Appendix to “How Do Training Programs Assign Participants to Training? Characterizing the Assignment Rules of Government Agencies for Welfare-to-Work Programs in California”

Oscar A. Mitnik

December 2009

In the paper I sketched a simple returns-maximization model for the decision making process of a county’s welfare agency that has to assign welfare recipients to different treatment alternatives. Here I present fully the model and characterize testable implications with respect to the effects of changes in the budget constraint, in local economic conditions and in the initial skills distribution of the welfare entrants. I also discuss how the basic model can be extended to allow for a role of the local political environment. I suggest extensions in which the welfare agency has a richer set of preferences over participants than implied by the simple returns-maximization model, for example across types of training or across individuals with different initial skills. These preferences can arise from the aggregation of preferences of the different stakeholders, for example through the policy and local political process. I show that these preferences can affect the assignment to training decisions.

1 The returns-maximization model

The county’s problem Assume a county needs to make training decisions for an entering cohort of welfare recipients. Each individual’s characteristics, which are denoted by a random vector \( X_i \), include her demographic information, educational attainment, previous employment and earnings histories, and, in some cases, evaluations (either objective or subjective) made by county officers on the individual’s potential.\(^1\) I assume that this information is aggregated by the county in a single-dimension index \( \theta = \omega(X_i) \), where \( \omega \) is a known weighting function. Furthermore, I assume that this index is continuous, that it takes on values in the interval \([\theta, \theta]\), and that it follows a county-specific probability distribution function, \( f(\theta) \). I assume that the ability to generate earnings, or to find employment, is positively correlated with \( \theta \), which can be interpreted as a measure of an individual’s initial skills level. I refer to \( \theta \) simply as skills.

The county has three treatment options for each individual: 1) to not provide any training services (treatment \( N \)); 2) to provide LFA training services (treatment \( L \)); or 3) to provide HCD

\(^1\) As discussed in the paper, some elements of the vector \( X_i \) are observed and some are not.
training services (treatment $H$). The county evaluates the effect of each treatment option on an individual with skills $\theta$, by looking at the outcome variable $Y(\theta)$. This variable can represent different outcomes of interest for the county (e.g., earnings, or the probability of finding a job).

Treatments $N$, $L$ and $H$ have associated, at each period, outcome generating functions $Y^N(\theta)$, $Y^L(\theta)$ and $Y^H(\theta)$, respectively. The outcome functions for treatments $L$ and $H$ can also be rewritten as $Y^L(\theta) = Y^N(\theta) + \Delta^L(\theta)$ and $Y^H(\theta) = Y^N(\theta) + \Delta^H(\theta)$, where $\Delta^L$ and $\Delta^H$ represent the treatment effects of training $L$ and $H$, respectively, and they are assumed to be non-negative functions of $\theta$.\(^2\)

Because training takes time, the present value of the outcome under treatment $N$ could be higher than the present value of the outcome under treatment $L$, or treatment $H$ for that matter, for some values of $\theta$. I assume that each individual, at the time of treatment, has a future labor life of $T$ periods (fixed). Treatment $L$ takes place over $\tau_L > 0$ periods, while training $H$ takes place over $\tau_H > 0$ periods. As HCD training takes longer than LFA training, this implies that $\tau_H > \tau_L$ (and $\tau_N \equiv 0$).

The expected present value of the future stream of benefits (in each period $t$) associated with receiving treatment $i$ is represented by the benefit functions $V^i(\theta) = \sum_{t=1}^{T} \rho^t E[Y^i_t(\theta)]$, where $\rho$ is the discount factor, $i = N, L, H$. To simplify the setup, I assume that the expected outcome is constant over time, i.e. $E[Y^i_t(\theta)] = E[Y^i(\theta)]$ for all $t$, which implies that the benefit functions can be re-expressed as $V^i(\theta) = \kappa_i E[Y^i(\theta)]$, where $\kappa_i$ is a constant. Without loss of generality it can be assumed that $\rho = 1$, in which case $\kappa_N = T$, $\kappa_L = T - \tau_L$, and $\kappa_H = T - \tau_H$ (and it is easy to see that $\kappa_N > \kappa_L > \kappa_H$).\(^3\)

Since the cost of training is borne by the county there is no direct cost of receiving training $L$ or training $H$ for the individuals, other than the opportunity cost of attending such training. The cost to the county associated with the provision of treatment $N$ (no training) is normalized to zero, i.e. $c_N \equiv 0$. Thus, the county faces additional direct costs per person $c_L$ and $c_H$ associated with the provision of treatments $L$ and $H$ respectively, where $c_H > c_L$, because HCD training is more expensive than LFA training. Finally, the county has a fixed budget $B$ which it must use to cover the expenses associated with providing the training services.

**The county’s problem** The county seeks to assign individuals to the different treatments such that it maximizes the expected returns to the investment. This problem implies that the county has to assign a treatment for each individual. To make some progress, I impose minimal conditions on the outcome generating functions $Y^N$, $Y^L$ and $Y^H$, that make the problem both analytically simpler and, more importantly, informative in terms of the implications associated with the changes in the economic environment faced by the county. Assuming that the outcome generating functions are concave and satisfy a standard “single crossing” property, which ensures that the benefit functions cross at most once, the decision problem of the county is simplified greatly. Formally, the following assumptions are made.

\(^2\)However, note than the differential treatment effect $\Delta^L(\theta) - \Delta^H(\theta)$ can be positive or negative.

\(^3\)Note that if $\rho < 1$, then $\kappa_N = \frac{1 - \rho^{T+1}}{1 - \rho}$, $\kappa_L = \frac{\rho T_L - \rho^{T+1}}{1 - \rho}$, $\kappa_H = \frac{\rho T_H - \rho^{T+1}}{1 - \rho}$ and it still holds that $\kappa_N > \kappa_L > \kappa_H$. 
**Assumption 1**

The outcome generating functions $Y^i(\theta)$ ($i = N, L, H$) are:

i) twice differentiable

ii) concave: $Y^i_\theta \leq 0$

iii) strictly increasing: $Y^i_\theta > 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$

iv) bounded from below and above: $Y^i(\theta) \geq 0$ and $Y^i(\bar{\theta}) < K$, where the constant $K < \infty$.

Assumption 1 implies that there are always positive but decreasing returns to skills for any treatment and that the problem will be well specified. Note that with $Y^N$ concave, it is sufficient that the treatment effects $\Delta^L$ and $\Delta^H$ are concave, to have outcome functions $Y^L$ and $Y^H$ that are concave.\(^4\)

**Assumption 2**

There is an ordering of the (present value of) expected marginal returns to skills for the three outcome functions $Y^N$, $Y^L$ and $Y^H$ that is unchanged for all $\theta \in [\underline{\theta}, \overline{\theta}]$. That is, for all triplets of outcome functions $(Y_i, Y_j, Y_k)$ ordered by their marginal returns, it is satisfied that

$$\kappa_i Y^i_\theta < \kappa_j Y^j_\theta < \kappa_k Y^k_\theta \quad \forall \theta, \ i,j,k = N, L, H, \ i \neq j \neq k.$$ 

Assumption 2 has a clear economic interpretation: if a particular treatment is better for individuals of higher skills, then the marginal returns to skills are also higher for these individuals. This is a standard single crossing property, and is crucial to be able to simplify the decision problem of the county.\(^5\)

**Assumption 3**

The cumulative distribution function of $\theta$ is differentiable and strictly increasing in $\theta$: $F_\theta(\theta) \equiv f(\theta) > 0$, for $\theta \in [\underline{\theta}, \overline{\theta}]$.

Assumption 3 guarantees that there are no regions of $\theta$ in which the density function is not defined, and it is necessary to assure that the first order conditions of the problem are well defined.

It is straightforward to prove that under Assumptions 1 and 3 a solution always exists.\(^6\) The solution implies that the interaction of the functions $V^N(\theta)$, $V^L(\theta) - c_L$ and $V^H(\theta) - c_H$ divide the support of $\theta$, into (at most) three regions, in which all the individuals belonging to a region receive

\(^4\)Although this is not necessary as long as $Y^N_\theta > \Delta^L_\theta$, $Y^N_\theta > \Delta^H_\theta$ and $|Y^N_\theta| > |\Delta^L_\theta|$, $|Y^N_\theta| > |\Delta^H_\theta|$.  

\(^5\)This condition is analog to what is known in the incentives literature as “Spence-Mirrlees” condition or “constant sign” condition (see for example Salanié, 1997, pp. 31).  

\(^6\)Assumptions 1 and 3 guarantee that the objective function $W$ is continuous and that the budget set is compact.
the same treatment.

I denote these three regions over the distribution of $\theta$ as $R_l$ (low), $R_m$ (medium) and $R_h$ (high), and use the region subscript to denote the treatment provided in that region (i.e. $l, m$ and $h$, where each of them can assume values $N, L$ or $H$). Thus, the county’s problem reduces to choosing two critical values for $\theta$, that define the regions’ limits. Designating $\theta_l$ as the critical value that separates regions $R_l$ and $R_m$, and $\theta_h$ as the critical value that separates regions $R_m$ and $R_h$, the county will choose the $\theta_l$ and $\theta_h$ that solve the following problem:

$$\max_{\{\theta_l, \theta_h\}} W = \int_{\theta_l}^{\theta_h} [V^l(\theta) - c_l]dF(\theta) + \int_{\theta_l}^{\theta_h} [V^m(\theta) - c_m]dF(\theta)$$

$$+ \int_{\theta_h}^{\theta_m} [V^h(\theta) - c_h]dF(\theta) - \int_{\theta_h}^{\theta_m} V^N(\theta)dF(\theta) \quad (P1)$$

$$s.t. F(\theta_l)c_l + [F(\theta_h) - F(\theta_l)] c_m + [1 - F(\theta_h)]c_h \leq B,$$

where $c_l$, $c_m$, or $c_h$ will assume the value 0 when valued at treatment $N$ (i.e. $c_N \equiv 0$).

The formulation of $(P1)$ implies that the county only needs to set a decision rule based on the two critical values; assignment to each treatment follows directly from this decision rule. This allows to study the effects of changes in the environment faced by the county welfare agency simply by analyzing the effects on the critical values for $\theta$.

**Solution** The first order necessary conditions for $(P1)$ (using a Lagrange multiplier $\lambda$ for the budget constraint) are the following:

$$\theta_l : [V^l(\theta_l) - c_l]f(\theta_l) - [V^m(\theta_l) - c_m]f(\theta_l) + \lambda f(\theta_l)(c_m - c_l) \leq 0$$

$$\theta_h : [V^m(\theta_h) - c_m]f(\theta_h) - [V^h(\theta_h) - c_h]f(\theta_h) + \lambda f(\theta_h)(c_h - c_m) \leq 0 \quad (A1)$$

$$\lambda : B + F(\theta_l)(c_m - c_l) + F(\theta_h)(c_h - c_m) - c_h \geq 0.$$

When the budget constraint is binding, it is easy to show that the county will choose the $\theta_l$ and $\theta_h$ that satisfy the following condition (where the superindexes stands for “constrained”):

$$\frac{V^m(\theta^l_\lambda) - V^l(\theta^l_\lambda)}{c_m - c_l} = \frac{V^h(\theta^h_\lambda) - V^m(\theta^h_\lambda)}{c_h - c_m}. \quad (A2)$$

That is, the treatment regions are chosen in a way that equalizes the ratio of marginal benefits to the ratio of marginal costs of the treatments across regions. The second order conditions are presented in the Proofs Section below. Depending on which type of individual benefits more from each type of treatment, there are six possible interior solutions. All these solutions imply that the support of the skills distribution is divided in three regions (low, medium, high skills), and a different treatment is offered to the individuals who fall in each region. Thus, I characterize those six cases by the treatment received by the individuals in the regions of low, medium and high skills, in this order. Denoting them by the notation $[R_l, R_m, R_h]$, where the subscripts refer to the
treatment, the cases are:

1. \([R_N, R_L, R_H] \implies \text{No Training, Training } L, \text{ Training } H\]
2. \([R_N, R_H, R_L] \implies \text{No Training, Training } H, \text{ Training } L\]
3. \([R_L, R_N, R_H] \implies \text{Training } L, \text{ No Training, Training } H\]
4. \([R_L, R_H, R_N] \implies \text{Training } L, \text{ Training } H, \text{ No Training}\]
5. \([R_H, R_N, R_L] \implies \text{Training } H, \text{ No Training, Training } L\]
6. \([R_H, R_L, R_N] \implies \text{Training } H, \text{ Training } L, \text{ No Training}\].

**Budget constraint**  Here there are two possibilities: If the county has enough funds, such that the budget constraint is not binding, then the welfare agency provides training to all the individuals for which the (net) returns to training are positive. The allocation between training \(L\) and training \(H\) is determined by the marginal returns to each pair of treatments compared with the marginal costs of treatment. Hence, inside each skill region the individuals receive the treatment with the greater (net) benefit compared to the other two. Appendix Figure 1 depicts this situation, where panels A to F display Cases 1 to 6 specified above.\(^7\)

If the budget constraint is binding, it is easy to show that the county will choose the \(\theta_l\) and \(\theta_h\) that equalize the ratio of marginal benefits to the ratio of marginal costs associated with the treatments across regions. Hence, the county resorts to substituting between treatments until that condition is satisfied. Note that the county no longer equates just private marginal benefits with private marginal costs, but now has to consider the “social” cost implicit in the fact that providing training to some individuals imply that training has to be denied to other individuals.

To describe exactly how the county attains the optimality condition consider Case 1 as an example. Faced with a binding budget constraint, the county can train fewer people (with respect to the unconstrained case); the issue is how to decide which individuals to train, and which training to offer them. There are two effects. First, because training \(H\) is the most expensive treatment, the county chooses to offer fewer people this treatment. As it is shown in Appendix Figure 1.A, the individuals “denied” treatment \(H\) (that is, individuals who in the unconstrained case would have received training \(H\)) will be offered treatment \(L\). But, as more and more individuals around \(\theta_h^n\) are offered treatment \(L\) instead of treatment \(H\), the lost marginal benefits of training increase (as measured by the vertical distance between the curves \(V^H(\theta) - c_H\) and \(V^L(\theta) - c_L\)). Second, if the county denies training to individuals that otherwise would have received treatment \(L\), starting with the individuals around \(\theta_l^n\), the lost marginal benefits are also small at the beginning and

\(^7\)Note that if the individuals do take into account the direct costs of training, \(c_L\) and \(c_H\), the solution is the same as would be obtained in the decentralized problem where the individuals are allowed to choose the treatment strategy themselves. For example, individuals could be offered “vouchers”, valid to use in any training of their choice, allowing them to keep any difference between the value of the voucher and the cost of the training. In this way, they would completely internalize the cost of training. As such, the model shares the essential features of the Roy Model of self-selection in the labor market (Roy, 1951), in which heterogeneous agents self-assign themselves to occupations according to a principle of comparative advantage (e.g., Willis, 1986; Heckman and Honore, 1990).
increase as more and more people are denied training. The new equilibrium critical values are represented in the Figure by $\theta^*_i$ and $\theta^*_h$ and the shaded areas roughly indicate the lost “social” benefit of substituting for one treatment or the other. Here, note that even if the county has equal concern for every individual (weights them equally), it will choose not to offer every individual her optimal treatment option, because it takes into account the trade-off generated by the resource constraint.

Testable implications of the returns-maximization model  The objective of the model is to study how changes in its parameters affect the proportions of people trained (total and in each type of training). These proportions are denoted by $P_N$, $P_L$, and $P_H$ for treatments $N$, $L$ and $H$ respectively. Note that $P_N = 1 - P_T$, where $P_T = P_L + P_H$ represents the total proportion of individuals receiving any type of training.

To study the changes in $P_N$, $P_L$, and $P_H$ it is clear that the key is to analyze the reaction of the optimal critical values $\theta^*_l$ and $\theta^*_h$ (where the superscripts will be dropped hereafter to simplify notation) to changes in the environment faced by the welfare agency.\(^8\) Of interest are the effects of changes in the budget $B$, the effects of changes in a parameter vector $\Gamma$ that affects the opportunity cost of training $Y^N(\theta; \Gamma)$, and the effect of changes in a parameter vector $\Psi$ that affects the distribution of skills $F(\theta; \Psi)$.

The discussion above should make clear the testable implications for changes in the budget constraint. If the budget increases, the total proportion of people not trained should decrease, and the total proportion of people receiving training $H$ (the more expensive treatment) should increase, because these individuals are the first affected by the binding budget constraint. The effects on the proportion of people receiving training $L$ are ambiguous, however, depending on whether or not training $L$ was rationed (Cases 3 and 5, positive), or was being used as substitute for training $H$ (Cases 2 and 5, negative) or a combination of both (Cases 1 and 6, ambiguous).

The following proposition indicates formally the effects of changes in available budget to the county, $B$, on $P_N$, $P_L$ and $P_H$.

**Proposition 1**  If the budget $B$ increases then:

a) the total proportion of people not trained, $P_N$, decreases;
b) the proportion of people receiving training $H$, $P_H$, increases;
c) the change in the proportion of people receiving training $L$, $P_L$, is ambiguous in Case 1 and Case 6, negative in Case 3 and Case 5, and positive in Case 2 and Case 4.

Regarding the effects of changes in the opportunity cost of training, one could concentrate just on $Y^N$, the potential outcome under no training. If this potential outcome increases, with the training effects held constant, then one would expect that the proportion of individuals trained should decrease ($P_N$ would increase). How $P_L$ and $P_H$ change is more complex to analyze. Given\(^8\) The two critical values are enough to determine the proportions of each treatment: using the general notation for Cases 1 through 6, the proportions are $P_i = F(\theta_i)$, $P_m = F(\theta_h) - F(\theta_i)$, and $P_h = 1 - F(\theta_h)$. 

6
that the budget is fixed and that $P_N$ increases, with everything else held equal, the county has a greater budget available for the people that are going to be trained. Given that training $H$ is always rationed, there should be a substitution of training $L$ for training $H$ (therefore making $P_L$ decrease and $P_H$ increase). However, because training $H$ is the training that takes longer to be completed, the opportunity cost of this training with respect to training $N$ increases more than for training $L$, making training $H$ less desirable. If this effect is big enough, then the substitution will be from training $H$ to training $L$, which in turn might even imply (if the savings are big enough, given that training $H$ is more expensive) that the increase in $P_L$ makes $P_N$ actually decrease.

Formally, one can parameterize $Y^N$ with a vector $\Gamma$ that measures changes in the outcome generating function under no treatment in two ways: by increasing $Y^N$ in a constant manner, and by increasing it more for more able individuals. Then, $Y^N(\theta; \Gamma) = \gamma_0 + \gamma_1 Y^N(\theta)$, where the vector $\Gamma$ is formed by two elements: $\gamma_0$ which implies a parallel translation of the original function $Y^N(\theta)$ and $\gamma_1$ which implies a proportional increase of the returns to skills. Note that a change in $Y^N$ affects also $Y^L$ and $Y^H$ (the outcome under training $L$ and $H$), but just through the effect of $Y^N$. This rules out the possibility that changes in the economic situation of the county also affect the treatment effects $\Delta^L$ or $\Delta^H$.

The effects of a change in $\gamma_0$, a parallel translation in $Y^N$, are specified in the following proposition, and depend on a condition that is shown in the Proofs Section below.\footnote{The effects of a change in $\gamma_1$, a proportional translation in $Y^N$, are the same as the effects of a change in $\gamma_0$. The proposition proving this, available upon request, is not presented in the interest of space.}

**Proposition 2**

a) In general the effects of a parallel translation in the outcome generating function $Y^N$ are ambiguous;
b) If the ratio of the opportunity cost to its direct cost of training $H$ is “small enough” relative to the same ratio for training $L$, then an increase in $\gamma_0$ will reduce $P_L$ and will increase $P_N$ and $P_H$.

If the distribution of skills associated with a given cohort of individuals to be treated changes, it would be useful to understand how the county would change its optimal decisions. This happens to be a particularly difficult issue to explore. Below in the Proofs Section I show that for two distributional assumptions (Uniform and Normal), the effects of changes in the distribution of skills are ambiguous. The intuition is that depending on the (relative) cost of the treatment towards which the distribution moves, the county might be able to increase the proportion of people receiving that particular treatment, for the cheaper treatments, but not for the expensive treatments. Therefore, changes that would be easy to analyze if the budget constraint was not binding, become extremely complicated with a binding budget constraint. See the Proofs Section for details.

## 2 The role of the local political environment

The simple returns-maximization model assumes that only factors associated with returns to training enter the county’s decision-making process. However, there are reasons why this might not be
the case. As welfare agencies are public bureaucracies there are many ways in which their decisions could reflect other objectives that arise due to the interaction of the objectives and preferences of the involved parties.

The local political environment could be reflected in different ways in the objective function of the county. For example, one could assume that the county has preferences for a particular type of training. Alternatively, one could assume that the preferences are over particular types of individuals or groups (i.e. based on skills, race/ethnicity, etc.).

In what follows I show how two such alternative objective functions could affect the assignment to training decisions, namely the case where the county has a preference for a particular type of training and an alternative case, where the county has a preference for individuals in a particular position in the skills distribution.\(^\text{10}\)

In the particular case of LFA and HCD training, it seems reasonable to assume that one of the two types of training could be preferred. The two types of training represent two distinctive approaches with differences that can be deemed philosophical. For example, Hamilton et al. (2001) judge that there has been disagreement over which type of training should be used as the best way of fostering the goal of self-sufficiency. They characterize the differences between LFA and HCD as: “[LFA training] emphasizes quick employment, reflecting the belief that individuals can best build their employability and improve their skills, eventually achieving self-sufficiency, through actual work, even if their initial jobs are minimum wage and without fringe benefits”, while “[HCD training] emphasizes initial investments in short-term education and, in some cases, training, reflecting the view that these investments will eventually enable individuals to obtain higher-wages, long-lasting jobs with health insurance coverage” (Hamilton et al., 2001, pp. 4). Ideological attitudes towards the role of work and education can also explain the preferences for one type of training or another. In particular, the evidence shows that in the specific case of training policies, and attitude towards welfare reform in general, more liberal individuals favor an education-approach, and more conservative individuals favor a work-first approach (e.g. Nathan and Gais, 1998; Weaver, 2002). In addition, these preferences could reflect heterogeneity in the discount rates associated with the benefits of training by policy makers; it has been shown that LFA training benefits are more immediate, while HCD training benefits appear more in the medium- and long-run (Hotz et al. 2006; Dyke et al., 2006).

Introducing preferences for a particular type of training in the model, considers the case of a county that cares about expected returns from the assignment to training policies, but also has a preference for one type of training over the other. To make the notation simpler, suppose that the shape of the outcome generating functions are such that lower skills individuals do not benefit from training, middle skills individuals benefit from training $L$, and upper skills individuals benefit from training $H$ (Case 1 in the previous Section). Then, denoting by $\alpha$ the weight (preference) given to training $L$ by the county, the problem to solve is as follows:

\(^{10}\)This would arise in a specification where the county’s objective function presents inequality aversion; this type of objective function is proposed by Dehejia (2005). Heckman et al. (2002) also allow for the possibility in their model that there are preferences for a particular group, that is not explained by returns to training.
In the Proofs Section I show that under “moderate” values of $\alpha$ (i.e. such that both types of training are still offered), the empirical implications regarding changes in the budget and opportunity costs of training of the returns-maximization model still hold. However, the optimal treatment proportions change and are affected by $\alpha$. The effects of $\alpha$ are clear: higher preference for training $L$ increases $P_L$ and decreases $P_H$ and $P_N$ (see Proposition 3 in the Proofs Section). If the preferences for training $L$ are extreme ($\alpha$ is “large”), then the testable implications of the returns-maximization model regarding changes in budget and local economics conditions will not hold.

An alternative objective function is one where the program administrator exhibits inequality aversion. It can be expressed using a Constant Elasticity of Substitution (CES) welfare function. Here the decision maker cares about the after-treatment outcome distribution and solves the following problem:

$$\max_{\{\theta_t, \theta_h\}} W = \alpha \int_{\theta_t}^{\theta_h} [V^L(\theta) - V^N(\theta) - c_m] dF(\theta)$$

$$+ (1 - \alpha) \int_{\theta_h}^{\theta} [V^H(\theta) - V^N(\theta) - c_h] dF(\theta)$$

$$s.t. [F(\theta_h) - F(\theta_t)] c_L + [1 - F(\theta_h)] c_H \leq B.$$  \hspace{1cm} (P2)

where $\varepsilon$ is the inequality aversion parameter. When $\varepsilon = 1$ there is no inequality aversion, if $\varepsilon = 0$ the utility function becomes logarithmic and implies unitary inequality aversion, and if $\varepsilon \to -\infty$ then the there infinite aversion to inequality and the welfare function becomes Rawlsian where only the welfare of the individual worst off matters. Solving for the first order conditions of this problem (see Proofs Section) it is straightforward to show that the first order condition is analogous to the first order condition of $(P1)$, equation (A2). In fact $(P1)$ is a special case of $(P3)$ when $\varepsilon = 1$, i.e. there is no aversion to inequality. It can be shown that for some low levels of inequality aversion the results of $(P1)$ will hold, but for higher levels of inequality aversion the results can actually be very different. If the first case occurs, though, it would not be possible to differentiate between preferences towards the lower earnings group introduced by the inequality aversion parameter, from the preferences for a type of training (as in $(P2)$). If the inequality aversion is high, then depending which training is more beneficial for the lower skills individuals, the results from $(P1)$ could be even stronger, or could become completely the opposite. In particular, if training $H$ favors lower skills individuals, the results from $(P3)$ will be similar with respect to changes in budget and local economic conditions to the results from $(P1)$. Empirically, this type of situation (Case 6) is the one suggested by the results in the paper.
Proofs

Second order conditions for \((P1)\)  The second order conditions that need to be satisfied around a maximum for \(P1\), denoting by \(\mathcal{L}\) the Lagrangian, are:

\[
\text{det}(M) \equiv \text{det} \begin{bmatrix}
0 & \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta_h} \\
\frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \theta} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \theta_h} \\
\frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \theta} & \frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \theta_h}
\end{bmatrix} > 0 \quad (A3)
\]

where (valued at the critical values \(\theta_i^*\) and \(\theta_h^*\) that solve \((P1)\)):

\[
\frac{\partial^2 \mathcal{L}}{\partial \theta \partial \lambda} = V_i^l(\theta_i^*) - V_{\theta}^m(\theta_i^*);
\]
\[
\frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \theta} = V_{\theta}^m(\theta_h^*) - V_{\theta}^h(\theta_h^*);
\]
\[
\frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \theta} = 0;
\]
\[
\frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \theta} = f(\theta_i^*)(c_m - c_l);
\]
\[
\frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \lambda} = f(\theta_i^*)(c_h - c_m).
\]

Then, \(\text{det}(M) > 0 \iff -\left(\frac{\partial^2 \mathcal{L}}{\partial \theta \partial \lambda}\right)^2 + \left(\frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \theta} - \frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \lambda}\right) > 0\). Therefore, the sufficient conditions for the critical values that solve \((P1)\) are,

\[
f(\theta_i^*)^2(c_m - c_l)^2[V_{\theta}^l(\theta_h^*) - V_{\theta}^m(\theta_h^*)] > -f(\theta_i^*)^2(c_h - c_m)^2[V_{\theta}^m(\theta_i^*) - V_{\theta}^l(\theta_i^*)], \quad (A4)
\]

and will always be satisfied because under Assumption 2 \(V_{\theta}^h(\theta_h^*) > V_{\theta}^m(\theta_h^*)\) and \(V_{\theta}^m(\theta_i^*) > V_{\theta}^l(\theta_i^*)\).

Necessary conditions for interior solutions to \((P1)\)  The following conditions are necessary for an “interior solution” of \((P1)\) in which all three treatments \(l, m\) and \(h\) (in ascending order of the distribution of \(\theta\)) are offered:

i) \(V^l(\theta) - c_l > V^m(\theta) - c_m > V^h(\theta) - c_h\);

ii) \(V^l(\theta) - c_l < V^m(\theta) - c_m < V^h(\theta) - c_h\);

iii) \(V^h(\theta_i) - c_h < V^l(\theta_i^*) - c_l\).

iv) \(V^h(\theta_i) - c_h < V^m(\theta_i^*) - c_m\).

Changes in budget (Proposition 1)  Proof. Equations (A2) and the third line of (A1) form a system of implicit equations

\[
\Pi_1(\theta_i, \theta_h, B, \Gamma, \Psi) = \frac{V_{\theta}^h(\theta_h; \Gamma) - V_{\theta}^m(\theta_h; \Gamma)}{c_h - c_m} - \frac{V_{\theta}^m(\theta_i; \Gamma) - V_{\theta}^l(\theta_i; \Gamma)}{c_m - c_l} = 0 \quad (A5)
\]
\[
\Pi_2(\theta_i, \theta_h, B, \Gamma, \Psi) = B + F(\theta_i; \Psi)(c_m - c_l) + F(\theta_h; \Psi)(c_h - c_m) - c_h = 0 \quad (A6)
\]
Differentiating (A5) and (A6) with respect to $\theta_l, \theta_h$ and $B$, and reordering terms, the following system needs to be solved to obtain the effect of changes of $B$ on $\theta_l$ and $\theta_h$:

\[
\begin{bmatrix}
d\theta_l \\
d\theta_h
\end{bmatrix} = A^{-1} \begin{bmatrix} 0 \\
dB \end{bmatrix},
\]

where

\[
A = \begin{bmatrix}
- V^m_\theta(\theta_l) - V^l_\theta(\theta_l) & V^h_\theta(\theta_h) - V^m_\theta(\theta_h) \\
-f(\theta_l)(c_m - c_l) & -f(\theta_h)(c_h - c_m)
\end{bmatrix}.
\]

Note that

\[
A^{-1} = \frac{1}{\det(A)} \begin{bmatrix}
-f(\theta_h)(c_h - c_m) & \frac{V^h_\theta(\theta_h) - V^m_\theta(\theta_h)}{c_h - c_m} \\
-f(\theta_l)(c_m - c_l) & \frac{V^l_\theta(\theta_l) - V^m_\theta(\theta_l)}{c_m - c_l}
\end{bmatrix}
\]

where

\[
\det(A) = \frac{V^m_\theta(\theta_l) - V^l_\theta(\theta_l)}{c_m - c_l} f(\theta_l)(c_h - c_m) + \frac{V^h_\theta(\theta_h) - V^m_\theta(\theta_h)}{c_h - c_m} f(\theta_l)(c_m - c_l).
\]

The sign of this determinant is key to evaluate the effects of changes in $B$, and it will depend on the particular case under which the county is operating. Because of Assumption 2 it always holds that $V^m_\theta(\theta_l) - V^l_\theta(\theta_l) > 0$ and that $V^h_\theta(\theta_h) - V^m_\theta(\theta_h) > 0$. Therefore, the sign of the determinant will depend on the differences of costs in each case. Given that $c_H > c_L > c_N$, it is easy to see that under Case 1 $c_h > c_m > c_l$; under Case 2 and Case 4 $c_m > c_l$ and $c_h < c_m$; under Case 3 and Case 5 $c_m < c_l$ and $c_h > c_m$; and under Case 6 $c_h < c_m < c_l$. In this way $\det(A)$ will be positive in Cases 1 and 6, and negative in Cases 2, 3, 4, and 5.

Hence, using (A7), (A9) and (A10), the effect of changes in $B$ on the critical values $\theta_l$ and $\theta_h$ will be:

\[
\frac{\partial \theta_l}{\partial B} = -\frac{1}{\det(A)} \frac{V^h_\theta(\theta_h) - V^m_\theta(\theta_h)}{c_h - c_m},
\]
\[
\frac{\partial \theta_h}{\partial B} = -\frac{1}{\det(A)} \frac{V^m_\theta(\theta_l) - V^l_\theta(\theta_l)}{c_m - c_l}
\]

where, given the sign of the cost differences and of $\det(A)$ in each case, it is clear that $\frac{\partial \theta_l}{\partial B}$ will be negative in Cases 1, 2 and 4, and it will be positive in Cases 3, 5 and 6. Likewise $\frac{\partial \theta_h}{\partial B}$ will be negative in Cases 1, 3 and 5, and it will be positive in Cases 2, 4 and 6.

Finally, analyzing for each case $P_l$, $P_m$, and $P_h$, the proportion of individuals receiving treatment $l, m,$ and $h$ respectively, it is easy to see that:

\[
\frac{\partial P_l}{\partial B} = f(\theta_l) \frac{\partial \theta_l}{\partial B},
\]
\[
\frac{\partial P_m}{\partial B} = f(\theta_l) \frac{\partial \theta_l}{\partial B} - f(\theta_l) \frac{\partial \theta_l}{\partial B},
\]
\[
\frac{\partial P_h}{\partial B} = -f(\theta_h) \frac{\partial \theta_h}{\partial B}.
\]
Using (A11) and (A12) in (A13), (A14) and (A15), it can be seen that the proportion of people receiving treatment \( N \) will always decrease (i.e. total training will increase) \((\frac{\partial P_N}{\partial B} < 0 \iff \frac{\partial (P_L + P_H)}{\partial B} > 0)\), and that the proportion of people receiving training \( H \) will always increase \((\frac{\partial P_H}{\partial B} > 0)\). The change in the proportion of people receiving training \( L \) will be negative in Cases 2 and 4 (where training \( H \) is offered to the individuals in the middle of the distribution of \( \theta \)) and positive in Cases 3 and 5 (where treatment \( N \) is offered to the individuals in the middle of the distribution of \( \theta \)). However, it will ambiguous in Cases 1 and 6 (where treatment \( L \) is offered to the individuals in the middle of the distribution of \( \theta \)), because in the equation \( \frac{\partial P_L}{\partial B} = f(\theta_h) \frac{\partial \theta_h}{\partial B} - f(\theta_l) \frac{\partial \theta_l}{\partial B} \) (even without taking into account \( f(\theta_l) \) and \( f(\theta_h) \)), it is ambiguous whether \( \frac{\partial \theta_l}{\partial B} > \frac{\partial \theta_h}{\partial B} \) or \( \frac{\partial \theta_l}{\partial B} < \frac{\partial \theta_h}{\partial B} \). \( \square \)

**Effects of changes in the opportunity cost of training (Proposition 2)**

Proof. Differentiating (A5) and (A6) with respect to \( \theta_l, \theta_h \) and \( \gamma_0 \), and reordering terms, the following system

\[
\begin{bmatrix}
\frac{d\theta_l}{d\theta_h}
\end{bmatrix} = A^{-1} \begin{bmatrix}
\frac{(\kappa_m - \kappa_l)}{c_m - c_l} - \frac{(\kappa_h - \kappa_m)}{c_h - c_m}
\end{bmatrix} \frac{d\gamma_0}{D},
\]

(A16)

where \( A \) is defined as in (A8), needs to be solved to obtain the effects of changes of \( \gamma_0 \) on \( \theta_l \) and \( \theta_h \). Note that \( \kappa_m - \kappa_l \equiv -(\tau_m - \tau_l) \), \( \kappa_h - \kappa_m \equiv -(\tau_h - \tau_m) \), for \( l, m, h = N, L, H \), and \( \tau_N \equiv 0 \).

Then, replacing these terms and using (A9), (A10) and (A16), the effects can be expressed as

\[
\frac{\partial \theta_l}{\partial \gamma_0} = - \frac{1}{\det(A)} f(\theta_h)(c_h - c_m) \begin{bmatrix}
\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l}
\end{bmatrix},
\]

(A17)

\[
\frac{\partial \theta_h}{\partial \gamma_0} = \frac{1}{\det(A)} f(\theta_l)(c_m - c_l) \begin{bmatrix}
\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l}
\end{bmatrix},
\]

(A18)

where, by the analysis made in the proof of Proposition 1, \( \det(A) \) will be positive in Cases 1 and 6 and negative in Cases 2 through 5. It is clear that the sign of the derivatives will depend crucially on the sign of the term between square brackets. Each ratio in the bracketed term represents the ratio of relative opportunity costs of two treatments (the higher \( \tau \), the longer the number of periods in which the individual can not participate in the labor market) versus the relative direct costs of these two treatments (as \( c_H > c_L > c_N \) and \( \tau_H > \tau_L > \tau_N \), the two ratios will be always positive).

To analyze the changes in the proportions, use expressions analogous to (A13), (A14) and (A15), and introduce (A17) and (A18) to get (rearranging terms):

\[
\frac{\partial P_l}{\partial \gamma_0} = - \frac{1}{\det(A)} f(\theta_l)f(\theta_h)(c_h - c_m) \begin{bmatrix}
\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l}
\end{bmatrix},
\]

(A19)

\[
\frac{\partial P_m}{\partial \gamma_0} = \frac{1}{\det(A)} f(\theta_l)f(\theta_h)(c_m - c_l) \begin{bmatrix}
\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l}
\end{bmatrix},
\]

(A20)

\[
\frac{\partial P_h}{\partial \gamma_0} = - \frac{1}{\det(A)} f(\theta_l)f(\theta_h)(c_m - c_l) \begin{bmatrix}
\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l}
\end{bmatrix}.
\]

(A21)

\[\text{In the case in which the discount factor is not one, the terms will be a little bit different, but all the analysis holds the same. Specifically, with } \rho < 1, \kappa_m - \kappa_l \equiv -\frac{e^{\rho m - \rho^2 l}}{1 - p}, \kappa_h - \kappa_m \equiv -\frac{e^{\rho h - \rho^2 l}}{1 - p}.\]
From (A19), (A20) and (A21), it is clear that each proportion change according to a weight given by the difference in direct costs of the other two treatments, times the difference in ratios of opportunity costs. The sign of these differences will be different in each case, but it is straightforward to see that if
\[
\frac{\tau H}{c_H} < \frac{\tau L}{c_L},
\]
the bracketed expression will be negative in Cases 1, 2 and 3, and it will be positive in Cases 4, 5 and 6. This implies that in every case \( \frac{\partial P_N}{\partial \theta_0} > 0, \frac{\partial P_L}{\partial \theta_0} < 0 \) and \( \frac{\partial P_H}{\partial \theta_0} > 0. \)

Effects of change in skills distribution

Let’s call \( \Psi \) the vector of parameters that define the distribution of skills, and represent this distribution as \( F(\theta; \Psi) \). Using the definitions of the proportions of individuals receiving treatment \( l, m \) and \( h \), it is easy to see that the effect on these proportion of a change in \( \Psi \) will be composed by two parts: the change in the optimal \( \theta_l \) and \( \theta_h \) because of the distributional change, and the change in the mass of the distribution for given \( \theta_l \) and \( \theta_h \):

\[
\frac{\partial P_l}{\partial \Psi} = f(\theta_l) \frac{\partial \theta_l}{\partial \Psi} + F_\Psi(\theta_l), \tag{A22}
\]

\[
\frac{\partial P_m}{\partial \Psi} = \left[ f(\theta_h) \frac{\partial \theta_h}{\partial \Psi} + F_\Psi(\theta_h) \right] - \left[ f(\theta_l) \frac{\partial \theta_l}{\partial \Psi} + F_\Psi(\theta_l) \right], \tag{A23}
\]

\[
\frac{\partial P_h}{\partial \Psi} = -\left[ f(\theta_h) \frac{\partial \theta_h}{\partial \Psi} + F_\Psi(\theta_h) \right]. \tag{A24}
\]

The effects of a change in \( \Psi \) can be analyzed in the same way as with changes in \( B \) or \( \gamma_0 \). In particular, differentiating (A5) and (A6) with respect to \( \theta_l, \theta_h \) and \( \Psi \), and reordering terms, the following system

\[
\begin{bmatrix}
\frac{d\theta_l}{d\theta_h} \\
\frac{d\theta_h}{d\theta_l}
\end{bmatrix} = A^{-1} \begin{bmatrix}
0 \\
[(c_m - c_l)F_\Psi(\theta_l) + (c_h - c_m)F_\Psi(\theta_h)]d\Psi
\end{bmatrix}, \tag{A25}
\]

where \( A \) is defined as in (A8), needs to be solved to obtain the effects of changes of \( \Psi \) on \( \theta_l \) and \( \theta_h \). Then, using (A9), (A10), (A11), (A12) and (A25), the effect of changes in \( \Psi \) on the critical values \( \theta_l \) and \( \theta_h \) can be expressed as:

\[
\frac{\partial \theta_l}{\partial \Psi} = \frac{\partial \theta_l}{\partial B} [(c_m - c_l)F_\Psi(\theta_l) + (c_h - c_m)F_\Psi(\theta_h)] \tag{A26}
\]

\[
\frac{\partial \theta_h}{\partial \Psi} = \frac{\partial \theta_h}{\partial B} [(c_m - c_l)F_\Psi(\theta_l) + (c_h - c_m)F_\Psi(\theta_h)]. \tag{A27}
\]

The term in the square bracket shows that the interaction of the relative costs of the treatments and the changes in the mass of the distribution of \( \theta \), will determine the effect on \( \theta_l \) and \( \theta_h \). However, (A26) and (A27) are not enough to characterize the effects of changes in \( \Psi \) on the proportions,\footnote{With a discount factor \( \rho < 1 \), the analogous condition (which has the same economic implications) will be \( \frac{1 - \rho^{\gamma_2}}{c_2} < \frac{1 - \rho^{\gamma_1}}{c_1} \).}
because as it is shown in (A22), (A23) and (A24), there will be an extra term, \( F_\Psi \) that will make the final effect in most of the cases ambiguous (many of the signs would be ambiguous even without this extra term).

I make two different distributional assumptions. The first one assumes that \( F(\theta) \) is distributed Uniform \((\underline{\theta}, \overline{\theta})\), and defines \( \Psi \equiv \overline{\theta} \). This implies studying the effect of the entrance (or exit) of more able individuals (given the properties of the Uniform distribution, an increase in \( \overline{\theta} \) imply an increase in both \( E(\theta) \) and \( V(\theta) \)). The results are ambiguous in most cases: \( F_\Psi \) is always negative, but \( \frac{\partial F_\Psi}{\partial \overline{\theta}} \) and \( \frac{\partial F_\Psi}{\partial \underline{\theta}} \) are some times positive and some times negative (depending on the case). This reflects the fact that depending on the (relative) cost of the treatment towards which the distribution moves, the county would be able to increase the proportion of people receiving that particular treatment. The results below show that in the cases (except Case 1, ambiguous) in which training \( L \) is offered to individuals of relative less skills than the ones offered training \( H \), \( P_L \) decreases unambiguously if \( \overline{\theta} \) increases. Intuitively, given that the budget remains constant, one should think that this would imply a less than proportional increase in \( P_H \) and an increase in \( P_N \). In the same way, intuition suggests that the opposite effects should occur for cases in which training \( L \) is offered to (relatively) more able individuals than individuals offered training \( H \).

The second assumption is that \( F(\theta) \) is distributed Normal \((\mu, \sigma^2)\), and \( \Psi \equiv [\mu, \sigma^2] \). Below I show that analyzing the effects of changes in \( \mu \) and in \( \sigma^2 \) gives also ambiguous results. However, because an increase in \( \mu \) with \( \sigma^2 \) fixed implies that the new distribution stochastically dominates the old one, the intuition would be equivalent to changing \( \overline{\theta} \) in the Uniform case.

- **Uniform case** Assume \( \theta \) is distributed Uniform \((\underline{\theta}, \overline{\theta})\), and define \( \Psi \equiv \overline{\theta} \), then \( F(\theta) = \frac{\theta - \underline{\theta}}{\overline{\theta} - \underline{\theta}} \), \( f(\theta) = \frac{1}{\overline{\theta} - \underline{\theta}} \), which implies that \( F_\Psi(\theta) = -\frac{\theta - \underline{\theta}}{(\overline{\theta} - \underline{\theta})^2} = -f(\theta)F(\theta) < 0 \) and that \( \frac{\partial F_\Psi(\theta)}{\partial \overline{\theta}} = -f(\theta)^2 < 0 \) (that is, an increase in \( \overline{\theta} \) implies that the new distribution stochastically dominates the old one, and that the distance between the two cumulative distribution functions increases with \( \overline{\theta} \)). The bracketed term in (A23) and (A24) will be negative in Cases 1 and 3, positive in Cases 4 and 6, and will have an ambiguous sign in Cases 2 and 5, which implies that in Cases 1 and 6 \( \frac{\partial \Psi}{\partial \overline{\theta}} > 0 \) and \( \frac{\partial \Psi}{\partial \underline{\theta}} > 0 \), in Cases 3 and 4 \( \frac{\partial \Psi}{\partial \overline{\theta}} < 0 \) and \( \frac{\partial \Psi}{\partial \underline{\theta}} > 0 \) and in Cases 2 and 5 the derivatives will have an ambiguous sign. Using (A22), (A23) and (A24), the only unambiguous signs will be \( \frac{\partial F_\Psi}{\partial \overline{\theta}} < 0 \) in Cases 3 and 4.

- **Normal case** Assume \( \theta \) is distributed Normal \((\mu, \sigma^2)\), and define \( \Psi \equiv [\mu, \sigma^2] \). Then for a particular value of \( \theta \), say \( \overline{\theta} \), \( F_\mu(\overline{\theta}) = \int_{\overline{\theta}}^\infty f(\theta)(\frac{\mu - \theta}{\sigma^2})d\theta = \int_{\frac{\mu - \overline{\theta}}{\sigma}}^\infty \phi(z)dz < 0 \), where \( \phi \) is the pdf of a standard Normal, and \( z = \frac{\theta - \mu}{\sigma} \). Note that although \( F_\mu(\overline{\theta}) \) is negative, it will attain a minimum at \( \overline{\theta} = \mu \) and then it will increase asymptotically towards zero with \( \theta \). Therefore, when evaluating \( F_\mu(\theta_1) \) and \( F_\mu(\theta_2) \) in (A22) and (A23), it is not possible to determine which value is higher. However, still an increase in \( \mu \) implies that the new distribution stochastically dominates the old one, and intuitively results are similar to changes in \( \overline{\theta} \) in the Uniform case.
Changes in $\sigma^2$ are even more difficult to evaluate because $F_{\sigma^2} (\bar{\theta}) = \frac{1}{2\pi\sigma} \int_{-\infty}^{\bar{\theta}} f(\theta)(((\frac{\theta - \bar{\theta}}{\sigma})^2 - 1) d\theta = \frac{1}{2\pi\sigma^2} \Phi(\frac{\bar{\theta} - \mu}{\sigma})z^2dz - F(\bar{\theta}) \leq 0$ (it is $> 0$ when $\bar{\theta} < \mu$, $= 0$ when $\bar{\theta} = \mu$, and $< 0$ when $\bar{\theta} > \mu$). Hence, it is not possible to evaluate $F_{\sigma^2}(\theta_l)$ versus $F_{\sigma^2}(\theta_h)$ in (A22) and (A23), given that the position of $\theta_l$ and $\theta_h$ with respect to $\mu$ is unknown.

**Preference for one type of training** The solution to $(P2)$ is similar to the solution to $(P1)$. The first order conditions from $(P2)$ are:

\[
\begin{align*}
\theta_l : &\ - \alpha[V^L(\theta_l) - V^N(\theta_l) - c_L]f(\theta_l) + \lambda f(\theta_l)c_L \leq 0 \quad \text{(A28)} \\
\theta_h : &\ \alpha[V^L(\theta_h) - c_L] - (1 - \alpha)[V^H(\theta_h) - c_H] + (1 - 2\alpha)V^N(\theta_h)f(\theta_h) \\
&\ \quad + \lambda f(\theta_h)(c_H - c_L) \leq 0 \quad \text{(A29)} \\
\lambda : &\ B - [F(\theta_h) - F(\theta_l)]c_L - [1 - F(\theta_h)]c_H \geq 0. \quad \text{(A30)}
\end{align*}
\]

Combining (A28) and (A29), and defining $a \equiv \frac{1-\alpha}{\alpha}$ as the preference factor for training $H$, the agency chooses $\theta_l$ and $\theta_h$ such that the following condition is satisfied:

\[
\frac{V^L(\theta_l) - V^N(\theta_l)}{c_L} = aV^H(\theta_h) - V^L(\theta_h) - (a - 1)[V^N(\theta_h) - c_H] \frac{c_H - c_L}{c_H - c_L} \quad \text{(A31)}
\]

That is, with $(P1)$ the ratio of marginal benefits to marginal costs of each training are equated, but with a weight factor to benefits given by $a$. It is easy to see that when $\alpha = 1/2$ (i.e. there is equal weight given to each type of training), then $a = 1$, and (A31) collapses to the optimality condition $(A2)$.

If $\alpha$ is such that an interior solution occurs in which (A31) is satisfied, then the empirical implications regarding budget and opportunity cost of training under $(P1)$ still hold, and empirical implications can be derived for changes in $\alpha$. For that is necessary that the preference for training $L$ $\alpha$ be in a certain range, as specified in the following proposition:

**Proposition 3** If the preference for training $L$ is such that $\alpha < \frac{V^H(\theta_h) - V^N(\theta_h)}{V^H(\theta_h) + V^L(\theta_h) - 2V^N(\theta_h)}$ then:

a) the results in Proposition 1 hold;

b) if also $\alpha > \frac{\kappa_N - \kappa_H}{\kappa_H - c_L} \frac{c_L}{\kappa_N - \kappa_H + c_L}$ then the results of Proposition 2 hold;

c) an increase in $\alpha$ increases $P_L$ and decreases $P_N$ and $P_H$.

**Proof.** The condition $\alpha < \frac{V^H(\theta_h) - V^N(\theta_h)}{V^H(\theta_h) + V^L(\theta_h) - 2V^N(\theta_h)}$ is a sufficient condition for the second order conditions of $(P2)$ to be satisfied (because it assures that $[aV^H(\theta_h) - V^L(\theta_h) - (a - 1)V^N(\theta_h)] > 0$, which is a condition analogous to the one imposed in Assumption 2). Then, (A30) and (A31) can
be used to form the following system of equations:

\[
\Pi_1(\theta_l, \theta_h, B, \gamma_0) = \frac{V^L(\theta_l) - V^N(\theta_l)}{c_L} - \frac{aV^H(\theta_h) - V^L(\theta_h) - (a-1)[V^N(\theta_h) - c_H]}{c_H - c_L} = 0 \quad (A32)
\]

\[
\Pi_2(\theta_l, \theta_h, B, \gamma_0) = B - [F(\theta_h) - F(\theta_l)]c_L - [1 - F(\theta_h)]c_H = 0. \quad (A33)
\]

Part a) can be proved by differentiating (A32) and (A33) with respect to \(\theta_l, \theta_h\) and \(B\), and reordering terms to form the following system

\[
\begin{bmatrix}
\frac{d\theta_l}{d\theta_h}
\end{bmatrix} = A^{-1} \begin{bmatrix}
0 \\
\frac{dB}{dB}
\end{bmatrix}, \quad (A34)
\]

where

\[
A^{-1} = \frac{1}{\det(A)} \begin{bmatrix}
-f(\theta_h)(c_H - c_L) & -[aV^H(\theta_h) - V^L(\theta_h) - (a-1)V^N(\theta_h)] \\
f(\theta_l)c_L & \frac{V^L(\theta_l) - V^N(\theta_l)}{c_L}
\end{bmatrix} \quad (A35)
\]

and

\[
\det(A) = \frac{V^L_\theta(\theta_h) - V^N_\theta(\theta_l)}{c_L}f(\theta_h)(c_H - c_L) + [aV^H_\theta(\theta_h) - V^L_\theta(\theta_h) - (a-1)V^N_\theta(\theta_h)]f(\theta_l)c_L. \quad (A36)
\]

The sign of the determinant in (A36) is positive, and it is straightforward to see that the same results with respect to changes in \(B\) as under (P1) hold (see (A11) through (A15)).

Part b) can be proved by differentiating (A32) and (A33) with respect to \(\theta_l, \theta_h\) and \(\gamma_0\), and using (A35) and (A36) in the system

\[
\begin{bmatrix}
\frac{d\theta_l}{d\theta_h}
\end{bmatrix} = A^{-1} \begin{bmatrix}
\frac{\kappa_L - \kappa_N}{c_L} - \frac{\kappa_H - \kappa_L - (a-1)\kappa_N}{c_H - c_L} \\
0
\end{bmatrix} \gamma_0 \quad (A37)
\]

where under the condition imposed in part b) it can be shown that \(\frac{\kappa_L - \kappa_N}{c_L} < \frac{\kappa_H - \kappa_L - (a-1)\kappa_N}{c_H - c_L}\) and all the results in (A17) through (A21) hold.

Finally, part c) of the proposition can be proved by differentiating (A32) and (A33) with respect to \(\theta_l, \theta_h\) and \(a\), and forming the system

\[
\begin{bmatrix}
\frac{d\theta_l}{d\theta_h}
\end{bmatrix} = A^{-1} \begin{bmatrix}
\frac{V^H(\theta_h) - V^N(\theta_h) - c_H}{c_H - c_L} \\
0
\end{bmatrix} da \quad (A38)
\]
where (A35) and (A36) still hold. Solving (A38), it is clear that

$$\frac{\partial \theta_l}{\partial a} = \frac{1}{\det(A)} f(\theta_h)(c_H - c_L) \frac{V^H(\theta_h) - V^N(\theta_h) - c_H}{c_H - c_L}$$

(A39)

$$\frac{\partial \theta_h}{\partial a} = -\frac{1}{\det(A)} f(\theta_l)c_L \frac{V^H(\theta_h) - V^N(\theta_h) - c_H}{c_H - c_L}$$

(A40)

will be positive and negative respectively. This clearly implies that as $a$ increases $P_L$ decreases and $P_N$ and $P_H$ increase (see for example (A13)-(A15)). Given that $a = \frac{1-\alpha}{\alpha}$, then $\alpha$ and $a$ will go in opposite directions, and an increase in $\alpha$ will increase $P_L$ and decrease $P_N$ and $P_H$. ■

Objective function with inequality aversion Solving for the first order conditions of this problem it is straightforward to show that the optimality condition will be

$$\frac{[V^h(\theta_h)]^\varepsilon - [V^m(\theta_h)]^\varepsilon}{[V^m(\theta_l)]^\varepsilon - [V^l(\theta_l)]^\varepsilon} = \frac{c_h - c_m}{c_m - c_l}$$

(A41)

which is analogous to the first order condition of (P1), equation (A2). In fact (P1) is a special case of (P3) when $\varepsilon = 1$, i.e. there is no aversion to inequality. ■

References


Appendix Figure 1. Interior Solution Cases

A. Case 1
B. Case 2
C. Case 3
D. Case 4
E. Case 5
F. Case 6

No Training  LFA Training  HCD Training
No Training  HCD Training  LFA Training
LFA Training  No Training  HCD Training
LFA Training  HCD Training  No Training
HCD Training  No Training  LFA Training
HCD Training  LFA Training  No Training