Optimal Income Taxation and Job Choice

by

Robin Boadway[†]

Zhen Song[††]

Jean-François Tremblay[†††]

July, 2013

Abstract

This paper studies optimal income taxation when there are different types of jobs for workers of different skills. Each type of job has a given feasible range of incomes from which workers can choose by varying their labour supply. Workers are more productive than all others in the jobs that suit them best. The model combines features of the classic optimal income tax literature with labour variability along the intensive margin with those of the extensive-margin approach where workers make discrete job choices and/or participation decisions. Some specific results are as follows. First-best maximin levels of utility can be achieved in the second-best. Marginal tax rates below the top can often be negative or zero. When there are more than two skill-types of workers and jobs, incentive constraints are not necessarily binding on adjacent types as in the standard intensive-margin model. When participation decisions are allowed, the intensive margin and the extensive margin tend to have opposite effects on the level of participation taxes.

Keywords: optimal income tax, job choice, intensive margin, extensive margin

JEL classification: H21, H23, H24

Acknowledgments: We are grateful to Laurence Simard for careful comments on an earlier version of this paper. Research support was provided by the Social Sciences and Humanities Research Council of Canada.

[†] Queen's University, Kingston, Canada

[††] CEMA & CIAS, Central University of Finance & Economics, Beijing, China

[†††] University of Ottawa, Ottawa, Canada

1 Introduction

In models of optimal income taxation, the focus has been largely on the labour supply side. The nature of jobs offered has not been explicitly modeled. Two extreme cases have been studied. The standard intensive-margin approach, following Mirrlees (1971), effectively assumes that workers of different skills are perfect substitutes in production. The amounts of effective labour per hour of work reflects skills, and workers receive a fixed wage rate equal to their skill. In extensive-margin models, jobs are offered that are suited for each type of skill, and that pay a fixed wage for a given amount of effort (Diamond 1980). In some versions of the extensive-margin model, workers of a given skill can choose the job suited for a less-skilled worker, but if they do they receive the same fixed wage (Saez 2002).

These two approaches have led to important insights into the structure of optimal income taxes and how they are affected by the nature of labour supply decisions, but they have set aside characteristics of jobs that are potentially relevant for income tax design. For one thing, workers have different job-specific skills, which may make them more productive than others in the job that suits them best. Given the division of labour in a modern economy and the specialization that entails, a worker tends to perform a narrow set of tasks within a work unit. Each set of tasks requires a corresponding set of special skills that workers possess in different degrees, such as innate ability, education and training, and work experience. Workers will have different comparative advantages at jobs requiring different sets of skills. Moreover, some workers may have an absolute advantage over workers with different skills at those jobs that suit them best. This may be the case even if some jobs require higher skills than others in some aggregate sense. Workers who are most suited for higher skill jobs may be less productive in jobs requiring lesser skills.

For another thing, the assumption used in extensive-margin models that hours worked and incomes are absolutely fixed seems a bit restrictive, while that used in intensive-margin models that workers can choose any income by freely varying their labour supply may also be restrictive. The spirit of extensive-margin models can be captured in less inflexible ways, while at the same time offering workers some flexibility of effort, by assuming that different jobs allow different income ranges. This possibility seems clear in comparing high-income and low-income jobs. One expects that these will differ at least in their lower income bound and perhaps also in their upper bound. Even between jobs that belong to a similar broad income category, income ranges available from them may still differ. For example, sales jobs in agency/brokerage firms can be said to belong to the same broad income category but often offer different minimum, guaranteed salaries. In this case, the income ranges of different jobs differ at least in their lower bounds. More generally, it is reasonable to suppose that some minimum discrete quantity of work-time per day is necessary just to settle into the required routine in a given job, and beyond some maximum, concentration or strength may deteriorate significantly.

In this paper, we explore the consequences for optimal nonlinear income taxation of these two features of jobs: the absolute advantage workers have in jobs suited to them, and the fact that different jobs may have different feasible income ranges. These features are incorporated in a stylized way into the discrete-types model of Stiglitz (1982) and Guesnerie and Roberts (1982). Workers are assumed to differ in a single characteristic, which we refer to as skill. There are different jobs in the economy, each one most suited for a given skill level, and at least the lower limits of the income ranges of the jobs differ. Workers obtain the highest wage rate in the job matching their skill. They can choose other jobs, but will earn a lower wage. As well, each job pays the highest wage to the worker with the matching skill, and jobs requiring higher skill levels pay higher wage rates. Note the important feature that a worker with a higher skill level who chooses a job requiring lower skills will earn a lower wage rate than the worker with the matching skills.

We first analyze a basic model with two worker types, high- and low-skilled, and two job types, also high- and low-skilled. Workers in this model make two sequential decisions. They choose one of the jobs, and then they decide how much labour to supply, subject to the income bounds associated with the job. The government is assumed to have non-negative aversion to inequality, so it redistributes from the higher skilled to the lower skilled workers. We then extend the model to allow for more worker- and job-types, albeit restricting their numbers to be the same. Subsequently, we introduce a fixed cost of labour market participation to the multiple-type model, thereby adding a third labour decision margin along with job choice and labour supply.

The results in our basic model contrast sharply to those in the standard Mirrlees-

Stiglitz model. The marginal tax rate for workers below the top will be negative if the incentive constraint is binding and if both workers are in the job that matches their skills, unless one of the income bounds is binding. The marginal tax rate at the top will be zero unless an income constraint is binding at the upper bound. With maximin social preferences, first-best levels of utility can be achieved through nonlinear income taxation in the second-best, and that may be true with less redistributive social welfare functions as well. With utilitarian social preferences, outcomes in both the first- and the second-best involve more redistribution than under maximin preferences. Moreover, a higher government revenue requirement and a lower redistributive preference tend to result in the low-skilled worker working at the high-skilled job, which does not match his skill level.

In the multiple-type model, negative marginal tax rates for workers below the top again obtain under similar conditions as in the basic model. Moreover, since a worker's type in our model is fully described by the various job-specific wage rates, it is multi-dimensional. Without a more concrete specification, this makes it difficult to determine the set of binding incentive constraints. We thus focus on the more manageable 3-worker-3-job case. We find that, as is common in multi-dimensional screening models, global as well as local incentive constraints may be binding. Due to this, worker types below the top can also face a zero marginal tax rate in some optima.

When the participation margin is introduced into the multiple-type model via a fixed participation cost, the negative marginal tax rate result still holds. We also find that the participation margin tends to increase the absolute value of the marginal tax rates of all workers below the top. As well, some features of the participation taxes found in the literature appear, which we derive under quasilinear-in-consumption preferences. Intensive-margin and job choices tend to have opposite effects to the participation margin on the level of participation taxes, with the directions of their influences varying with skill levels. Specifically, under some assumptions on welfare weights, participation taxes for workers below a certain type tend to be higher in a model with both margins than in a model with only an extensive margin, while for workers above that type, the opposite is true. As a special case of this, participation taxes at the bottom can be of a negative income tax (NIC) type or an earned income tax credit (EITC) type, depending on the opposing influences from the two margins.

The intuition for the possibility of a negative marginal tax rate below the top is straightforward. A higher skilled worker mimicking a lower skilled worker may have to choose the lower skilled job. If so, the mimicking high skilled worker will be earning a lower wage than the lower skilled worker. This effectively makes the incentive constraint upward-binding. Relaxing the incentive constraint to facilitate redistribution involves distorting the labour supply of the low skilled worker upwards, which requires a negative marginal tax rate.

This paper is related to several strands of literature. The seminal paper in the optimal income tax literature is Mirrlees (1971), who assumes workers are drawn from a continuous skill distribution. (See also Tuomala (1990) and Ebert (1992) for useful elaborations.) We follow the discrete skill-type case, proposed initially by Stern (1982) and Stiglitz (1982) for the two-type case, and Guesnerie and Seade (1982) for multiple types, and pursued subsequently by Weymark (1986), Homburg (2001) and Hellwig (2007), among others. The extensive-margin model of participation began with Diamond (1980) and was extended to incorporate job choice by Saez (2002). There have been various applications of the optimal tax problem using extensive-margin or discrete labour supply assumptions, including Laroque (2005) and Choné and Laroque (2011), who study general cases of participation tax rates in extensive-margin models, including for the maximin case; Hungerbühler et al (2006), who add involuntary unemployment from search frictions to an extensive-margin model; Jacobsen Kleven, Kreiner and Saez (2009), who analyze the effect of participation decisions of secondary earners for family income taxation; and Diamond and Spinnewijn (2011), who consider a two-period job choice model with fixed earnings to allow for capital income taxation. These models often give negative participation tax rates at the bottom of the skill distribution.

Negative marginal tax rates can also arise in an intensive-margin setting. This can result from general equilibrium effects as in Stiglitz (1982), who argued that by stimulating high-skilled labour supply, a negative marginal tax rate at the top can improve relative wages for the low-skilled. Negative marginal tax rates can also arise if individuals have both differences in skills and in preferences for leisure, so upward-binding incentive constraints apply if the government wants to redistribute from those with high skills and preferences for leisure to those with low skills and preferences for leisure (Boadway et al 2000; Choné and Laroque 2010). Beaudry, Blackorby and Szalay (2009) assume workers have differing productivities in market and non-market activities, and show that when the government cannot observe non-market activities, marginal tax rates at the bottom will be negative. Krause (2009) obtains the possibility of negative marginal tax rates in a two-period model where second-period wages depend on first-period labour supply, via a learning-by-doing effect. It may be optimal to subsidize low-skilled labour supply in the first period, since this will increase the low-skilled wage rate in the second period which both improves redistribution and weakens the second period incentive constraint.

Finally, the approach which bears some similarity to ours is Jacquet, Lehmann and Van der Linden (2013), who add a participation choice to an intensive-margin Mirrlees-type model. In their model, skills are perfect substitutes in production, so there is no job choice dimension. Adding an extensive-margin decision reduces marginal tax rates. As mentioned, this differs from our finding that (at least with quasilinear-in-consumption preferences) the extensive margin increases the absolute value of marginal tax rates, whether they are positive or negative.

The rest of the paper is organized as follows. Section 2 sets up the model by introducing job-specific wage rates and job-specific income ranges into a standard, discrete-type model of optimal income taxation. Section 3 analyzes this model in the 2-worker-2-job basic case. Section 4 generalizes the basic model by allowing for more worker types and more jobs. In section 5, we add a fixed cost of participation to the utility function and introduce the extensive margin. Section 6 considers briefly a model with job-specific disutility-of-work. Section 7 contains some concluding remarks.

2 Basic Setting

The model combines features of the classic optimal income tax literature with labour variability along the intensive margin (Mirrlees 1971; Stiglitz 1982) with those of the extensive-margin approach where workers make discrete job choice and/or participation decisions (Diamond 1980; Saez 2002). There are a given number of types of jobs, indexed by the superscript $j = 1, \cdots, N$, each matching a given level of skills. By convention, job j matches more productive skills than job $j-1$, and offers a higher wage rate if the job is filled with the most suitable worker. There is a population of workers who differ in their aptitude and ability to fill jobs of different skills. For simplicity, we assume that there are also N worker types, which will be indexed by a subscript $i = 1, \cdots, N$. Workers can in principle fill any type of job. However, they will earn the highest wage rate if they choose the job for which they are most suited, which by convention is the job for which $j = i$. The wage rates reflect worker productivities and are fixed for a given type of worker in a given job. Let w_i^j be the wage of a type-i worker in a type-j job. The pattern of wage rates satisfies the following assumption:

Assumption 1. Wage rates satisfy:

(i) $w_j^j > w_j^k$ for all $k \neq j$

(ii) $w_j^j > w_k^j$ for all $k \neq j$

(iii) $w_j^j > w_{j-1}^{j-1}$

Assumption 1(i) states that a worker makes the highest wage in the job for which he is best suited. It applies whether a worker of given skill (worker j, say) chooses a job suited for a lower-skilled worker (job k, with $k < j$) or for a higher-skilled worker (job k, with $k > j$). Assumption 1(ii) says that a worker most suited for any job will earn a higher wage in that job than any other type of worker (including those suited for higher skilled jobs). Assumption 1(iii) indicates that workers of higher skills earn higher wages than those of lower skills if both are employed in their most suited job. These assumptions do not fully characterize all possible wage patterns, but are the most relevant ones for our purposes.[1]

Workers can choose jobs at will, and can change jobs without cost. They can also choose how much labour to supply, which is determined both by their preferences and by some restrictions we impose below on permissible earnings in each type of job. (They

[1] For example, these assumptions do not fully specify the relation between wages of workers who are not in jobs for which they are best suited (i.e., $w_{j+1}^j \gtreqless w_j^{j+1}$). In our base case, we consider two types of jobs, $j = 1, 2$, and two types of workers $i = 1, 2$, where worker 2 is high-skilled and worker 1 is low-skilled. Assumption 1 implies that either $w_2^2 > w_1^1 > w_1^2 > w_2^1$ or $w_2^2 > w_1^1 > w_2^1 > w_1^2$.

may also choose whether to participate in the labour market, but we defer consideration of that until later.) All workers share the same preferences, represented by the utility function $U(C,L)$, where C is aggregate consumption and L is labour supply. We assume that $U(C,L)$ is increasing in C and decreasing in L, and is strictly concave. Moreover, leisure is assumed to be non-inferior, and leisure and consumption are gross substitutes. Utility does not depend directly on the type of job, but the wage rate does.

Let $Y_i^j \equiv w_i^j L_i^j$ be income of a type-i worker in a type-j job. Following standard practice, it is useful to rewrite utility in terms of consumption and income as follows:

$$V_i^j(C_i^j, Y_i^j) \equiv U\Big(C_i^j, \frac{Y_i^j}{w_i^j}\Big)$$

If $j = i$, the worker is working in the most-suited job. A worker of a given type could mimic the consumption-income bundle of a different type of worker. For a worker of type i who mimics a worker of type k when the latter is in a type$-k$ job, utility will be written $\widehat{V}_i^k(C_k^k, Y_k^k/w_i^k)$. Note that this expression assumes that mimicking involves choosing the job of the other type, though that need not be the case depending on the permissible levels of income that can be earned in each job, to which we now turn.

We assume that for each job-type, income is restricted to be within a given range. Let $\overline{Y}^j$ and $\underline{Y}^j$ be the upper and lower limits of income that can be earned in job j. We generally assume that $\overline{Y}^j > \overline{Y}^{j-1}$ and $\underline{Y}^j > \underline{Y}^{j-1}$, so it is possible that the two ranges are disjoint or partially overlap. We consider both cases below, as well as the case where there is no upper bound. We assume that all workers can reach all income levels in jobs that they might be tempted to take. This simplifies the analysis, and can be relaxed to some extent without affecting our main insights. The assumption that feasible incomes are bounded in each job can be viewed as a reasonable generalization of the pure extensive-margin job choice model of Saez (2002), where incomes are completely fixed. It allows us to combine features of the extensive and intensive margins of labour supply, albeit at the cost of keeping track of these constraints. If there were no bounds on incomes, workers would always choose the job-type in which they are most productive.

The assumptions of the basic setting are illustrated in Figure 1, which depicts the preferences for a given type of worker, assumed to be type 2, who chooses among three job-types: one for which he is most suited (type 2), one for a lesser-skilled worker (type 1) and one for a more-skilled worker (type 3). The feasible income ranges for the three jobs are assumed to be disjoint for illustrative purposes. The case where the income ranges overlap will be drawn when we consider that case. The wage rate paid to the type-2 worker in job 2 exceeds that in both other jobs, $w_2^2 > w_2^1, w_2^3$. This implies that if a worker works in jobs of type 1 or type 3, more effort is required to earn a given level of income, so indifference curves will be steeper for any given income. This is reflected in the three indifference curves shown in the figure in each job. Those labeled i represent the same level of utility in all jobs, indifference curves ii represents a lower level of utility, and those labeled iii represent an even lower level. If $\overline{Y}^1$ is sufficiently close to $\underline{Y}^2$, indifference curves in job 1 at $\overline{Y}^1$ will be steeper than those in job 2 at $\underline{Y}^2$ and vertically higher. Similarly, for $\overline{Y}^2$ sufficiently close to $\underline{Y}^3$, the indifference curves in job 2 will be flatter at $\overline{Y}^2$ than those for job 3 at $\underline{Y}^3$ and vertically lower. This is the case shown in the figure.

Figure 1 shows two more things. First, in the laissez-faire situation, workers must choose a point on a budget constraint that is at 45^o to the origin. In the circumstances of Figure 1, worker 2 will choose the indifference curve i in job 2, the most-suited job. We assume in what follows that workers always choose their most suited job in the laissez-faire.[2] Second, the figure shows expansion paths for the worker in jobs 1 and 2 from lump-sum changes in income. Note in particular that such changes will eventually lead to a corner solution at an income bound (unless preferences are quasilinear in consumption so expansion paths are vertical). Thus, for some levels of utility, a worker's indifference curve may have a slope everywhere below unity or above unity in the income range of some job. This will have implications for optimal marginal tax rates below.

Figure 2 allows for both type 1's and type 2's on the same diagram, and considers three types of jobs, 1, 2 and 3, that each of the two workers can choose. Three indifference curves labeled i, ii and iii are shown for the type-1 worker, while those for type-2's are labeled $\boldsymbol{i}$, $\boldsymbol{ii}$ and $\boldsymbol{iii}$. These curves are drawn on the assumption that worker 1 obtains a lower wage

[2] It is possible that workers might choose a job for which they are not best suited in the laissez-faire. This can occur for certain configurations of preferences and feasible income ranges. For simplicity, we assume this case away.

rate in job 3 than in job 2 ($w_1^2 > w_1^3$), and worker 2 earns a higher wage in job 3 than does worker 1 ($w_2^3 > w_1^3$). This figure can be used to illustrate some important properties of the model that are relevant for the optimal income tax structure. The first is that if there were only two workers and two jobs, say, jobs 1 and 2, the single-crossing property applies for these disjoint indifference curves. That is, if indifference curves of worker 1 and worker 2 intersect in the range of job 1, these same indifference curves cannot intersect in the income range of job 2.[3] However, this single-crossing property does not hold more generally. For example, in Figure 2, indifference curves for workers 1 and 2 can intersect in both income ranges 1 and 3 (or in both 2 and 3). The second is that the single-crossing property works in the opposite direction in different income ranges. In the income range of job 1, type 1's have flatter indifference curves than type 2's at an intersection point since they have higher wages. At the same time, at intersections in job 2's income range, type 2's have flatter indifference curves than type 1's for the same reason.

Finally, in the laissez-faire, higher-skilled workers are better off than lower-skilled ones, given that we have assumed that the laissez-faire is not income-constrained. This is illustrated by the points ℓ_1 and ℓ_2 in Figure 2 for the case of workers 1 and 2 when both are assumed to choose points within the feasible income ranges of their preferred jobs. Worker 2 has a higher wage rate than worker 1, and they are both on 45^o budget lines. As in the standard case, worker 2 is better off than worker 1. The implication is that a government with non-negative aversion to inequality will want to redistribute from type 2's to type 1's. We now turn to government policy in the simplest case where there are only two types of workers and two types of jobs.

3 Optimal Income Taxation in the Two-Type Case

There are two types of workers, $i = 1, 2$, and two types of jobs, $j = 1, 2$. For simplicity, assume there is one worker of each type. The government levies a non-linear income tax on workers, or equivalently chooses consumption-income bundles for the two types of workers, to maximize a weighted utilitarian social welfare function, subject to a) the

[3] To see this, suppose that indifference curves i and $\boldsymbol{i}$ intersect twice, once in income range 1 and again in income range 2. By Assumption 1(ii), the former intersection implies that worker 1 is better off than worker 2, while the latter implies the opposite. This is a contradiction.

requirement of raising a given level of revenue, b) the workers' incentive constraints, and c) the income bounds associated with each job. The incentive constraints reflect the fact that the government cannot observe worker types, only the income they earn. We also assume the government does not observe the workers' jobs, although in the case where income ranges are distinct this can be inferred. We assume from the outset that the optimum involves a separating equilibrium. It is straightforward to show that this will be optimal unless revenue requirements are so high that both workers pool at the upper bound of the income range of job 2. However, unlike in the laissez-faire and in a first-best allocation, we cannot assume that it is optimal for each type of worker to be in their most-suited job. To allow for different possibilities, we drop the job superscript from consumption-income bundles for the two types.

More formally, we write the social welfare function as $\gamma_1 V_1(C_1, Y_1) + \gamma_2 V_2(C_2, Y_2)$, where γ_1 and γ_2 are the social welfare weights for the low-skilled and high-skilled workers. We assume that $\gamma_1 \geq \gamma_2 \geq 0$ given that the government is averse to inequality. Under utilitarianism, we have $\gamma_1 = \gamma_2 > 0$, while with maximin, $\gamma_1 > \gamma_2 = 0$.[4] The government maximizes social welfare subject to several constraints. One is the resource or budget constraint, $(Y_1 - C_1) + (Y_2 - C_2) \geq K$, where $K \geq 0$ by assumption. Next, there are incentive constraints on the two types of workers, $V_2(C_2, Y_2) \geq \widehat{V}_2(C_1, Y_1)$ and $V_1(C_1, Y_1) \geq \widehat{V}_1(C_2, Y_2)$, which ensure that each worker prefers his own consumption-income bundle to that of the other worker. Unlike in the standard case, mimicking may require workers changing jobs, for example, when the income ranges are disjoint and workers are in different income ranges in the optimum. Finally, optimal incomes may be constrained by the bounds on income ranges in the two jobs: $\underline{Y}^j \leq Y_i \leq \overline{Y}^j$ for $i, j = 1, 2$. The specific form these take will depend on whether income ranges are disjoint or overlap.[5]

We begin first with the case where the income ranges of the two job-types are disjoint.

---

[4] We could instead have used a quasi-concave social welfare function. Using the weighted utilitarian form simplifies the notation slightly without affecting the results.

[5] In principle, we should also impose non-negativity constraints on consumption, $C_i \geq 0$ for $i = 1, 2$. These can become binding when the revenue requirements of the government are sufficiently large. Although this can lead to some unusual results, they would take us too far afield, so we assume the non-negative consumption constraints are always slack.

Later we take up the case of overlapping income ranges.

3.1 Disjoint Feasible Income Ranges

Let $A_1 \equiv (C_1, Y_1)$ and $A_2 \equiv (C_2, Y_2)$ denote the allocations intended for type-1 and type-2 workers respectively. The income ranges are assumed to be disjoint and non-contiguous, so $\overline{Y}^1$ is strictly less than $\underline{Y}^2$.[6] In principle, both A_1 and A_2 can be in either job, so there are four possibilities: (i) A_1 is in income range 1 and A_2 is in income range 2, (ii) both A_1 and A_2 are in income range 2, (iii) A_1 is in income range 2 and A_2 is in income range 1, and (iv) both A_1 and A_2 are in income range 1. In fact, case (iii) cannot arise since the two incentive constraints cannot both be satisfied in this case.[7] As well, case (iv) cannot arise because it will be possible to move A_2 to job 2 so as to keep worker 2 as well off while increasing government revenue without violating any incentive constraint.[8] Cases (i) and (ii) are possible and we consider them in turn.

3.1.1 Case (i): Both Workers in Most-Suited Jobs

In this case, we can show that, in contrast to the standard intensive-margin case, when the optimal bundles A_1 and A_2 are in the interior of the income ranges of jobs 1 and 2 and the incentive constraint on the high-skilled worker is binding, the marginal tax rate

---

[6] If income ranges were disjoint but contiguous, so $\overline{Y}^1 = \underline{Y}^2$, a discontinuity would arise leading to the possibility of non-existence of the optimum. Because the income level $Y = \overline{Y}^1 = \underline{Y}^2$ could be earned at both jobs 1 and 2 and each worker has different wage rates at different jobs, each worker's utility would change discontinuously with income around $Y = \overline{Y}^1 = \underline{Y}^2$. In this case, the existence of an optimum would not be guaranteed.

[7] Proof: By the single-crossing property, indifference curves for workers 1 and 2 intersect at most once. Suppose worker 2's incentive constraint is satisfied, so $U(C_2^1, Y_2^1/w_2^1) \geq U(C_1^2, Y_1^2/w_2^2)$. Also, $w_1^1 > w_2^1$ implies $U(C_2^1, Y_2^1/w_1^1) > U(C_2^1, Y_2^1/w_2^1)$ and $w_2^2 > w_1^2$ implies $U(C_1^2, Y_1^2/w_1^2) < U(C_1^2, Y_1^2/w_2^2)$. Therefore, $U(C_2^1, Y_2^1/w_1^1) > U(C_1^2, Y_1^2/w_1^2)$, so worker 1's incentive constraint is violated. The same reasoning implies that if worker 1's incentive constraint is satisfied, worker 2's will be violated. ■

[8] The reason is that in this case, indifference curves will intersect in income range 1 when incentive constraints are satisfied. In range 2, the indifference curve for worker 1 will be everywhere above that for worker 2, and the 45^o line tangent to the indifference curve of worker 2 in job 1, say, I_2^1, is above the 45^o line tangent to that in job 2, I_2^2. Therefore, a movement of A_2 from some point on I_2^1 to the unit-slope point of I_2^2 is a movement of less than 45^os and will increase government revenue without violating an incentive constraint. This is so even if the optimum is such that there is no unit-slope point on I_2^2.

on the low-skilled worker is negative, while that on the high-skilled worker is zero. In this case, the high-skilled worker is worse off than the low-skilled one. If the upper income bound becomes binding on worker 1, his marginal tax rate can take any sign. If it becomes binding on worker 2, his marginal tax rate is positive. Moreover, unlike in the standard case, the incentive constraint on type-2 workers need not be binding in the optimum, even though the government is averse to inequality. In this case, the optimum will be first-best. The following analysis demonstrates these results.

We continue to suppress job superscripts. We know that type-1 will never mimic type-2 in the optimum since as we have seen, in the laissez-faire the type-1 worker is worse off than the type-2, so a government with non-negative aversion to inequality will want to redistribute from the type-2 to the type-1 as in the standard model. The government then maximizes $\gamma_1 V_1(C_1, Y_1) + \gamma_2 V_2(C_2, Y_2)$ subject to its budget constraint, $(Y_1 - C_1) + (Y_2 - C_2) \geq K$, the incentive constraint on the high-skilled, $V_2(C_2, Y_2) \geq \widehat{V}_2(C_1, Y_1)$, and the income bounds, $\underline{Y}^1 \leq Y_1 \leq \overline{Y}^1$ and $\underline{Y}^2 \leq Y_2 \leq \overline{Y}^2$.

The Lagrange expression can be written:

$$\mathcal{L} = \gamma_1 V_1(C_1, Y_1) + \gamma_2 V_2(C_2, Y_2) + \lambda[(Y_1 - C_1) + (Y_2 - C_2) - K] + \mu[V_2(C_2, Y_2) - \widehat{V}_2(C_1, Y_1)]$$
$$+ \underline{\delta}^1 (Y_1 - \underline{Y}^1) + \overline{\delta}^1 (-Y_1 + \overline{Y}^1) + \underline{\delta}^2 (Y_2 - \underline{Y}^2) + \overline{\delta}^2 (-Y_2 + \overline{Y}^2)$$

The first-order conditions on C_1 and Y_1 for worker 1, and C_2 and Y_2 for worker 2 are:

$$\gamma_1 \frac{\partial V_1}{\partial C_1} - \lambda - \mu \frac{\partial \widehat{V}_2}{\partial C_1} = 0; \qquad \gamma_1 \frac{\partial V_1}{\partial Y_1} + \lambda + (\underline{\delta}^1 - \overline{\delta}^1) - \mu \frac{\partial \widehat{V}_2}{\partial Y_1} = 0 \tag{1}$$

$$\gamma_2 \frac{\partial V_2}{\partial C_2} - \lambda + \mu \frac{\partial V_2}{\partial C_2} = 0; \qquad \gamma_2 \frac{\partial V_2}{\partial Y_2} + \lambda + (\underline{\delta}^2 - \overline{\delta}^2) + \mu \frac{\partial V_2}{\partial Y_2} = 0 \tag{2}$$

The interpretation of these results will differ depending on whether the income constraints are binding. Consider first the case where they are not, which corresponds with the standard case.

Income Constraints not Binding

In this case, $\underline{\delta}^j = \overline{\delta}^j = 0$ for $j = 1, 2$, so (1) and (2) reduce to the first-order conditions of the standard case (Stiglitz 1982). However, there are three important differences here compared with the standard case.

First, it is possible that the incentive constraint is slack in the optimum, and related to this, the first-best maximin outcome can be achieved in the second-best. In the standard case, the incentive constraint on high-skilled workers must bind in an optimum when the social welfare function has non-negative aversion to inequality. The reason is that when the incentive constraint is satisfied, type-2's must be better off than type-1's.[9] In our model, the mimicking type-2 need not be better off than the type-1 worker for the incentive constraint to be satisfied. That is because the mimicker earns a lower wage rate in job 1 than does the type-1 worker by Assumption 1(ii) (i.e., $w_1^1 > w_2^1$). Therefore, it is possible that the incentive constraint is not binding in the optimum, depending on the relative wage rates w_1^1 and w_2^1 and the relative values of social welfare weights γ_1 and γ_2.

This outcome can be pictured by starting from the laissez-faire outcome given by points ℓ_1 and ℓ_2 in Figure 2, and redistributing income lump-sum from the type-2 to the type-1 worker. The 45^o budget line of the type-2 falls and that of the type-1 rises, and both segments of their indifference curves shift as well. The optimum might be achieved without the incentive constraint becoming binding as illustrated in Figure 3, where now I_i^j denotes the indifference curve of a type-i worker in job j. In this case, the Lagrange multiplier on the incentive constraint μ is zero, implying by (1) and (2) that $(\partial V_1/\partial Y_1)/(\partial V_1/\partial C_1) = (\partial V_2/\partial Y_2)/(\partial V_2/\partial C_2) = 1$. Thus, marginal tax rates are zero and a first-best social outcome is achieved.

Note in particular that a maximin outcome can be achieved as long as revenue requirement is not too extreme, making the maximin technologically infeasible. This is because, when the two types have the same level of utility, I_2^1 is above I_1^1 and I_1^2 is above I_2^2 due to Assumption 1(ii). In fact, since these indifference curves are some distance apart, first-best outcomes "near" the equal-utility maximin outcome can also be achieved in the second best.

Second, when the incentive constraint is binding, so $\mu > 0$, the optimal marginal tax rate on the type-1 worker will be negative. Figure 4 illustrates this case. For the high-skilled worker, (2) implies that $(\partial V_2/\partial Y_2)/(\partial V_2/\partial C_2) = 1$ when income constraints are

[9] That is, $V_2(C_2, Y_2) \geq \widehat{V}_2(C_1, Y_1) > V_1(C_1, Y_1)$ since the mimicker supplies less labour to earn the same income as a type-1 person.

not binding, so the marginal tax rate at the top is zero as in the standard case. For the low-skilled worker, (1) can be written:

$$-\frac{\partial V_1/\partial Y_1}{\partial V_1/\partial C_1} = \frac{\lambda - \mu(\partial \widehat{V}_2/\partial Y_1)}{\lambda + \mu(\partial \widehat{V}_2/\partial C_1)} = \frac{1 - k(\partial \widehat{V}_2/\partial Y_1)/(\partial \widehat{V}_2/\partial C_1)}{1+k} \tag{3}$$

where $k = (\mu \partial \widehat{V}_2/\partial C_1)/\lambda > 0$. As Figure 4 shows, $-(\partial \widehat{V}_2/\partial Y_1)/(\partial \widehat{V}_2/\partial C_1) > 1$, implying that $-(\partial V_1/\partial Y_1)/(\partial V_1/\partial C_1) > 1$. Therefore, the marginal tax rate for the type-1 worker is negative, contrary to the standard case.[10]

Third, whether or not the incentive constraint is binding, the type-1 worker can be better off than the type-2 worker in an optimum. If the incentive constraint binds, type-2 worker earns a lower wage than type-1 when he is mimicking the latter's income. Therefore, $V_1(C_1, Y_1) > \widehat{V}_2(C_1, Y_1) = V_2(C_2, Y_2)$.

The intuition for the negative marginal tax rate for worker 1 is clear. When each worker works at his most productive job, worker 2 has a higher wage rate than worker 1. However, by Assumption 1(ii) worker 2 has a lower wage rate when he mimics worker 1 by working in job 1. Since the incentive constraint prevents a less-productive worker from mimicking a more-productive worker, it is optimal for the tax system to encourage, rather than discourage, worker 1's labour supply, as long as his income remains in the income range for job 1. Clearly, it is important that worker 1 is more productive in job 1 than is the more highly skilled worker 2. If that were not the case, the standard result of a positive marginal tax rate at the bottom would apply.

Income Constraints Binding

The fact that incomes are bounded for each type of job leads to the possibility that in an optimum, income constraints will be binding. To explore the possibilities, assume that in the laissez-faire both worker-types are in the interior of the income range for their most-suited job as in Figure 2. Redistribution from type-2's to type-1's will initially cause the allocation for the type-2's to move in a southeast direction, and for the type 1's to move

[10] More formally, rewrite (3) as $x = (1+ky)/(1+k)$, or $x - 1 = k(y - x)$, where x and y are the marginal rates of substitution at (C_1, Y_1) of workers 1 and 2. By the single-crossing property in the feasible income range for job 1, $y > x$. Therefore, $x > 1$, which implies a negative marginal tax rate.

northwest. That implies that if the income constraint binds for type-2's, it will be at $\overline{Y}^2$, while for type-1's it will be at $\underline{Y}^1$, at least as long as the government is purely redistributive. If the government requires positive revenue, it is possible that $Y_1 \leq \overline{Y}^1$ becomes binding if revenue requirements are large enough. The following cases are then possible.

i) $Y_2 \leq \overline{Y}^2$ is binding: If the income of the type-2 worker is strictly constrained by the type-2 job upper bound, $\overline{\delta}^2 > 0$,[11] while $\underline{\delta}^2 = 0$. First-order conditions (2) then yield:

$$-\frac{\partial V_2/\partial Y_2}{\partial V_2/\partial C_2} = \frac{\lambda - \overline{\delta}^2}{\lambda} < 1 \tag{4}$$

Therefore, the marginal tax rate at the top will be positive. This reflects the fact that type-2's indifference curve will have a slope less than unity everywhere in the income range $[\underline{Y}^2, \overline{Y}^2]$.

ii) $Y_1 \geq \underline{Y}^1$ is binding: Similar reasoning applies if the type-1 worker is strictly income-constrained at $\underline{Y}^1$. Then, the multipliers satisfy $\underline{\delta}^1 > 0$ and $\overline{\delta}^1 = 0$, and first-order conditions (1) reduce to:

$$-\frac{\partial V_1/\partial Y_1}{\partial V_1/\partial C_1} = \frac{\lambda + \underline{\delta}^1 - \mu(\partial \widehat{V}_2/\partial Y_1)}{\lambda + \mu(\partial \widehat{V}_2/\partial C_1)} = \frac{1 + \underline{\delta}^1/\lambda - k(\partial \widehat{V}_2/\partial Y_1)/(\partial \widehat{V}_2/\partial C_1)}{1+k} \tag{5}$$

The additional term in the numerator makes the marginal tax rate on the type-1's even more negative.

iii) $Y_1 \leq \overline{Y}^1$ is binding: With positive revenue requirements, both workers may pay positive taxes and the maximum income constraint on type-1 may bind (and that on type-2 may as well). In this case, first-order condition (1) becomes:

$$-\frac{\partial V_1/\partial Y_1}{\partial V_1/\partial C_1} = \frac{\lambda - \overline{\delta}^1 - \mu(\partial \widehat{V}_2/\partial Y_1)}{\lambda + \mu(\partial \widehat{V}_2/\partial C_1)} = \frac{1 - \overline{\delta}^1/\lambda - k(\partial \widehat{V}_2/\partial Y_1)/(\partial \widehat{V}_2/\partial C_1)}{1+k} \tag{6}$$

The presence of $-\overline{\delta}^1/\lambda$ in the numerator implies that worker 1's marginal tax rate may take either sign.

Note that, as in the case without binding income constraints, the incentive constraint may still be slack. First-best maximin utility levels and those near the maximin can still

[11] Technically, it is possible for $\overline{\delta}^2 = 0$ and $Y_2 = \overline{Y}^2$ simultaneously. For simplicity, we ignore all such boundary cases in what follows.

be achieved in the second-best, even when some worker is income constrained. The only difference is that the marginal tax rates of the income-constrained worker is no longer zero, as shown for the three possible cases above.

Also note that, when preferences are quasilinear-in-consumption, the type-1 worker's marginal tax rate will always be negative as long as the incentive constraint is binding, whether or not $\overline{Y}^1$ binds. Because we assumed that each worker works at his most productive job and is not income-constrained in the laissez-faire, the slope at $\overline{Y}^1$ of the type-1 worker's indifference curve through his laissez-faire allocation is greater than unity. With quasilinear-in-consumption preferences, this will also be true for any of his indifference curves, including the one through his optimal allocation.

3.1.2 Case (ii): Both Workers in Job 2

This case is similar to the standard model if no income constraints are binding, and the standard government problem applies. The type-2 worker is more productive than type-1 in job 2, and redistribution also goes from type-2 to type-1. Figure 5 depicts an optimum at bundles A_1 and A_2 in this case. Standard results apply, with zero marginal tax rates at the top and positive ones at the bottom.

The existence of separate jobs with income bounds, however, leads to two modifications of the standard model. First, if A_1 is in the interior of income range 2 but is close enough to the lower income bound $\underline{Y}^2$, moving it into income range 1 may increase revenue without violating any constraint, so the proposed Case (ii) optimum is not globally optimal. Figure 5 illustrates this possibility. If A_1 were on the dotted segment, it may be dominated by some allocations on the same indifference curve near or at the upper bound of income range 1. This occurs when the movement from A_1 to such allocations is at more than 45 degrees and therefore can increase government revenue without violating any constraint.

Second, if one or both of the optimal bundles A_1 and A_2 hits the upper income bound $\overline{Y}^2$, both types' marginal tax rate will be positive. To see this, note first that A_2 cannot hit the lower income bound. Our assumptions that each worker works at his most productive job and is not income-constrained in the laissez faire, together with a non-negative revenue requirement, imply that the slopes of type-2 worker's indifference curves at the lower income bound $\underline{Y}^2$ are always less than unity in an optimum. So moving A_2 up the indifference curve increases revenue without violating any constraint. It is, however, possible for A_2 to be located at the upper income bound $\overline{Y}^2$. In that case, type-2's indifference curve has a slope less than unity at that bound.

By the same token, A_1 may also be located at the upper income bound, where the slope of type-1's indifference curve is less than unity. When this is the case, pooling will occur if type-2's incentive constraint binds in an optimum.[12] A_1 may also be at the lower bound, where type-1's indifference curve has a slope less than unity. Thus, no matter where A_1 is in income range 2, type-1's marginal tax rate is positive. When A_2 is at the upper bound, type-2's marginal tax rate is positive.

3.1.3 Tax Implementation

Let us now briefly discuss the implementation of optimal allocations with a tax function $T(Y)$, or equivalently, a consumption function $C(Y)$. Since Case (ii) is similar to a standard model, we only discuss Case (i), focusing on the interior case where the incentive constraint binds, depicted in Figure 4. For convenience, we assume that indifference curves never touch the horizontal axis.

Following the literature (e.g., Homburg 2002), one can focus on the lower envelope of the indifference curves through optimal allocations when finding the relevant consumption function $C(Y)$. Then, to implement the optimum allocation (A_1, A_2) in Figure 4, the $C(Y)$ will coincide with I_2^1 up to A_1, kink at that point, coincide with or run below I_1^1 to reach the lower endpoint of I_2^2, rise along I_2^2 up to A_2, and then go below or coincide with I_2^2 afterwards. Note that the kink at A_1 remains as long as A_1 does not hit the upper income bound $\overline{Y}^1$. This pattern is obviously more complicated than in the standard case. In particular, we can make the following two observations about $C(Y)$.

First, unlike in the standard case, as the type-distribution becomes dense and eventually continuous, a continuous consumption function may still not become smooth. As can be seen from Figure 4, consumption at the upper end of I_2^1 may be higher or lower than

[12] For the standard reason, pooling cannot occur in the interior of income range 2.

that at the lower end of I_2^2. If A_1 is near or at $\underline{Y}^1$, a continuous consumption function implementing the optimal allocations needs to have a decreasing segment between $Y = \overline{Y}^1$ and $Y = \underline{Y}^2$, in order to extend from A_1 to the lower end of I_2^2. Thus, optimal consumption levels may not monotonically increase with income, depending on the shape of indifference curves, differences in job-specific wage rates of a given worker, the sizes of gaps between income ranges, the government revenue requirement, and the welfare weights.

Second, in a standard model, optimal tax structures exist for which the left derivative of the tax function is equal to one minus the marginal rate of substitution (Homburg 2002), which is the definition of the marginal tax rate we used. However, in our model, this may not be the case. In Figure 4, for example, it is the right derivative of such tax functions that is equal to the marginal tax rate for worker 1, which as we have seen is negative.

3.1.4 Utility Possibilities Frontier (UPF)

The second-best UPF for Case (i) has an interesting feature: a utilitarian optimum is more redistributive than a maximin optimum in terms of utility levels. That is, starting from the laissez-faire, there is more redistribution from the type-2 worker to the type-1 worker under utilitarianism than under maximin. Figure 6 illustrates. The lighter curves are the first-best and second-best UPF's.

The reason for this result is the following. First, for the same reason as in the standard model, a first-best utilitarian optimum involves worker 1 having higher utility than worker 2. Second, as shown in section 3.1.1, incentive constraints are still slack in or "near" a maximin optimum. Thus, starting from a maximin, further redistribution toward worker 1 along the first-best UPF is still possible. Therefore, a second-best utilitarian optimum will be more redistributive than a maximin optimum in terms of the relative level of utility.

This contrasts with the second-best outcome in a standard model, where a utilitarian optimum is always less redistributive than a maximin one because worker 2 can never be made worse off than worker 1.

3.1.5 Comparing Case (i) and Case (ii) Optima

It is instructive to characterize the circumstances in which each case is globally optimal. This involves a comparison between discretely different allocations, which is difficult to do analytically. However, graphical and numerical approaches suffice to suggest the following finding.

When the income range of job 1 is not too small relative to that of job 2, Case (i) typically dominates Case (ii), for various levels of revenue requirement and degrees of intended redistribution.[13] However, when income range 1 is small compared to income range 2, Case (i) is no longer always superior. It is more likely to be overall optimal, the lower is the revenue requirement and the more redistributive is the government objective. Conversely, Case (ii) is more likely to be optimal when the revenue requirement is very high and redistribution is less important.

Begin with the graphical approach and refer to the laissez-faire in Figure 2 (points ℓ_1 and ℓ_2), and to Cases (i) and (ii) in Figures 4 and 5, respectively. First, we can show that a purely redistributive optimum necessarily belongs to Case (i) rather than Case (ii).

To see this, recall the single-crossing property of the two workers' indifference curves. In the laissez-faire, I_1 and I_2 either intersect in income range 2 or do not intersect at all, and in the latter case, I_1^2 lies above I_2^2. Also note that, in a purely redistributive optimum, A_1 must be on an I_1 that is higher than the laissez-faire allocation ℓ_1. Next, for those I_1's lying above the laissez-faire I_1, a movement from any point on I_1^2 to the upper endpoint of I_1^1 — that is, the point where indifference curve I_1^1 hits the income bound $\overline{Y}^1$ — is more than 45 degrees.[14] Finally, by the single-crossing property, I_1^1 lies below I_2^1 in any Case (ii) optimum when the incentive constraint is binding and I_1^2 and I_2^2 are intersecting.

With these observations, we deduce that starting from a Case (ii) purely redistribu-

[13] A higher degree of intended redistribution here means a higher γ_1/γ_2 ratio, that is, steeper social indifference curves of the weighted utilitarian social welfare function. In our model, in both the first-best and the second-best solutions for Case (i), utilitarianism turns out to be more redistributive in terms of relative utility levels than maximin. So "more redistributive" does not correspond to higher aversion-to-inequality or more concavity in a social welfare function.

[14] To see this, imagine a figure similar to Figure 5 but drawn for the purely redistributive case. First, note that $w_1^1 > w_1^2$ according to Assumption 1(i). So if we extend I_1^2 into income range 1, it will be above the upper endpoint of I_1^1. By our assumptions that both workers works in their most suited jobs and are not income-constrained in the laissez-faire, for those I_1's lying above the laissez-faire one, the slope at the upper end point of I_1^1 and the slope at the lower end point of I_1^2 are both greater than unity, and the latter is greater than the former. Therefore, the above movement is more than 45 degrees.

tive optimum, moving A_1 to income range 1 will increase revenue without violating any constraint and will thus allow for further redistribution. Moreover, starting from a Case (i) optimum, one cannot find a movement of A_1 into income range 2 that allows for further redistribution. By the single-crossing property, I_1^2 lies above I_2^2 in a Case (i) optimum when the incentive constraint is binding. So moving A_1 up I_1 into income range 2 will violate both worker 2's incentive constraint and the government budget constraint. Thus, an overall optimum must belong to Case (i) when the revenue requirement is zero so the government problem is a purely redistributive one. This is the case no matter how redistributive the government's preferences are. Therefore, a Case (i) optimum will tend to be overall optimal when the revenue requirement is not too high.

On the other hand, the effects of other configurations of parameter values are difficult to derive analytically, so we resort to simulations. The following simulations are representative. The utility function takes the Cobb-Douglas form, $U(C, L) = C^{0.64}(\overline{H} - L)^{0.36}$, where C is consumption, $\overline{H} = 1.00001$ is time endowment, and L is labour supply. Job-specific wage rates are $w_1^1 = 0.6$, $w_1^2 = 0.45$, $w_2^1 = 0.5$, and $w_2^2 = 1$. The income bounds are $\underline{Y}^1 = 0.1$, $\overline{Y}^1 = 0.3999$, $\underline{Y}^2 = 0.4$, and $\overline{Y}^2 = 1$, and for comparison, we also used other values. Revenue requirement increases from the purely redistributive level of zero to 1.15. We used grid search to compute the utility possibilities frontiers for both cases in both the first- and the second-best.

Figure 6 shows an example. Under this model specification, the highest technologically possible revenue is around 1.3, so a revenue requirement of 0.5 is in the lower middle range. As can be seen in the figure, Case (i) second-best UPF, AM_sNU_sN', dominates Case (ii) second-best UPF, am_sn', except at the upper left segment. A pure utilitarian social welfare function corresponds to $\gamma_1/\gamma_2 = 1$ and has contour lines with slopes of minus one. So a Case (i) utilitarian outcome will be at U_s, while a Case (ii) one will be at u_s. Clearly, social welfare is higher in Case (i). A similar argument applies in the maximin case. In fact, for all values of γ_1/γ_2 in the range from (minus) the slope of line aA in Figure 6 to close to infinity, social welfare under a weighted utilitarian social welfare function will be higher in Case (i) than in Case (ii).

As revenue requirement increases, the UPF's shrink towards the origin. However, the

Case (i) UPF's may or may not shrink more quickly than the Case (ii) ones, depending on the relative size of income ranges 1 and 2. We considered three sets of income ranges. In all three, $\underline{Y}^1 = 0.1$ and $\overline{Y}^2 = 1$. However, the division points are ($\overline{Y}^1 = 0.1999$, $\underline{Y}^2 = 0.2$), ($\overline{Y}^1 = 0.2999$, $\underline{Y}^2 = 0.3$), and ($\overline{Y}^1 = 0.3999$, $\underline{Y}^2 = 0.4$), respectively. In the latter two,Case (i) UPF's largely maintain the same relative position to Case (ii) UPF's as that in Figure 6. So, Case (i) typically dominates Case (ii). However, in the first case, where income range 1 is relatively small as compared to income range 2, Case (i) UPF's shrink faster than Case (ii) ones and eventually become dominated. Then, Case (ii) will be the global optimum.

3.2 Overlapping Feasible Income Ranges

Suppose that higher incomes of job 1 overlap with lower incomes of job 2. In particular, let the income range limits satisfy $\underline{Y}^1 < \underline{Y}^2 < \overline{Y}^1 < \overline{Y}^2$, and assume that there are three intervals: $[\underline{Y}^1, \underline{Y}^2 - \varepsilon]$, $[\underline{Y}^2, \overline{Y}^1]$, and $[\overline{Y}^1 + \varepsilon, \overline{Y}^2]$, where $\varepsilon > 0$ is a small constant.[15] Call these the 'low', 'middle', and 'high' income ranges. Then, job 1's income ranges are the first two, and job 2's are the last two. The middle range is the overlapping income range. All other assumptions in the disjoint income range case apply. Figure 7 depicts some typical indifference curves. Single-crossing again applies. In the laissez-faire in particular, I_1 lies completely below I_2 in the low income range. In the middle and the high income ranges, they may or may not intersect; if they do not, I_1 is completely above I_2. It turns out that the optima parallel those in the basic case, so we simply describe the qualitative features of solutions.

Three cases can be distinguished. In the first case, the optimal A_1 is in the low income range, and A_2 is in one of job 2's income ranges. This parallels Case (i) in the disjoint ranges model above. Worker 1's marginal tax rate is negative if the upper income bound of the low income range does not bind too much. Worker 2's is zero if he is not income-constrained at the optimum. Otherwise, it can take either sign. In case two, A_1 is in the middle income range, and A_2 is in the job 2 income ranges. In case three, both A_1 and

[15] We could abstract from upper income limits by simply assuming that $\overline{Y}^1$ and $\overline{Y}^2$ are large. Then, cases two and three discussed below will become the same, but the rest of the analysis remains unchanged.

A_2 are in the high income range. These two cases parallel Case (ii) of the disjoint ranges model and resembles a standard model optimum. Worker 1's marginal tax rate is positive. Worker 2's is zero if he is not income-constrained, and can be positive otherwise.

Similar to the disjoint ranges model, first-best maximin utility levels can be attained in the second-best by an anonymous tax schedule which induces A_1 to be in the low income range and A_2 to be in the middle or the high income range. Second-best utilitarian optimum again tends to be more redistributive than second-best maximin.

3.3 Summary of Results in the Two-Type Case

The following summarizes the main results obtained when the income ranges of the two jobs are disjoint. Most of these results also apply when the income ranges of the two jobs (partially) overlap.

i) In both the first-best and the second-best, the utilitarian optimum is more redistributive than the maximin optimum.

ii) It is possible that first-best levels of utilities be achieved with a non-linear income tax schedule in the second-best, in which case the incentive constraint is not binding.

iii) If the incentive constraint binds in the second-best, an optimum may be one with each worker working in his most suited job (Case i) or one with both workers in the high income job (Case ii). A lower revenue requirement and a higher redistributive preference tend to favour Case (i), and vice versa.

iv) In the Case (i) optimum, the low-skilled worker's marginal tax rate is negative if his upper income bound is not binding; otherwise, it may take either sign. The high-skilled worker's marginal tax rate is zero if his upper income bound does not bind; otherwise, it is positive. Moreover, the low-skilled worker's optimum consumption may be higher than the high-skilled worker's, implying that consumption may not increase with income.

v) The Case (ii) optimum is qualitatively the same as that of a standard model, with the exception that if the upper income bound binds for the high-skilled worker, his marginal tax rate will be positive.

4 Optimal Income Taxation in the Multiple-Type Case

In this section, we generalize the basic model by allowing for N worker-types and N job-types. To facilitate discussion, we consider only disjoint income ranges and focus on optima where each worker works at his most productive job.[16] All other assumptions are retained.

Section 4.1 shows that the possibility of a negative marginal tax rate also exists in this multiple-type model. Section 4.2 explores possible patterns of binding incentive constraints and the corresponding marginal tax rates in a specific 3-by-3 example.

4.1 Some General Results on Marginal Tax Rates

As in the previous section, suppose for simplicity that there is one representative worker of each type. The government objective is the weighted utilitarian social welfare function, $\sum_{i=1}^{N} \gamma_i V_i(C_i, Y_i)$, where $\gamma_1, \cdots, \gamma_N$ are positive weights. The government maximizes social welfare subject to the budget constraint $\sum_{i=1}^{N}(Y_i - C_i) \geq K$, the incentive constraints, $V_i(C_i, Y_i) \geq \widehat{V}_i(C_{i'}, Y_{i'}),\ \forall i, i' \neq i$, and the income bounds, $\underline{Y}^i \leq Y_i \leq \overline{Y}^i,\ \forall i$.

We show that the main result of the basic model generalizes. A worker type's marginal tax rate will be negative if (1) at least one incentive constraint preventing some other worker type from mimicking him binds and (2) his income is not upward-constrained too much in the sense that the Lagrange multiplier associated with the relevant upper income bound is not too large.

The Lagrange expression for the government problem can be written:

$$\mathcal{L} = \sum_{i=1}^{N} \gamma_i V_i(C_i, Y_i) + \lambda \Big[\sum_{i=1}^{N} (Y_i - C_i) - K \Big] + \sum_{i=1}^{N} \sum_{i' \neq i} \mu_{i \to i'} \Big[V_i(C_i, Y_i) - \widehat{V}_i(C_{i'}, Y_{i'}) \Big]$$

$$+ \sum_{i=1}^{N} \underline{\delta}^i (Y_i - \underline{Y}^i) + \sum_{i=1}^{N} \overline{\delta}^i (-Y_i + \overline{Y}^i)$$

where $\mu_{i \to i'}$ is the multiplier for the incentive constraint that prevents worker i from mimicking worker i'. Note that both upward and downward incentive constraints are included,

[16] Needless to say, the latter assumes away the possibility of occupational distortions by the tax system and removes potentially interesting optima from our consideration. However, when N is not small, there can be many ways in which workers' occupational choices are distorted away from their most productive ones, and as we have seen, making discrete comparisons when occupations can be distorted is difficult. Exploring properties of an optimum when this restriction is dropped is left for future research.

because no restriction has been imposed on the social welfare weights. Also, global as well as local ones are included, for reasons discussed below. We shall write $\mu^d_{i\to i'}$ when the relevant incentive constraint is a downward one and $\mu^u_{i\to i'}$ when it is an upward one. To simplify notation, we denote $V_i(C_i, Y_i)$ by V_i and $\widehat{V}_k(C_i, Y_i)$ by $\widehat{V}_k$ in what follows.

The first-order conditions on C_i and Y_i are

$$\Big(\gamma_i + \sum_{k>i} \mu^u_{i\to k} + \sum_{k<i} \mu^d_{i\to k}\Big) \frac{\partial V_i}{\partial C_i} - \lambda - \sum_{k<i} \mu^u_{k\to i} \frac{\partial \widehat{V}_k}{\partial C_i} - \sum_{k>i} \mu^d_{k\to i} \frac{\partial \widehat{V}_k}{\partial C_i} = 0 \tag{7}$$

$$\Big(\gamma_i + \sum_{k>i} \mu^u_{i\to k} + \sum_{k<i} \mu^d_{i\to k}\Big) \frac{\partial V_i}{\partial Y_i} + \lambda + \underline{\delta}^i - \overline{\delta}^i - \sum_{k<i} \mu^u_{k\to i} \frac{\partial \widehat{V}_k}{\partial Y_i} - \sum_{k>i} \mu^d_{k\to i} \frac{\partial \widehat{V}_k}{\partial Y_i} = 0 \tag{8}$$

These can be rearranged to give

$$\begin{aligned} -\frac{\partial V_i/\partial Y_i}{\partial V_i/\partial C_i} = 1 + \frac{1}{\lambda}(\underline{\delta}^i - \overline{\delta}^i) &+ \frac{1}{\lambda} \sum_{k<i} \mu^u_{k\to i} \frac{\partial \widehat{V}_k}{\partial C_i} \left[-\frac{\partial \widehat{V}_k/\partial Y_i}{\partial \widehat{V}_k/\partial C_i} + \frac{\partial V_i/\partial Y_i}{\partial V_i/\partial C_i} \right] \\ &+ \frac{1}{\lambda} \sum_{k>i} \mu^d_{k\to i} \frac{\partial \widehat{V}_k}{\partial C_i} \left[-\frac{\partial \widehat{V}_k/\partial Y_i}{\partial \widehat{V}_k/\partial C_i} + \frac{\partial V_i/\partial Y_i}{\partial V_i/\partial C_i} \right] \end{aligned} \tag{9}$$

Except for $-\overline{\delta}^i$, all other terms on the right-hand side are non-negative; in particular, the two differences in marginal rates of substitution in square brackets are positive because of Assumption 1(ii) that worker i is more productive at job i than other workers $k \neq i$.

We can infer from (9) that the slope of worker i's indifference curve at his optimal allocation will be strictly greater than unity and his marginal tax rate will be negative, if the following two conditions are satisfied:

(1) the constraint on income $Y_i \leq \overline{Y}^i$ is slack so that $\overline{\delta}^i = 0$, or does not bind too much so that $\overline{\delta}^i$ is sufficiently small, and

(2) at least one incentive constraint preventing some other type k from mimicking type i is binding (or, roughly equivalently, there is some redistribution towards type i).

When condition (2) holds but condition (1) does not, $\overline{\delta}^i > 0$ may be large enough to make type i's marginal tax rate positive. As noted in the basic model, this is likely to occur when revenue requirement is high and the tax system is set to encourage more work by worker i. An exception is when preferences are quasilinear-in-consumption, in which case the marginal tax rate will always be negative.

In any case, a negative marginal tax rate comes from Assumption 1(ii) that a worker is more productive than others at the job that suits him best. With this assumption, redistribution towards any type of worker calls for an upward distortion of his labour supply in order to provide an incentive to, and extract revenue from, the types from which the tax system redistributes.

4.2 Further Analysis

The derivation above does not allow one to determine which incentive constraints will bind in an optimum. To explore this issue, we use the 'utility curve' tool of Matthews and Moore (1987) and analyze a simple 3-by-3 example. A by-product of this analysis is the finding that the marginal tax rate for types below the highest may be zero, due to a binding global incentive constraint.

To simplify the analysis and make the results more transparent, we introduce some further assumptions for use in subsections 4.2.1 and 4.2.2.

Assumption 2

(i) Worker i's wage at job j is given by $w_i^j = w_i^i - \alpha|j-i|$, $\alpha > 0$, $j = 1, \cdots, N$.

(ii) Workers' highest wage rates satisfy $w_i^i = a + bi$, $a > b > 0$, $i = 1, \cdots N$.

Assumption 2(i) restricts each worker's wage profile to be piecewise linear and single-peaked at his most suitable job. Since α is the same for all workers, this assumption also implies that, for any jobs j and k, the difference in wage rates $w_i^j - w_i^k$ does not depend on worker type i or on the distance between i and j or k. It only depends on the distance between j and k, that is, $w_i^j - w_i^k = \alpha|k-j|$. Assumption 2(ii) states that all the workers' highest wage rates are related in an increasing, linear fashion. Note that to be consistent with Assumption 1(ii), α needs to be greater than b.

We make the following assumption about preferences. Recall that we are deleting job superscripts, so C_j refers to the consumption intended for a type-j worker in a type-j job.

Assumption 3 Preferences are quasilinear in consumption and given by $U(C_j, Y_j/w_i^j) = C_j - h(Y_j/w_i^j)$, where $h(\cdot)$ is strictly increasing and strictly convex.[17]

[17] This assumption is made mainly to simplify analysis in the 3-by-3 example below. Combined

4.2.1 Utility Curve and a Single-Crossing Property

In the 2-by-2 model of section 2, the standard single-crossing property (SCP) holds; that is, indifference curves intersect only once. However, in the N-by-N case, this standard SCP may not hold due to Assumption 1, and we need to deal with global incentive constraints. We will make use of the utility curve tool developed by Matthews and Moore (1987) to deal with incentive constraints between nonadjacent consumer types in a multidimensional screening problem.

A utility curve plots the utility that different types of agents get from a given allocation (not necessarily an optimal one). In our context, it plots the utility levels to workers of different skill-types $i = 1, \cdots, N$ from a consumption-income bundle associated with job j, $U(C_j, Y_j/w_i^j)$. Figure 8 shows two representative utility curves out of the many possible ones, one for a job-2 bundle and one for a job-3 bundle, for the $N = 8$ case. The utility curves only consist of the N points because types are discrete in our model. For visual clarity, however, we connect points on the same utility curve. Because of Assumption 1, a utility curve has its peak at the most productive worker-type for the given bundle.

The SCP in utility curves, as defined by Matthews and Moore (1987), is satisfied for a set of bundles $\{(C_1, Y_1), \cdots, (C_N, Y_N)\}$ if no pair of the utility curves associated with those bundles intersect at more than one point, and they actually cross at any point of intersection. The two utility curves in Figure 8 satisfy this SCP. In our model, when the SCP in utility curves holds and some standard restrictions on welfare weights are imposed, only downward adjacent incentive constraints bind in an optimum, much the same as in a standard model. However, when this SCP does not hold, global incentive constraints need to be explicitly considered, even under the same kind of restrictions on welfare weights. In that case, the utility curves can help determine which constraints bind. It should be noted that two particular utility curves may not intersect at all, in which case the relevant incentive constraint is either not satisfied or strictly satisfied.

We begin by showing that the utility curves for two arbitrary allocations may cross

with the assumption on welfare weights made there that $\gamma_1 > \gamma_2 > \gamma_3$, it ensures that starting from an optimum and ignoring incentive constraints, the government wants to further transfer consumption from the higher types to the lower types.

more than once, even under the simplifying Assumptions 2 and 3 above. Specifically, the rising segments of any two utility curves in our model always cross at most once, but their declining segments may cross again, violating the SCP. We then derive a sufficient condition that rules out the latter crossing.

Consider Figure 8. Suppose the two utility curves are for some arbitrary bundles in jobs j and j', $j > j'$. As worker type increases from some arbitrary $i-1$ to i, the change in the utility curve for the job-j bundle is $h(Y_j/w^j_{i-1}) - h(Y_j/w^j_i)$, and that for the job-$j'$ bundle is $h(Y_{j'}/w^{j'}_{i-1}) - h(Y_{j'}/w^{j'}_i)$. When $i < j' < j$, the utility curves are increasing; when $i > j > j'$, they are decreasing. We will show that the change in the job-j utility curve is bigger than that for job j' when both curves increase, while the comparison is ambiguous when they both decrease.

First, when both curves increase ($i < j' < j$), we have

$$\frac{Y_j}{w^j_{i-1}} - \frac{Y_j}{w^j_i} = Y_j \frac{w^j_i - w^j_{i-1}}{w^j_{i-1}w^j_i} > Y_{j'} \frac{w^{j'}_i - w^{j'}_{i-1}}{w^{j'}_{i-1}w^{j'}_i} = \frac{Y_{j'}}{w^{j'}_{i-1}} - \frac{Y_{j'}}{w^{j'}_i}, \tag{10}$$

because $Y_j > Y_{j'}$, $w^{j'}_{i-1} > w^j_{i-1}$, $w^{j'}_i > w^j_i$, and by Assumption 2(i), $w^j_i - w^j_{i-1} = w^{j'}_i - w^{j'}_{i-1}$. Together with Assumption 3 that $h(\cdot)$ is a strictly increasing and strictly convex function and the fact that $Y_j/w^j_i > Y_{j'}/w^{j'}_i$, the above inequality implies that $h(Y_j/w^j_{i-1}) - h(Y_j/w^j_i) > h(Y_{j'}/w^{j'}_{i-1}) - h(Y_{j'}/w^{j'}_i) > 0$, That is, curve j rises faster with worker type i than curve j'. Thus, the two increasing segments of the utility curves can cross at most once. However, when both curves decrease ($i > j > j'$), inequality (10) becomes ambiguous. We still have $Y_j > Y'_j$. However, now $w^j_{i-1} > w^{j'}_{i-1}$ and $w^j_i > w^{j'}_i$, so the comparison can go in either direction.

If, for all i and $i-1$ with $i > i-1 \geq j > j'$, we have $h(Y_j/w^j_{i-1}) - h(Y_j/w^j_i) > h(Y_{j'}/w^{j'}_{i-1}) - h(Y_{j'}/w^{j'}_i)$,[18] then utility curve j falls more slowly than utility curve j'. Then, the two utility curves satisfy the SCP overall. However, if the opposite occurs, the decreasing segments of the utility curves may also intersect when their increasing segments already intersect,[19] and the SCP in utility curves does not hold. Therefore, a sufficient

[18] Note that, when $i > i-1 \geq j > j'$, both sides of the inequality are negative.

[19] The decreasing segments may intersect more than once, depending on the values of Y_j, $Y_{j'}$, and the wage rates.

condition for the SCP in utility curves to be satisfied is that inequality (10) holds for all $i > i-1 \geq j > j'$, which seems restrictive.

Given that the SCP in utility curves may not hold, we may need to take into account both global and local incentive constraints. We turn to a simple three-type example to show how the utility curve tool can be used to analyze the pattern of binding incentive constraints and therefore the nature of the optimum.

4.2.2 A 3-by-3 Example

We continue to focus on the disjoint income range case and restrict attention to interior allocations where each worker works at his most productive job. We assume that the government values some redistribution, so $\gamma_1 > \gamma_2 > \gamma_3$, given that utility is quasilinear in consumption.

Figures 9 to 12 depict different configurations for utility curves for three worker-types and three job-types case. These curves characterize four different situations. In Figure 9, the three bundles are not incentive-compatible. Worker 2 would prefer the bundle of job 1, (C_1, Y_1), to that of his own job, (C_2, Y_2). Figure 10 satisfies incentive-compatibility as well as the SCP, since no pairs of indifference curves intersect more than once. Figures 11 and 12 both violate SCP. In Figure 11, although utility curves all intersect once, that for job 2 does not cross that of job 1, so SCP is not satisfied. In Figure 12, the utility curve for job 3 intersects that of job 1 twice. We use these figures in the following analysis. For illustrative purpose, the allocations underlying Figures 10 through 12 are depicted in Figures 14 through 16 in the C-Y space. As well, Figure 13 depicts the utility curves corresponding to Figures 3 (dashed lines) and 4 (solid lines) for the 2-by-2 basic case.

To analyze the qualitative features of an optimum using these utility curves, it is convenient to distinguish between two cases: those in which feasible allocations satisfy the SCP and those in which they do not.

Case (i) Best allocation for which the SCP applies

In this case, the pattern of binding incentive constraints is the same as in our basic model. Therefore, there will be no distortion at the top and an upward distortion of labour supply below the top.

To see this, note that when the SCP holds, only local incentive constraints need to be considered. First, from the shape of utility curves, it is easy to see that at most one incentive constraint can bind between any pair of workers. Second, in an optimum, it will be the incentive constraints on higher types that bind. That is, $IC^d_{3\to 2}$ and $IC^d_{2\to 1}$ will bind.[20] If they do not bind, one can redistribute consumption from higher type workers to lower type workers to increase social welfare without violating any constraint, because utility is quasilinear in consumption, $\gamma_1 > \gamma_2 > \gamma_3$, and we restrict attention to strictly interior allocations.

Figure 10 depicts the utility curves for a best allocation in this case. Combining this with the result on marginal tax rates in the general analysis for the N-by-N case in Section 4.1, we know that the marginal tax rates are $T'(Y_1) < 0$, $T'(Y_2) < 0$, and $T'(Y_3) = 0$. Figure 14 shows these in the C-Y space. Similar to the basic case, utility falls with skills, and the tax schedule may exhibit declining consumption.

Case (ii) Best allocation for which the SCP does not apply

In this case, we can show that depending on the pattern of binding incentive constraints, worker 2's marginal tax rate can be zero or negative. Those of worker 1 and worker 3 are qualitatively the same as in the case above.

To see this, note first that the SCP can be violated only between utility curves 1 and 2, as is clear from Figure 9. Because of their shapes, utility curves 1 and 3 can intersect at most once. So will utility curves 2 and 3: by our analysis in the previous section, between $i = 1$ and $i = 2$, the rising segment of curve 2 is flatter than the rising segment of curve 3. However, between $i = 2$ and $i = 3$, the decreasing segments of curves 1 and 2 may or may not intersect, when their rising segments have already intersected. So the SCP may be violated between these two curves.

Second, because $\gamma_1 > \gamma_2 > \gamma_3$ and preference is quasilinear in consumption, some incentive constraint preventing other workers to mimic worker 1 would bind. That is, one or both of $IC^d_{2\to 1}$ and $IC^d_{3\to 1}$ will bind in an optimum. Similarly, at least one constraint

[20] IC stands for 'incentive constraint'. The superscript and subscript have the same interpretation as those for the Lagrange multiplier μ in section 4.1.

preventing worker 3 to mimic others will bind in an optimum. That is, one or both of $IC^d_{3\to 2}$ and $IC^d_{3\to 1}$ will bind.

Combining these observations and noting that the SCP does not hold, we can deduce that the pattern of optimum is one of the two depicted in Figures 11 and 12. Call these cases a and b, and consider each in turn.

a: In the Figure 11 optimum, $IC^d_{3\to 1}$ and $IC^d_{2\to 1}$ are binding. Worker 3's local downward incentive constraint $IC^d_{3\to 2}$ cannot bind when $IC^d_{2\to 1}$ binds, because the SCP does not apply. Instead, the global incentive constraint $IC^d_{3\to 1}$ binds. From the general result for the N-by-N case, we know that the marginal tax rates satisfy $T'(Y_1) < 0$ and $T'(Y_2) = T'(Y_3) = 0$. Worker 2 also faces a zero marginal tax rate because no incentive constraint preventing others to mimic him binds in the optimum. Figure 15 shows these in the C-Y space.

b: In the Figure 12 optimum, $IC^d_{3\to 2}$ and $IC^d_{3\to 1}$ are binding. Again, because SCP does not apply, $IC^d_{2\to 1}$ cannot bind when $IC^d_{3\to 2}$ binds. The marginal tax rates satisfy $T'(Y_1) < 0$, $T'(Y_2) < 0$, and $T'(Y_3) = 0$. Figure 16 shows these in the C-Y space.

Comparing the two optima, we see that $IC^d_{3\to 1}$ binds in both cases a and b, and that $IC^d_{2\to 1}$ binds in case a while $IC^d_{3\to 2}$ binds in case b. Thus, relatively speaking, there is more redistribution from worker 2 in case a than in case b, and moving from a case b optimum to a case a optimum can be thought of as further redistributing from worker 2 to the other two workers. So which case is likely to be optimal for Case (ii) would depend on the relative magnitude of the social welfare weights. For example, if $\gamma_2 < (\gamma_1 + \gamma_3)/2$, then starting from a case b optimum and taking one unit of consumption from worker 2 and giving one half to each of workers 1 and 3 will increase social welfare, and case a will be the overall optimum for Case (ii).

The overall optimum would be the solution in the two cases above that gives a higher social welfare. The comparison is a discrete one, and analytical techniques shed little light on the problem.

4.3 Summary of Results in the Multiple-Type Case

In this section, we have analyzed optimal income taxes when there are multiple types of workers and a corresponding number of jobs, under the restriction that each worker's optimal allocation is in his most suitable job. The main results are as follows.

i) As in the two-type model, a type's marginal tax rate will be negative when at least one incentive constraint preventing some other type from mimicking him binds and his income is not upward-constrained. It can take either sign when the same kind of incentive constraint binds but he is upward income constrained. Again, the negative marginal tax rate result mainly comes from Assumption 1(ii), that each worker is more productive than others at his most suitable job. This also leads to higher skilled workers having lower utility when their incentive constraint binds in an upward direction on lower skilled workers who have higher wages. As in the basic case, it is also possible for first-best optimum levels of utility to be achieved in a second-best.

ii) Due to the fact that worker types are multi-dimensional, it is generally difficult to determine which incentive constraints will bind in an optimum without concrete model specification. In a 3-by-3 example with quasilinear-in-consumption preferences, we show that an optimum may be one of three possible patterns.

iii) In the first case, the optimal allocations are such that utility curves satisfy the single-crossing property defined by Matthews and Moore (1987). In this case, only local incentive constraints need to be considered. Given the assumption that less skilled worker receives a higher welfare weight, we find that all adjacent downward incentive constraints ($IC^d_{3\to2}$ and $IC^d_{2\to1}$) bind and that workers 1 and 2 face negative marginal tax rate ($T'(Y_1) < 0$ and $T'(Y_2) < 0$) and worker 3 faces zero marginal tax rate ($T'(Y_3) = 0$). Workers with higher skills have lower utility in the optimum.

iv) In both the second and the third cases, the optimal allocations are such that utility curves do not satisfy the single-crossing property. In the second case, the binding incentive constraints are worker 2's and worker 3's incentive constraints preventing them to mimic worker 1 ($IC^d_{3\to1}$ and $IC^d_{2\to1}$). Worker 1 alone faces a negative marginal tax rate $T'(Y_1) < 0$, while workers 2 and 3 both face zero marginal tax rates ($T'(Y_2) = T'(Y_3) = 0$). In the third case, the binding incentive constraints are worker 3's incentive constraints preventing him to mimic workers 1 and 2 ($IC^d_{3\to2}$ and $IC^d_{3\to1}$). Worker 3 faces a zero marginal tax rate ($T'(Y_3) = 0$), and workers 1 and 2 face negative marginal tax rates ($T'(Y_1) < 0$ and $T'(Y_2) < 0$).

5 Labour Market Participation

In this section, we incorporate a participation decision in the multiple-type model under Assumption 1 and, later, under Assumptions 1 and 3, following Diamond (1980) and, more closely, Jacquet *et al* (2013).

There are N types of workers and N jobs with disjoint income ranges. Job-specific wage rates satisfy Assumption 1. There is now more than one worker of each type. Let f_i be the fraction of type i in the population, with $\sum_{i=1}^{N} f_i = 1$. We again focus on optima where all workers choose to work at their most suitable jobs. Workers make three choices: job choice, participation (the extensive margin), and labour supply (the intensive margin).

An individual who chooses to work incurs a fixed cost of participation, which varies across workers. Let the cost of participation of a worker of type i, denoted by the random variable m_i, be drawn from the distribution function $G_i(m_i)$, with density $g_i(m_i)$, with support $[\underline{m}_i, \overline{m}_i]$, where both $\underline{m}_i$ and $\overline{m}_i$ can take either sign. An individual who works also incurs the usual disutility of labour, which varies with the amount of labour supplied. Let C_0 be consumption of the non-participants, which must be the same for all types. Then, the utility of an individual of skill type i is

$$\begin{cases} V^i(C_j, Y_j) - m_i & \text{if working at job } j, \\ V^i(C_0, 0) & \text{if not working and receiving a transfer} \end{cases}$$

For a type i individual to be indifferent between working and not working, his fixed cost of participation, denoted as $\tilde{m}_i$, needs to be such that $V^i(C_i, Y_i) - \tilde{m}_i = V^i(C_0, 0)$, from which we can write

$$\tilde{m}_i(C_0, C_i, Y_i) = V^i(C_i, Y_i) - V^i(C_0, 0). \tag{11}$$

Type i individuals with fixed cost lower than $\tilde{m}_i$ will choose to work, and those with fixed cost higher than $\tilde{m}_i$ will choose not to work. We have

$$\frac{\partial \tilde{m}_i}{\partial C_i} = \frac{\partial V_i(C_i, Y_i)}{\partial C_i}, \qquad \frac{\partial \tilde{m}_i}{\partial Y_i} = \frac{\partial V_i(C_i, Y_i)}{\partial Y_i}, \qquad \frac{\partial \tilde{m}_i}{\partial C_0} = -\frac{\partial V_i(C_0, 0)}{\partial C_0}. \tag{12}$$

In what follows, we assume that there are both workers and non-participants for each skill type in the optima.

Two points will emerge from the analysis below. First, the participation margin tends to increase the absolute value of the marginal tax rates for working individuals, whether they are positive or negative. Second, under some assumptions on welfare weights, the influence of the extensive margin and the combined influence of the intensive margin and the job-choice margin tend to have opposite effects on the levels of the participation taxes. For lower-skill workers, extensive margin effects call for lower (and possibly negative) participation taxes while intensive margin effects call for higher (and possibly positive) ones, and vice versa for higher-skill workers. This finding is similar to that found in Jacquet *et al* (2013).

Below, we first solve the government maximization problem. Then, we analyze the marginal tax rates and the participation taxes in turn.

5.1 The Government Problem

The government objective is again assumed to be weighted utilitarian. Let $\underline{\gamma}_i$ and $\overline{\gamma}_i$ be the social welfare weights given to type-i workers and type-i non-workers. Similar to the multiple-type case, the analysis of marginal tax rates below does not require one to specify the relative magnitude of the welfare weights, but to obtain more concrete results for participation taxes, we do make Assumption 4 below on these weights. The incentive constraints are

$$V_i(C_i, Y_i) - m_i \geq \widehat{V}_i(C_{i'}, Y_{i'}) - m_i, \quad \text{or} \quad V_i(C_i, Y_i) \geq \widehat{V}_i(C_{i'}, Y_{i'}), \quad \forall i, i' \neq i,$$

and they apply to working individuals only. This simple form of incentive constraints comes in part from our assumption that m_i does not vary with a worker's job choice. Note that given our focus below on optima with workers working in their respective matching jobs, mimicking implies working in the other worker's job. The extensive margin is captured by the functions $\tilde{m}_i(C_0, C_i, Y_i)$ as in (11).

The Lagrangian function $\mathcal{L}$ is

$$\sum_{i=1}^{N} f_i \Big\{ \underline{\gamma}_i \int_{\underline{m}_i}^{\tilde{m}_i(C_0,C_i,Y_i)} [V_i(C_i, Y_i) - m_i] g_i(m_i)\, dm_i + \overline{\gamma}_i \int_{\tilde{m}_i(C_0,C_i,Y_i)}^{\overline{m}_i} [V_i(C_0, 0)] g_i(m_i)\, dm_i \Big\}$$

$$+\lambda \Big\{ \sum_{i=1}^{N} f_i \left[\int_{\underline{m}_i}^{\tilde{m}_i(C_0,C_i,Y_i)} (Y_i - C_i) g_i(m_i)\, dm_i + \int_{\tilde{m}_i(C_0,C_i,Y_i)}^{\overline{m}_i} (C_0) g_i(m_i)\, dm_i \right] - K \Big\}$$

$$+\sum_{i=1}^{N}\sum_{i'\neq i}\mu_{i\to i'}[V_i(C_i,Y_i)-\widehat{V}_i(C_{i'},Y_{i'})]+\sum_{i=1}^{N}\underline{\delta}^i(Y_i-\underline{Y}^i)+\sum_{i=1}^{N}\overline{\delta}^i(-Y_i+\overline{Y}^i)$$

Using (12), we can write the first-order conditions on C_i, Y_i and C_0 as

$$\begin{aligned}
[C_i]:\ & \frac{\partial V_i(C_i,Y_i)}{\partial C_i}\{f_i[(\underline{\gamma}_i-\overline{\gamma}_i)V_i(C_0,0)g_i(\tilde{m}_i)+\underline{\gamma}_iG_i(\tilde{m}_i) \\
& +\lambda(Y_i-C_i+C_0)g_i(\tilde{m}_i)]+\sum_{t\neq i}\lambda_{i\to t}\}=\lambda f_iG_i(\tilde{m}_i)+\sum_{k\neq i}\mu_{k\to i}\frac{\partial\widehat{V}_k(C_i,Y_i)}{\partial C_i}, \\
[Y_i]:\ & -\frac{\partial V_i(C_i,Y_i)}{\partial Y_i}\{f_i[(\underline{\gamma}_i-\overline{\gamma}_i)V_i(C_0,0)g_i(\tilde{m}_i)+\underline{\gamma}_iG_i(\tilde{m}_i)-(\underline{\delta}^i-\overline{\delta}^i) \\
& +\lambda(Y_i-C_i+C_0)g_i(\tilde{m}_i)]+\sum_{t\neq i}\mu_{i\to t}\}=\lambda f_iG_i(\tilde{m}_i)-\sum_{k\neq i}\mu_{k\to i}\frac{\partial\widehat{V}_k(C_i,Y_i)}{\partial Y_i}, \\
[C_0]:\ & \sum_{i=1}^{N}f_i\frac{\partial V_i(C_0,0)}{\partial C_0}\{\overline{\gamma}_i(1-G_i(\tilde{m}_i))-(\underline{\gamma}_i-\overline{\gamma}_i)V_i(C_0,0)g_i(\tilde{m}_i) \\
& -\lambda(Y_i-C_i+C_0)g_i(\tilde{m}_i)\}=\lambda\sum_{i=1}^{N}f_i(1-G_i(\tilde{m}_i))
\end{aligned}$$

5.1.1 Marginal Tax Rates

From $[C_i]$ and $[Y_i]$, we get

$$-\frac{\partial V_i/\partial Y_i}{\partial V_i/\partial C_i}=1+\frac{\underline{\delta}^i-\overline{\delta}^i}{\lambda f_iG_i(\tilde{m}_i)}+\frac{1}{\lambda f_iG_i(\tilde{m}_i)}\sum_{k\neq i}\mu_{k\to i}\frac{\partial\widehat{V}_k}{\partial C_i}\left[-\frac{\partial\widehat{V}_k/\partial Y_i}{\partial\widehat{V}_k/\partial C_i}+\frac{\partial V_i/\partial Y_i}{\partial V_i/\partial C_i}\right].\quad(13)$$

The marginal tax rate for type i will be negative, provided that $\overline{\delta}^i$ is not too large. This expression differs from its counterpart in the N-by-N-type case without a participation margin only by the factor $f_iG_i(\tilde{m}_i)$ on the right-hand side. Since this factor is less than unity, its presence tends to increase the absolute value of the marginal tax rate, whether the latter is positive or negative. This verifies our first claim above.

5.1.2 Participation Taxes

Participation taxes refer to $Y_i-C_i+C_0$ and are the tax/transfer for a type i worker in excess of that for a non-worker. In this section, we make Assumptions 1 and 3.

Preferences are quasilinear in consumption under Assumption 3, so first-order condition $[C_i]$ can be written as

$$f_i\lambda(Y_i - C_i + C_0)g_i(\tilde{m}_i) = f_i(\overline{\gamma}_i - \underline{\gamma}_i)C_0 g_i(\tilde{m}_i) + (\lambda - \underline{\gamma}_i)f_i G_i(\tilde{m}_i) + \sum_{k\neq i}\mu_{k\to i} - \sum_{t\neq i}\mu_{i\to t}. \tag{14}$$

The first two terms on the right-hand side relate to the participation margin, and the last two relate to the combined effects of the intensive, hours-of-work margin and of the job choice margin.

With the following assumption, we can further characterize the participation tax.

Assumption 4

(i) The welfare weights for workers decrease with $i : \underline{\gamma}_1 > \underline{\gamma}_2 > \cdots > \underline{\gamma}_N$,

(ii) The welfare weights for non-workers $\overline{\gamma}_i$ are all equal to $\overline{\gamma}$ and satisfy $\underline{\gamma}_1 > \overline{\gamma} > \underline{\gamma}_N$, that is, the least productive workers get a higher welfare weight than non-workers.

Begin with the least productive workers. Consider condition $[C_1]$.

i) Assumption 4(ii) implies that $\overline{\gamma}_1 = \overline{\gamma} < \underline{\gamma}_1$, so the first term on the righthand side of (14) is negative.

ii) Note that adding up the $[C_i]$'s and $[C_0]$ and rearranging give

$$\lambda = \sum_{i=1}^{N} f_i G_i(\tilde{m}_i)\underline{\gamma}_i + \sum_{i=1}^{N} f_i[1 - G_i(\tilde{m}_i)]\overline{\gamma}_i.$$

Since $\sum_{i=1}^{N} f_i G_i(\tilde{m}_i) + \sum_{i=1}^{N} f_i[1 - G_i(\tilde{m}_i)] = 1$, λ is a weighted average of the $2N$ numbers, $\underline{\gamma}_i$ and $\overline{\gamma}_i$. Assumption 4(ii) and the fact that λ is the above weighted average imply that the second term on the righthand side of (14) is negative.

iii) However, by Assumption 4(i), there will likely be some redistribution towards the lower types, including worker 1. Then, $\sum_{k\neq 1}\mu_{k\to 1} - \sum_{t\neq 1}\mu_{1\to t}$ in (14) is likely to be positive, weakening the case of a negative $Y_1 - C_1 + C_0$.

From (i) to (iii), we see that at the bottom, the extensive margin influences captured by the first two terms and the intensive margin and job-choice margin influences captured by the last two terms work in opposite directions. Their net effect is likely to be a NIC for the least productive workers ($Y_1 - C_1 + C_0 > 0$) if the incentive constraints preventing higher type workers to mimic type 1 bind and make $\sum_{k\neq 1}\mu_{k\to 1} - \sum_{t\neq 1}\mu_{1\to t}$ positive and large; otherwise, it is likely to be an EITC ($Y_1 - C_1 + C_0 < 0$).

iv) As i increases, the first two terms on the righthand side of (14) eventually become positive (if they are negative at $i = 1$), due to the decrease in $\underline{\gamma}_i$. However, the last two terms tend to be negative, because Assumption 4(i) implies redistribution from the more productive to the less productive workers and therefore a tendency for the last two terms to decrease and become more negative.

Therefore, the two sets of margins have opposite effects on the participation taxes. For the lowest types, extensive margin consideration calls for a negative participation tax, while intensive margin and job-choice margin considerations call for a positive one; and vice versa for the highest types.

5.2 Summary of Results in the Model with an Extensive Margin

In this section, we characterized marginal tax rates and participation taxes when a labour market participation decision is introduced into the multiple-type model. We have two main findings.

i) The introduction of the extensive margin tends to increase the absolute value of the marginal tax rates of all workers below the top, whether they are positive or negative.

ii) Intensive margin considerations and extensive margin considerations tend to have opposite effects on the level of the participation tax, with the direction of their interaction depending on skill levels. Under some assumptions about welfare weights, participation taxes tend to be higher for workers below a certain type when the model includes a participation decision, while the opposite is true for workers above that type.

iii) Participation taxes at the bottom can be of a negative income tax (NIC) type or an earned income tax credit (EITC) type, depending on the opposing influences of the two margins.

6 Job-Specific Disutility-of-Work

A key assumption of the paper, Assumption 1(ii) on absolute advantage in workers' job-specific productivity, may not always hold. Some workers may be more/less productive than others at all jobs and therefore do not have an absolute advantage. In fact, the pattern of individuals' comparative advantage and absolute advantage can be complex. Some may do very well at any job. Others may be experts in certain jobs only. Still others may be able to do many jobs but are not the most productive at any job.

This does not necessarily affect our main results. For example, it does not necessarily mean that negative marginal tax rates cannot arise below the top. Consider the case where there is a worker who is more productive than others at any job. As long as some worker below the top is more productive than his potential mimickers and the top worker is not among those potential mimickers, the former still faces a negative marginal tax (as long as his upper income bound does not bind too much). In this sense, Assumption 1(ii) can be relaxed to some extent.

In reality, different workers may also have different disutility-of-work in different jobs. Similarly as in Diamond (2006), we can amend our model to introduce job-specific disutility, in addition to job-specific productivities. For example, suppose the utility function is $U(C,L) = u(C) - \alpha_i^j h(L)$, where $h(L)$ is the usual disutility-of-work and α_i^j is worker i's job-specific disutility parameter at job j. For simplicity, we further assume quasilinear-in-leisure preferences, so $U(C,L) = u(C) - \alpha_i^j L$. Then, the type-specific utility function becomes

$$V_i^j(C,Y) = u(C) - \alpha_i^j \frac{Y}{w_i^j}.$$

Thus, workers differ in both productivities and preferences. Optimal income taxation in such a model has been analyzed by Boadway *et al* (2002), and their focus was on the consequences of giving various welfare weights to individuals with different preferences.

For our purpose, having job-specific disutility beside job-specific productivity makes our model more flexible. By varying parameter values, the standard model and our basic model above can emerge as special cases.

Following Boadway *et al* (2002), denote by $\widehat{w}_i^j \equiv w_i^j/\alpha_i^j$. Then we have $V_i^j(C,Y) = u(C) - Y/\widehat{w}_i^j$. If $\widehat{w}_i^j$ satisfy Assumption 1, the model will be equivalent to the basic model in terms of results. Note that only $\widehat{w}_i^j$ are required to satisfy Assumption 1, so each worker's wage rates can actually be constant across jobs ($w_i^1 = w_i^2$, $i = 1, 2$), just as in a standard model. Therefore, introducing job-specific disutilities gives us an extra degree of freedom that enables us to completely do away with the assumptions on job-specific productivities. On the other hand, if Assumptions 1(i) (comparative advantage) and 1(iii) (ordering of highest wage rates) but not 1(ii) (absolute advantage) hold for $\widehat{w}_i^j$, the model resembles a standard model.

7 Concluding Remarks

This paper characterized optimal income taxation in a setting where workers of different skills are better suited for different jobs. It was shown that, relative to more standard settings, optimal taxation may involve more redistribution, if jobs differ in the range of available income they offer and if individuals are more productive than others at the jobs that suit them best. For those to whom the tax system redistributes, negative marginal tax rates can often arise, because mimickers are locally less productive than these individuals. For those from whom the tax system redistributes, even those who are not the most skilled may also face zero marginal tax rates. Moreover, first-best maximin outcomes may be attained using second-best nonlinear income tax policies. As well, when the model includes a labour market participation decision, the extensive and intensive margins tend to have opposite effects on participation taxes.

In future research, we plan to consider the following issues.

First, it may be desirable to model jobs or, more generally, the demand side of the labour market, more explicitly. In the present paper, the production technology is linear, and labour of different workers are perfect substitutes. Although there are different "jobs", firms are absent. Moreover, general equilibrium effects are not considered.

Second, due to the difficulty associated with discrete comparisons, this paper has only offered limited results on whether and how a tax system should distort individuals' job choices. It might be useful to explore this issue further, in light of the contributions of Bhagwati and Srinivasan (1977) and Grossman (1983), for example.

A third issue relates to the assumption that workers have an absolute advantage at their most suitable job. In the basic 2-by-2 case, the absence of this assumption will invalidate the negative marginal tax result. In the multiple-type cases of Sections 4 and

5, as long as there are at least some worker types that satisfy this assumption, negative marginal tax rates may still hold. However, it would be potentially interesting to study distortions in job choice under more general assumptions.

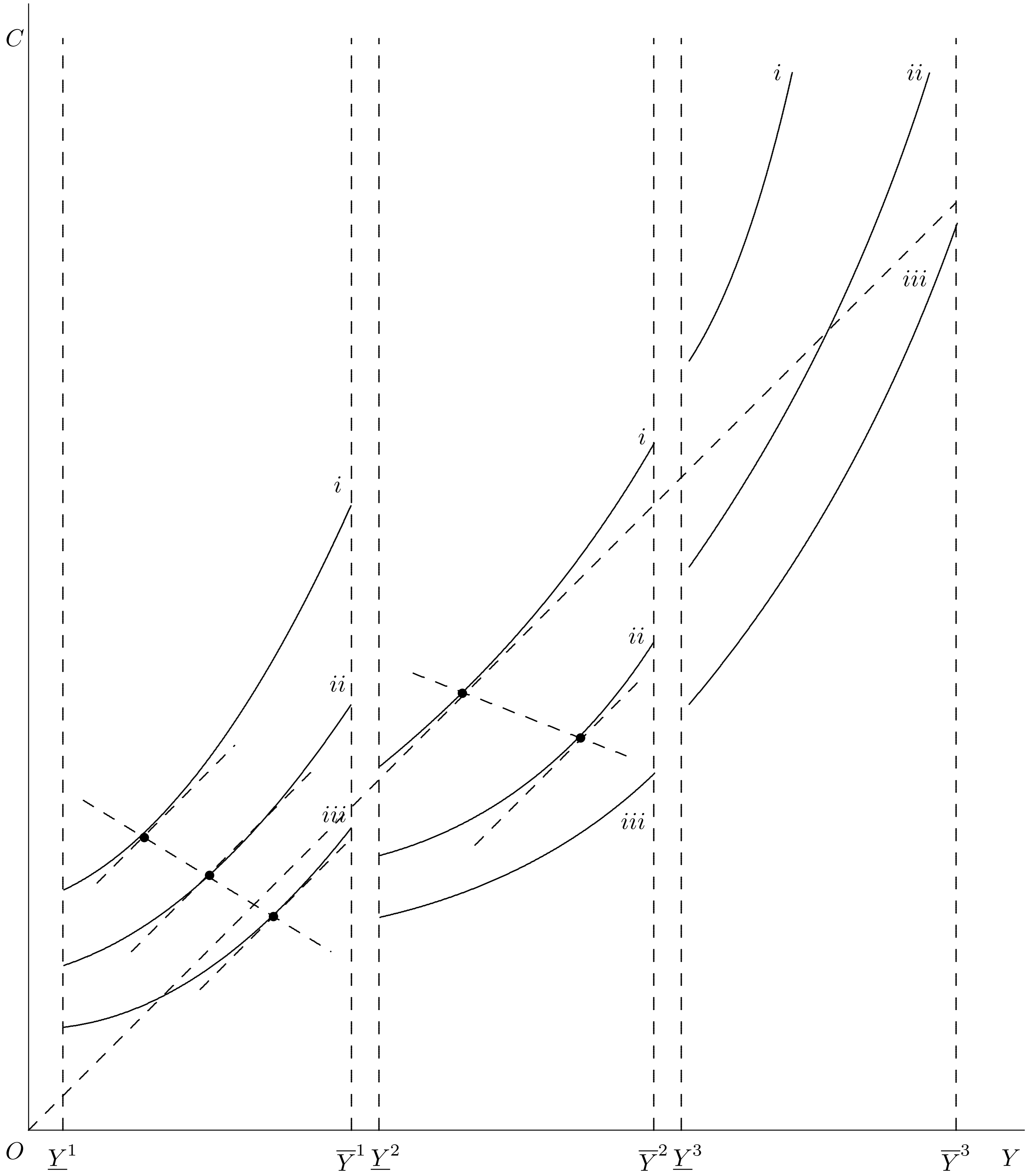


Figure 1
Three Indifference Curves of Worker 2

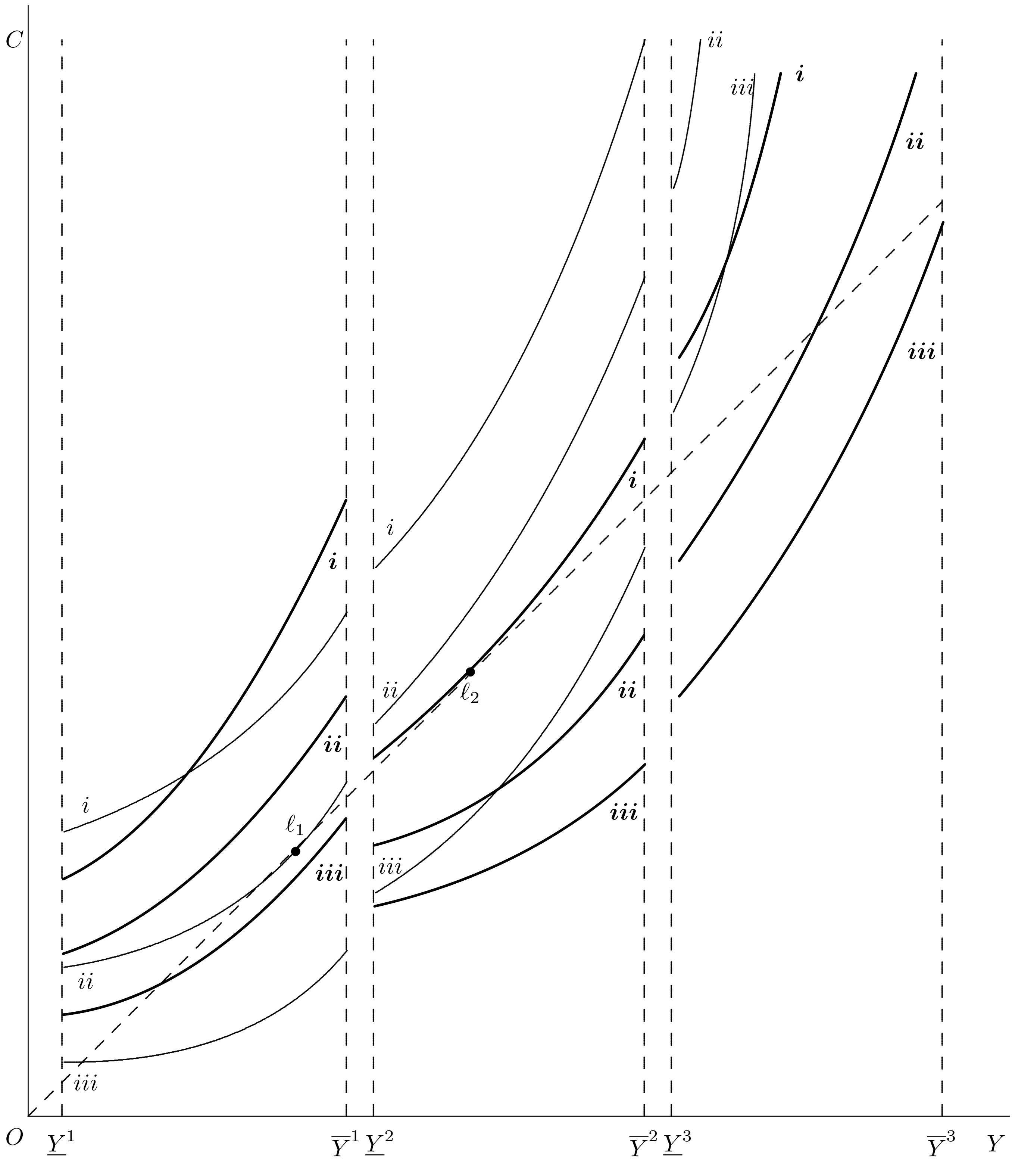


Figure 2
Indifference Curves of Worker 1 and Worker 2

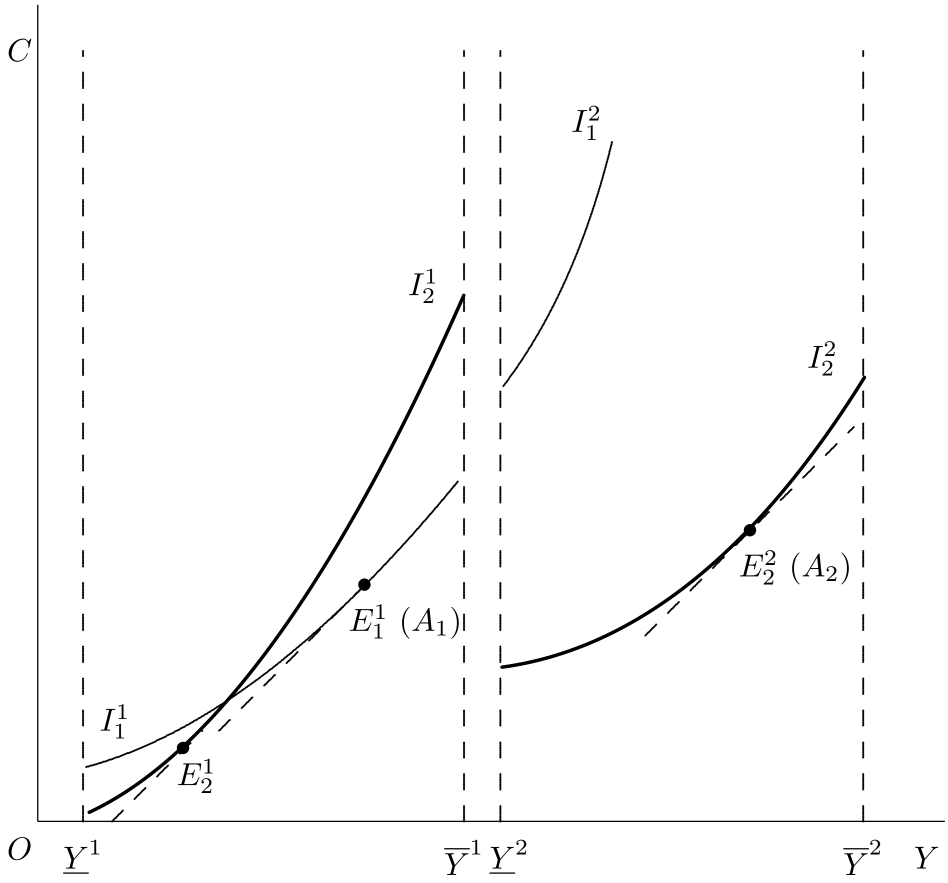


Figure 3

Case (i): *A Fully Revealing Optimum*

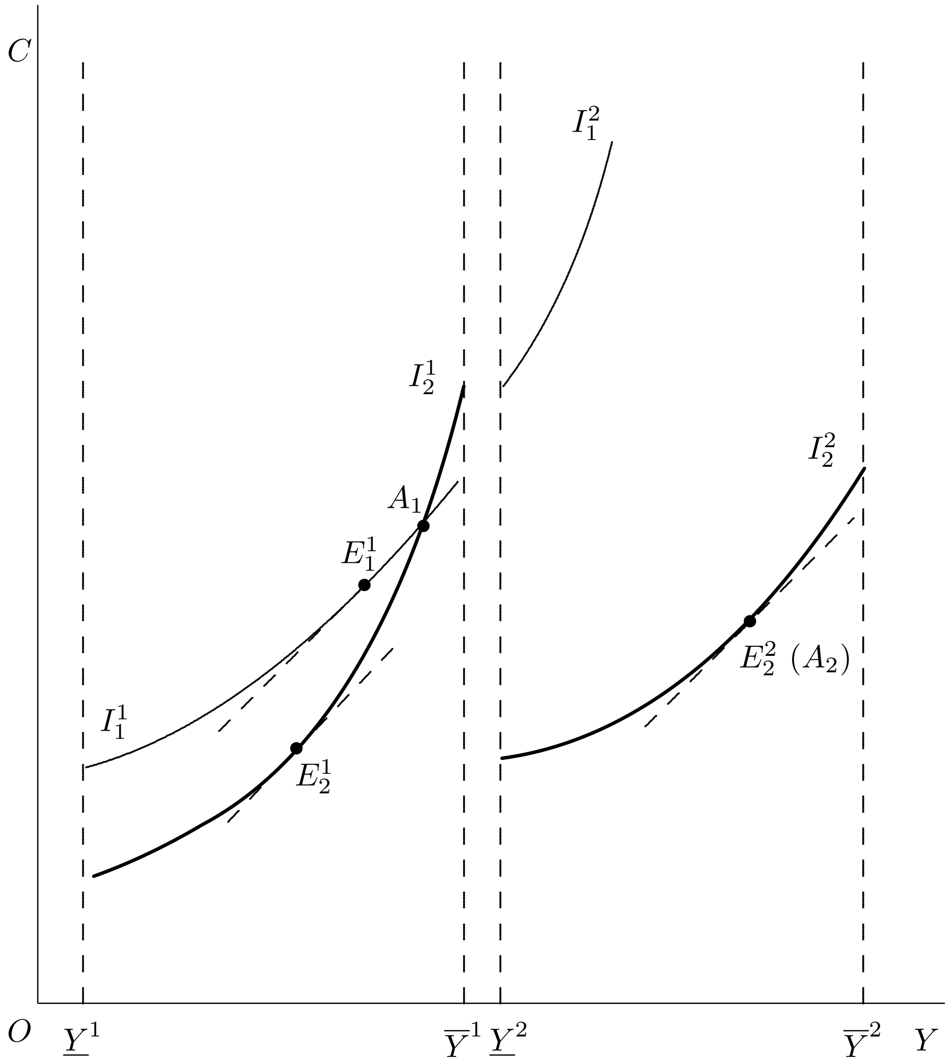


Figure 4

Case (i): *An Interior Optimum*

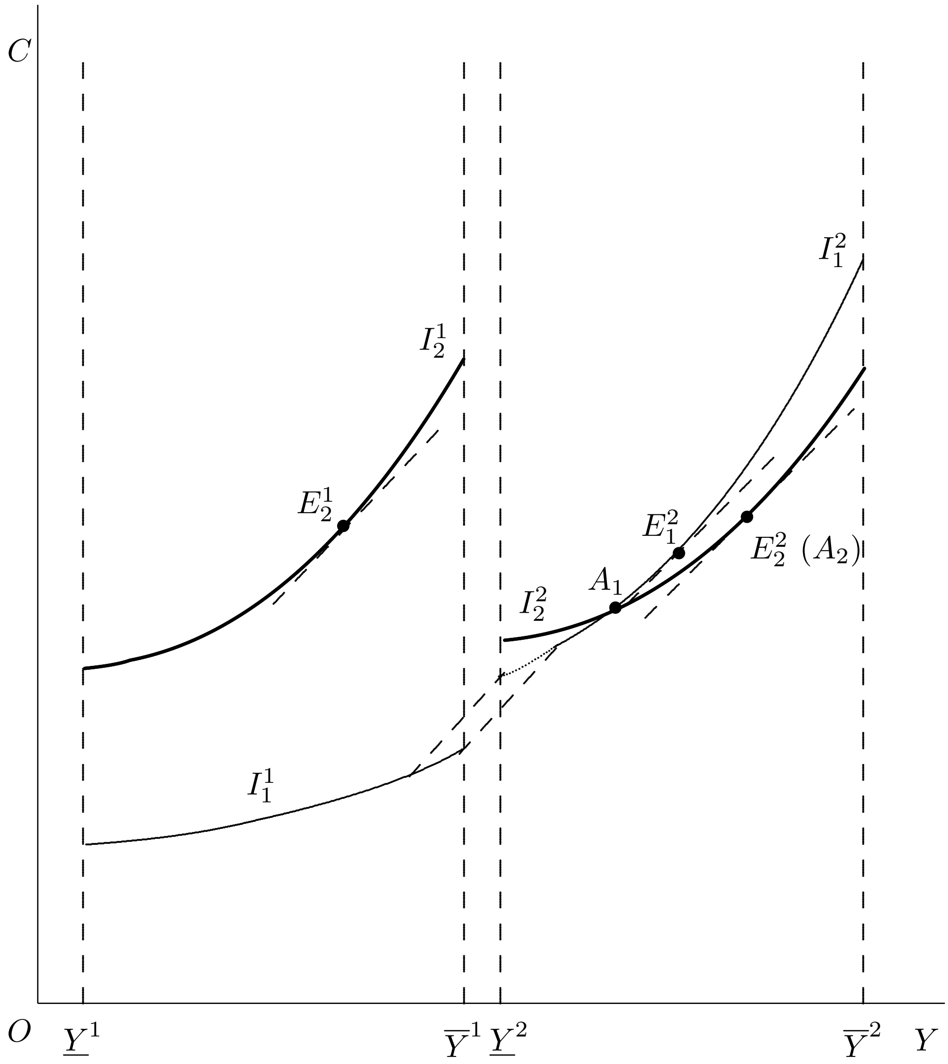
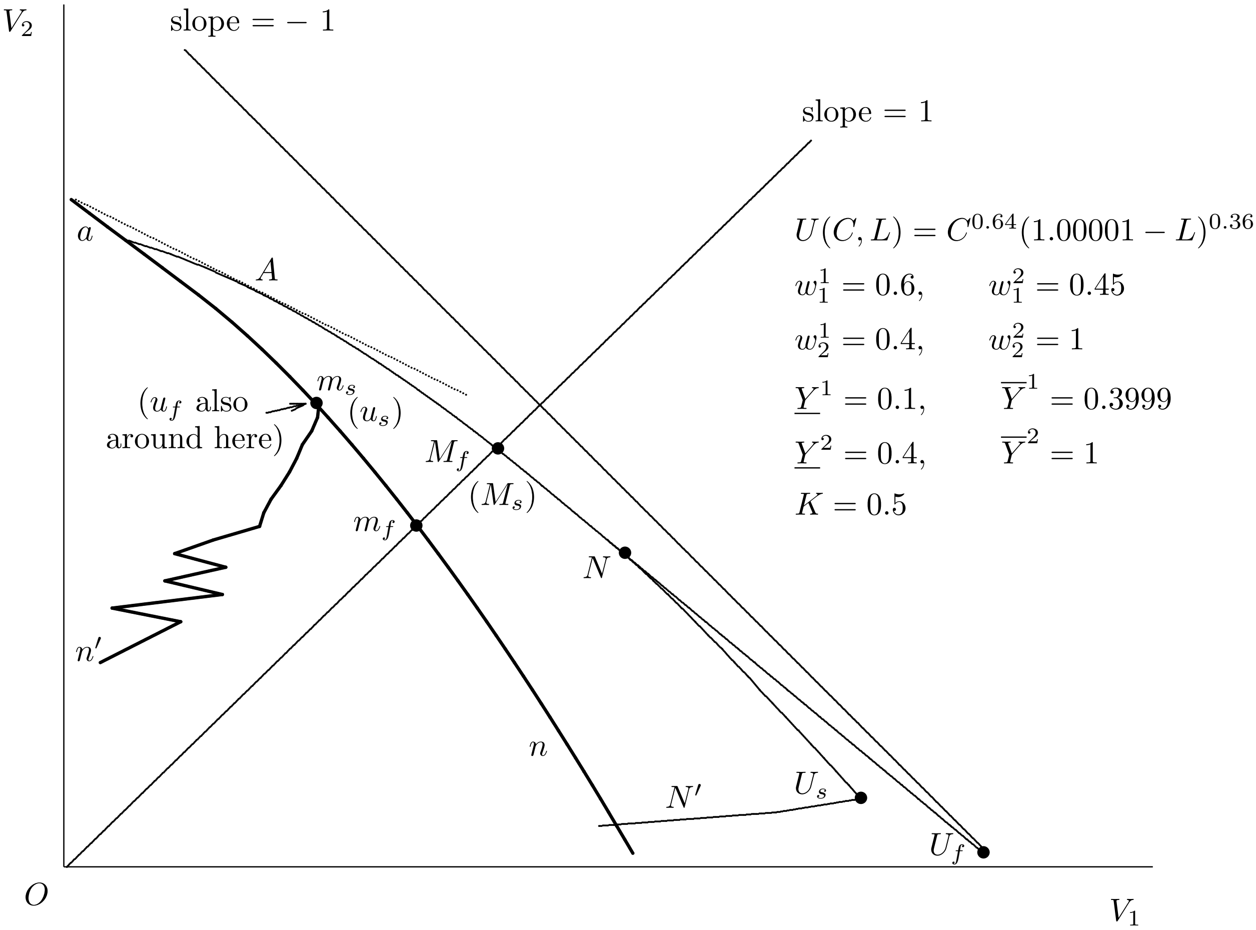


Figure 5

Case (ii): *Both Workers in Job 2*



AM_fNU_f : first-best, Case (i) UPF; $\quad AM_sNU_sN'$: second-best, Case (i) UPF

am_fn : first-best, Case (ii) UPF; $\quad am_sn'$: second-best, Case (ii) UPF

$U_f, u_f \,/\, U_s, u_s$: first-/second-best utilitarian outcome

$M_f, m_f \,/\, M_s, m_s$: first-/second-best maximin outcomes

(|slope| of am_fn around 45^o line greater than 1, due to lower income bound of job 2)

Figure 6
First- and Second-Best UPF's in Case (i) and Case (ii)

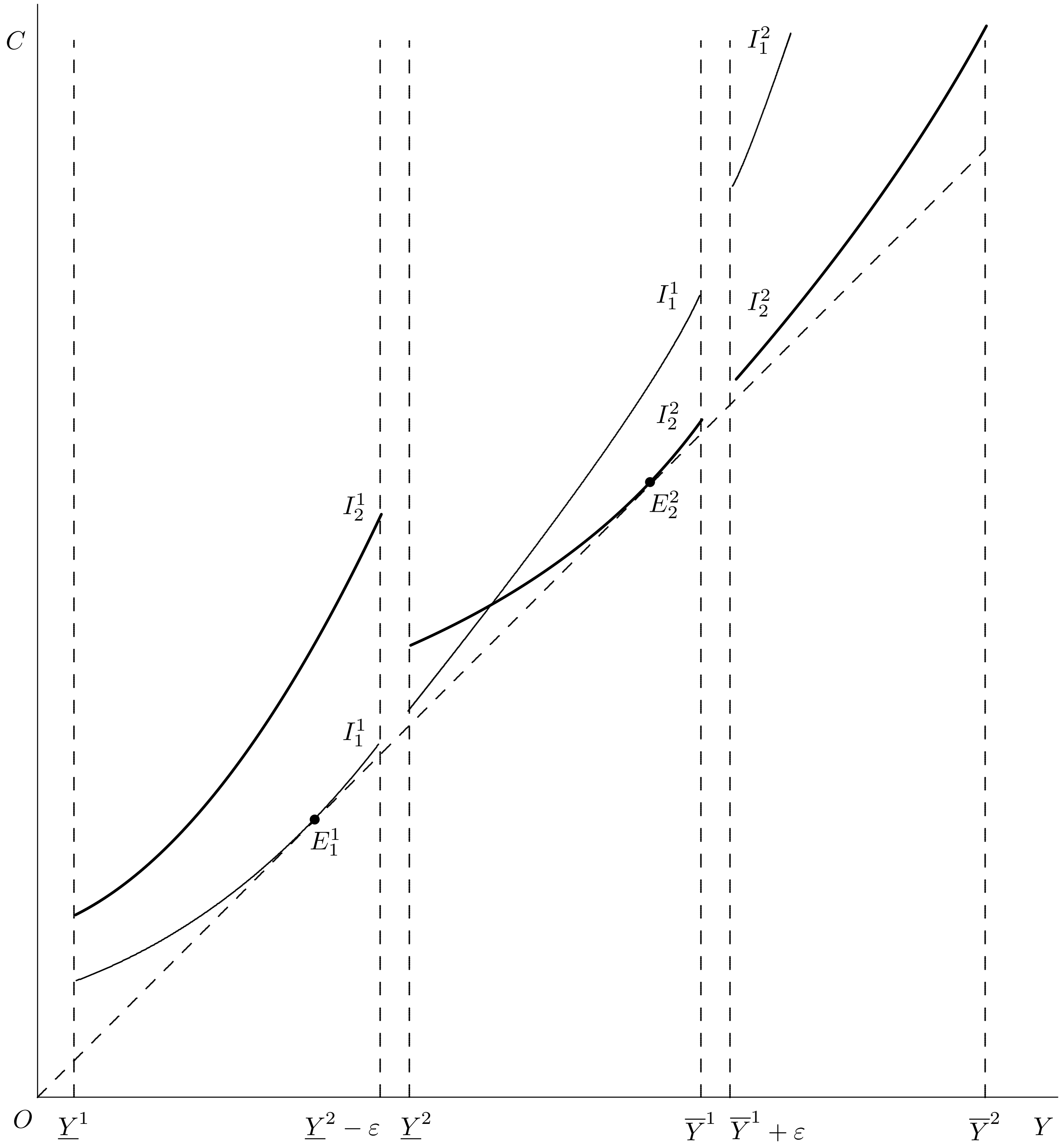


Figure 7
Partially Overlapping Income Range Case: Indifference Curves in the Laissez-Faire

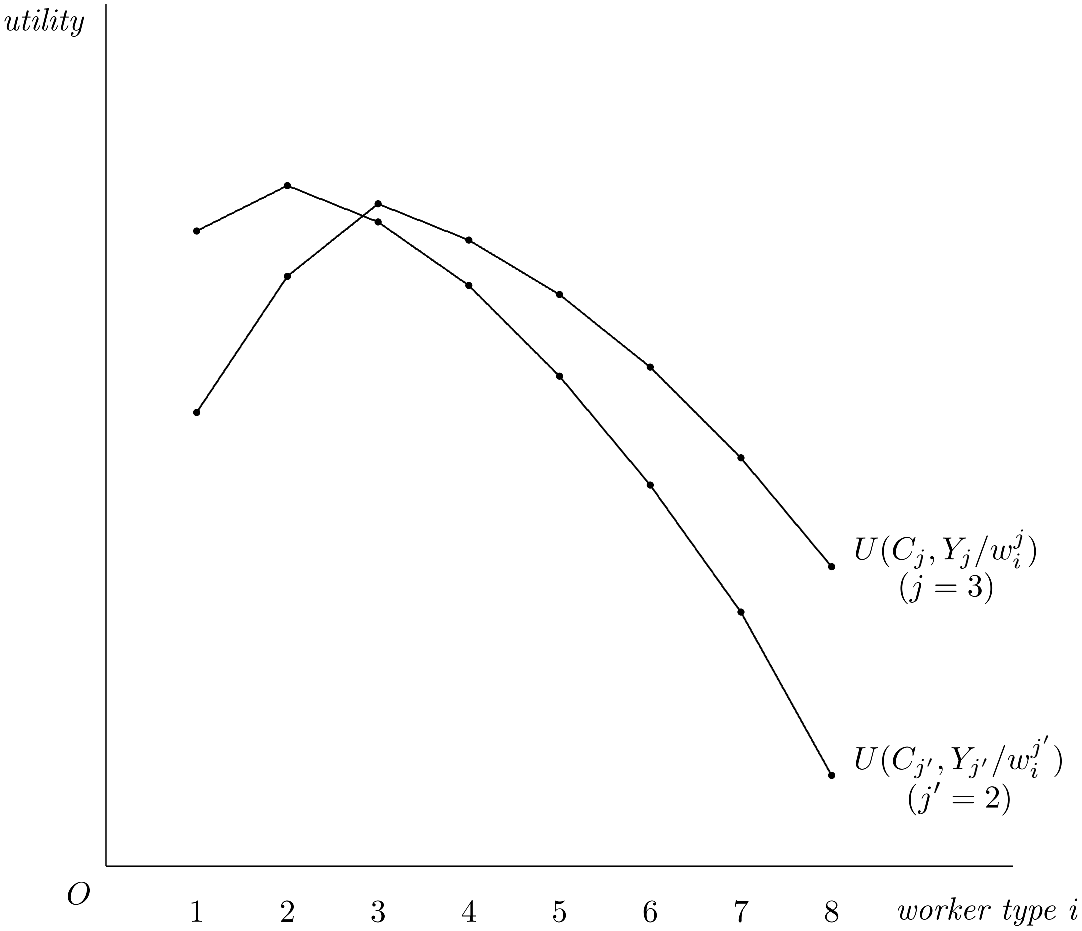


Figure 8

Utility Curves when SCP Holds

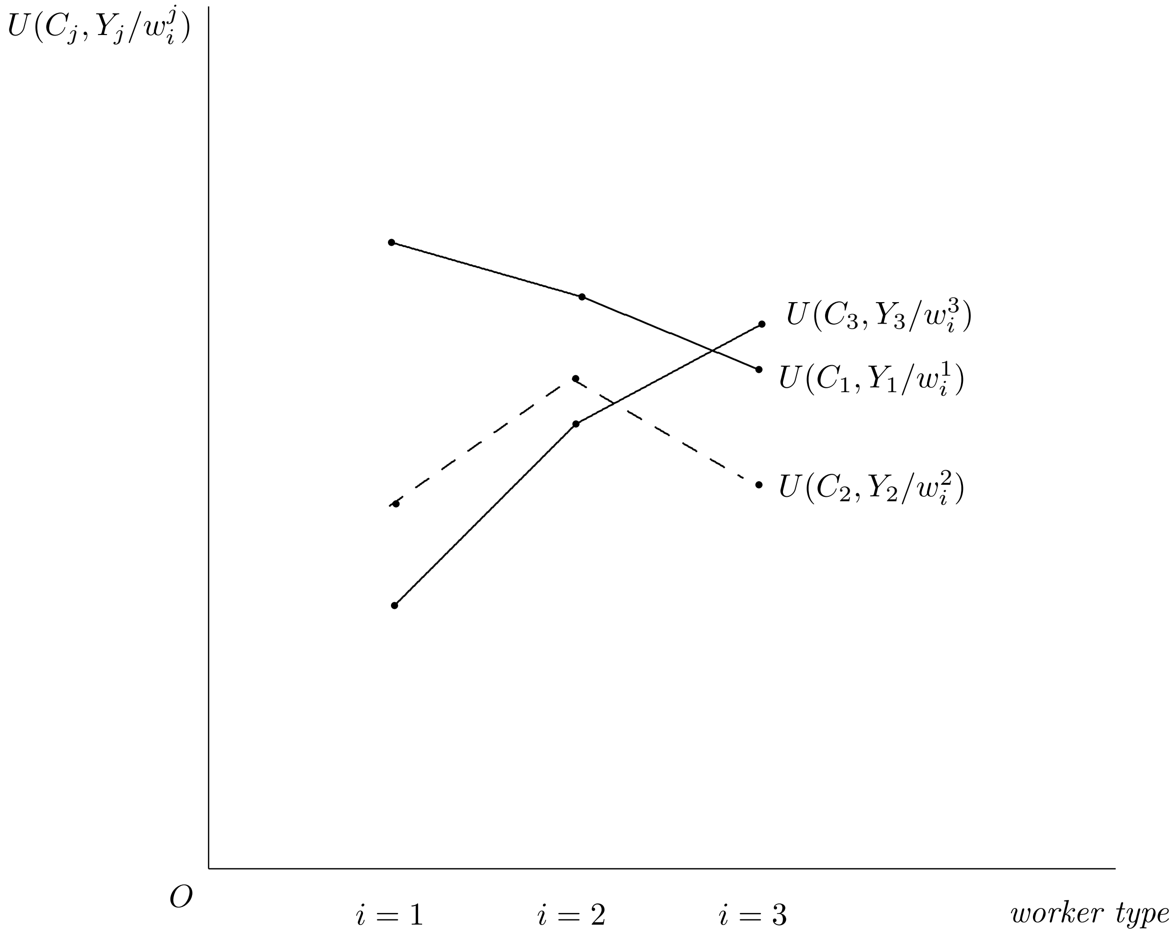


Figure 9

3-by-3 Case: Utility Curves that Violate Incentive Constraint

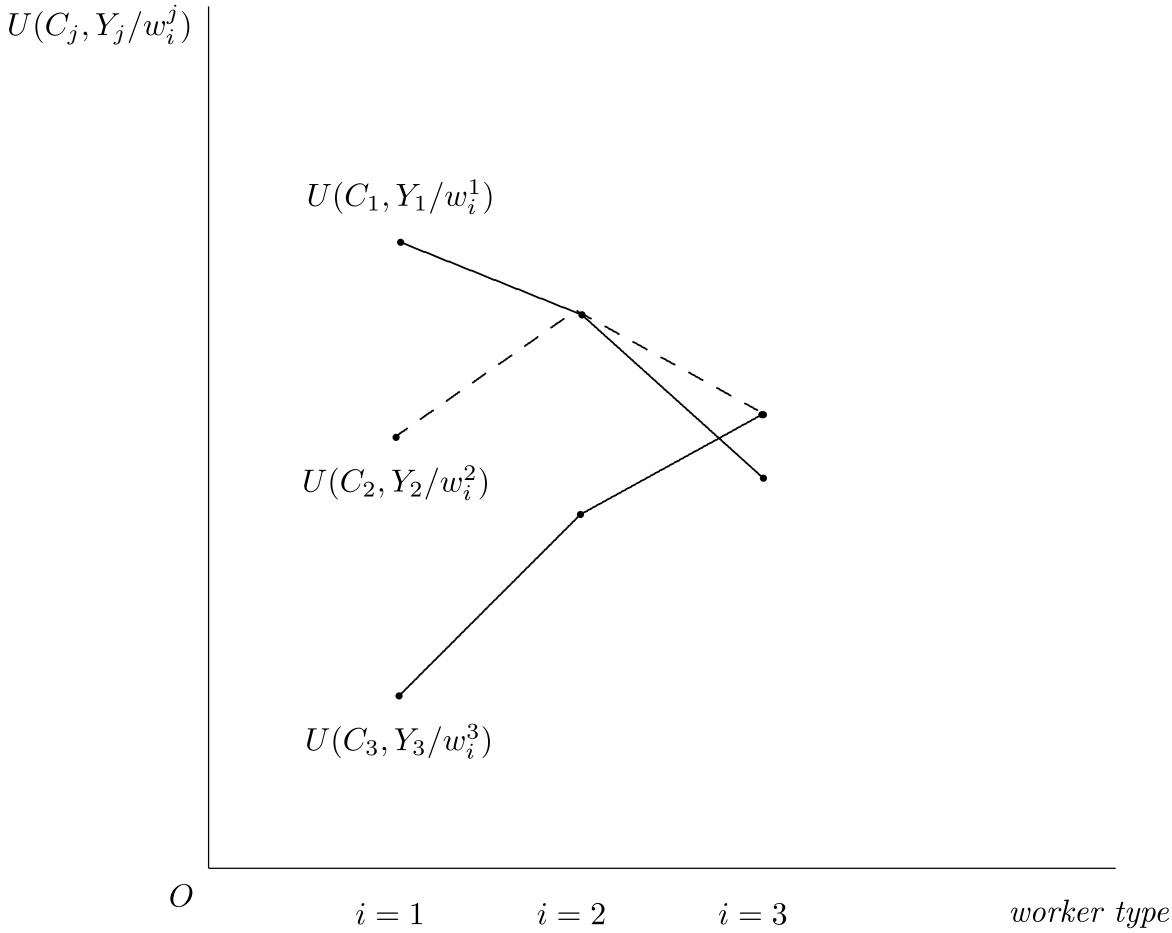


Figure 10

3-by-3 Case: Solution when SCP holds

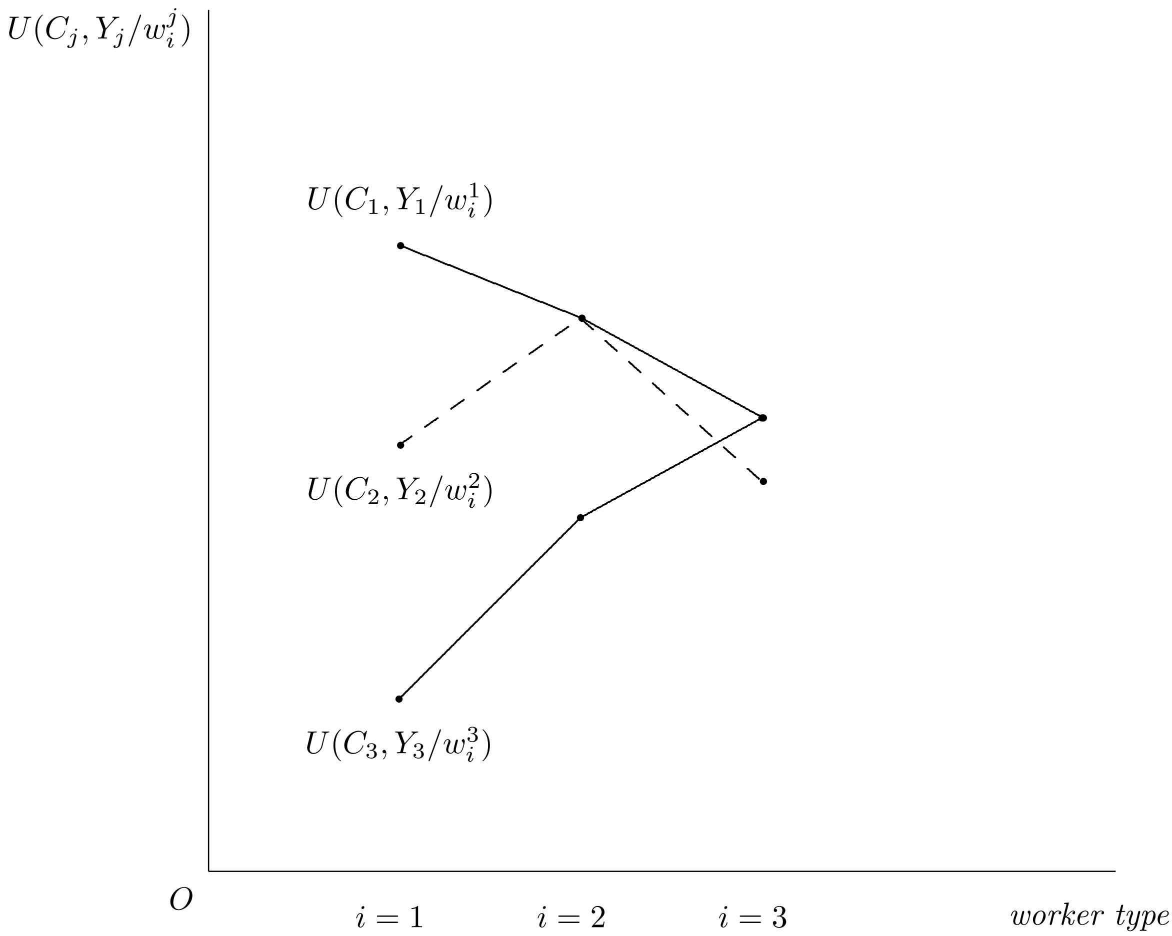


Figure 11

3-by-3 Case: Solution when SCP does not hold

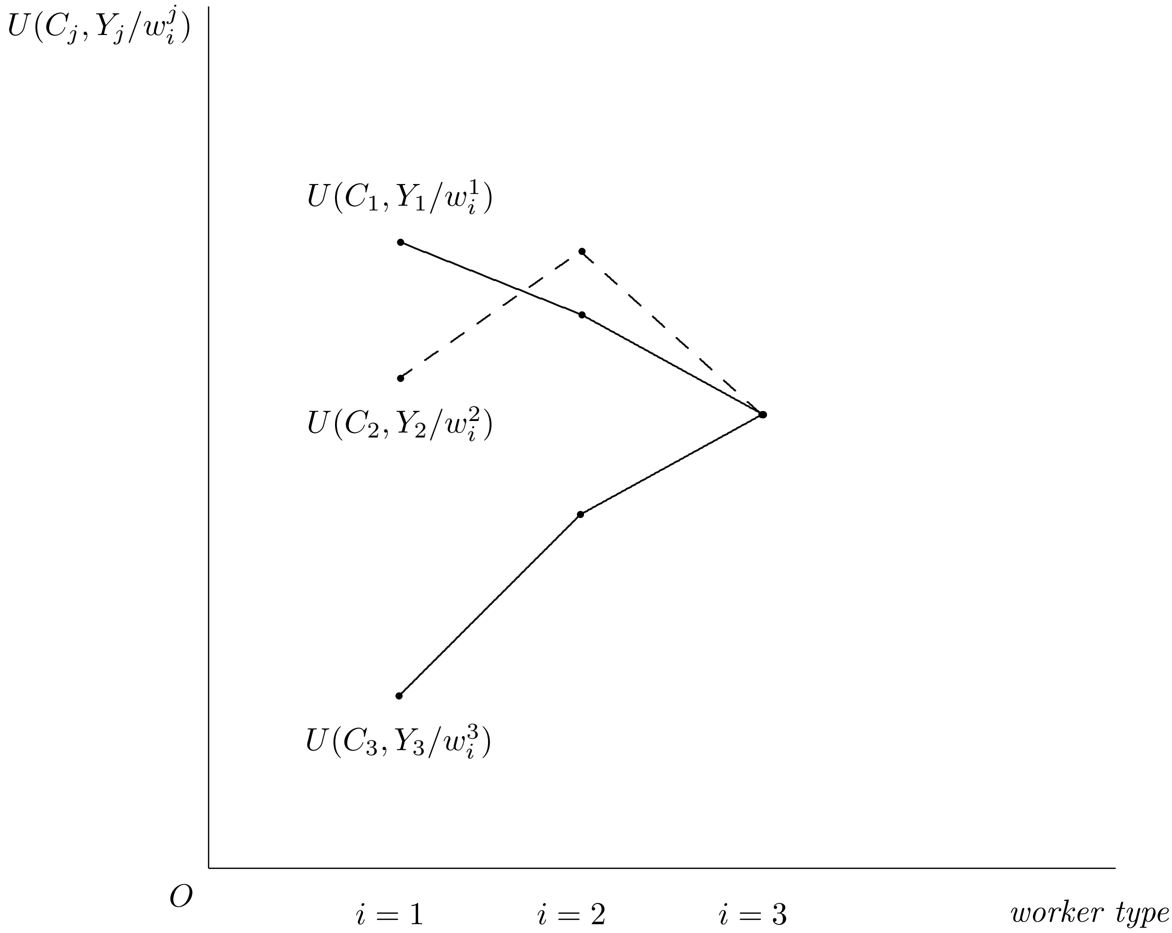


Figure 12

3-by-3 Case: solution when SCP does not hold

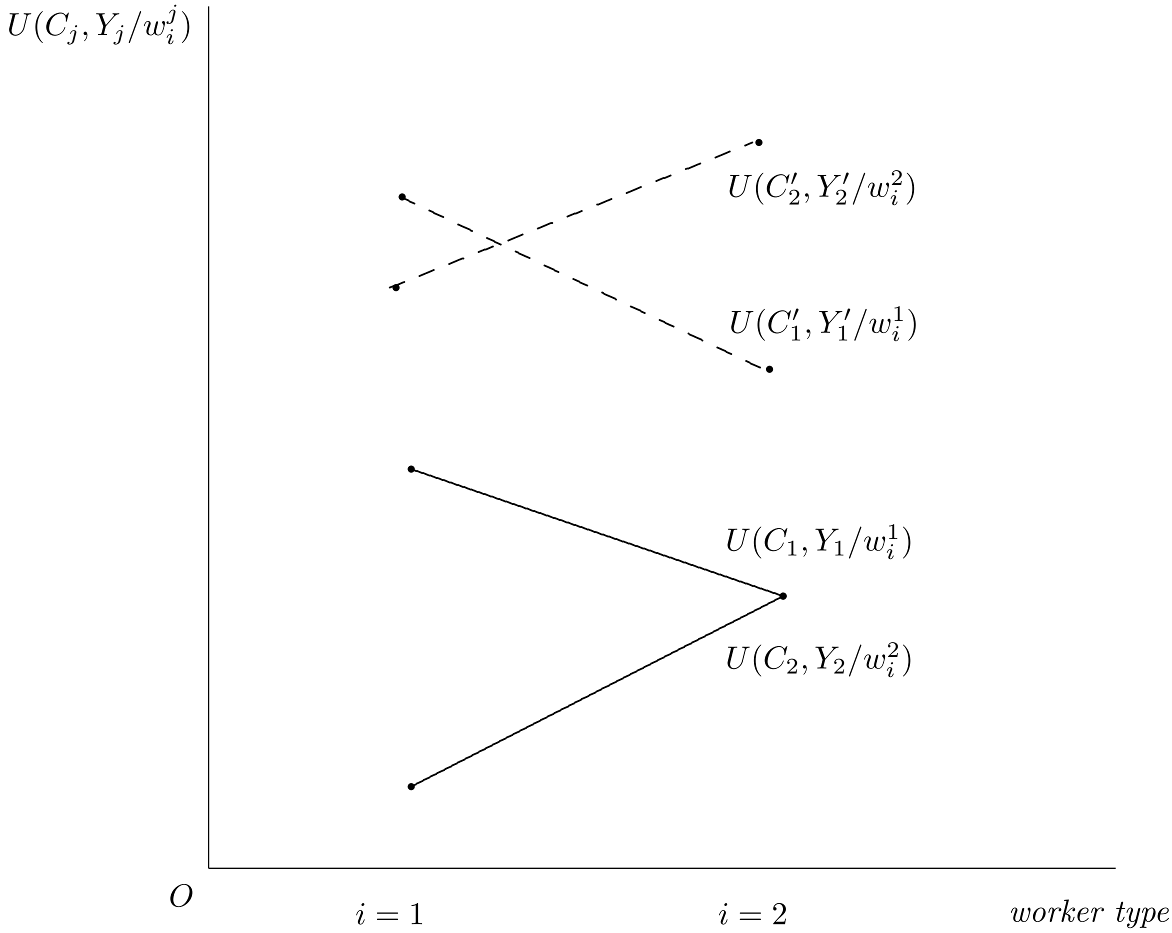


Figure 13 (corresponds to Figures 3 and 4)

Utility Curves in the 2-by-2 Case

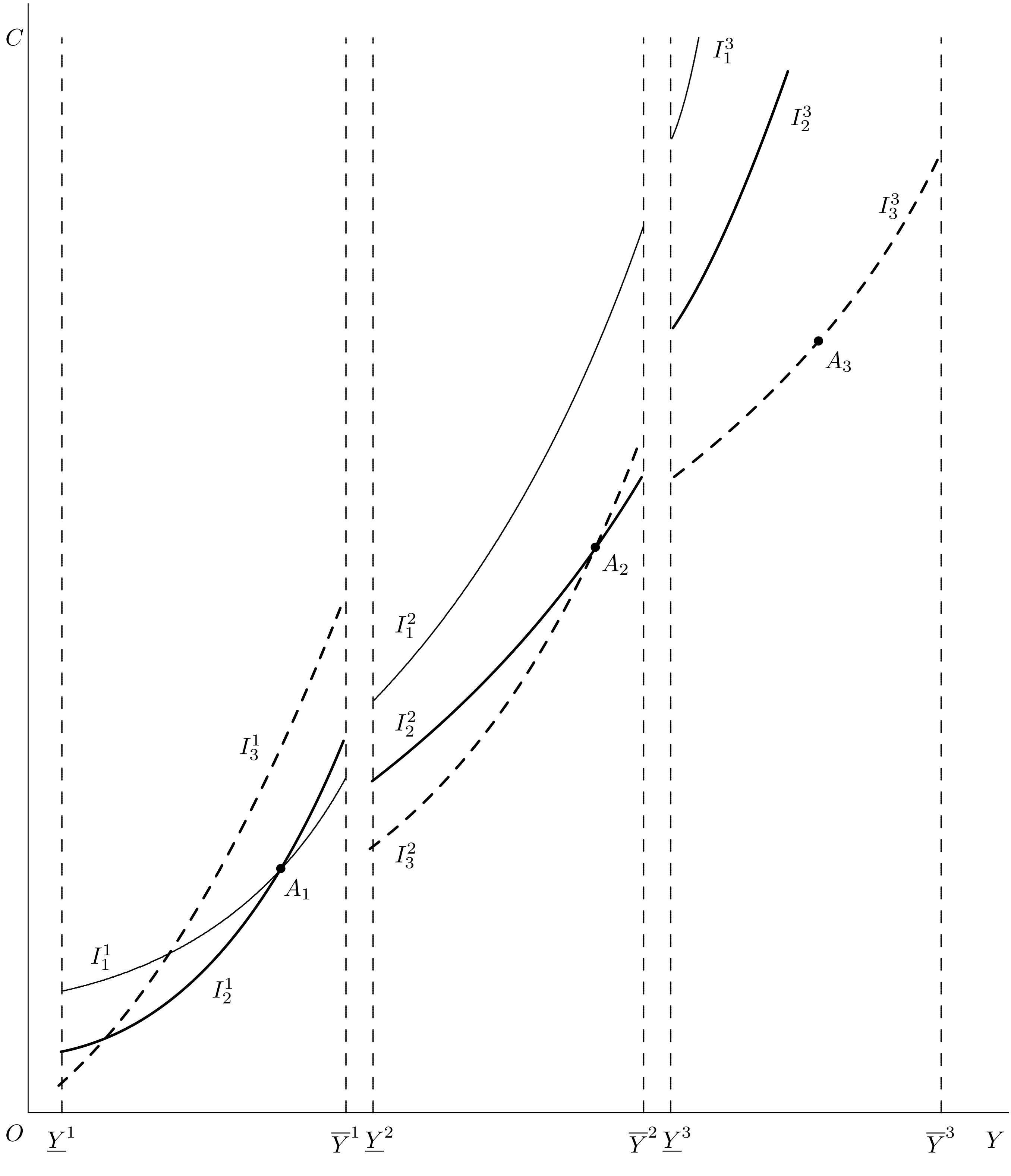


Figure 14 (corresponds to Figure 10)

3-by-3 example: optimum in C-Y space when SCP holds

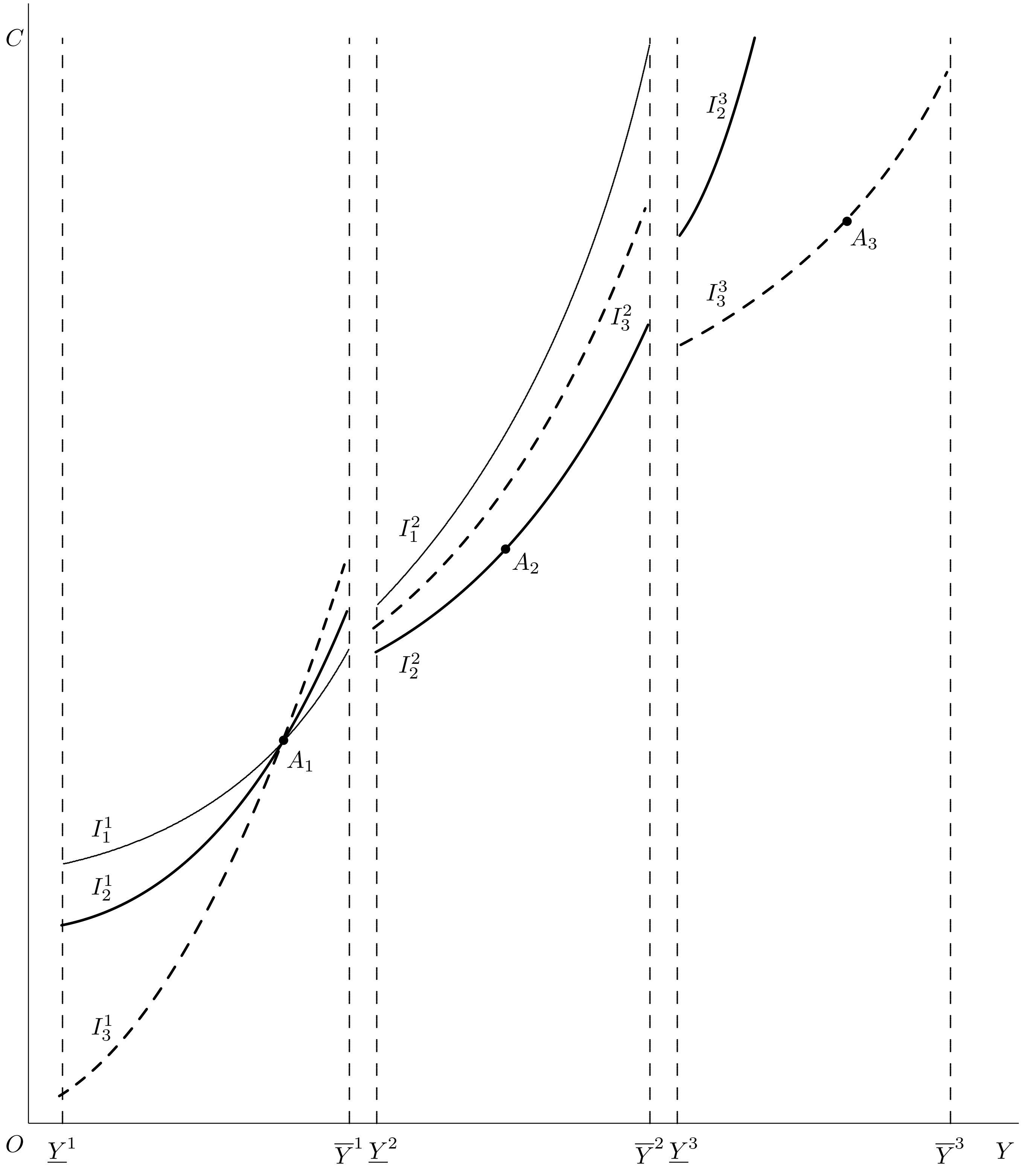


Figure 15 (corresponds to Figure 11)
3-by-3 example: optimum 1 in C-Y space when SCP does not hold

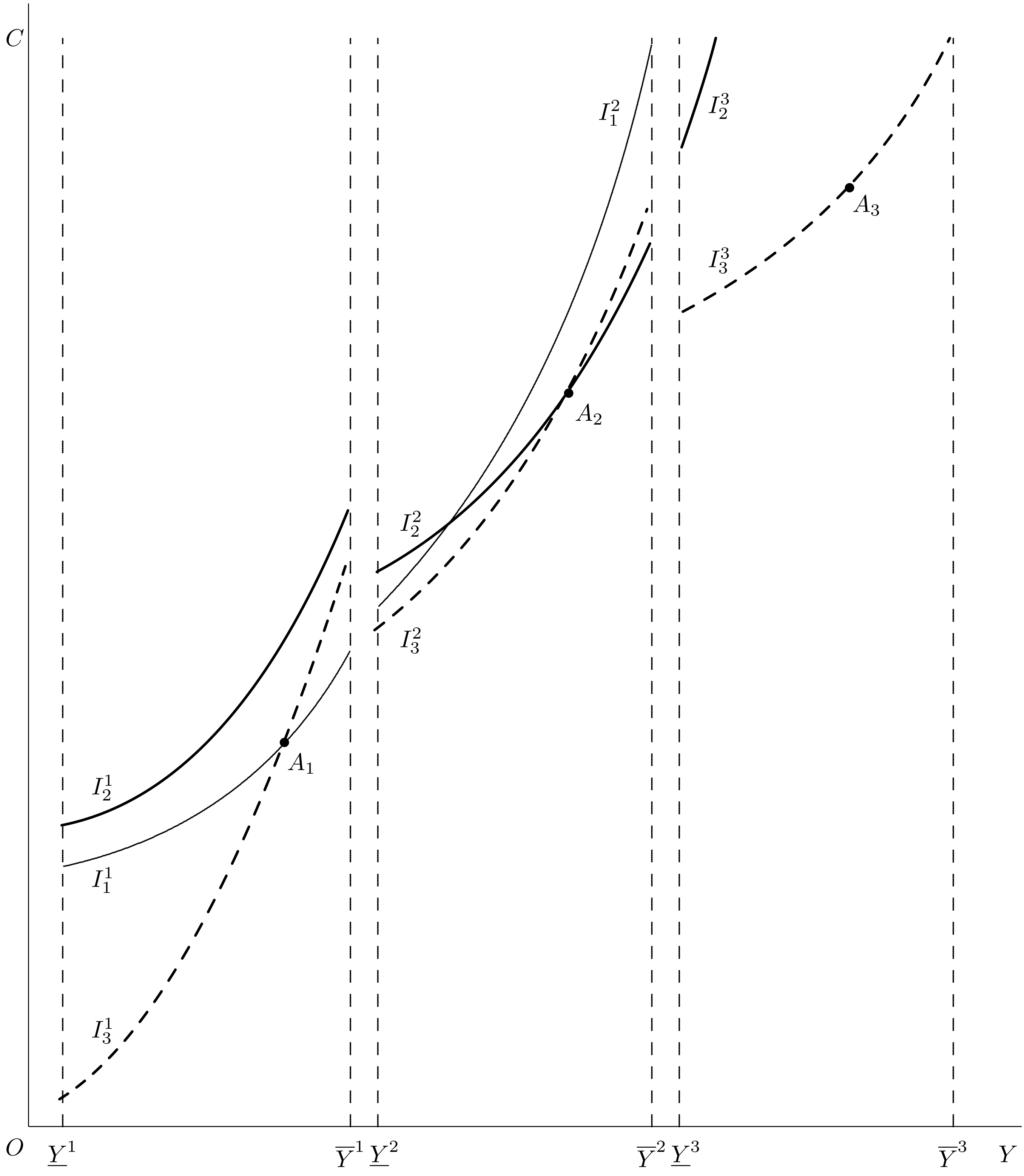


Figure 16 (corresponds to Figure 12)

3-by-3 example: optimum 2 in C-Y space when SCP does not hold

References

Beaudry, Paul, Charles Blackorby, and Dezso Szalay (2009), 'Taxes and Employment Subsidies in Optimal Redistribution Programs', *American Economic Review* **99**, 216–242.

Bhagwati, Jagdish N. and T.N. Srinivasan (1977), 'Education in a 'Job Ladder' Model and the Fairness-in-Hiring Rule', *Journal of Public Economics* **7**, 1–22.

Boadway, Robin, Maurice Marchand, Pierre Pestieau, and Maria Del Mar Racionero (2002), 'Optimal Redistribution with Heterogeneous Preferences for Leisure', *Journal of Public Economic Theory* **4**, 475–98.

Choné, Philippe and Guy Laroque (2010), 'Negative Marginal Tax Rates and Heterogeneity', *American Economic Review,* **100**, 2532–47.

Choné, Philippe and Guy Laroque (2011), 'Optimal Taxation in the Extensive Model', *Journal of Economic Theory* **146**, 425–453.

Diamond, Peter (1980), 'Income Taxation with Fixed Hours of Work', *Journal of Public Economics* **13**, 101–110.

Diamond, Peter and Johannes Spinnewijn (2011), 'Capital Income Taxes with Heterogeneous Discount Rates', *American Economic Journal: Economic Policy* **3**, 52–76.

Ebert, Udo (1992), 'A Reexamination of the Optimal Nonlinear Income Tax', *Journal of Public Economics* **49**, 47?73.

Grossman, Gene M. (1983), 'Partially Mobile Capital: A General Approach to Two Sector Trade Theory', *Journal of International Economics* **14**, 1–17.

Guesnerie, Roger and Jesus Seade (1982), 'Nonlinear Pricing in a Finite Economy', *Journal of Public Economics* **17**, 157–179.

Hellwig, Martin F. (2007), 'A Contribution to the Theory of Optimal Utilitarian Income Taxation', *Journal of Public Economics* **91**, 1449–1477.

Homburg, Stefan (2002), 'The Optimal Income Tax: Restatement and Extensions', *Finanzarchiv: Public Finance Analysis* **58**, 363–395.

Hungerbühler, Mathias, Etienne Lehmann, Alexis Parmentier, and Bruno Van der Linden (2006), 'Optimal Redistributive Taxation in a Search Equilibrium Model', *Review of Economic Studies* **73**, 743–67.

Jacobsen Kleven, Henrik, Claus Thustrup Kreiner, and Emmanuel Saez (2009), 'The Optimal Income Taxation of Couples', *Econometrica* **77**, 537–60.

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden (2013), 'Optimal Redistributive Taxation with both Extensive and Intensive Responses', *Journal of Economic Theory*, forthcoming..

Krause, Alan (2009), 'Optimal Nonlinear Income Taxation with Learning-by-Doing', *Journal of Public Economics* **93**, 1098–1110.

Laroque, Guy (2005), 'Income Maintenance and Labor Force Participation', *Econometrica* **73**, 341–76.

Matthews, Steven and John Moore (1987), 'Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening', *Econometrica* **55**, 441–467.

Mirrlees, James A. (1971), 'An Exploration in the Theory of Optimum Income Taxation', *The Review of Economic Studies* **38**, 175–208.

Saez, Emmanuel (2002), 'Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses', *The Quarterly Journal of Economics,* **117**, 1039–1073.

Stern, Nicholas H. (1982), 'Optimum Taxation with Errors in Administration', *Journal of Public Economics* **17**, 181–211.

Stiglitz, Joseph E. (1982), 'Self-Selection and Pareto Efficient Taxation', *Journal of Public Economics* **17**, 213–240.

Tuomala, Matti (1990), *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.

Weymark, John A. (1986), 'A Reduced-Form Optimal Nonlinear Income Tax Problem', *Journal of Public Economics* **30**, 199–217.