

How does the Optimal Two-Bracket Income Tax Depend on Wage Inequality[®]

Li Gan[†] Jijun Tan[‡]

[®] This paper is edited for 第七届中国经济学年会. It is a paper in Public Economics. All rights reserved.
[†] Texas A&M University, Southwest University of Finance and Economics and NBER
[‡] Shanghai University of Finance and Economics
We are grateful to Don Fullerton, Daniel Hamermesh, Jennifer Huang, Wojciech Kopczuk, Kim Ruhl, Joel Slemrod, Shlomo Yitzhaki, and Robert Williams. We are also grateful to Shutao Cao, Ye Feng, Laura Juarez, and Ping Yan.

Abstract

Using the CES utility function, this paper numerically examines the relationship between the optimal tax-and-transfer systems and inequality of earnings under major alternative social welfare functions. In a one-bracket linear tax system, both the optimal income tax rate and the government transfer increase when earning inequality expands. In the two-bracket case, the optimal lower bracket rate and income threshold do not change in a way that is monotonic. The optimal upper bracket rate and government transfer increase with the wage spread. The lower bracket rate is greater than the upper bracket one when the spread is small, but it is larger when the spread is large. With a large elasticity of substitution between consumption and leisure, the two-bracket tax structure converges to the one-bracket case when the wage spread becomes large.

Key words: optimal income taxation, two-bracket linear income tax, wage inequality

1. Introduction

Rising inequality of earnings is one of the most remarkable characteristics of both the American economy and many other economies. In the past two decades, as described by Rosen (2004) in his popular public economics text book, Americans have seen both higher relative earnings for those at the top compared to the median as well as lower relative earnings for those at the bottom compared to the median. During this period, more serious inequalities have also been seen in many other countries such as China. Also during this period, some countries have tightened up eligibility requirements for welfare recipients. For example, the U. S. in 1996 replaced AFDC (Aid to Families with Dependent Children) by TANF (Temporary Aid for Needy Families). At the other end of the income spectrum, the U. S. has cut the top personal income tax rate to a relatively very low level. In essence, both the enlarging of inequality and the shrinking of redistribution programs have occurred since the 1980's. These two changes seem at odds with each other. One might think that welfare programs should be designed so that when earning dispersion rises, support for those at the bottom would also rise.

The economics literature has not addressed this question completely. Mirrless (1971), Stern (1976) and Cooter and Helpman (1974) show that the optimal rate of the one-bracket income tax should increase when ability is more unequally distributed among people. However, the research of Helpman and Sdaka (1978) argues that the optimal rate does not always increase. With respect to the relation between growing earning inequality and the optimal multiple-bracket income tax, economists have not paid much attention.

The purpose of this paper is to analyze how a tax-and-transfer system featuring a

lump-sum transfer and a one-bracket linear income tax or a two-bracket piecewise-linear income tax might optimally be altered in response to changes in the underlying inequality of earnings. Using the same constant elasticity of substitution (CES) utility function that is commonly used in the literature, we numerically examine the relationship between the optimal tax-and-transfer systems and inequality of earnings under major alternative social welfare functions (SWF) such as the Bentham SWF and the Nash SWF. If the tax-and-transfer system has two income brackets, comparison of the magnitudes of both marginal tax rates is of interest.

In the one-bracket case, we find that both the optimal income tax rate and the negative intercept (government transfer) become larger when earning inequality becomes more serious. In the two-bracket case, we find that the marginal rate of the lower bracket is greater than that of the upper bracket when the spread of the wage is relatively small, but it is larger when the spread is relatively large. Beyond that, surprisingly, we find that with a relatively large elasticity of substitution between consumption and leisure in the consumer's utility function, the two-bracket tax structure converges to the one-bracket structure when the wage spread becomes relatively quite large. Furthermore, though the optimal lower bracket rate and income threshold do not show monotonicity, the optimal upper bracket rate and government transfer are increasing with the wage spread.

Our paper is different from the literature in the following respects. First of all, we examine continuous changes of earning inequality in the one-bracket case, while the existing literature has investigated only a few discrete cases of earning inequality. Our research thus clarifies any ambiguity in the comparison among those discrete cases. Second, we also focus on effects of enlarging earning inequality on the optimal two-

bracket income tax, effects not addressed in the literature. We find some new and interesting results.

The paper proceeds as follows. Section 2 reviews the optimal income tax literature; Section 3 investigates the relationship between earning inequality and the one-bracket optimal income tax using theoretical distributions of ability; Section 4 similarly studies the relationship between earning inequality and the two-bracket optimal income tax, finally, Section 5 presents our conclusions.

2. Literature Review

Early literature calculates the optimal income tax rate assuming a linear income tax, a log utility functional form, a lognormal distribution of ability, and a particular social welfare function (SWF) to be maximized. For example, Mirrless (1971) shows that the optimal one-bracket income tax rate is increased when ability is more unequally distributed. He uses only two discrete cases of ability dispersion to address this tendency, however, one representing a moderate level of dispersion and the other representing an extremely high level of dispersion. Also, he does not investigate this problem based on a mean preserving process, so the change in mean wage may have confounded his result.

Stern (1976) reexamines Mirrless' simulation results for the optimal one-bracket income tax using the same distributions of ability as used by Mirrless (1971) but with a different utility functional form, a CES utility function with a smaller elasticity of substitution between consumption and leisure. He also shows that the optimal one-bracket income tax rate tends to increase when the spread of the ability distribution rises.

He still does not control the mean, and he uses exactly the same two discrete cases used by Mirrless.

Cooter and Helpman (1974) also use a CES utility function to examine the relationship between the optimal one-bracket income tax and ability dispersion. They use three different types of ability distributions to represent low, medium, and high ability inequality. Thus they have three cases instead of two, and they do keep the mean unchanged in their simulations. Still, however, these distributions differ from each other not only in ability dispersion, but also in many other respects. Their simulation results show that the optimal one-bracket income tax rate tends to increase under any of the seven social welfare functions used in their paper.

All three of the above studies find that the optimal one-bracket income tax rate increases when ability is more widely distributed, but Helpman and Sadka (1978) argue that theoretically this rate does not always increase and is not determined in general. Slemrod and Bakijia (2000) survey the papers mentioned above and tend to discount the conclusion of Helpman and Sadka (1978). They agree that the optimal one-bracket income tax should rise when earning inequality becomes more serious.

Besides the one-bracket income tax, economists have also studied more complicated taxation structures, such as the multiple-bracket income tax. Simulation results of Mirrless (1976) find that rates of the optimal marginal income tax including the rate of the top bracket should be greater than zero. In theoretical work, however, Phelps (1973), Sadka (1976), and Seada (1977) argue that the optimal income tax rate of the very top person should be zero (because to change the rate from any positive number to zero is a Pareto improvement). Stiglitz (1982, 1987) also agrees that the person with

highest ability should have a zero marginal income tax rate, and the person with lowest ability should have a positive marginal income tax. Basically, economists suggest a zero marginal tax rate of the top person because such a tax rate can encourage the richest person to work more and thus to improve total social welfare. The debate is not over. Sheshinski (1989) stands out, saying that a smaller upper bracket tax rate with a larger lower bracket tax rate is not optimal. However, Slemrod, Yitzhaki and Mayshar, and Lundholm (1994) point out that Sheshinski's proof is not reasonable. Furthermore, they show in their many simulations that the optimal marginal income tax rate of the upper bracket is smaller than that of the lower bracket.

Whereas that body of work employs two or three levels of ability dispersion, this paper investigates the whole spectrum to see how the optimal tax rate is affected by each increment to the variance of wages (holding the mean constant). In addition, whereas that body of work looks at the effect of wage dispersion on the one-bracket rate, this paper looks at effects on both rates of a two-bracket income tax. Whereas Slemrod *et al* (1994) consider only one level of wage dispersion and find that the second-bracket rate is lower than the first-bracket rate, this paper shows that the reverse pattern occurs for higher wage dispersion. Moreover, evidence suggests that the higher level of wage dispersion is now more relevant for the U. S., and especially other countries. Thus the optimal second-bracket rate is likely higher than the low-bracket rate.

3. Earning Inequality and the Optimal One-Bracket Linear Income Tax

A. The Model

Consider a simple model with N agents who have identical preferences given by the utility function:

$$U(c_i, 1-h_i) \quad (1)$$

where c_i is the consumption of individual i , and h_i is her labor supply. This function is nicely behaved in the sense that $U_1 > 0$, $U_2 > 0$, $U_{11} < 0$, and $U_{22} < 0$. Each individual has a time endowment of one and may split it between leisure and labor. Each individual randomly gets ability and corresponding wage w according to a probability distribution with $f_w(w)$ as its p.d.f. and $F_w(w)$ as its C.D.F.. Wages accepted by individuals are independent of one another. A government maximizes a particular social welfare function using a one-bracket linear tax-and-transfer system that has a lump-sum benefit b to all individuals and a constant marginal tax rate t . We assume that each individual uses all her income to consume and does not save, no matter whether she receives wage income or government transfer.

The individual's budget constraint is $b+(1-t) w_i h_i = c_i$. Thus, given w_i , b , and t , individual i maximizes:

$$U [b+(1-t) w_i h_i, 1-h_i] \quad (2)$$

by choosing her labor supply, h_i . This generates her labor supply function $h_i(w)$. It is straightforward to see that individual i participates in the labor market as long as:

$$U [b+(1-t)w_i h_i(w_i), 1-h_i(w_i)] \geq U(b,1) \quad (3)$$

where $U(b,1)$ is the utility that individual i can get if she does not provide any labor.

Let w_i^* be the wage at which individual i is indifferent between working and not working, i.e.:

$$U [b+(1-t)w_i^* h_i(w_i^*), 1-h_i(w_i^*)] = U(b,1) \quad (4)$$

This w_i^* indicates the no-envy wage developed by Foley (1967) and Varian (1974). Let:

$$P_w = \Pr (w_i \geq w_i^*) \quad (5)$$

be the probability that individual i works. Given that w_i has a C.D.F., $F_w(w_i)$, the probability can be written as:

$$P_w = 1 - F_w (w_i^*) \quad (6)$$

Because individuals are *ex ante* identical and their wages are independent of one another, subscript i in (1) to (6) can be ignored, so that all individuals have exactly the same equations.

By choosing b and t , the government maximizes a particular social welfare function (SWF) subject to a balanced government budget constraint:

$$\sum_i t w_i h(w_i) = N b \quad (7)$$

where the left side of the equation is the revenue of the government and the right side is the expenditure of the government. When the population is big enough, equation (7) can be written as:

$$E [t w h(w)] = b \quad (8)$$

Substituting w^* into (8), we get:

$$P_w E [t w h(w) | w > w^*] = b \quad (9)$$

Given $F_w(w)$ and $f_w(w)$, the balanced government budget constraint can be rewritten as:

$$t \int_{w^*}^{\infty} w h(w) f_w(w) dw = b \quad (10)$$

Social welfare functions of the government could include the Bentham SWF and the Nash SWF, both of which are utilitarian social welfare functions. Under each different SWF, the optimizing problem of the government is different:

1. The Bentham SWF. Under this criterion, the government maximizes the un-weighted sum of everybody's utility. CES utility $U(c_i, 1-h_i)$ is homothetic, but the marginal utility of consumption c declines with the amount of consumption, and so even the un-weighted sum of utilities can be raised by redistribution from a person with high c to a person with low c .¹ So, the government is averse to unequal consumption. The government maximizes the expected utility of a single person when the population is big enough, because individuals are *ex ante* identical and their wages are independent of one another. So, the government chooses t and b to maximize:

$$E \{ [U[(1-t)wh(w), 1-h(w)] \} \quad (11)$$

subject to (8). Given $F_w(w)$ and $f_w(w)$, the government chooses t and b to maximize:

$$\int_{w^*}^{\infty} U[(1-t)wh(w), 1-h(w)] f_w(w) dw + F_w(w^*) U(b, 1) \quad (12)$$

subject to (10).

2. The Nash SWF. Under this criterion, the government is averse to unequal utility itself; it maximizes the un-weighted product of the utility of all individuals. When the population is big enough, the government maximizes the expectation of the log of utility of a single person, choosing t and b to maximize:

$$E (\log \{ U[b+(1-t)wh(w), 1-h(w)] \}) \quad (13)$$

subject to (8). Given $F_w(w)$ and $f_w(w)$, the government chooses t and b to maximize:

$$\int_{w^*}^{\infty} \log \{ U[b+(1-t)wh(w), 1-h(w)] \} f_w(w) dw + F_w(w^*) \log [U(b, 1)] \quad (14)$$

¹ If income were used to buy two goods X and Y , where $U(X, Y)$ is a CES or other homothetic utility functions, then the marginal utility of income is constant, and redistribution of income cannot raise the un-weighted sum of utilities. In our case, however, no redistribution ($t=b=0$) would mean that each person uses endowment $w_i \cdot 1$ to maximize $U(w_i h_i, 1-h_i)$. To see that some redistribution can increase welfare in this case, consider the simple example where preferences involve inelastic demand for leisure ($1-h_i$). Then $U(w_i h_i, 1-h_i)$ can mean every unequal distribution of consumption $c_i = w_i h_i$, and concavity in c means that $b > 0$ can help raise total welfare.

subject to (10).

B. The Utility Functional Form

For comparability to the literature, we choose the CES utility function following Cooter and Helpman (1974), Stern (1976), and Slemrod, *et al* (1994). Let the functional form of (1) be:

$$[\alpha c_i^{(\sigma-1)/\sigma} + (1-\alpha)(1-h_i)^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)} \quad (15)$$

where σ is the elasticity of substitution between consumption and leisure, and α is the weight on consumption. Given this function, the individual chooses h_i to maximize:

$$\{\alpha[b+(1-t)w_i h_i]^{(\sigma-1)/\sigma} + (1-\alpha)(1-h_i)^{(\sigma-1)/\sigma}\}^{\sigma/(\sigma-1)} \quad (16)$$

By solving (16), we get:

$$h_i(w_i) = \{1-b[(1-\alpha)/\alpha]^\sigma [(1-t)w_i]^{-\sigma}\} / \{1+[(1-\alpha)/\alpha]^\sigma [(1-t)w_i]^{1-\sigma}\} \quad (17)$$

Inequality (3) becomes:

$$\{\alpha[b+(1-t)w_i h_i(w_i)]^{(\sigma-1)/\sigma} + (1-\alpha)[1-h_i(w_i)]^{(\sigma-1)/\sigma}\}^{\sigma/(\sigma-1)} \geq U(b, 1) \quad (18)$$

where $U(b, 1) = [\alpha b^{(\sigma-1)/\sigma} + (1-\alpha)]^{\sigma/(\sigma-1)}$ is individual i 's utility when she stays outside the labor market. The wage rate that makes inequality (18) into an equation is:

$$w_i^* = \{b^{1/\sigma} [(1-\alpha)/\alpha]\} / (1-t) \quad (19)$$

Individual i will work if and only if $w_i \geq w_i^*$. Again, since individuals are *ex ante* identical and their wages are independent of one another, subscript i in (15) to (19) can be ignored, which means that all individuals can have exactly the same equations.

C. Simulation Results with a Relatively Small σ

In order to find how the optimal tax-and-transfer system depends on earning

inequality, we first show how the values of the optimal one-bracket income tax rate t and government transfer b change with a mean-preserving spread of earning inequality. Our interpretation is that the increase of the standard deviation of a particular wage distribution describes an increased dispersion of earnings only, with no other changes (such as the mean wage). So, each particular value of the standard deviation has at least one corresponding pair of values for the optimal tax rate and transfer. By investigating those values, we may see the relationship between the tax-and-transfer program and earning inequality.

Table 1: Key Elasticities for Labor Supply of the Mean Person ²
(evaluated at $t = 0.224$, $b = 0.057$, and $\alpha = 0.6136$)

	$\sigma = 0.4$	$\sigma = 1.0$
Uncompensated Labor Supply Elasticity	-0.141	0.133
Compensated Labor Supply Elasticity	0.232	0.649
Income Elasticity	-0.373	-0.517

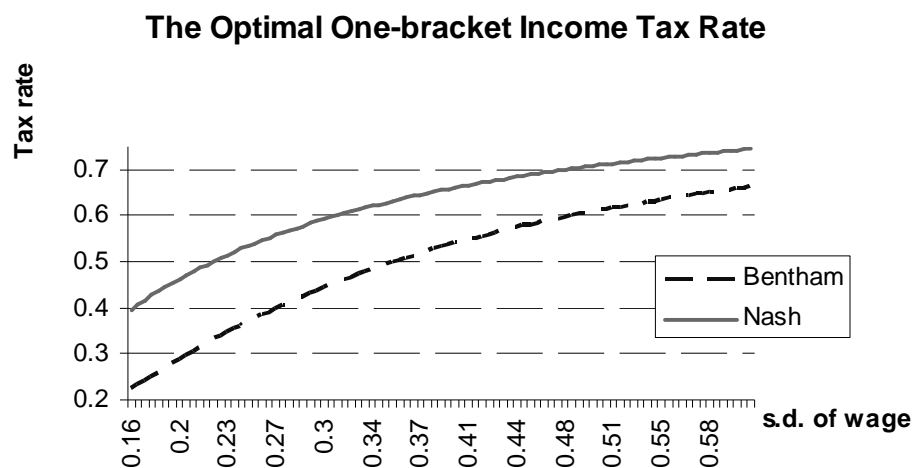
In our simulation, we first assume that the wage distribution is lognormal with a mean of 0.3969 as found by Lydall (1968) and used by Mirrless (1971) and Stern (1976) in their simulations.³ This mean wage rate represents the labor income of the person with mean wage who uses all her time endowment to work and does not rest. Mirrless (1971) says that the lognormal distribution is “intended to represent a realistic distribution of skills within the population”. Following Stern (1976), we set the elasticity of substitution

² The income elasticity is calculated by $(\partial h / \partial b) \{ [b + (1-t)wh] / h \}$. The uncompensated labor supply elasticity is not zero when σ is set to 1.0 (Cobb-Douglas utility) is because our model has non-labor income (the government transfer). The compensated labor supply elasticity is calculated by Slutsky equation: compensated elasticity = uncompensated elasticity – income elasticity.

³ The 0.3969 is the mean of the lognormal distribution used by Mirrless (1971), Stern (1976), and Slemrod *et al* (1994), the corresponding normal distribution of which has a mean of -1 and a variance of 0.39. Mirrless (1971) uses this value first. He, however, does not indicate what the real meaning of the values is and only says it is derived from a table of Lydall (1968).

between consumption and leisure in the CES utility function at $\sigma = 0.4$, and the consumption weight at $\alpha = 0.6136$. Stern (1976) argues that $\sigma = 0.4$ is a more realistic value than $\sigma = 1$ used by Mirrless (1971).⁴ Changes of σ cause changes of the elasticities for labor supply as shown by Table 1.

Figure 1



In our simulation, we change the standard deviation (s.d.) of the wage gradually from 0.1609 to 0.6109 by increments of 0.005, and for each value we calculate the optimal tax rate and transfer, keeping the mean wage constant at 0.3969. So, the coefficient of variation (c.v.) of wage changes from 0.405 to 1.539.⁵ The 0.1609 represents very moderate earning inequality, as used by Mirrless (1971) and Stern (1976),

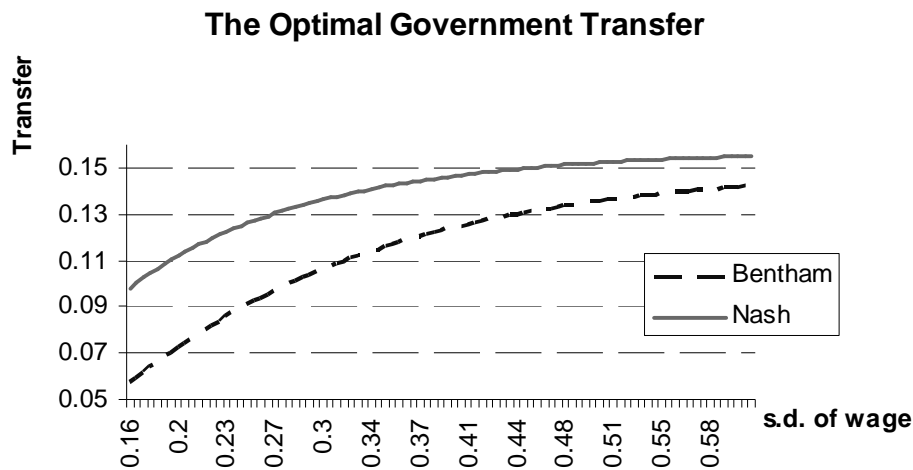
⁴ The reason that 0.6136 is chosen as the value of α by Stern (1976) is because when σ and α are set to 0.5 and 0.6136, a person facing no tax and transfer would to use two thirds of her time endowment to work.

⁵ The coefficient of variation of the wage rate in the U. S. varies from 0.590 to 0.888 during the period between 1979 and 2004 (by data from CPS MORG 1979-2004, NBER). That of Mexico varies from 1.561 to 2.721 during the period between 1995 and 1999 (by data from INEGI). Assuming that income inequality is highly correlated with wage inequality, we expect to see even larger values from most other developing countries due to the famous Kuznets Curve (Kuznets, 1955) that says income inequality increases when a country starts to be industrialized but finally decreases when it becomes a developed country. Glaeser (2005) confirms this relationship. An updated Kuznets Curve with 1998 data from the World Bank can be found in his paper.

while the 0.6109 represents quite serious earning inequality.⁶

Figure 1 shows how the optimal one-bracket income tax rate reacts under both the Bentham SWF and the Nash SWF when earning inequality changes from the moderate level to the serious level. We find that under both SWFs, the optimal rate is strictly increasing with the standard deviation of wage with no exceptions. When the spread is relatively low, such as 0.1609, the optimal rate is 0.224 under the Bentham SWF (0.397 under the Nash SWF).⁷ When the spread is extremely large, such as 0.6109, the optimal rate is as big as 0.664 under the Bentham SWF (0.745 under the Nash SWF). Intuition here is straightforward. When earning inequality becomes more serious, more individuals drop into the low income class and depend on government transfer to live. Thus, the government needs to collect more revenue from those working to subsidize the others.

Figure 2



Beyond this, the optimal rates under the Bentham SWF are always larger than

⁶ The 0.1609 is the s.d. of the lognormal distribution used by Mirrless (1971), Stern (1976), and Slemrod *et al* (1994), the corresponding normal distribution of which has a mean of -1 and a variance of 0.39.

under the Nash SWF. This is because the Nash SWF puts more weight on the utility of the poor than does the Bentham. Hence, the government needs to have higher tax rates that can collect more revenue to finance more transfers to the poor. With respect to the government transfer b , it is also strictly increasing with the wage spread under both SWFs as shown in Figure 2. The optimal transfer grows from 0.057 (roughly 14.4% of the mean wage) to 0.142 (35.8%) under the Bentham SWF (from 0.098 (24.7%) to 0.155 (39.1%) under the Nash SWF) when the standard deviation increases from 0.1609 to 0.6109. As expected, the optimal transfers under the Nash SWF are all bigger than under the Bentham SWF.

D. Simulation Results with a Relatively Large σ .

Though Stern (1976) believes that the small value of the elasticity of substitution between consumption and leisure ($\sigma = 0.4$) is more realistic than larger values, the value of $\sigma = 1$ used by Mirrless (1971) is still of interest at least for comparison. As shown in Figure 3 and 4, we repeat the simulations above with $\sigma = 1.0$, while holding other parameters unchanged. Basically, a larger σ means a larger uncompensated labor supply elasticity as shown by Table 1.

We find that the change of σ from 0.4 to 1.0 does not affect our conclusion that the optimal one-bracket linear income tax rate and government transfer are strictly increasing with the wage spread. However, the increase of σ shifts down both the optimal tax rate and the optimal transfer. For the s.d. = 0.1609 used by Mirrless (1971) and Stern (1976), as shown in Figure 3, the optimal rate drops from 0.224 to 0.126 under

⁷ Stern (1976) gets an optimal tax rate of 0.223 under the Bentham SWF. The rest of the horizontal axis in Figure 1 is new.

the Bentham SWF. As shown in Figure 4, the optimal transfer drops from 0.057 to 0.029 under the Bentham SWF (The optimal rate drops from 0.397 to 0.229, while the optimal transfer drops from 0.098 to 0.050 under the Nash SWF). Obviously, increases of the uncompensated labor supply elasticity force the government to implement smaller and smaller tax rates. By using lower tax rates, the government encourages elastic workers to work, so that enough revenue can be collected from them to finance government transfers.

Figure 3

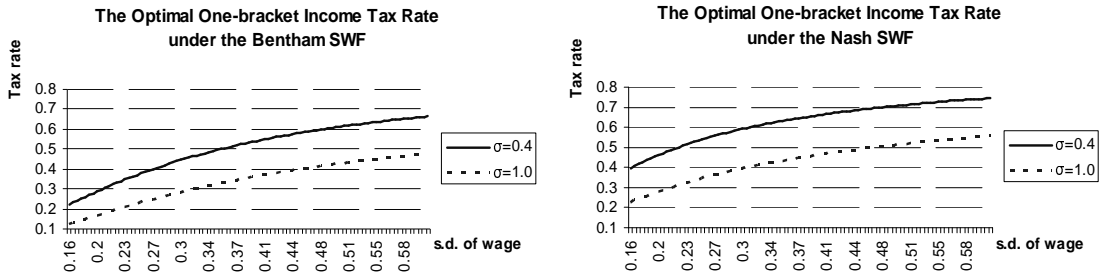
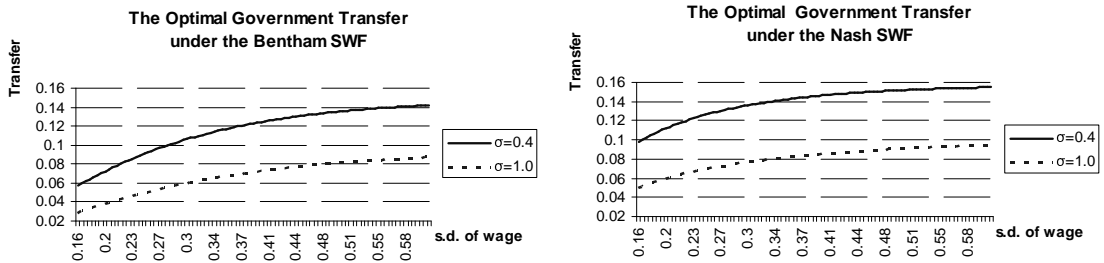


Figure 4



4. Earning Inequality and the Optimal Two-Bracket Linear Income Tax

A. The Model

We continue to use the same SWFs and lognormal distributions of ability as used in the previous section, and in this section we investigate the relationship between earning inequality and the optimal two-bracket linear income tax. Compared to the one-bracket case, the two-bracket case is more complicated. Though the preferences of

individuals have not changed, the budget constraint has changed because of the introduction of the second marginal tax rate. Explicitly, individual i now chooses h_i to maximize:

$$U [b+(1-t_1) \min(w_i h_i, \hat{Y})+(1-t_2) \max(w_i h_i - \hat{Y}, 0), 1-h_i] \quad (20)$$

where h_i is the labor supply of individual i , t_1 is the marginal tax rate of the first bracket, t_2 is the marginal tax rate of the second bracket, and \hat{Y} is the threshold between the first income bracket and the second income bracket. We still use the CES utility function (15) as the utility functional form. Then (20) becomes:

$$\{\alpha [b+(1-t_1) \min(w_i h_i, \hat{Y})+(1-t_2) \max(w_i h_i - \hat{Y}, 0)]^{(\sigma-1)/\sigma} + (1-\alpha)(1-h_i)^{(\sigma-1)/\sigma}\}^{\sigma/(\sigma-1)} \quad (21)$$

Again, because all individuals are *ex ante* identical, and their wages are independent of one another, subscript i can be ignored in (20) and (21). The government now has four policy tools instead of two: one government transfer, one income threshold and two marginal income tax rates. Therefore, under Bentham's additive SWF, the government chooses t_1 , t_2 , b , and \hat{Y} to maximize:

$$\int_{w^*}^{\infty} U\{b+(1-t_1) \min[wh(w), \hat{Y}] + (1-t_2) \max[wh(w) - \hat{Y}, 0], 1-h(w)\} f_w(w) dw + F_w(w^*) U[b, 1] \quad (22)$$

subject to the balanced budget constraint:

$$\int_{w^*}^{\infty} \{t_1 \min[wh(w), \hat{Y}] + t_2 \max[wh(w) - \hat{Y}, 0]\} f_w(w) dw = b \quad (23)$$

where $f_w(w)$ is the p.d.f. of ability, $F_w(w)$ is the C.D.F of ability, and w^* is the labor market participation condition that fulfills:

$$U \{b+(1-t_1) \min[w^* h(w^*), \hat{Y}] + (1-t_2) \max[w^* h(w^*) - \hat{Y}, 0], 1-h(w^*)\} = U(b, 1) \quad (24)$$

If and only if $w_i < w^*$, individual i stays outside the labor market. Under the multiplicative Nash SWF, the government chooses t_1 , t_2 , b , and \hat{Y} to maximize:

$$\int_{w^*}^{\infty} \log \{ U[b+(1-t_1)\min(wh(w), \hat{Y})+(1-t_2)\max(wh(w)-\hat{Y}, 0), 1-h(w)] \} f_w(w) dw + F_w(w^*) \log[U(b,1)] \quad (25)$$

subject to (23).

Since both the individual's and the government's problem are highly non-differentiable, we follow Slemrod, *et al* (1994) by using approximating methods to simulate the relationship between the optimal two-bracket linear income tax and earning inequality. We draw 2000 points from each lognormal distribution of ability used in this section to represent a wage distribution. Each point accounts for a 0.0005 increase in the cumulative frequency of the wage. The lowest cumulative frequency is 0.0005 while the highest is 0.9995. Without losing generality, we assume only 2000 individuals live in the economy, and each is exclusively assigned a wage from the 2000 wages drawn. In the approximation, the individual's problem does not change at all. Each individual still maximizes (20) by choosing labor supply h . However, the government's problem changes a little bit. Under the Bentham SWF, the government now chooses t_1 , t_2 , b , and \hat{Y} to maximize:

$$\sum_i U_i \{ b+(1-t_1) \min[w_i h_i(w_i), \hat{Y}] + (1-t_2) \max[w_i h_i(w_i) - \hat{Y}, 0], 1-h(w_i) \} \quad (26)$$

subject to the balanced government budget constraint:

$$\sum_i \{ t_1 \min[w_i h_i(w_i), \hat{Y}] + t_2 \max[w_i h_i(w_i) - \hat{Y}, 0] \} = 2000 b \quad (27)$$

where i ranges from 1 to 2000, and w_i is the wage of individual i . Under the Nash SWF, the government chooses t_1 , t_2 , b , and \hat{Y} to maximize:

$$\sum_i \log \{ U_i [b+(1-t_1) \min(w_i h_i(w_i), \hat{Y}) + (1-t_2) \max(w_i h_i(w_i) - \hat{Y}, 0), 1-h(w_i)] \} \quad (28)$$

subject to (27). Actually, the government has only three free choices from the four tools, because the fourth tool can be solved out by the balanced government constraint (27). In our simulations, government transfer b is solved out, leaving t_1 , t_2 , and \hat{Y} as the

chosen variables.

B. Simulation Results with a Relatively Small σ

In this section, we simulate a case where the elasticity of substitution between consumption and leisure (σ) is set to as small as 0.4.⁸ This value is from Stern (1976) and followed by Slemrod, *et al* (1994). In this simulation, α is set to 0.6136 following Stern (1976) as in the one-bracket cases.⁹ Table 2 shows the key elasticities for labor supply of the mean person when she faces a two-bracket income tax. Figure 5 shows the effect of earning inequality on optimal tax rates under both the Bentham SWF and the Nash SWF.

Table 2: Key Elasticities for Labor Supply of the Mean Person¹⁰
(evaluated at $t_1 = 0.234$, $t_2 = 0.200$, $b = 0.059$, $\hat{Y} = 0.315$, and $\alpha = 0.6136$)

	$\sigma = 0.4$	$\sigma = 1.0$
Uncompensated Labor Supply Elasticity	-0.139	0.140
Compensated Labor Supply Elasticity	0.236	0.666
Income Elasticity	-0.375	-0.526

When the wage standard deviation is relatively small, we find the result of Slemrod *et al* (1994) that the optimal lower bracket rate is greater than the upper bracket rate.¹¹ When the standard deviation is relatively large, however, then we find the

⁸ We change σ around 0.4 from 0.3 to 0.7 by 0.1 to check the sensitivity of our simulation. Figure A1 in Appendix A shows results when σ is set to 0.3, 0.5, 0.6 and 0.7. These alternatives do not change our conclusion at all.

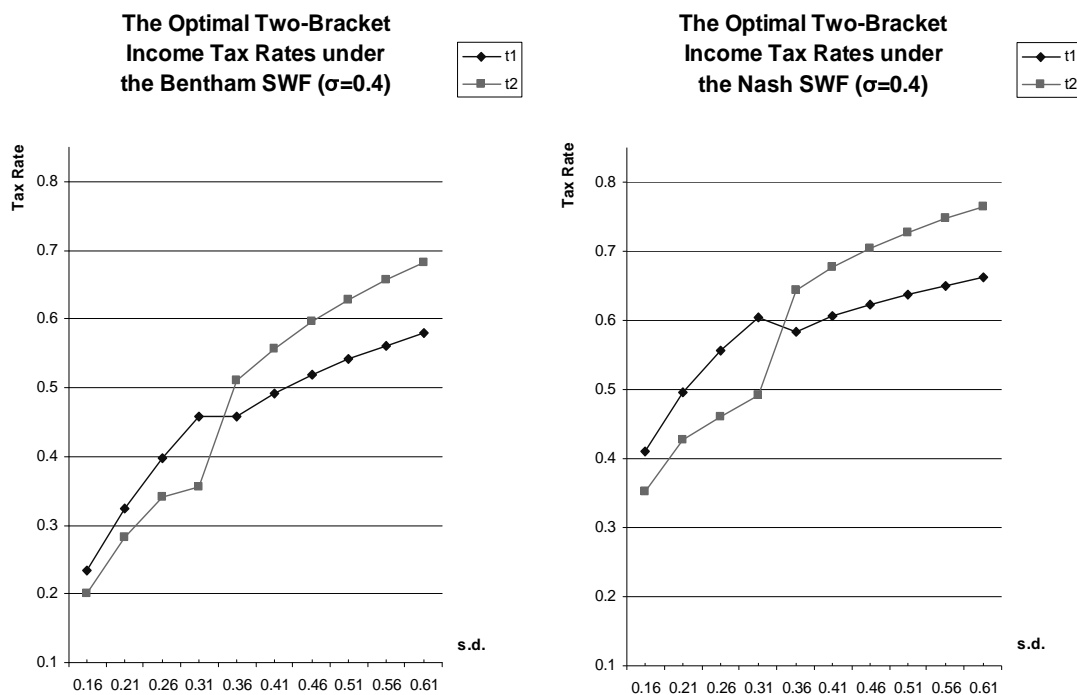
⁹ Slemrod *et al* (1994) use varied α in their simulations. Particularly, they use 0.41 in their $\sigma = 0.4$ case, which put a less-than-half weight to consumption.

¹⁰ The income elasticity is calculated by $(\partial h / \partial b) \{ [b + (1 - t_1) \min(wh, \hat{Y}) + (1 - t_2) \max(wh - \hat{Y}, 0)] / h \}$. The uncompensated labor supply elasticity is not zero when σ is set to 1.0 (Cobb-Douglas utility) is still because of non-labor income (the government transfer). The compensated labor supply elasticity is calculated by Slutsky equation: compensated elasticity = uncompensated elasticity - income elasticity

¹¹ In the case where $\sigma = 0.1609$ and the SWF is the Bentham SWF, Slemrod *et al* (1994) find that t_1 , t_2 , b , and \hat{Y} equal 0.234, 0.202, 0.058 and 0.300 respectively. We find that they are 0.234, 0.200, 0.059 and

opposite result.¹² This “switchover point” appears when the standard deviation is somewhere between 0.3109 and 0.3609, for both SWFs, roughly twice as big as the 0.1609 used by Slemrod, *et al* (1994) as their only earning inequality level. The coefficient of variation at the switchover point is between 0.783 and 0.909 (for both SWFs).¹³ In contrast, because Slemrod *et al* (1994) use only s.d. = 0.1609, they find that the optimal lower bracket rate is always greater than the upper bracket one. Surprisingly, when we allow for greater possible wage inequality, we show that their result does not always hold.

Figure 5



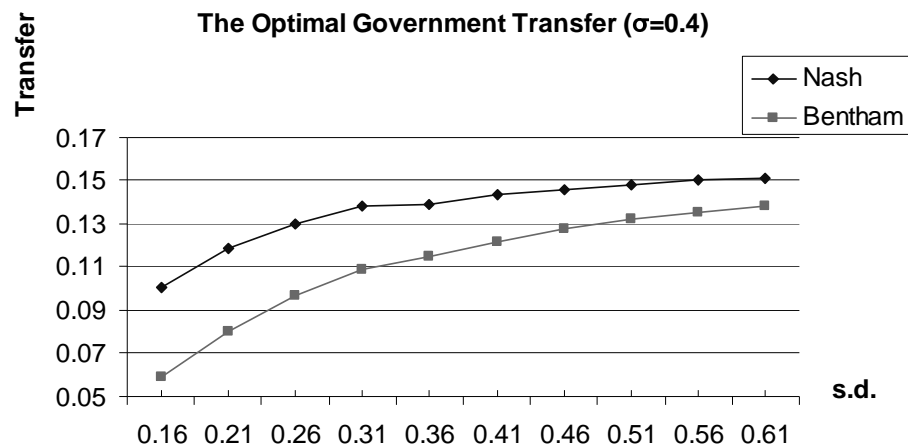
0.315. Those two groups of values differ from each other slightly because Slemrod *et al* use $\alpha = 0.41$ and we use $\alpha = 0.6136$ following Stern (1976).

¹² In all ten cases including σ is set to 0.3, 0.4, 0.5, 0.6 and 0.7, we find this switchover appearing between 0.2109 and 0.4109. Moreover, six out of ten times, the switchover appears between 0.3109 and 0.3609.

¹³ The highest three coefficients of variation of the wage rate of the U. S. between 1979 and 2004 are 0.888 (1993), 0.802 (2004), and 0.793 (1992), values that are in this interval. The s.d. = 0.1609 used by Slemrod *et al* (1994) yields a coefficient of variation equal to 0.405, which is outside the range of 0.590 to 0.888 witnessed in the U. S. from 1979 to 2004.

Furthermore, under both SWFs, the optimal upper bracket rate (t_2) is always increasing with the wage spread (for our parameters), whereas the optimal lower bracket rate (t_1) is increasing overall but not around the switchover point. For example, when earning inequality changes from a mild level where the wage spread is 0.1609 used by Mirrless (1971), Stern (1976) and Slemrod *et al* (1994) to the extreme level where the spread 0.6109, the optimal upper bracket rate increases monotonically from 0.200 to 0.682 under the Bentham SWF (from 0.351 to 0.764 under the Nash SWF). Surprisingly, the optimal lower bracket rate does not increase monotonically. Overall, it increases from 0.234 to 0.579 under the Bentham SWF (from 0.410 to 0.662 under the Nash SWF). In the Nash case, the lower bracket rate falls a bit when the standard deviation rises from 0.3109 to 0.3609 (where the rate decreases slightly from 0.604 to 0.583). Though the lower bracket rate does not have a setback in the Bentham case when $\sigma = 0.4$, it is quite flat in the switching area, changing from 0.45788 to 0.45849 (and it does fall near the switchover point when σ is set to 0.3 or 0.5).

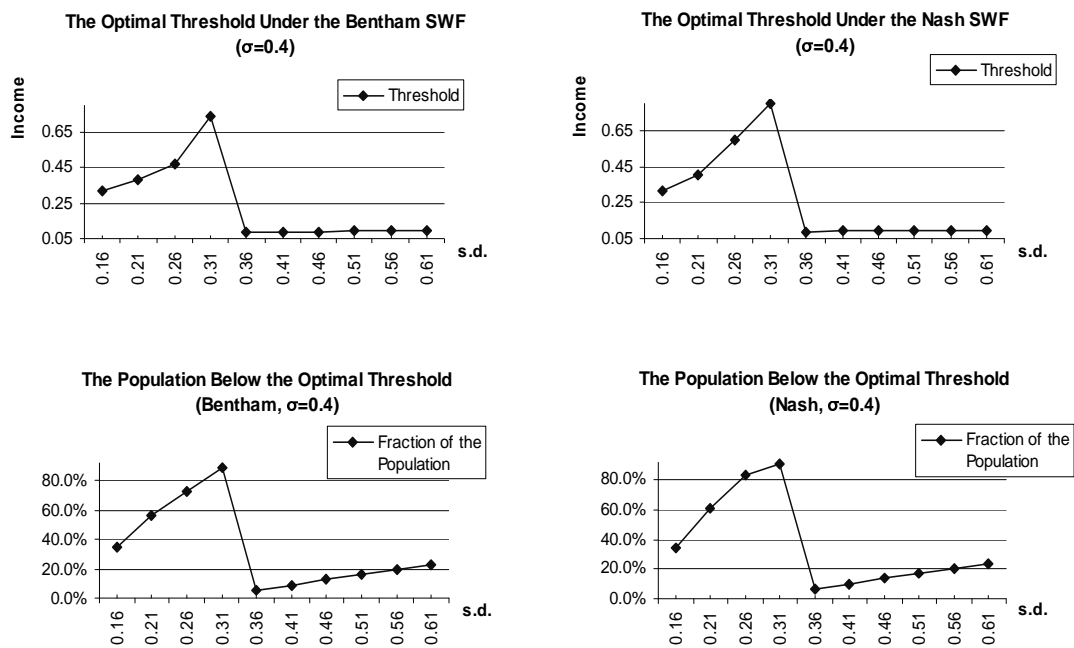
Figure 6



Regarding the optimal government transfer, as shown in Figure 6, it is strictly

increasing with the wage spread under both SWFs.¹⁴ When the wage standard deviation changes from 0.1609 to 0.6109, the optimal transfer grows from 0.059 (roughly 14.9% of the mean wage) to 0.138 (34.8%) under the Bentham SWF (from 0.101 (25.4%) to 0.151 (38.0%) under the Nash SWF). As shown in Figure 7, the optimal income threshold (\hat{Y}) that divides the two brackets does not show a monotone property. Moreover, both of the optimal rates (t_1 and t_2) under the Bentham SWF are larger than under the Nash SWF, while the optimal transfer (b) under the Bentham SWF is less than under the Nash SWF, a result that is similar to the one-bracket case. It is still because the Nash SWF puts more weight on the poor.

Figure 7



Several reasons together explain our conclusions regarding an increase in wage spread. First, the reason for the overall increases of both optimal rates is that the government has to increase both rates to collect necessary revenue to help support the

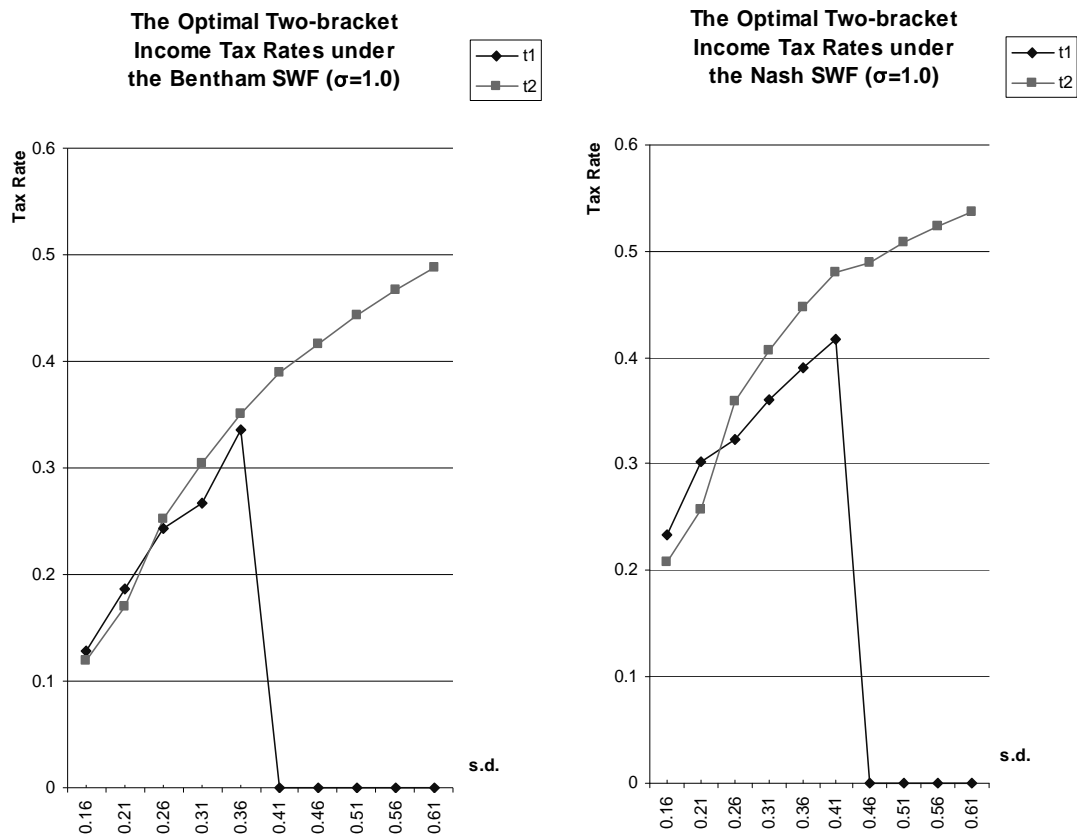
¹⁴ Please also see simulation results for other values of σ in Figure A2 of Appendix A.

poor when earning inequality become more serious. This is comparable to the one-bracket case. Second, when earning inequality is relatively mild, the population of the “middle class” that pays positive net taxes but whose income is still less than or equal the threshold \hat{Y} is relatively large. For example, the optimal threshold, as shown in Figure 7, can be as high as 0.737 under the Bentham SWF (0.796 under the Nash SWF), which are almost twice as big as the mean wage, 0.3969. The result is that 91.2% of the population does not pay the upper bracket tax under the Bentham SWF (89.3% under the Nash SWF). The large population of the middle class means that the taxable income of this class is also large. Thus the government is able to raise a substantial amount of revenue from the middle class in the first bracket (facing t_1). As a result, the government does not need to raise substantial revenue from the rich who earn more than the threshold. Without losing revenue, the government can implement smaller upper bracket rates, to encourage labor supply of the rich, those who are the most productive workers in the economy. However, as earning inequality rises to a high level, the middle class shrinks rapidly. As shown in Figure 7, the optimal threshold rises gradually and then decreases dramatically from 0.737 to 0.082 under the Bentham SWF (from 0.796 to 0.088 under the Nash SWF). The population below the threshold cut from 91.2% to 5.0% under the Bentham SWF (from 89.3% to 6.0% under the Nash SWF). The majority of the tax base is then shifted from the lower bracket to the top bracket. Finally, since the government is then treated the rich as the major target, it is able to give those with low income a smaller lower bracket rate that encourages them to work and improve their welfare and therefore total social welfare as well.

C. Simulation Results with a Relatively Large σ

In addition to the simulation with a small σ , we repeat the simulation approach, but changing σ from 0.4 used by Stern (1976) to 1.0 used by Mirrless (1971).¹⁵ Again, as shown in Table 2, the increase of σ means the increase of the uncompensated labor supply elasticity.

Figure 8

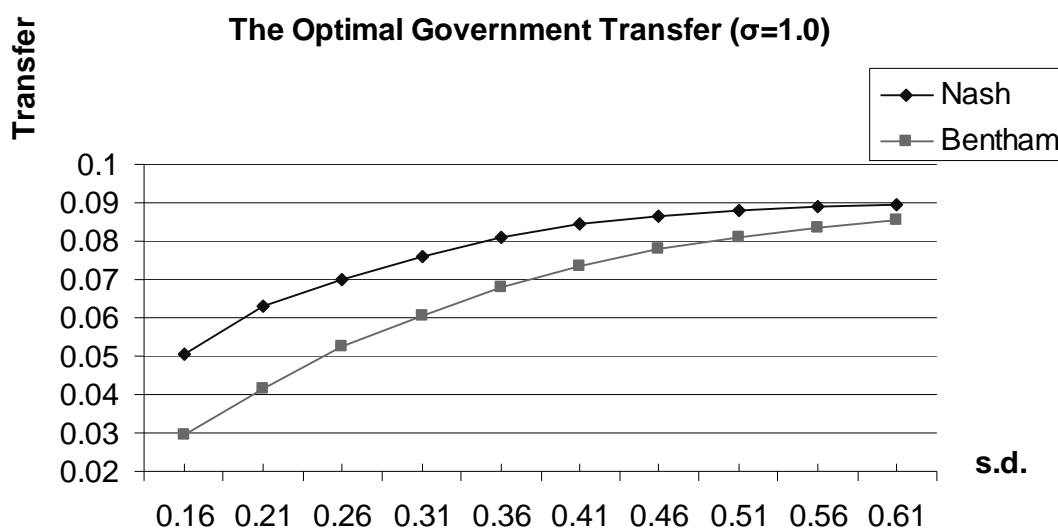


As shown in Figure 8 and 9, we find that the optimal upper bracket rate and government transfer are increasing with the wage spread under both SWFs. The optimal lower bracket rate, however, is not monotonic. Figure 10 shows the optimal threshold is

¹⁵ Other values of σ , such as 0.9 and 1.1 that are around 1.0, are also simulated to check the sensitivity of our simulation. To set σ to be 0.9 or 1.1 does not change our conclusion at all. Please see the simulation results in Table 1 and Table 2 in Appendix B.

not monotonic either. In addition, we still find a switchover point where t_2 rises above t_1 under both SWFs (when the standard deviation is between 0.2109 and 0.2609 as shown in Figure 8).¹⁶ The coefficient of variation of the switchover point is between 0.531 and 0.657.¹⁷ Before the switch, the optimal lower bracket rate is greater than the optimal upper bracket one. After the switch, the upper bracket rate is higher.

Figure 9



Furthermore, we find that the two-bracket tax structure converges completely to the one-bracket case under both SWFs when earning inequality becomes quite serious.¹⁸ As shown in Figure 8, the optimal upper bracket rate is always greater than zero and is increasing with the wage spread under both SWFs, whereas the optimal lower bracket rate stays positive only before the point where $s.d. = 0.4109$ and $c.v. = 1.035$ under the

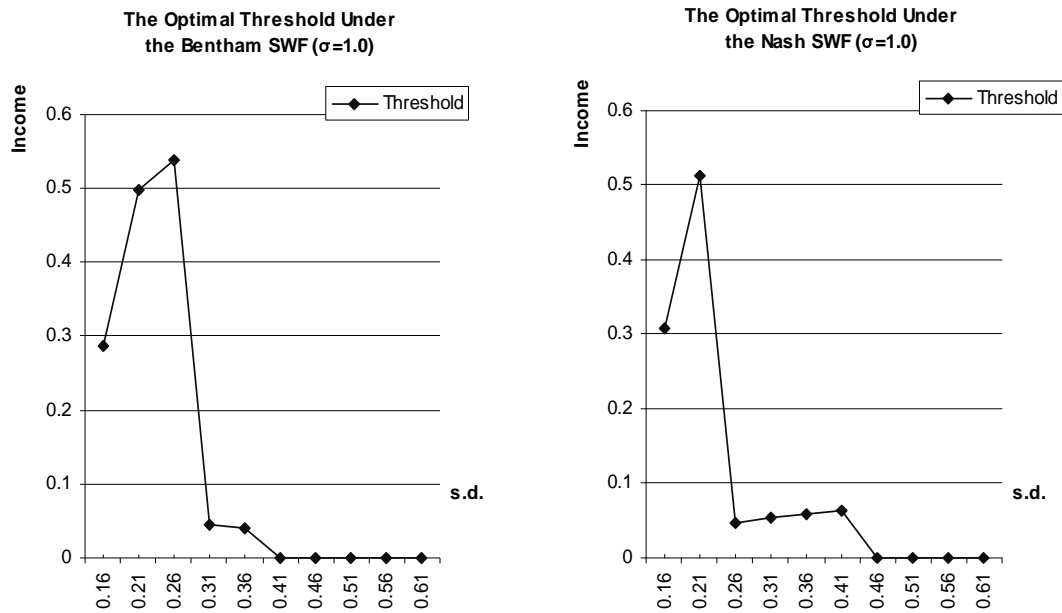
¹⁶ In all six cases including $\sigma = 0.9, 1.0,$ and $1.1,$ we find this switchover appearing between 0.2109 and 0.3609. Please see these switchovers in Table B1 and B2 of Appendix B.

¹⁷ Most coefficient of variations of wage of the U.S. ranging from 0.590 to 0.888 from 1979 to 2004 are included in this interval, while 0.405 generated by $s.d. = 0.1609$ is still not included.

¹⁸ We find this in all six cases where $\sigma = 0.9, 1.0$ and $1.1.$ Please see Table B1 and B2 of Appendix B.

Bentham SWF (s.d. = 0.4609 and c.v. 1.161 under the Nash SWF).¹⁹ It then drops to zero after that. In addition, when the lower bracket rate drops to zero, the threshold drops to zero also. This means that the two-bracket structure converges to one-bracket.²⁰

Figure 10



This convergence is actually a special case of the switchover, in which the optimal lower bracket rate falls to as small as zero, and the threshold drops to zero also. First, as explained in the previous section, when earning inequality is quite serious, the optimal lower bracket rate is smaller than the optimal upper bracket rate to help the middle class. Second, workers of the middle class are very elastic because large σ means a large uncompensated labor supply elasticity. This forces the government to use an even smaller lower bracket rate to keep the middle class working and to prevent them

¹⁹ The coefficients of variation of Mexico from 1995 to 1999 are all larger than the 1.035 of the Bentham case (larger than the 1.161 of the Nash case also). Given that most developing countries have similar c.v. with that of Mexico, and most developed countries have similar c.v. with that of the U. S., a one-bracket income tax could be more suitable for developing countries than developed countries.

²⁰ Moreover, the two-bracket structure converges earlier as σ becomes larger from 0.9 to 1.1 as shown in Table B1 and Table B2 of Appendix B.

from becoming net welfare recipients. Last, the pressure of raising revenue drives the government to decrease the threshold, and to enlarge the population in the higher bracket, to collect more taxes that can be used to finance government transfers. All these causes interacting together imply that the middle class disappears while the two-bracket structure of taxation becomes one bracket.

5. Conclusions

Consistent with the results of Mirrless (1971), Stern (1976), and Cooter and Helpman (1974), our simulations generally favor the conclusion for the one-bracket case that both the optimal income tax rate and the government transfer increase when earnings become more unequally distributed. Moreover, we go on to show that the tax rate and transfer are strictly increasing with the wage spread. This conclusion does not depend on whether a relatively small or large elasticity of substitution between consumption and leisure is used in the simulation. A larger value of the elasticity changes only the magnitude but not the trend.

In the two-bracket case, we similarly find that the optimal upper bracket rate and government transfer are also always increasing with the wage spread. When the substitution elasticity is relatively small, the optimal lower bracket rate is increasing with wage disparity overall, but not in the area near the switchover point. It is not monotonic when the elasticity is large. We confirm results of Slemrod *et al* (1994) for a relatively low wage disparity that the upper bracket rate is less than the lower bracket rate. With a wage spread close to that of the U. S. in recent years, however, the result is reversed. Beyond this, we also find an interesting phenomenon. With a relatively large elasticity of

substitution between consumption and leisure in the individual's utility function, the optimal two-bracket income tax structure converges to the one-bracket case when earning inequality becomes serious. Though this can be treated as a special case of the switchover, it is still surprising that the lower bracket rate and the income threshold can be as low as zero. Furthermore, this theoretically simulated result may indicate that developing countries with serious income inequality may need to implement the one-bracket income tax structure instead of the multiple-bracket structure.

Reference:

Cooter, R. and Helpman, E. (1974), "Optimal Income Taxation for Transfer Payments under Different Social Welfare Criteria", *Quarterly Journal of Economics* 88: 656-670.

Glaeser, E. (2005), "Inequality", NBER working paper series #11511.

Foley, D. (1967), "Resource Allocation and the Public Sector", *Yale Economic Essays* 7: 45-98.

Helpman, E. and Sadka, E. (1978), "The Optimal Income Tax", *Journal of Public Economics* 9: 383-393.

Kopczuk W., J. Slemrod and S. Yitzhaki (2005), "The Limitations of Decentralized World Redistribution: An Optimal Taxation Approach" , *European Economic Review*, 49: 1051-1079.

Kuznets, S. (1955), "Economic Growth and Income Inequality", *American Economic Review* 65: 1-28.

Lydall, H. F. (1968), *The Structure of Earnings*, Oxford, U.K..

- Mirrless, J. (1971), “An Exploration in the Theory of Optimum Income Taxation”, *Review of Economic Studies* 38: 175-208.
- Instituto Nacional De Estadística, “Encuesta Nacional de Empleo Urbano (National Urban Employment Survey) 1995-1999”, Geografica e Informatica (National Statistics Institute), Mexico.
- National Bureau of Economic Research, Inc. (NBER), “CPS Merged Outgoing Rotation Groups (CPS MORG) 1979 – 2004”.
- Phelps, E. (1973), “The Taxation of Wage Income for Economic Justice”, *Quarterly Journal of Economics* 3: 331-354.
- Sadka, E. (1976), “On Income Distribution, Incentive Effects, and Optimal Income Taxation”, *Review of Economic Studies* 9: 261-268.
- Seade, J. (1977), “On the Shape of Optimal Tax Schedules”, *Journal of Public Economics* 7: 203-236.
- Sheshinski, E. (1989), “Note on the Shape of the Optimum Income Tax Schedule”, *Journal of Public Economics* 40: 201-215.
- Stiglitz, J. (1982), “Self-selection and Pareto Efficient Taxation”, *Journal of Public Economics* 17: 213-240.
- Stiglitz, J. (1987), “Pareto Efficient and Optimal Taxation and the New Welfare Economics”, *Handbook of Public Economics* V2, North-Holland, Amsterdam.
- Slemrod, J., Yitzhaki S. and Mayshar J., and Lundholm, M. (1994), “The Optimal Two-Bracket Linear Income Tax”, *Journal of Public Economics* 53:269-290.
- Slemrod, J. and Bakija, J. (2000), “Does Growing Inequality Reduce Tax Progressivity? Should It?”, NBER working paper series #7576.

Stern, N. H. (1976), "On the Specification of Models of Optimum Income Taxation",

Journal of Public Economics 6: 123-162.

Varian, H. (1974), "Equity, Envy, and Efficiency", *Journal of Economic Theory* 9: 63-91.

Appendix A

Figure A1

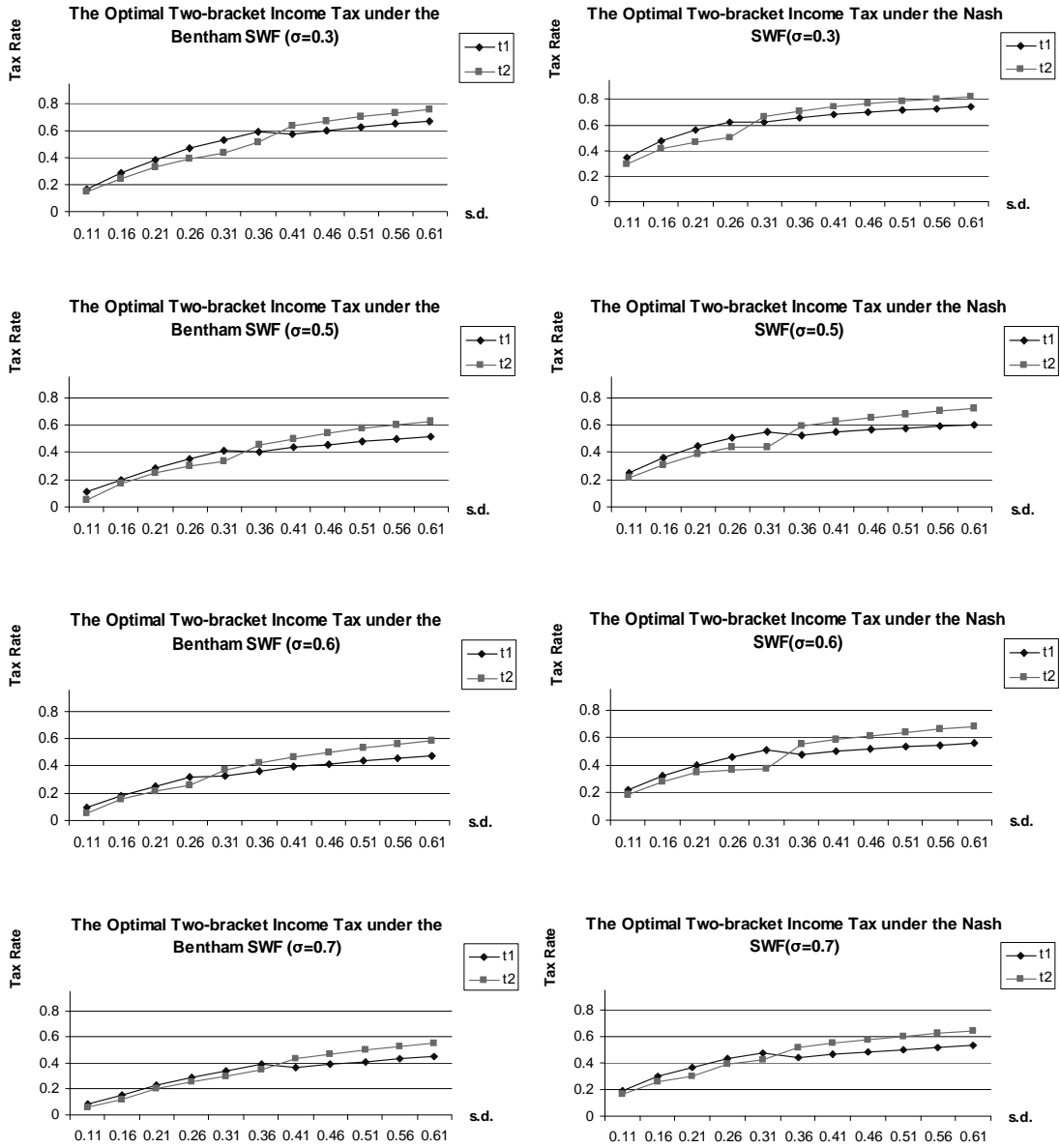
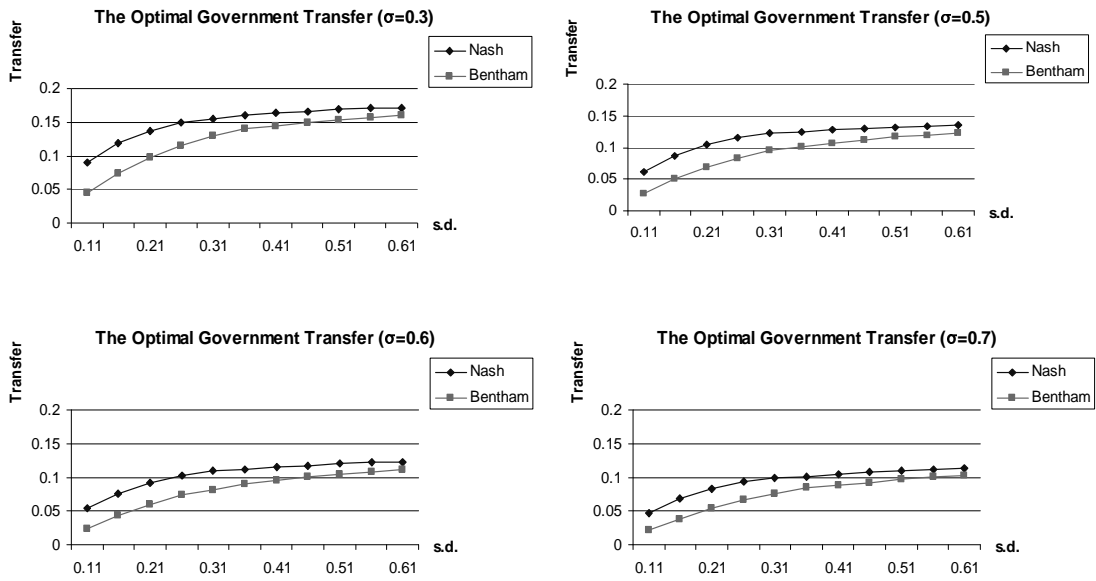


Figure A2



Appendix B

Table B1

The Optimal Two-bracket Income Tax with Large σ under the Bentham SWF

σ \ s.d.	0.1609			0.2109			0.2609			0.3109			0.3609		
	T_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}
0.9	0.142	0.119	0.397	0.196	0.171	0.492	0.224	0.252	0.044	0.272	0.304	0.051	0.308	0.350	0.053
1.0	0.134	0.118	0.287	0.187	0.164	0.498	0.244	0.220	0.539	0.267	0.293	0.045	0.336	0.336	0.041
1.1	0.128	0.113	0.316	0.181	0.167	0.387	0.235	0.209	0.632	0.286	0.251	0.830	0.000	0.328	0.000
σ \ s.d.	0.4109			0.4609			0.5109			0.5509			0.6109		
	T_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}
0.9	0.338	0.390	0.056	0.000	0.417	0.000	0.000	0.444	0.000	0.000	0.467	0.000	0.000	0.487	0.000
1.0	0.000	0.374	0.000	0.000	0.406	0.000	0.000	0.434	0.000	0.000	0.458	0.000	0.000	0.478	0.000
1.1	0.000	0.366	0.000	0.000	0.399	0.000	0.000	0.428	0.000	0.000	0.453	0.000	0.000	0.475	0.000

Table B2

The Optimal Two-bracket Income Tax with Large σ under the Nash SWF

σ \ s.d.	0.1609			0.2109			0.2609			0.3109			0.3609		
	T_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}
0.9	0.251	0.223	0.287	0.323	0.287	0.375	0.380	0.344	0.471	0.373	0.426	0.055	0.403	0.466	0.059
1.0	0.234	0.208	0.307	0.303	0.257	0.514	0.323	0.359	0.047	0.361	0.407	0.053	0.391	0.447	0.058
1.1	0.221	0.201	0.273	0.288	0.256	0.461	0.344	0.302	0.659	0.351	0.391	0.052	0.000	0.421	0.000
σ \ s.d.	0.4109			0.4609			0.5109			0.5609			0.6109		
	T_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}	t_1	t_2	\hat{Y}
0.9	0.429	0.500	0.062	0.451	0.528	0.065	0.470	0.552	0.069	0.487	0.573	0.072	0.503	0.591	0.075
1.0	0.417	0.480	0.063	0.000	0.489	0.000	0.000	0.508	0.000	0.000	0.524	0.000	0.000	0.538	0.000
1.1	0.000	0.450	0.000	0.000	0.473	0.000	0.000	0.492	0.000	0.000	0.508	0.000	0.000	0.522	0.000

Li Gan,

Department of Economics

Texas A&M University

4228 TAMU

College Station, TX 77843-4228

Telephone: 1-979-862-1667

Fax: 1-979-847-8757

Email: gan@econmail.tamu.edu

Jijun Tan

School of Economics and Institute for Advanced Research

Shanghai University of Finance and Economics

777 Guoding Road

Shanghai, China 200433

Telephone: 86-21-65902211

Fax: 86-21-65904198

Email: tanjijun@mail.shufe.edu.cn

