (15)(15'),

$$y^*(t)/y(t)=(1-t)\{y_t(t)/y(t)-\alpha/(1-t)\}/(1-\alpha) \qquad (C4)$$

Hence by differentiation

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

because, by Lemma 4,5, $\partial\{\tilde{\tilde{y}}(t)/y(t)\}/\partial\sigma<0$ without the condition $\alpha\leqq 1/2$, and by definition, $\zeta^3(t)<0$, and we also have,

$$\partial\{y^*(t)/y(t)\}/\partial\sigma=(1-t)t^2\{y_t(t)/y(t)\}^2\phi(t)>0 \qquad (C5)$$

where

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

In view of Lemma 4, (C5) provides a <u>proof to Lemma 6</u>.

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

$$(1-\alpha)E(y^{\alpha-1})\zeta^2_\sigma(t)/\alpha=\int^{y(t)}y(t)^{\alpha-1}g(y)\,dy\{(1-\alpha)^2\sigma+(1-\alpha)y_\sigma(t)/y(t)\}$$
$$+\int^{y(t)}\{y^{\alpha-1}-y(t)^{\alpha-1}\}g_\sigma(y)\,dy-(1-\alpha)^2\int^{y(t)}y^{\alpha-1}g(y)\,dy \qquad (C6)$$

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 these two graphs shift in the t,y plane, as the σ-parameter increases. We wish to know what signs the differences $\zeta^1_\sigma(t^*)-\zeta^2_\sigma(t^*)$

and $\zeta_t^1(t^*)-\zeta_t^2(t^*)$ will take, respectively, when they satisfy, for t*,

$$\zeta_\sigma^1(t^*)-\zeta_\sigma^2(t^*)=-\{\zeta_t^1(t^*)-\zeta_t^2(t^*)\}t_\sigma^* \qquad (C7)$$

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

Differentiation gives

$$\zeta_t^2(t)/\zeta^2(t)=\{\tilde{\tilde{y}}_{tt}(t)/\tilde{\tilde{y}}_t(t)-y_{tt}(t)/y_t(t)\}$$

$$=\alpha y(t)^{\alpha-2}\{\textstyle\int^{y(t)} g(y)dy\}y_t(y)/E(y^{\alpha-1})>0 \qquad (C8)$$

and

$$(1-t)\zeta_t^1(t)/\alpha=\{-\zeta^2(t)-\tilde{\tilde{y}}(t)/y(t)\}-\zeta^3(t)\partial\{y^*(t)/y(t)\}/\partial t/y_t(t)/y(t) \qquad (C9)$$

(B9) may reduce

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

which is negative if

$$(1+\alpha)/(1-\alpha)>\partial\{y^*(t^*)/y(t^*)\}/\partial t \qquad (C10)$$

(Condition of Lemma 9).

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

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

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

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

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

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 (C 5).

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

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

$$\Delta^1(t)=(1-\alpha)\sigma E(y^{\alpha-1})(1-\alpha)\int^{w(t)}\exp\{-w^2/2+(\alpha-1)\sigma(w(t)-w)\}dw$$
$$-\exp-w(t)^2/2\ ,\ \Delta^1(0)=\Delta^2(t)=0$$
$$\Delta^1_t(t)=\{w_t(t)\exp-w(t)^2/2\}\{x(t)-2(\alpha-1)\sigma\}>0 \leftrightarrow x(t)>-2(1-\alpha)\sigma.$$

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.

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

$$\{V_\sigma(t)/V(t)\}\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\}$$
$$=\{(1-\alpha)\tilde{\tilde{y}}_\sigma(t)+\alpha y_\sigma(t)\}-\tilde{\tilde{y}}_\sigma\{(1-\alpha)\tilde{\tilde{y}}(t)+\alpha y(t)\}/\tilde{\tilde{y}} \quad (C11)$$

Let Γ denote the R.H.S.of (C11), then, we show the t-derivative of $\Gamma(t)$ positive in the interval. Note $\Gamma(0)=0$. In

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

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.