Optimal Unemployment Insurance with Worker Profiling*

[Click here for the latest version]

Sergio Cappellini†

November 12, 2021

ABSTRACT

In unemployment assistance programs, the government profiles recipients according to their traits with the twofold goal of facilitating their reemployment and eliminating overpayments. To this purpose, a profiling program establishes (i) which recipients to profile, (ii) when and (iii) how accurately, and (iv) the transfers to be paid after it. This paper provides criteria to rank existing profiling programs, as well as an estimate of the welfare gains from the adoption of the optimal one. Two types of programs are possible at the optimum. The first type are generous programs in which high costs of search-effort compensation make it too costly to delegate the job search to recipients, who thus receive full consumption insurance. The second type, instead, are less generous programs in which recipients who are profiled as highly employable are incentivized to search with lower transfers to reduce effort-compensation costs. Moreover, if poorly employable recipients are better compensated for search effort than left at rest, a fraction of them may be persuaded via profiling to revise their reemployment expectations upward and search at lower incentive costs.

JEL classification: D82, D83, J65, J68

Keywords: Bayesian Persuasion, Job-Search Assistance, Non-Contractible Effort, Social Assistance, Unemployment Insurance, Worker Profiling

*I would like to thank for helpful comments Nicola Pavoni, Nenad Kos, Antonella Trigari and participants at the 19th Annual EEFS Conference, the 16th Annual Symposium on Economic Theory, Policy and Applications (ATINER), the 15th GRASS Workshop, the 2021 EALE Conference, and the 42nd ASSET Annual Meeting.

†PhD candidate, Bocconi University, e-mail: sergio.cappellini@unibocconi.it

1 Introduction

A renewed interest in optimal design of active labor-market policies (ALMPs) started in 2007 amid the financial crisis. Nowadays, following the outbreak of the Covid-19 pandemic, welfare support to the poor and the jobless is at the core of the political agenda of many governments worldwide. Nonetheless, the unprecedented increase in unemployment rates and the contemporaneous economic recession have led to a disproportion between public resources and the need for social security, which ultimately results in a push for optimizing public spending.[1] The trade-off between income support, incentive provision to job search and cost minimization for the public provider has led to policies tailored to recipients' characteristics. As a consequence, tracing a profile of any jobseeker who requests public financial support constitutes an aspect of first-order importance for the design of an effective welfare program. Profiling of welfare claimants is present in most OECD countries[2] and is usually employed as a tool to support and improve the design of existing ALMPs.

In the US, Worker Profiling and Reemployment Services (WPRS) and Reemployment and Eligibility Assessment (REA)[3] are two Federal-funded programs that profile welfare claimants. All workers who request access to public welfare support are asked to report their personal traits, such as education, past working experiences, family background, etc. This information allows for an *early assessment* of reemployment expectations, based on the statistical evidence provided by historical data on claimants' unemployment spells. In addition, both WPRS and REA may implement an *in-depth assessment* of the human capital of each claimant, in the form of one-on-one interviews and/or skill tests, to better tailor the assistance program to their needs.

The two programs generate savings for the provider through distinct channels. First, by improving upon the fit between workers and job-search methods. For instance, in WPRS "UI claimants who are identified through profiling methods as likely to exhaust benefits and who are in need of reemployment services to transition to new employment participate in reemployment services, such as job search assistance" (US Dept. of Labor[4]). Second, by designing transfers based on recipients' needs during the unemployment spell. This holds especially for REA[5], that is devoted to "enhance the rapid reemployment of unemployed workers, identify existing and

[1] Public unemployment spending in the US reached \$622 billions in 2021, accounting for 6.7% of the annual Federal budget (USASpending.gov, https://www.usaspending.gov/explorer/agency).

[2] Some examples are given by Worker Profiling and Reemployment Services and Reemployment and Eligibility Assessment programs (US), the Suivi Mensuel Personnalisé (France), 4-Phase Model (Germany) and Work Programme (UK).

[3] In 2015, REA has been replaced by the REemployment Services and Eligibility Assessment (RESEA) program, which provides greater access to reemployment services. I will nonetheless refer to the former version of the program, as it provides a clearer distinction between profiling and reemployment services which eases the exposition.

[4] https://www.dol.gov/agencies/eta/american-job-centers/worker-profiling-remployment-services

[5] One of the several purposes of RESEA is to "[...] Strengthen UI program integrity" (US Dept. of Labor, https://www.dol.gov/agencies/eta/american-job-centers/RESEA). Hence, the two versions of the program have similar targets.

eliminate potential overpayments, and realize cost savings for UI trust funds" (Poe-Yamagata et al., 2011).[6]

Profiling is complementary to welfare policies, which instead deal with income support and provision of search incentives and assistance. US welfare assistance is funded partly by the Federal government and partly by single States, while the organization and design is mainly deferred to the latter. Profiling programs thus greatly differ along many dimensions, namely (i) who should be profiled, (ii) when and (iii) how accurately, (iv) whether profilees should be requested to search or rest in the meantime and/or upon it, based on the new information obtained, and (v) which payments should accompany it. All these dimensions must therefore be taken into account in the analysis of an optimal profiling program.

The first objective of this paper is to develop a framework suitable to study the main complementarities between profiling and welfare policies. Optimal welfare provision solves the problem of a risk-neutral public welfare provider (hereafter, 'the government'), who needs to maximize the welfare of a risk-averse recipient (hereafter, 'the worker'), subject to a budget constraint and to non-contractible job-search effort of the latter. Following Shavell and Weiss (1979) and Hopenhayn and Nicolini (1997), the design of a welfare program can be formalized as a dynamic principal-agent problem, where the state of the problem is composed by the current utility of the worker/agent, implicit in the structure of future payments, and the level of her expected reemployment skills. Keeping track of the state allows for a recursive formulation of the problem. Job search failures are themselves informative about hidden reemployment perspectives, and cause a revision of expectations. Like in Pavoni et al. (2013), policy instruments arise as the combination of (i) job-search recommendation to workers ('Search' or 'Rest'), (ii) a transfer scheme, made of current consumption and continuation utilities, indexed to future employment status and profiling outcome, and (iii) technologies adopted. The technologies available to the planner are job search assistance and profiling, and can be implemented jointly. The optimal program arises as a sequence of policies over time.

The paper finds that reemployment expectations and the generosity of payments toward workers are crucial determinants of optimal programs. The cost of search incentives being decreasing in the expected human capital, and the one of search-effort compensation being increasing in the generosity of transfers, make the government save on these costs by delegating the job-search to workers when expected human capital (resp., program's generosity) is high (resp., low). For this reason, the government lowers the payments to job seekers who are profiled as highly employable, in the attempt to further ease incentive provision. Therefore, a REA-like program contemplating this form of 'punishments' should optimally be adopted jointly with direct job search. Furthermore, profiling possibly does not fully detect employability at the op-

[6] The report was commissioned by the Employment and Training Administration of the US Dept. of Labor.

timum. For low program's generosity and absent search incentives, the government would find it optimal to require also poorly employable workers to search. If the additional cost of search incentives is not too large, a fraction of poorly employable workers may be persuaded of being more employable than expected and requested to search afterwards. The paper also states sufficient conditions on workers' utility function that guarantee that the number of poorly employable workers requested to search declines in the level of generosity. On the one hand, indeed, search-effort compensation becomes more expensive and the gains from persuading poorly employable workers shrink accordingly. On the other hand, instead, the assumption on workers' utility function causes the decline of search-incentive costs in response to an increase of expectations to be more sizable when the program is more generous. These two forces are jointly conducive to a more accurate profiling when programs are more generous.

The second objective of the paper is to provide a benchmark to evaluate existing programs. To this aim, an upper bound of returns is estimated by solving for the optimal program, conditional on reemployment expectations of each worker and the implicit generosity of transfers envisaged by the government's welfare policies. Performance assessments about WPRS and REA conducted in the past focus on specific margins and targets. The advantage and main contribution of this paper's approach, instead, lies in the absence of any arbitrary assumption, neither about the margins to focus on, nor about the design of the program (sequence of policies and transfers, timing and accuracy of profiling, job-search methods, etc.).

The rest of the paper is organized as follows. Section 2 contains the literature review. Section 3 presents the economic environment. Section 4 describes the welfare policies. Section 5 solves for the optimal program when worker profiling is performance-based. Section 6 solves for the optimal program when worker profiling is based on a statistical assessment. Section 7 conducts a quantitative analysis on REA program in the US. Section 8 extends the analysis to the case of private worker search. Section 9 concludes.

2 Literature Review

The main contribution of this paper is the development of a framework suitable to study worker profiling within a welfare program toward the jobless. The paper provides an analysis of the gains and losses of profiling, in conjunction with others labor-market policies, when workers' human capital is not *ex-ante* observable.

Attempts have been made in the past to estimate returns of profiling programs. Sullivan et al. (2007) and Poe-Yamagata et al. (2011) are examples of such attempts. The first paper ranks WPRS programs in US States according to the occurrence of type-I error (i.e., the probability that a highly employable worker is profiled as lowly so). My paper finds a rationale for such a choice. Indeed, optimal information design always signals low employability with full precision, but noisily detects high employability (i.e., positive probability of type-II error) whenever a share of poorly employable workers is persuaded to be highly so. In the second paper, instead, authors conduct a field study on REA initiative in Florida, Idaho, Nevada and Illinois, and evaluate it over multiple dimensions, such as duration and total amount of unemployment benefits received, likelihood of reemployment and quarterly wage amounts received. In particular, the authors measure a positive impact of REA on public spending in three out of four States.[7] Their and this paper's estimates of cost savings have the same order of magnitude of millions of US Dollars (see Section 7.3).

The existence of an agency problem in the contractual relationship between the welfare provider and the recipients has long been acknowledged by the literature. The provider has the possibility to tackle it either by providing recipients with incentives (Atkenson and Lucas, 1995; Wang and Williamson, 1996; Hopenhayn and Nicolini, 1997; Chetty, 2008; Shimer and Werning, 2008), by monitoring them (Pavoni and Violante, 2007; Setty, 2019), or else, by conducting the search on their behalf (Pavoni et al., 2013; 2016). In all cases, the job search produces an extra cost, which possibly outweighs the expected gain from re-employment. For this reason both active and passive policies coexist in a welfare program and only workers with better job opportunities are referred to the active ones. When job opportunities are allowed to deteriorate during the unemployment spell, workers are reassigned to different policies. Likewise, in this paper any transition to a different policy follows the deterioration in expected human capital. Yet, such a deterioration stems from a learning process which lead agents to revise their initial expectations and does not involve any depreciation of physical human capital. Gonzalez and Shi (2010) study unemployment-to-job transitions in a context where workers are heterogeneous in (unobservable) skills and get discouraged by long-lasting unemployment spells. Permanence in unemployment makes them more inclined to accept lower wage proposals. Therefore, the reemployment equilibrium wage is increasing in the perceived probability of being high-skilled. Similarly, in my framework the duration of unemployment spells has a discouragement effect on job-seekers. However, the need of larger search incentives for more discouraged workers produces a contrasting effect on net reemployment wages.

Differently from physical depreciation, expectation revision can also follow profiling. Profiling can thus become the mean used by the government to persuade jobless recipients to seek new jobs by manipulating their expectations. The paper is related to the vast and growing literature on information design initiated by Kamenica and Gentzkow (2011), that deal with the design of an optimal signaling strategy from a principal/sender to an agent/receiver. The peculiarity

[7]The absence of any positive impact of REA in Illinois is attributed by the authors to the small number of eligible participants (3,122 in 2009).

of the present framework is the 'hybrid' nature of the problem, which mixes the design of information with that of an effort-incentivizing contract. Consequently, the well-known result of Bayesian persuasion stating that the optimal signal delivers the concave closure of the pre-signal payoff function only holds when lotteries over continuation utilities are not allowed and the problem is genuinely one of information design. When no constraint of this type is imposed, instead, profiling can be used to ease incentive provision also by randomizing over continuation utilities. Boleslavsky and Kim (2021) extend the concavification result to a setting with three players (sender, agent and receiver) and incentive provision. The sender designs a signal about a hidden state, and determines the receiver's prior belief by convincing the agent to exert private effort that affects the state distribution. Rodina (2020) considers a similar setting where the agent effort is not private. Bloedel and Segal (2018), Habibi (2020) and Zapechelnyuk (2020) also study the tension between incentive and information provision in the Bayesian persuasion framework applied to situations of agent's rational inattention, agent's time-inconsistency and quality certification, respectively. However, to the best of my knowledge, no work has studied the relationship between information design and incentive provision in unemployment insurance so far.

3 Economic Environment

Players' Interaction. A risk-neutral government (principal, it) and a risk-averse worker (agent, she) populate the economic environment in discrete time. Each player is infinitely-lived and discount future utility at rate $\beta \in (0, 1)$. The worker can be employed or not, and the government observes her employment status. In period 0, (i) the two players are uncertain about the worker's human capital and hold common expectations about it, (ii) the worker is unemployed, and (iii) the government offers her a contract contingent on any possible future employment status and new information about human capital. The contract is so designed to minimize the expected discounted value of net transfers to the worker, conditional on delivering to her a given expected discounted utility. For each history node, the contract specifies the technology/ies adopted by the government (assisted-search and/or profiling), the effort recommendation to the worker ('Search' or 'Rest') and transfers. Uncertainty about worker's employment status clears at the beginning of each period.

Human Capital and Job Search. Worker's human capital can be high ($h = H$), or low ($h = L$). Workers with high (resp., low) human capital are labelled as high- (resp., low-)skilled. If unemployed, the worker can either rest ($a = 0$) or search for a job ($a = 1$). In the first case, her job-finding probability is null. In the second case, the high-skilled worker finds a job with probability π_H, while the low-skilled one with probability $\pi_L \in (0, \pi_H)$. The job search is public, but non-contractible, and makes any worker incur effort cost e, with worker's utility over consumption c and effort a being separable and given by $v(c, a) = u(c) - e \cdot a$. The first-order derivative of u^{-1}, $1/u'$, is convex.

Market-sector production. Labor productivity is increasing in human capital ($\omega_H > \omega_L$). In the economy there is one market sector only, populated by identical atomistic firms competing à la Bertrand over job offers, and paying wages equal to labor productivity. Reemployment is an absorbing status, since the worker faces no risk of any future lay-off.

Expectations. Any worker who applies to welfare support undergoes an early assessment. The assessment attaches to the worker a probability μ of being high skilled ($\mu = Prob(h = H)$), which is henceforth referred to as *expectation*.[8] In actual programs, welfare claimants report personal information (social background, past working experiences, education, etc.), according to which the welfare provider makes an initial evaluation of their human capital. The evaluation of any claimant is based on historical data that measure the reemployment frequency of claimants with same characteristics. Highly-educated and more experienced workers, for instance, are statistically more likely to exit unemployment than workers with less experience and/or lower educational attainment.

Assisted-search technology. The government can search on behalf of the worker at cost κ^{ja}. The cost includes the administrative expenses of the offices which are in charge of looking for vacancies, create a network with prospective employers and maintain contacts with them, circulate the worker's CV, etc.

Profiling technology. Profiling detects human capital with some accuracy, and returns a publicly observable outcome, at cost κ^{wp}[9]. Profiling can be thought of as a lottery that returns a binary outcome -'Pass' ($r = p$) or 'Fail' ($r = f$)-, with predetermined odds. The government can choose to profile with different levels of accuracy workers holding different expectations. This means that the lottery odds are indexed by expectation μ and program's generosity U

$$\left\{\sigma(r|h, \mu, U)\right\}_{r\in\{p,f\},\, h\in\{H,L\}}$$

4 Policies

Any policy arises as the composition of (i) recommended search effort, (ii) consumption contract, and (iii) adopted assisted-search and/or profiling technology/ies (if any). Combinations of search effort levels and technologies gives rise to eight ($2 \times 2 \times 2$) possible policy instruments. However, when the assisted search technology is implemented, it would be redundant to prescribe positive

[8]By the law of large numbers, such a probability is unbiased, meaning that the fraction of high-skilled workers among all workers with same expectation coincides with the expectation itself.

[9]The cost includes administrative expenses, as in the case of assisted search.

search effort to the worker, which reduces to six the number of policies. If no technology is implemented, the government can decide whether to recommend positive search effort and pay incentives ('Unemployment Insurance', $i = UI$), or not ('Social Assistance', $i = SA$). If only the assisted search technology is implemented, it gives rise to 'Job-Search Assistance' ($i = JS$). Profiling without any search gives rise to 'Assistance and Profiling' ($i = AP$), whereas 'Insurance and Profiling' ($i = IP$) arises when the technology is adopted together with worker's search. Finally, 'Search-Assistance and Profiling' ($i = SP$) originates if both technologies are jointly adopted.

No Profiling			
Recommendation	Assisted Search	Delegated Search	No Search
'Search'	x	Unemployment Insurance (UI)	x
'Rest'	Job-Search assistance (JS)	x	Social Assistance (SA)
Profiling			
Recommendation	Assisted Search	Delegated Search	No Search
'Search'	x	Insurance & Profiling (IP)	x
'Rest'	Search-assistance & Profiling (SP)	x	Assistance & Profiling (AP)

Table 1: Policy Instruments

At time $t = 0$, the planner offers the unemployed agent an insurance contract that minimizes transfers and guarantees her an expected discounted utility equal to U. The planner's problem can be written recursively by keeping track of worker's expected human capital and promised utility -henceforth, a proxy for program's generosity- along the unemployment spell. The consumption contract of policy i consists of a menu of today's consumption c^i and tomorrow's continuation utilities $U_i^{s,r}$, contingent on reemployment ($s = w$) or not ($s = u$), and 'Pass' ($r = p$) or 'Fail' ($r = f$) outcome, if job search and/or profiling are conducted. Current expectation μ and the program's generosity U jointly determine the choice of the policy instrument. The government chooses the optimal policy $i(\mu, U)$ by solving

$$V(\mu, U) = \max_{i \in \{SA,JS,UI,AP,SP,IP\}} V^i(\mu, U) \quad (1)$$

The planner is allowed to randomize over worker's utility U under the constraint that the promised utility must be delivered in expectation. To this end, the operator $\mathbf{V}$ is defined as

$$\mathbf{V}(\mu, U) = \max_{\{U(x)\}_{x\in[0,1]}} \int_0^1 V(\mu, U(x))dx \quad (2)$$

$$\text{sub:} \quad U = \int_0^1 U(x)dx$$

In the following, I introduce the problem of the welfare provider in case of re-employment and for all six instruments during unemployment. First, I define welfare-oriented policies (SA, JS and UI) and later the profiling ones (AP, SP and IP).

4.1 Welfare Policies

Wage Tax/Subsidy (W). In case of successful job search, the worker's productivity is revealed. Therefore, the market-sector value when human capital is equal to $h \in \{H, L\}$ reads

$$W(h,U) = \max_{\tau, U^w} \tau + \beta W(h, U^w) = \max_{c^w, U^w} \omega_h - c^w + \beta W(h, U^w)$$
$$\text{sub:} \quad U = u(c^w) + \beta U^w \qquad \text{(PK)}$$

Since reemployment is assumed to be an absorbing state (the separation rate between employees and firms is assumed null), the planner is sure to raise tax/pay subsidy also in the next period. The labor tax τ is the wedge between gross (ω_h) and net wage (c^w). The Promise-Keeping (hereafter, (PK)) constraint is the recursive expression of worker's utility. It guarantees that utility flow from current period $u(c^w)$ and continuation utility U^w are large enough to match current utility level U. The optimal contract prescribes constant continuation utility ($U^w = U$).[10] Hence from (PK) one can obtain the closed-form expression for consumption $c^w = u^{-1}\big((1-\beta)U\big)$. The expression for labor tax/subsidy thus is

$$W(h,U) = \frac{\omega_h - u^{-1}\big((1-\beta)U\big)}{1-\beta}$$

Social Assistance (SA). The planner's problem when neither job search, nor profiling is performed reads

$$V^{SA}(\mu, U) = \max_{c^{sa}, U^{sa}} -c^{sa} + \beta \mathbf{V}(\mu, U^{sa})$$
$$\text{sub:} \quad U = u(c^{sa}) + \beta U^{sa} \qquad \text{(PK)}$$

The planner transfers c^{sa} and pledges continuation utility U^{sa}, without requiring the worker to exert any effort. SA is a passive measure, fully devoted to income support, and does not envisage any form of job search. Thus, there is no chance of reemployment for the worker, nor any chance for the provider of raising a labor tax in the incoming period. Differently from the definition of wage tax/subsidy, where reemployment is an absorbing state, the planner can freely select the best policy instrument in the next period. However, the following holds.

[10] At the optimum, U^w solves

$$W_U(h,U) = -\frac{1}{u'(c^w)} = W_U(h, U^w) \Longrightarrow U^w = U$$

Proposition 1 (Absorbing SA). *Social Assistance is an absorbing policy and its continuation utility equals current utility ($U^{sa} = U$).*

Proof. See Appendix A: Properties of SA, JS and UI. ■

The proof follows the same steps as in Pavoni et al. (2016). The result implies that, once the worker enters SA, she is never reallocated to any other policy, neither she can exit unemployment, as no search is conducted. This result is admittedly quite extreme for policymakers, who may find it hard to politically defend a welfare program granting life-time financial support to people who will never have the chance of getting reemployed. Yet, the result is remarkable in that it establishes that any passive policy should be regarded as a policy of last resort, to target only to workers with low expected human capital. Current consumption solves (PK) with $U^{sa} = U$.[11] The value of SA is independent of μ and has a closed-form expression

$$V^{SA}(U) = -\frac{u^{-1}\big((1-\beta)U\big)}{1-\beta} \tag{3}$$

No revision of expectations occurs during SA, as no job search is conducted. When, instead, the search is unsuccessful, both the government and the worker downward revise their initial expectation μ, according to the formula

$$\mu' := \frac{\mu\big(1-\pi_H\big)}{\mu\big(1-\pi_H\big) + (1-\mu)\big(1-\pi_L\big)} \leq \mu \tag{4}$$

where μ' is the revised probability that worker's human capital is $h = H$. μ' is lower than the initial one, with equality holding only if human capital was already known ($\mu \in \{0, 1\}$). The reason lies in the unbiasedness of μ, that is equal to the actual share of high-skilled workers among those who hold that expectation. Thus, a fraction $\pi(\mu) := \mu\pi_H + (1-\mu)\pi_L$ of them manages to find a new employment, which implies that the high-skilled who remained unemployed after one period are a fraction $\mu(1-\pi_H)/(1-\pi(\mu))$ of the initial group. Therefore, in case of failed search, a higher probability is attached to realization $h = L$.[12]

[11] At the optimum, U^{sa} solves

$$V_U^{SA}(\mu, U) = -\frac{1}{u'(c^{sa})} = V^{SA}(\mu, U^{sa}) \implies U^u = U$$

[12] If the failed attempts to exit unemployment are t, one for each period, then initial expectation μ is updated t times according to the formula

$$\mu^{(t)} = \mu^{(t-1)\prime} = \frac{\mu\big(1-\pi_H\big)^t}{\mu\big(1-\pi_H\big)^t + (1-\mu)\big(1-\pi_L\big)^t} \tag{5}$$

where the convention that $\mu^{(0)} = \mu$ is used. It is easy to see that:

- $\mu = 0$ and $\mu = 1$ are the only two expectations such that $\mu^{(t)} = \mu$. When players know human capital, no update ever occurs;
- $\lim_{t\to\infty} \mu^{(t)} = 0$, if $\mu^{(0)} < 1$.

Job-Search assistance (JS). When resorting to assisted search, the government looks for employment on worker's behalf, an activity that costs him κ^{ja}. The value of JS reads

$$V^{JS}(\mu, U) = \max_{c^{js}, U^w_H, U^w_L, U^u} -c^{js} - \kappa^{ja} + \beta\big[\mu\pi_H W(H, U^w_H) + (1-\mu)\pi_L W(L, U^w_L) + (1-\pi(\mu))\mathbf{V}(\mu', U^u)\big]$$

$$\text{sub:} \quad U = u(c^{js}) + \beta\big[\mu\pi_H U^w_H + (1-\mu)\pi_L U^w_L + (1-\pi(\mu))U^u\big] \qquad \text{(PK)}$$

Two are the sources of risk related to the job search. The first risk is related to its outcome (success or failure). The second one, instead, is connected to human capital realization, conditional on finding a new job for the worker. While the government finds it optimal to insure the agent against the latter, due to her risk aversion ($U^w_H = U^w_L$), the same holds for the former only if no search incentive is to be paid ahead, i.e. if the worker will in no case be referred to UI during the spell.[13] In either case, then the optimal contract solves

$$-\frac{1}{u'(c^{js})} = W_U(\mu, U^w) = \mathbf{V}_U(\mu', U^u)$$

with

$$W(\mu, U) := \frac{\mu\pi_H}{\pi(\mu)} W(H, U) + \frac{(1-\mu)\pi_L}{\pi(\mu)} W(L, U)$$

being the expected wage tax/subsidy, conditional on reemployment.

Unemployment Insurance (UI). The planner may delegate the job search to the agent and provide her with incentives to conduct it. Incentive provision originates from the fact that worker's effort is non-contractible, and boils down to adding an Incentive Compatibility constraint (hereafter, (IC)) to the planner's problem.

$$U \geq u(c^{ui}) + \beta U^u \qquad \text{(IC)}$$

The (IC) constraint guarantees incentive compatibility of the contract against agent's deviation from recommended search effort. Promise Keeping in UI takes into account the effort cost e exerted by the job-seeker agent

$$U = u(c^{ui}) - e + \beta\big[\pi(\mu)U^w + (1-\pi(\mu))U^u\big] \qquad \text{(PK)}$$

[13] The proof is reported in Appendix A: Properties of SA, JS and UI.

The problem of the planner reads

$$V^{UI}(\mu, U) = \max_{c^{ui}, U^w, U^u} -c^{ui} + \beta\big[\pi(\mu)W(\mu, U^w) + (1-\pi(\mu))\mathbf{V}(\mu', U^u)\big]$$
$$\text{sub:} \quad \text{(PK) - (IC)}$$

(IC) and (PK) constraints imply the following condition on the difference in continuation utilities between successful (U^w) and failed (U^u) search

$$U^w - U^u \geq \frac{e}{\beta\pi(\mu)} \tag{6}$$

The condition in (6) is binding at the optimum and accounts for the planner's cost of incentive provision. Incentive cost can be defined by the difference in costs between the cases of contractible and non-contractible effort.[14] Incentive costs are increasing in the cost of effort and decreasing in the level of patience (β) and confidence (μ). Intuitively, it is less expensive to convince the agent to search when she expects larger return on search and weighs more the prospective reward ensuing from it. Condition (6) shows that search incentives have a convex hyperbolic shape in the space of expectations. Concavity of V^{UI} in μ follows from convexity of incentive costs and the linearity of returns. Lemma 1 proves these two facts.

Lemma 1 (**Slopes of the value functions with respect to μ and U**). *Every policy return V^i is concave increasing in expectations, and concave decreasing in promised utility. V^{UI} is supermodular (i.e, $V^{UI}_{\mu U}(\mu, U) \geq 0$). And so is V^{JS}, whenever $U^w_{JS} \geq U^u_{JS}$.*

Proof. See Appendix A: Properties of SA, JS and UI. ■

Incentive costs depending negatively on expectations through the utility dispersion generates a comparative advantage of UI for high-end expectations. On the contrary, convexity of $1/u'$ is a sufficient condition for the costs of incentive provision and effort compensation to be increasing in U, and so for UI to suffer a comparative disadvantage for high-end generosities. For this second reason, more generous programs are mainly focused on assistance, in the form of income support and assisted search, and less on search incentives. The following proposition establishes how policies are located over the (μ, U) space.

Proposition 2 (**Welfare Policies in the (μ, U) Space**). *Assume* $\mathbf{V}$ *is 'locally' supermodular, that is, for every (μ, U), there exist ϵ_μ, $\epsilon_U > 0$ such that*

$$\mathbf{V}_U(\tilde{\mu}, U) \geq \mathbf{V}_U(\mu, U) \quad \textit{and} \quad \mathbf{V}_\mu(\mu, \tilde{U}) \geq \mathbf{V}_\mu(\mu, U)$$

[14] On the same lines of Pavoni and Violante (2007), one can imagine the existence of a policy which delegates job search to workers whenever effort is contractible. If that is the case, the government will only need to compensate for the worker's effort. If so, the incentive cost is defined as the difference in cost of contract between UI and this new policy.

for every $\tilde{\mu} \in (\mu, \mu + \epsilon_\mu)$ *and* $\tilde{U} \in (U, U + \epsilon_U)$, *and that* $\mu' \in (\mu - \epsilon_\mu, \mu)$ *and* $U_i^u \in (U - \epsilon_U, U)$. *Then,*

$$V_U^{UI}(\mu, U) \leq V_U^{JS}(\mu, U) \leq V_U^{SA}(U) = W_U(\mu, U) < 0 \tag{7}$$

and

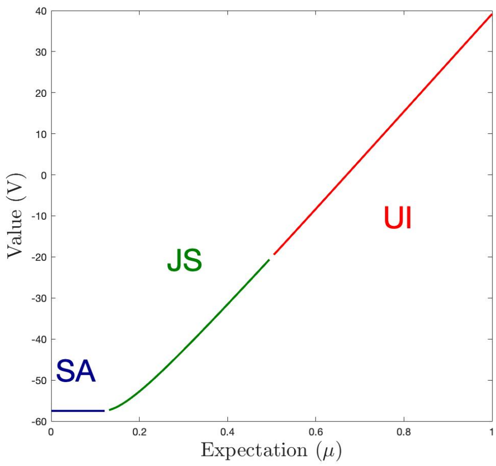
$$0 = V_\mu^{SA}(U) \leq V_\mu^{JS}(\mu, U) \leq V_\mu^{UI}(\mu, U) \tag{8}$$

where the last inequality holds (at least) at the crossing point. Lastly, if the program does not allow transition from JS to UI, then $V_{\mu\mu}^{JS}(\mu, U) = 0$ *and* $V_U^{JS}(\mu, U) = V_U^{SA}(U)$.

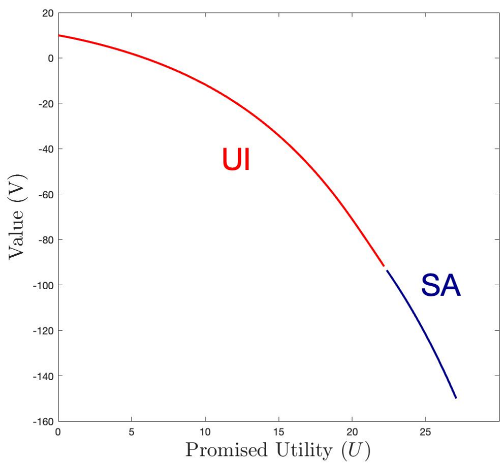
Proof. See Appendix A: Properties of SA, JS and UI. ∎

The assumption about local supermodularity of $\mathbf{V}$ descends from supermodularity of each V^i and requires the marginal gain of μ to be increasing in the level of generosity (i.e., $\mathbf{V}_{\mu U}(\mu, U) \geq 0$). A rise in $\pi(\mu)$ increases the return of the job search, no matter who between the government and the worker conducts it. However, the assumption only holds locally. Indeed, the shape of $\mathbf{V}$ is determined by two contrasting forces, within-policy supermodularity and between-policy submodularity. While the former dominates locally, the latter has a global impact on $\mathbf{V}$.

Prop. 2 can be read as follows. Fix μ and move U. Then, UI, JS and SA are optimal for low, intermediate and high U, respectively. Now, fix U and move μ. SA, JS and UI are optimal for low, intermediate and high μ, respectively. The marginal value of μ is increasing in the level of search intensity, duration and effort by the worker.



(a) Constant U ($\exp((1-\beta)U) = \$575$)



(b) Constant μ ($\pi(\mu) = 0.15$)

Figure 1: Value of welfare policies.[15]

[15]The parameter values and functional forms used in this Section are: $u(.) = \log(.)$, $\beta = 0.9$, $e = 0.5$, $\kappa^{ja} = 6$, $\kappa^{wp} = 1.5$, $\omega_H = 20$, $\omega_L = 3$, $\pi_H = 0.27$, $\pi_L = 0.14$. All monetary values are divided by 100.

Therefore, fixing generosity and spanning the space of expectations, one observes that policies with higher (resp., lower) marginal returns are optimal for higher (reps., lower) expectations. The upper envelope V thus displays a tendency toward between-policy convexity in the space of μ (see Fig. 1a). Such shape of V has deep implications on the choice of optimal profiling. Similarly, Fig. 1b plots V over U for constant μ. UI is optimal for low generosity and cost of effort. On the other end of the spectrum, SA is first-best policy for high generosity, as no search is conducted. The slope of JS lies in between the one of UI and SA and depends on the downstream program in case the worker is not reemployed in the next period. The following proposition clarifies the point.

Proposition 3 (Optimal Dynamics of Promised Utility and Benefits). *Continuation utility upon failed search is*

- *decreasing when UI is part of the policy sequence ahead;*
- *constant, otherwise.*

Unemployment benefits are constant in SA and JS, and decreasing in UI.

Proof. See Appendix A: Properties of SA, JS and UI. ■

Since incentive costs increase in the level of utility promised to the worker, such utility decreases during UI. Furthermore, whenever the welfare program implements assisted search first and worker's search later, the planner finds it optimal to start decreasing worker's promised utility while the worker is still in JS. This finding sheds light on the possible policy patterns that can arise as a function of worker's initial expectation and program's generosity. Fig. 2 shows two instances of optimal policy sequences, for same initial expectation ($\mu_0 = 0.9$) and different generosity. When generosity is higher ($\overline{U}_0 = 28.9$), the worker never enters UI and only moves from JS to SA. Thus, no effort exertion is requested her and so she is granted both consumption and utility insurance along the spell. Hence, she eventually exits unemployment or enters SA with the same utility level and consumption as the entry one. When, instead, generosity is lower ($\underline{U}_0 = 24.7$), the worker is relocated from JS to UI. For this purpose, her utility decreases over time until she finds a job and exits unemployment.

One may ask whether a worker in UI can be referred to JS. Two contrasting forces have an impact on the policy transition over the UI-JS frontier. First, the optimal contracts of UI and JS prescribe a decline in promised utility for values of the (μ, U) space that are close enough to the frontier. This produces a decrease in the difference in contract costs and makes UI more appealing *ceteris paribus*. Second, the downward revision of expectations causes an increase in the incentive cost of UI, which makes JS more appealing. The following proposition establishes a relationship between the concavity of V and the slope of the UI-JS frontier in the μ space that avoid the possibility of any transition from UI to JS.

Proposition 4 (**Policy Transitions**). *Let $\hat{U}(\mu)$ be the promised utility level that makes the government indifferent between administering UI or JS to a worker with expectation μ, i.e.*

$$V^{UI}(\mu, \hat{U}(\mu)) \equiv V^{JS}(\mu, \hat{U}(\mu))$$

In addition, define $\eta^i(\mu, U)$ as the real non-negative number such that

$$V_U^i(\mu, U) \equiv -g'((1-\beta)U + \eta^i(\mu, U)), \quad i = SA, UI, JS$$

where $g \equiv u^{-1}$ is the inverse utility function. Assume that $\eta^i(\mu, U)$ is non-increasing in U for every μ and that

$$\beta\big[\hat{U}(\mu) - \hat{U}(\mu')\big] \leq \eta^{UI}(\mu, \hat{U}(\mu)) \tag{9}$$

Then, any worker in UI either remains in UI, enters SA or exits unemployment. In particular, the optimal program never switches from UI to JS.

Proof. See Appendix A: Properties of SA, JS and UI. ■

The sufficient condition (9) establishes an upper bound on the slope of the UI-JS frontier in the μ space. Such slope is determined by the change in the difference of contract costs between UI and JS in response to an increase of μ and U. In particular, (i) incentive costs in UI fall in response to any increase of μ and (ii) both incentive and effort-compensation costs in UI rise in response to any increase in U.[16] An upper bound on the slope of the frontier thus requires that the first determinant does not have a major effect on the difference of contract costs, so that every variation of μ only requires a small variation of U (of equal sign) to reestablish the parity between the value of UI and JS. Condition (9) guarantees that the former effect prevails over the latter one.

5 Performance-Based Profiling

Profiling publicly discloses worker's human capital, up to a level of accuracy chosen by the government. For this purpose, profiling programs implement different strategies to infer the level of human capital of the worker. Some assign the profilee a given task and assess human capital based on how she performs. Therefore, profiling is designed as a test with two possible

[16] Assumption on convexity of $1/u'$ guarantees that incentive cost are positively related to promised utility.

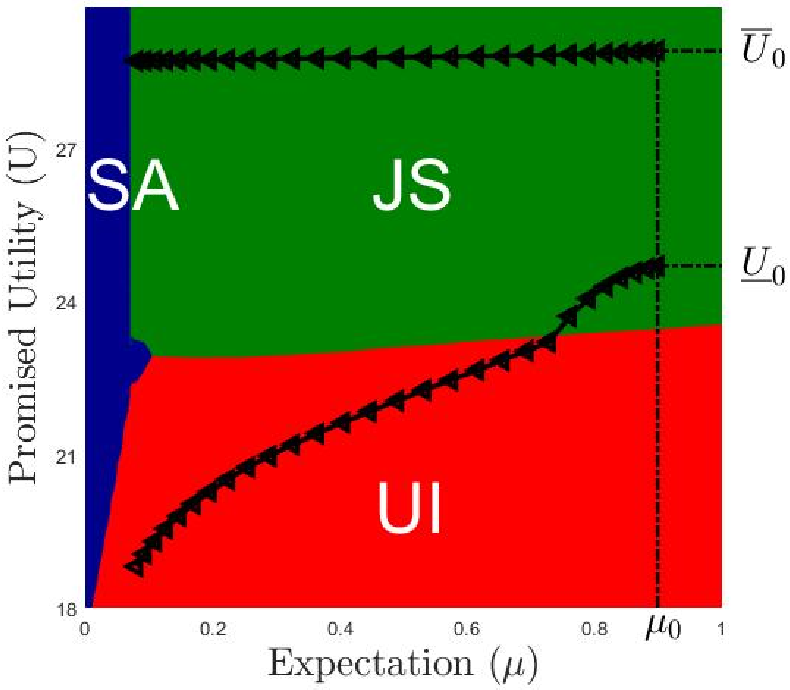


Figure 2: Optimal Welfare Policies in the Space of Expectation and Generosity.

outcomes, 'Pass' ($r = p$) or 'Fail' ($r = f$). Conditional on the outcome of the test, the worker is referred to a different policy. Increasing the difficulty of the task is a way to improve the accuracy of the test, as more low skilled people are failing it and receiving a 'Pass' is more indicative of high human capital. The probability schedule brings both the worker (i.e., the profilee) and the government (i.e., the profiler) to revise expectation μ upon observation of the public outcome r according to the formula

$$\mu_r = \frac{\mu\sigma(r|H,\mu)}{\mu\sigma(r|H,\mu) + (1-\mu)\sigma(r|L,\mu)}$$

A necessary and sufficient condition for profiling to induce a change in expectations is to avoid returning either outcome with the same probability, irrespective of underlying human capital realization (e.g., $\sigma(r|H,\mu) \neq \sigma(r|L,\mu)$). In addition, profiling does not create any type of bias in the aggregate, since expectations are correct on average. Which boils down to require that the revised expectations are equal in mean to the prior (so called Martingale Property, (MP) henceforth).

$$q\mu_p + (1-q)\mu_f = \mu, \quad \mu_f, \mu_p \in [0,1], \quad \mu_f \leq \mu \leq \mu_p \quad \text{(MP)}$$

(MP) can be interpreted as a restriction requiring profiling to be credible. Indeed, considering all workers who share the same expectation μ, inducing any of them to revise their expectation up to μ_p comes at the cost of inducing an expectation revision down to μ_f for someone else.[17] Given the nature of this profiling methodology, any profiled worker could pretend to be low-skilled by

[17] Without loss of generality, the posterior upon 'Fail' (μ_f) is set to be lower than the posterior upon 'Pass' (μ_p).

intentionally failing the test. It is thus necessary that the continuation utility upon 'Pass' is non smaller than the continuation utility upon 'Fail', in which case the worker retains the same expectation as before being profiled. Indeed, the worker who intentionally fails the test derives no new information about her human capital. Such a requirement, labelled No Discrimination constraint (ND, hereafter) imposes a restriction on the contract offered under either profiling outcome. In principle, the worker's utility of choosing the off-the-equilibrium action may be difficult to compute, as transfers are to be evaluated with the non-revised expectation. For instance, if the worker is requested to search right after, the consumption dispersion present in the UI contract is weighed differently as

$$U_f^u = u(c^{ui}) - e + \beta[\pi(\mu_f)U_{ui}^w + (1-\pi(\mu_f))U_{ui}^u(\mu'_f)] < u(c^{ui}) - e + \beta[\pi(\mu)U_{ui}^w + (1-\pi(\mu))U_{ui}^u(\mu')]$$

However, the following result applies.

Proposition 5 (**No Type I Error**). *Only low-skilled workers who undergo profiling receive 'Fail' (i.e., $\mu_f = 0$), and are referred to Social Assistance henceforth.*

Proof. Any disclosure of new information via profiling is equivalent to a randomization over the space of expectations with mean equal to initial μ. Therefore, the government finds it convenient to implement profiling whenever it can exploit the between-policy convexity generated by the different policy slopes as in (8). The linearity of V^{SA} in μ causes the concave closure of $\mathbf{V}$ (for constant U) to be obtained by referring failed profilees to SA with $\mu_f = 0$. Intuitively, reducing the likelihood of 'Fail' outcome increases the frequency of 'Pass' (q) more than one to one. This fact, joint with the linearity of V^{SA} in μ (see Fig. 1a), makes convenient to limit 'Fail' only to low-skilled people (i.e., zero probability of type I error) and induces as many workers as possible to upward revise expectations. ■

Therefore, since SA is an absorbing policy and no job search is conducted ever after, the (ND) constraint has an easy formulation.

The timing of profiling policies is as follows. In the current period, the profiled worker is paid current transfers and possibly asked to search. In the next period, the outcome of the job search (if any) is disclosed first, before the one of test. The timing implies that the government can not index worker's current (c^i) and future re-employment consumption (c_i^w) on the new information derived from profiling.

Assistance-and-Profiling (AP). AP does not envisage any job search. Thus, the planner's problem reads

$$
\begin{aligned}
V^{AP}(\mu,U) = & \max_{c^{ap},(U_r^u)_{r=\{p,f\}},\mu_p} -c^{ap} - \kappa^{wp} + \beta\big[q\mathbf{V}(\mu_p,U_p^u) + (1-q)\mathbf{V}(0,U_f^u)\big] \\
\text{sub:} \quad & U = u(c^{ap}) + \beta\big[qU_p^u + (1-q)U_f^u\big] \text{ (PK)}, \quad \text{(MP)} \\
& U \geq u(c^{ap}) + \beta U_f^u \text{ (ND)}
\end{aligned}
$$

The (ND) constraint can be rewritten as $U_p^u \geq U_f^u$. Absent (ND), the contract would equate the slopes of $\mathbf{V}$ across the different outcomes (i.e., $\mathbf{V}_U(\mu_p,U_p^u) = \mathbf{V}_U(0,U_f^u)$). By (7), this would boil down to set

$$U_f^u \geq U_p^u \tag{10}$$

Intuitively, if the government expects to incur an extra cost for incentive provision later during the unemployment spell, then it would save on this cost by decreasing the promised utility of the worker who passes the test and may at some point be referred to UI. Condition (10) and (ND) constraint bring the planner to insure the worker against the risk of profiling outcome ($U_f^u = U_p^u = U^u$).

Passing to the choice of μ_p, one may be tempted to adopt a reasoning similar to the one that leads to refer to SA only low-skilled workers ($\mu_f = 0$) and guess that 'Pass' outcome only targets high-skilled workers (e.g. $\mu_p = 1$) at the optimum. However, this is not always the case, due to concavity of $\mathbf{V}$ in μ for high-end expectations. Indeed, while concave returns cause the marginal gain of 'Pass' informativeness about high human capital to decline in the level of informativeness itself, reducing the frequency of 'Pass' and failing more workers cause a loss at the margin. Therefore, the planner trades off informativeness of 'Pass' against its frequency up to the point where the gain of higher informativeness equals the cost of lower frequency. In case the marginal gain exceeds the marginal cost for every μ, the test fully discloses high human capital. Otherwise, the internal solution satisfies

$$\mathbf{V}_\mu(\mu_p,U_{AP}^u) = \frac{\mathbf{V}(\mu_p,U_{AP}^u) - \mathbf{V}(0,U_{AP}^u)}{\mu_p} \tag{11}$$

Eq. 11 shows that the upper posterior does not depend on worker's initial expectations, which means that all profiled workers hold the same revised expectation after receiving a 'Pass'. The downside is that the value of information for the government is negative when workers' initial expectation is larger than μ_p, irrespective of the administrative cost of profiling, as the low-skilled ones among them are mistaken in a direction favorable to the government. Hence, disclosing any information about their actual human capital causes the planner a loss that outweighs the gain of informing high-skilled workers.

Search-Assistance-and-Profiling (SP). Whenever the planner jointly adopts assisted search and profiling technologies, its problem reads

$$
\begin{aligned}
V^{SP}(\mu, U) = \max_{c^{sp}, U^w, U^u_p, U^u_f, \mu_p} & -c^{sp} - \kappa^{wp} - \kappa^{ja} + \\
& + \beta\big[\pi(\mu)W(\mu, U^w) + q(1-\pi(\mu_p))\mathbf{V}(\mu'_p, U^u_p) + (1-q)(1-\pi_L)\mathbf{V}(0, U^u_f)\big] \\
\text{sub:} \quad & U = u(c^{sp}) + \beta\big[\pi(\mu)U^w + q(1-\pi(\mu_p))U^u_p + (1-q)(1-\pi_L)U^u_f\big] \quad \text{(PK)} \\
& U \geq u(c^{sp}) + \beta\big[\pi(\mu)U^w + (1-\pi(\mu))U^u_f\big] \;\; \text{(ND)}, \quad \text{(MP)}
\end{aligned}
$$

As for the case of AP, the planner insures the worker against the risks related to profiling, by committing to a constant continuation utility (i.e., $U^u_p = U^u_f$). About the informativeness of the profiling strategy, the posterior expectation μ_p induced by 'Pass' outcome, is either 1 or solves

$$
\mathbf{V}_\mu(\mu'_p, U^u_{SP}) = \frac{\mathbf{V}(\mu'_p, U^u_{SP}) - \mathbf{V}(0, U^u_{SP})}{\mu'_p} \tag{12}
$$

Indeed, if in case of AP the randomization in the space of expectations occurs over the upper envelope $\mathbf{V}$, now instead the randomization only occurs conditional on job-search failure. Therefore, optimal profiling in SP delivers the concave closure (net of cost κ^{wp}) of $(1-\pi(\mu))\mathbf{V}(\mu', U^u_{SP})$ in the space of expectations $\mu \in [0, 1]$.

Insurance-and-Profiling (IP). When profiling is implemented jointly with delegated search, the planner's problem reads

$$
\begin{aligned}
V^{IP}(\mu, U) = \max_{c^{ip}, U^w, U^u_p, U^u_f, \mu_p} & -c^{ip} - \kappa^{wp} + \\
& + \beta\big[\pi(\mu)W(\mu, U^w) + q(1-\pi(\mu_p))\mathbf{V}(\mu'_p, U^u_p) + (1-q)(1-\pi_L)\mathbf{V}(0, U^u_f)\big] \\
\text{sub:} \quad & U = u(c^{ip}) - e + \beta\big[\pi(\mu)U^w + q(1-\pi(\mu_p))U^u_p + (1-q)(1-\pi_L)U^u_f\big] \quad \text{(PK)} \\
& U \geq u(c^{ip}) + \beta\big[qU^u_p + (1-q)U^u_f\big] \quad \text{(IC)}, \quad U^u_p \geq U^u_f \;\; \text{(ND)}, \quad \text{(MP)}
\end{aligned}
$$

Similarly to SP, profiling delivers the concave closure of $(1-\pi(\mu))\mathbf{V}(\mu', U^u_{IP})$ by selecting a posterior upon 'Pass' which equal 1 or solves (12).

5.1 Optimal Welfare Program

I am now ready to characterize the optimal program. It is useful to first look at which policy is optimal in each region of the (μ, U) state space. Later, I will argue how the accuracy of profiling depends on the level of promised utility of the program. The following result places performance-based profiling policies in the (μ, U) space.

Proposition 6 (**Profiling Policies in the** (μ, U) **Space**)**.** *Fix* U*. No profiling policy is optimal for very high or very low expectations. Assistance-and-Profiling (AP) is the optimal profiling policy over low-end expectations, Search-assistance-and-Profiling (SP) and Insurance-and-Profiling (IP) over high-end ones. Now fix* μ*. IP is optimal for lower generosities, while AP and SP are optimal for higher generosities.*

Proof. See Appendix B: Properties of AP, SP and IP. ■

The first part of the proposition can be explained through gains and losses of profiling. Profiling generates savings for the government by delegating search to high-skilled workers with a lower cost of incentives. The losses are of two types. First, the government incurs administrative expenses (κ^{wp}). Second, it suffers a loss by passing any information to low-skilled workers who are overconfident about their human capital. Therefore, for very high and very low expectations, workers are on average efficiently matched with policies, and the gains from reallocation and/or transfer reduction are outweighed by the losses. The second part of Prop. 6 sheds light on the correspondence between profiling policies and their welfare counterparts. Indeed, each profiling policy dominates the other two in a region of the space of expectations where the first-best welfare policy is the one that conducts (or does not conduct) the job search with the same method.

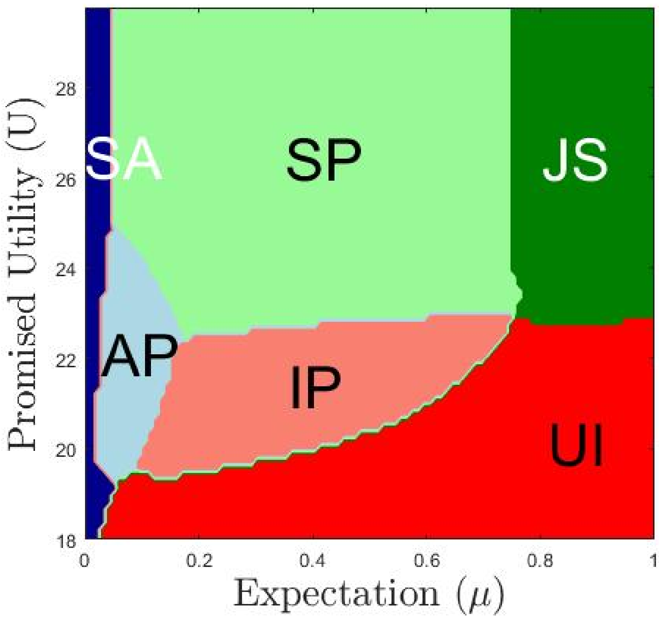


Figure 3: Optimal Policies in the Space of Expectation and Generosity.

Fig. 3 reports the optimal policies in the (μ, U) space. The complementarity of search effort and expectations is mirrored in the best profiling policy adopted. In particular, AP does not implement search and is thus optimal for lower-intermediate expectations. SP and IP, on the other hand, which contemplate different forms of search, are optimal for upper-intermediate expectations. As generosity rises (moving vertically from bottom to top of Fig. 3), the return of job search decreases due to higher costs of search-effort compensation and incentive provision (due to convex $1/u'$[18]), and worker's search (UI and IP) is replaced by assisted search (JS and SP) for high-end expectations, or by no search (SA and AP) for low-end expectations.

An important aspect of a profiling policy is the level of accuracy with which to detect the human capital of each profilee. Indeed, as outlined in Eq. 11 and 12, the value of information to the planner is possibly negative when the within-policy concavity of $\mathbf{V}$ prevails over its between-policy convexity in μ. If so, the optimal profiling strategy does not completely disclose high human capital upon 'Pass' ($\mu_p < 1$) and refers also a fraction of low-skilled workers to active policies upon it. Thus, any worker who receives a 'Pass' and is referred to an active policy can downward revise her expectation and reenter into IP, SP or AP at any later stage (unless she exits unemployment in the meantime). On the contrary, if profiling is fully accurate, workers do not revise expectation henceforth and never undergo profiling at any successive stage. The following result establishes a monotone relationship between accuracy and generosity.

Proposition 7 (**Optimal Profiling Accuracy**). *Fix μ. Then, the accuracy of profiling in detecting high human capital is increasing in the level of generosity ($\partial\mu_p/\partial U \geq 0$). In particular, profiling is fully accurate under Search-Assistance-and-Profiling (SP) when no worker's search (UI or IP) is ever implemented in the downstream policy sequence.*

Proof. See Appendix B: Properties of AP, SP and IP. ■

The intuition of the result hinges on the sensitivity of the slope of $\mathbf{V}$ in μ to changes of U. As anticipated in Prop. 2, $\mathbf{V}$ is locally within-policy supermodular and globally between-policy submodular. Looking at the determinants of μ_p in Eq. 11 and 12, within-policy supermodularity produces an increase in the left-hand side in response to a rise in U, which is accompanied by a decrease in the right-hand side due to between-policy submodularity. To reestablish equality between the two sides of the equation, μ_p must increase. In particular, for rather high generosity, Prop. 2 and 3 have shown that, when the optimal program never resorts to worker's search (for high-end generosities), worker's utility does not fall over time and the return of JS is linear in μ. Therefore, the concave closure of $\mathbf{V}$ delivered by SP is the one realized by the full randomization over expectations ($\mu_f = 0,\ \mu_p = 1$) and constant continuation utility over time ($U^u_{SP} = U$).

Government gains from information acquisition via profiling in two ways. First, by realizing an efficient match between policies and workers. And second, by providing (lower) search incentives only to workers who receive a 'Pass'. For high generosity levels, the gains from profiling exclusively originate from the first channel, as high-skilled workers are referred to Reemployment Services. This type of profiling is reminiscent of the WPRS program, where information

[18] See Lemma 1.

on human capital is used as a criterion to allocate services. On the contrary, for low generosity levels, information on human capital is used also to fine-tune transfers, like in REA program. Savings of this second type are even more relevant when no opportunistic behavior is possible for profilees and the program designer can thus index continuation utility to the profiling outcome, as now workers are possibly transferred to 'more generous' or 'less generous' programs based on their profiles (see Section 6).

6 Statistical Profiling

The presence of (ND) constraint in performance-based profiling prevents 'punishment' on workers who are referred to UI upon it. However, this restriction does not apply whenever human capital is assessed via statistical methods. In particular, some programs profile welfare claimants by conducting a background check and gathering observable data, and estimate a probability of reemployment in accordance with the information collected, on the base of a large number of past observations. Given that the outcome of statistical profiling does not rely on worker's commitment to it, the (ND) constraint does not apply in this new context. Consequently, the planner exploits the additional flexibility originating from the removal of 'no-punishment' restrictions as a leverage for incentive provision, with the target of reducing expected future transfers to recipients.

Profiling and reduction of transfers over time are two complementary instruments that open the way for sizable efficiency gains in the design of the optimal assistance program. Indeed, the planner now finds it optimal to index future transfers to the information detected during worker's profiling. Therefore, the contract of any profiling policy is not only consisting of the lottery odds of each outcome, but also of the schedule of continuation utilities depending on it.

Proposition 8 (Optimal Statistical Profiling). *When profiling refers workers to JS (for higher generosities), it is fully accurate. When, instead, profiling refers workers to UI (for lower generosities), the 'Pass' posterior is either 1 (i.e., full accuracy) or solves*

$$\mathbf{V}_\mu(\hat{\mu}, U_p^u) = \frac{\mathbf{V}(\hat{\mu}, U_p^u) - \mathbf{V}(0, U_f^u) + \mathbf{V}_U(0, U_f^u)(U_f^u - U_p^u)}{\hat{\mu}} \tag{13}$$

with $\hat{\mu} = \mu_p$ in AP and $\hat{\mu} = \mu_p'$ in SP/IP. As in case of performance-based profiling, $\hat{\mu}$ is increasing in generosity.

Proof. See Appendix C: Statistical Profiling. ■

The government sets different continuation utilities according to the profiling outcome. As shown in Prop. 2, the cost of incentive provision and effort compensation is increasing in generosity, which makes the marginal loss of higher generosity larger in UI than in SA. Hence, the government finds it optimal to lower the net discounted value of future payments upon 'Pass'. The result matches a characteristic of the actual REA program, where any worker who is found high-skilled is referred to minimum welfare support in the form of SNAP transfers up until reemployment. The criterion at the base of this rule is that any high-skilled worker does not need more generous transfers as she is likely to find reemployment soon.

The possibility to randomize over continuation utilities modifies the informativeness of the 'Pass' outcome. Eq. 13 strikes a new balance between incentive cost reduction of UI contract, the likelihood of being referred to it, and the new channel arising from the relaxation of the Incentive-Compatibility constraint.[19] Increasing informativeness, indeed, also increases the possibility of a 'Fail', conditional on which the planner pledges a larger utility. Hence, expected continuation utility for the agent is larger if the 'Pass' outcome is made more informative (and less likely) *ceteris paribus*, which allows the planner to further lower promised payments in order to restore contract efficiency (i.e., a binding (PK) constraint).

Proposition 9 (**Statistical Profiling Contracts**)**.**

- *If the policy sequence after 'Pass' includes UI, utility upon 'Pass' is lower than current utility. Otherwise, it remains constant.*
- *Unemployment benefits fall over time in IP, and remain constant otherwise. In particular, in IP benefits fall to a larger extent once workers receive a 'Pass'.*
- *In IP (resp., SP), the net wage upon reemployment is larger than (resp., equal to) current unemployment benefits.*

Proof. See Appendix C: Statistical Profiling. ■

Fig. 4 plots the patterns of policies, expectation, utility and unemployment benefits for a worker who enters the program with initial expectation of $\mu_0 = 0.95$ and promised utility of $U_0 = 22.4$. The worker is initially assisted in the search and profiled after 5 months. If she is found low-skilled, she is referred to SA with constant transfers. If, instead, she is found high-skilled, she is requested to search autonomously with transfers declining over time. As profiling is fully accurate and human capital entirely detected, any policy under either profiling outcome is absorbing.

[19] The first two forces where already at play in the problem with (ND) constraint (see Section 4).

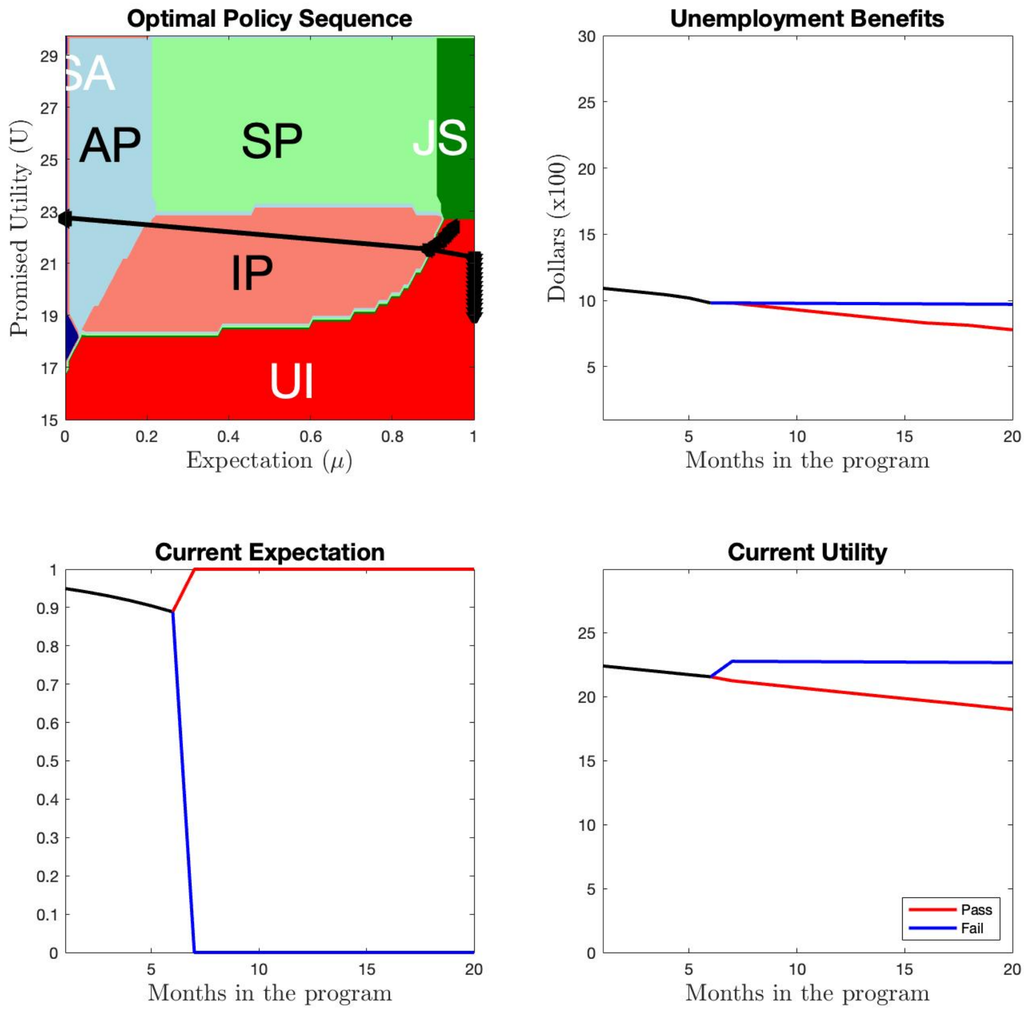


Figure 4: Consumption pattern upon profiling in a program with decreasing utility and $\mu_0 = 0.95$ and $c_0 = 100 \times exp((1-\beta)U_0) = \$939$.[20]

7 Quantitative Analysis

7.1 Parameterization

As anticipated in Section 1, many welfare programs worldwide combine UI benefits, profiling and job-search assistance, in the attempt to improve compliance to program requirements and the effectiveness of job search. In US, for example, two are the operating programs that profile workers: the Worker Profiling and Reemployment Services (WPRS) and the Reemployment and Eligibility Assessment (REA). WPRS is a federally-mandated program that supplies job-search assistance to welfare claimants who face a high risk of benefit exhaustion prior to reemployment. REA is, instead, a voluntary program each State can opt in, whose goal is to reduce fraud and

[20] Monetary values are scaled down by a factor of 100.

Parameter	Symbol	Value	Source
Preferences			
Discount Factor	β	0.9	
Search Effort Cost	e	0.27	various sources
Labor Market			
Job Search Hazard	$\{\pi_H, \pi_L\}$	{0.27, 0.14}	basic monthly CPS, y. 2019
Net Wage	$\{c^w_H, c^w_L\}$	{$2,498, $1,128}	Poe-Yamagata et al. (2011)
Wage Tax	$\{\tau_H, \tau_L\}$	{$178, -$224}	EIC, FICA
Assisted Search			
Administrative Cost	κ^{ja}	$430	Balducchi and O'Leary (2018)
Worker Profiling			
Administrative Cost	κ^{wp}	$50	Poe-Yamagata et al. (2011)
REA programs (FL, ID, IL, NV)	c^i,		
Generosity (consumption equivalent)	$i = FL, ID, IL, NV$	[$1,350,$2,301]	Nicholson and Needels (2011)

Table 2: Choice of Parameters Value

fund misallocation by excluding from UI benefits those recipients who either do not conduct any search activity, or do not need any form of welfare support (because they are highly re-employable). In other words, REA and WPRS differ in the use of information they collect with profiling, as with REA efficiency gains realizes via reduction of transfers, whereas with WPRS by implementing the job-search with proper methodology. To meet their target, both programs conduct an in-depth assessment of individual skills, based on which workers receive job-counseling, learn how to develop a resume and/or are directly referred to employers (see Manoli et al., 2018). Moreover, neither program allows workers enrolled in an employment or training program to access any of these services.

Poe-Yamagata et al. (2011) conduct a cost-benefit analysis of the REA programs in Florida, Idaho, Illinois and Nevada, which assisted a total of 134,550 claimants in 2009. Of all claimants, 58% were men, 66% were white and 13% black. The report distinguishes between high and low skilled workers. The weighted mean share of high skilled participants is 48%.[21]

Turning to to the choice of parameters (see Table 2), the parameters to be chosen are: the functional form of period utility ($u(.)$), the discount factor (β), the effort cost of searching (e), the on-the-job productivity (i.e., the gross wage) and reemployment hazard rates of high- and low-skilled workers ($\{\omega_h, \pi_h\}_{h\in\{H,L\}}$), and the cost of administering profiling (κ^{wp}). The unit of time is set to one month.

I use a logarithmic specification of utility and set the monthly discount factor equal to $\beta = 0.9$. Based on Pavoni et al. (2013), the working effort cost is 49% of the consumption equivalent for men and 62% for women, corresponding respectively to $\bar{e}^m = 0.67$ and $\bar{e}^w = 0.97$ given the logarithmic specification[22]. And given that the percentage of male participants within the four

[21]The relative weight assigned to each State depends on the number of participants it assisted. In 2009, Florida, Idaho, Illinois and Nevada supplied UI to 80,531, 18,156, 3,112 and 32,751 jobless workers, respectively (Poe-Yamagata et al., 2011). The report does not make a distinction between high- and low-skilled workers in Illinois. However, this is not a source of major concern, given the small number of welfare recipients in the State.

[22]Logarithm allows for separation of consumption utility from working disutility in a natural way, according

programs is 58%, the working effort cost of the average participant amounts to $\overline{e} = 0.58\overline{e}^m + 0.42\overline{e}^w = 0.8$. Krueger and Muller (2010) conduct an analysis on the cost of search effort based on the American Time Use Survey (ATUS) and find that jobseekers spend on average 160 minutes every day looking for a job. Following Pavoni et al. (2013), I target the search effort to 1/3 (160/480) of the working effort, hence $e = 0.8/3 = 0.27$. The value is consistent with Pavoni et al. (2013), who estimate a cost of effort of $e = 0.22$.

Poe-Yamagata et al. (2011) reports data about net wages earned in the last 10 quarters prior to the start of UI claim. Quarterly wages in all States display a hump-shaped pattern, which increases until it reaches a peak three quarters before displacement and steadily declines later on. The decline is consistent with the *Ashenfelter's dip*, suggesting that wages fall in the pre-layoff period (Ashenfelter, 1978). Preventing this effect from distorting estimates requires to exclude the last three quarters of pre-layoff wage. However, the paper does not consider human capital depreciation along the unemployed spell, which is instead well documented by the empirical literature (Keane and Wolpin, 1997; Neal, 1995) and requires to lower the last wage, in accordance with the duration of unemployment spell. As the two effects tend to offset each other, I simply consider the wage earned in the last quarter. As a consequence, the monthly net wages of Florida, Idaho and Nevada are \$1,833, \$1,367 and \$1,900, respectively.[23] The report, however, does not distinguish between wages of high- and low-skilled workers. Thus, I exploit the cross-sectional variation in wages and the share of high-skilled participants across States. Given that there are two unknowns and three States, I compute $\{c_H, c_L\}$ as the pair that minimizes the loss function

$$\Lambda(\hat{c}_H, \hat{c}_L) = \sum_{i=1}^{3} \varphi_i(\theta_i \hat{c}_H + (1-\theta_i)\hat{c}_L - c_i)^2, \quad i = \{FL, ID, NV\}$$

with φ_i being the fraction of all welfare recipients in country i. The computation delivers monthly wages equal to $c_H^w = \$2,498$ and $c_L^w = \$1,128$. In order to compute their gross counterpart, I reverse engineer the gross labor income by computing the tax and deductibles that led to net amounts. In US, employees are subject to the Federal Insurance Contribution Act (FICA) tax, which is comprehensive of Social Security and Medicare tax. FICA tax is a net payroll tax which is levied half on employers and half on employees, and amounted to 15.3% of Adjusted Gross Income (AGI) in 2009. Moreover, taxpayers with an AGI lower than a certain amount, that depends on their marital status and number of children, are entitled to an Earned Income Credit (EIC). Since no data on the marital status or the number of children of recipients is available, I assume that the representative recipient is married and has two children. Under

to the formula

$$log((1-\xi)c) = \log(c) + \log(1-\xi) = \log(c) - \overline{e}$$

with $\xi \in \{0.49, 0.62\}$ being the consumption equivalent of working disutility.

[23] Poe-Yamagata et al. (2011) does not report the percentage of high-skilled recipients in Illinois, which makes their data on wages useless for the estimation of $\{c_H, c_L\}$.

2009 FICA and EIC tax schemes, fiscal neutrality for a married couple with two children is achieved at a gross annual income of $26,250, with the couple paying a tax (resp., receiving a subsidy) for an income above (resp., below) that threshold. Therefore, low-skilled recipients, whose net annual income is $13,536, receive a tax credit under EIC, making their gross income lower than the net one, and precisely equal to $10,844. High-skilled recipients, instead, have a gross income of $32,112 and a net one of $29,976.[24] Therefore, monthly gross wages are equal to $\omega_H = \$2,676$ and $\omega_L = \$904$.

I estimate the hazard rates $\{\pi_H, \pi_L\}$, using data from the basic monthly Current Population Survey (CPS). Following the method-of-moments estimation, the probability of reemployment after t periods is computed as the fraction of workers who exit unemployment at that time. Reemployment probabilities are chosen as the ones that minimize the distance between the probabilities of reemployment so computed and the expected hazard rates, with weights given by the fraction of high- and low-skilled workers in the sample (for a more detailed description, see Appendix E: Estimation of hazard rates). The estimated monthly hazard rates are $\pi_H = 0.27$ and $\pi_L = 0.14$. The assumption that the worker can exit unemployment only upon search is quite extreme. I therefore assume the rate of reemployment in case no search is conducted either by the worker or by the government to be equal to half the after-search rate of low-skilled workers, i.e. $\hat{\pi} = 0.07$. A positive hazard rate in case of no search has a positive impact on the return of passive labor-market policies, like SA and AP, and a negative impact on the return of effort-incentivizing ones, like UI and IP, due to the increase in incentive costs. The value of off-the-equilibrium zero-effort action, indeed, is higher and the incentive constraint now requires satisfying a tighter condition

$$U^w - U^u \geq \frac{e}{\beta(\pi(\mu) - \hat{\pi})}$$

Passing to the choice of κ^{wp}, the estimates of average per-capita cost of REA in 2009 contained in the report range from $12 (Idaho) to $134 (Illinois) and include cost of personnel and operative costs of centers supplying REA services (e.g., State Workforce Agencies and One-Stop Career Centers). I, therefore, set the administrative cost of profiling equal to the weighted average of REA per-capita cost among the four State programs, that is, $\kappa^{wp} = \$50$. Instead, the cost of assisting any worker in the job search is based on Balducchi and O'Leary (2018), who estimate $\kappa^{ja} = \$430$. Such a figure is consistent with other estimates ($\kappa^{ja} = \$500$ in Pavoni et al. (2016)), as well as cost estimates of programs that perform different activities (for instance, search monitoring), but feature a similar set of operations (regular meetings with personnel at One-Stop Career Centers, phone calls to employers, etc.). For instance, Pavoni and Violante

[24]The net annual income of high- and low-skilled workers is $\$2,498 \times 12 = \$29,976$ and $\$1,128 \times 12 = \$13,535$, respectively. Low-skilled workers pay $1,659 under FICA, i.e. the 15.3% of their gross income, but receive $4,350 under EIC, hence receiving an annual subsidy of $2,691. High-skilled workers, instead, pay a FICA tax of $15.3\% \times \$32,112 = \$4,913$, and are given a tax rebate of $2,774, that account for an annual tax of $2,139.

(2007) estimate a monthly cost of monitoring of \$478 per claimant.

The generosity of any program depends both on the amount of flow endowments and the duration. Poe-Yamagata et al. (2011) collects data about the average maximum and weekly benefit in each State, as well as the distribution of benefit duration among participants. The weekly benefit amount ranges from \$234 in Florida to \$299 in Nevada, suggesting a substantial variability in generosity of State programs. In the US, unemployment benefits are paid under four distinct schemes, which are activated in succession, depending on the current labor market situation of each US State. Unemployment Insurance (UI) benefits last up to 26 weeks in all States. Workers who are still unemployed at the end of the 26th week, are entitled to additional 53 weeks under the Emergency Unemployment Compensation (EUC) scheme. Moreover, States pay additional benefits up to 20 more weeks under the Extended Benefits (EB) scheme, if their unemployment rate exceeds 8.5%, which was the case for all four States in 2009. Exhaustees of UI, EUC and EB are finally referred to the Supplemental Nutrition Assistance Program (SNAP). This constitutes the typical instance of a purely income-support measure of last resort, consisting of a constant allowance for the purchase of food, with no eligibility assessment or time limit. Transfers decline over time, as claimants move from one program to another. WPRS and REA never profile workers after they have exhausted UI, EUC and EB, as no assisted search or further transfer reduction is possible once the worker enters SNAP. I assume that workers who are entitled to 26 weeks of regular UI benefits are assisted under EUC and EB programs for the whole prospective duration of the programs, i.e. 73 weeks, and that exhaustees who are still unemployed at the end of UI+EUC+EB receive an endowment from the Supplemental Nutrition Assistance Program (SNAP), which replaced the Food Stamps Program in 2008. Average total payment was \$7,930 under EUC and \$3,844 under EB (Nicholson and Needels, 2011), hence constituting a monthly endowment of $c^{EUC/EB} = \$645$, while a family of four people was receiving a \$501 monthly benefit from SNAP.[25] The program's generosity for each of the four States is computed backward from the moment the welfare recipient enters into SNAP or finds reemployment, up until the first month when she receives regular UI benefits. Worker's utility of reemployment in case she is high-(resp., low-)skilled amounts to[26]

$$U_H^w = \frac{u(c_H^w)}{1-\beta} = \frac{\log(24.98)}{1-0.9(1-)} = 32.18 \quad U_L^w = \frac{u(c_L^w)}{1-\beta} = \frac{\log(11.28)}{1-0.9} = 24.23$$

while in SNAP with no search it is equal to

$$U_{e=0}^{SNAP}(\mu) = \frac{u(c^{SNAP}) + \beta\hat{\pi}[\mu U_H^w + (1-\mu)U_L^w]}{1-\beta(1-\hat{\pi})} = 22.32\mu + 19.25(1-\mu)$$

[25] See SNAP Data Tables at the following link: https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap.

[26] All monetary amounts are normalized so that 1 consumption unit corresponds to \$100.

Condition $U_L^w > U_{e=0}^{SNAP}(0) + \frac{e}{\beta(\pi_L - \hat{\pi})}$ implies that the worker always finds it convenient to search also during SNAP. Hence, the value of SNAP can be rewritten as function of entry expectation μ

$$U^{SNAP}(\mu) = \mu \frac{u(c_{SNAP}) - e + \beta \pi_H U_H^w}{1 - \beta(1 - \pi_H)} + (1 - \mu) \frac{u(c_{SNAP}) - e + \beta \pi_L U_L^w}{1 - \beta(1 - \pi_L)}$$

If the worker is entitled to regular UI, EUC and EB, then her assistance program lasts for 26+53+20=99 weeks, that is, around 25 months. Starting from the last month, the following recursion is implemented

$$U_{i,j,t} = u(c_t^j) - e + \beta[\mu_i^{t-1} \pi_H U_H^w + (1 - \mu_i^{t-1}) \pi_L U_L^w + (1 - \pi(\mu_i^{t-1})) U_{i,j,t+1}], \ 1 \leq t \leq 25,$$
$$j = \{FL, ID, IL, NV\}, \ i = \{< HS, HS, < CD, CD, GD\}$$

with $U_{i,j,26} = U^{SNAP}(\mu_i^{26})$, j indexing States and i indexing education. The initial probability of being high-skilled, μ_i^0, equals the share of high-skilled individuals with same educational attainment, θ_i. The generosity levels of each program and educational attainment, expressed in consumption-equivalent terms,[27] are reported in Table 3. Unsurprisingly, the generosity of the program is increasing in the level of educational attainment, due to higher initial expectations and $U_H^w > U_L^w$. Among the four States, Illinois (resp., Idaho) is the most (resp., least) generous one for all levels of education.

States	Less Than HS	HS Diploma	Some College	College	Graduate
Florida	\$1,350	\$1,536	\$1,580	\$1,748	\$1,811
Idaho	\$1,141	\$1,282	\$1,315	\$1,440	\$1,487
Illinois	\$1,666	\$1,920	\$1,981	\$2,212	\$2,301
Nevada	\$1,362	\$1,550	\$1,595	\$1,763	\$1,827

Table 3: Program generosity for any State and educational level (consumption equivalent).

7.2 Optimal REA Program

Workers are assessed via in-person interviews with REA staff. When issuing the call for the interview, States target those who are predicted to be likely to exhaust their UI benefit. The assessment is based on interviews, that last 45 min on average. After the interview, workers profiled as high skilled suffer a reduction of unemployment benefits. Therefore, the profiling methodology adopted in the REA program of statistical type, as it allows 'punishment' on recipients, based on their skills.

[27] The expression of consumption equivalent of U is

$$c(U) = \exp((1 - \beta)U)$$

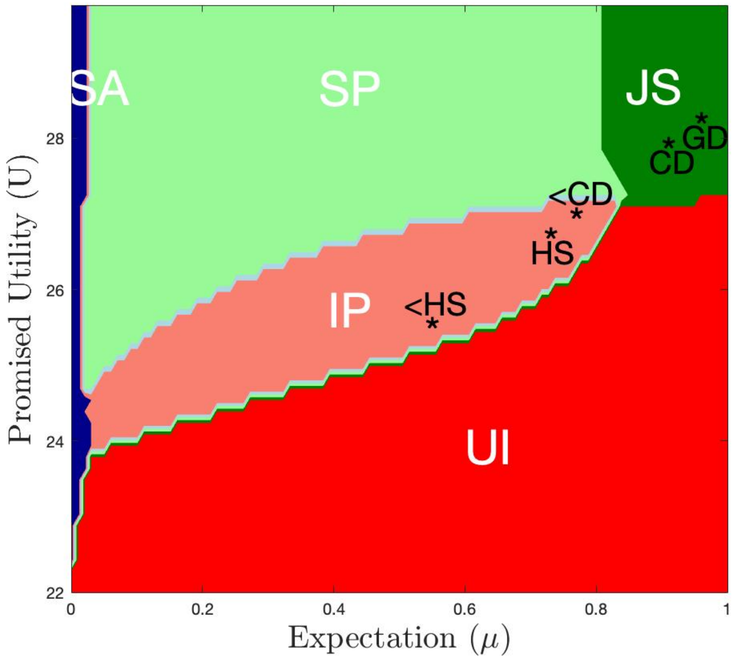


Figure 5: Optimal Policies for Florida's REA Recipients.
Note: <HS=Less Than High School, HS=High School Diploma, <CD=Some College, CD=College Degree, GD=Graduate Degree

Fig. 5 reports the optimal policies in the state-space of programs' generosity and initial expectation, and locates Florida's REA recipients over it according to their educational attainment.[28] Recipients with less than a college degree should better search and at the same time undergo profiling right upon entry into the program, while those with a higher educational attainment should be initially assisted in the search but not profiled. Quite surprisingly, the ones who are assisted in the search are those recipients whose search would have a larger expected return. The reason is that a larger $\pi(\mu_0)$ is correlated with larger consumption upon reemployment and so higher implicit utility U_0. Hence, graduates' effort is too expensive to compensate for the government.

7.3 Welfare Gains

A relevant question for policymakers is how large savings they can realize from designing an optimal profiling program. In order to estimate such value, I compare two distinct programs, one with only welfare policies SA, UI and JS ($\mathcal{W}$) and the other encompassing all six policies ($\mathcal{P}$). Fig. 6 reports the optimal patterns of promised utility, unemployment benefits and wage taxes/subsidies in the two programs for Florida's jobseekers with a high school degree (i.e., the most numerous group, accounting for 54% of all recipients in Florida in 2009), whose initial expectation and promised utility (in consumption equivalent terms) are $\mu_0 = 0.72$ and $c(U_0) =$

[28]The initial generosity of REA programs in the other three States is quite similar.

\$1,436, respectively. As shown in Fig. 5, this group of jobseekers are profiled under IP right after entering $\mathcal{P}$ and referred to UI and SA, while in the program with no profiling they never exit UI. The pattern of promised utility and unemployment benefits is thus falling over time only for high-skilled workers in $\mathcal{P}$ and for both high- and low-skilled ones in $\mathcal{W}$. However, the larger incentive costs in the latter case are conducive to a steadier decline in both benefits and utility. The reemployment tax displays a monotone increasing pattern in $\mathcal{P}$ (in the weak sense for low-skilled workers) due to the declining promised utility. In $\mathcal{W}$, instead, this component is contrasted by declining expected productivity (and gross wage thereof) of workers and increasing incentive costs, as expectations are revised downward. For this reason, the pattern of expected wage tax upon reemployment is non monotonic in time.

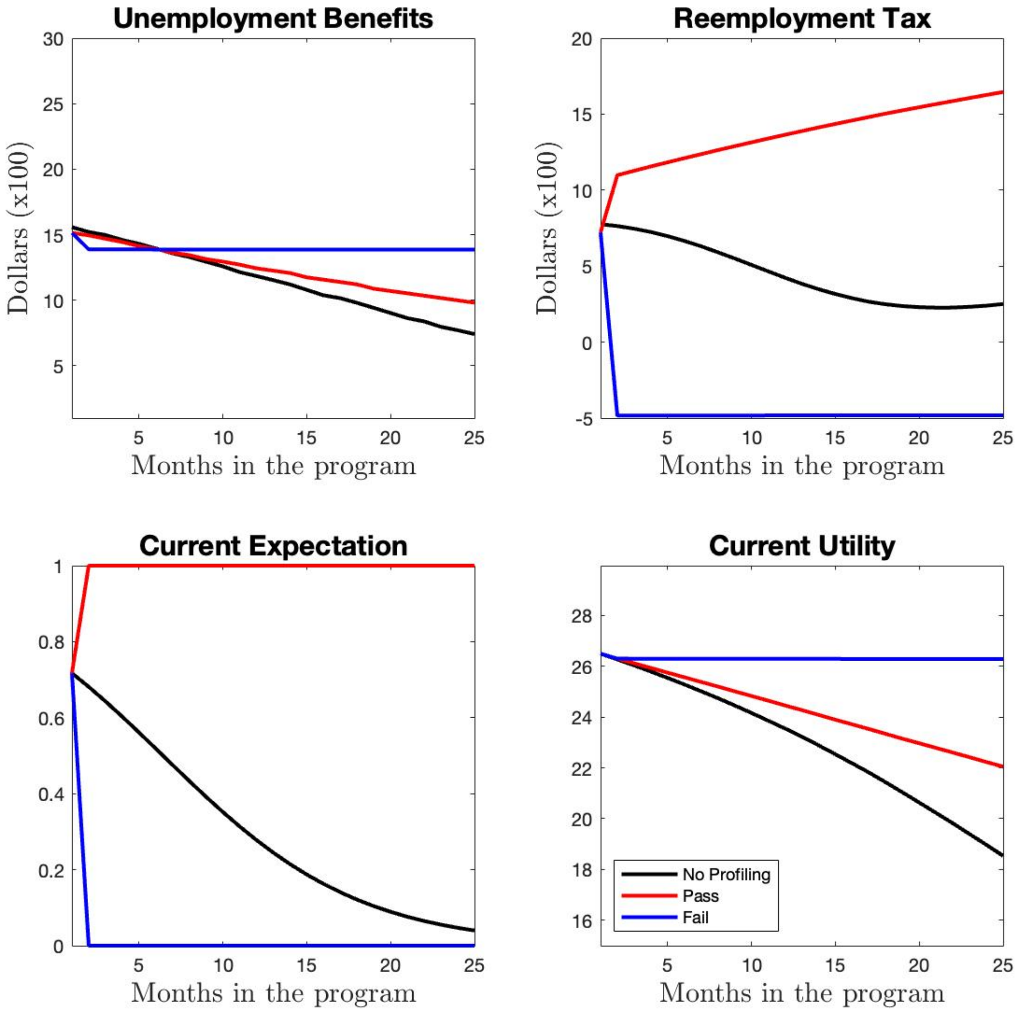


Figure 6: Optimal REA program of Florida for recipients with a high-school diploma over a 25-month horizon (UI+EUC+EB). Initial expectation and generosity are $\mu_0 = 0.72$ and $c(U_0) =$ \$1,436, respectively.

Table 4 reports the per-capita welfare gain of profiling for each educational group. Given the relative size of each group and the annual number (80,531) of REA recipients, the aggregate welfare gain of Florida in 2009 would have been \$1,817,200. This figure is consistent with the estimate provided by Poe-Yamagata et al. (2011), who estimate net per-capita savings of \$47 for UI recipients and overall savings of \$3,784,957.[29]

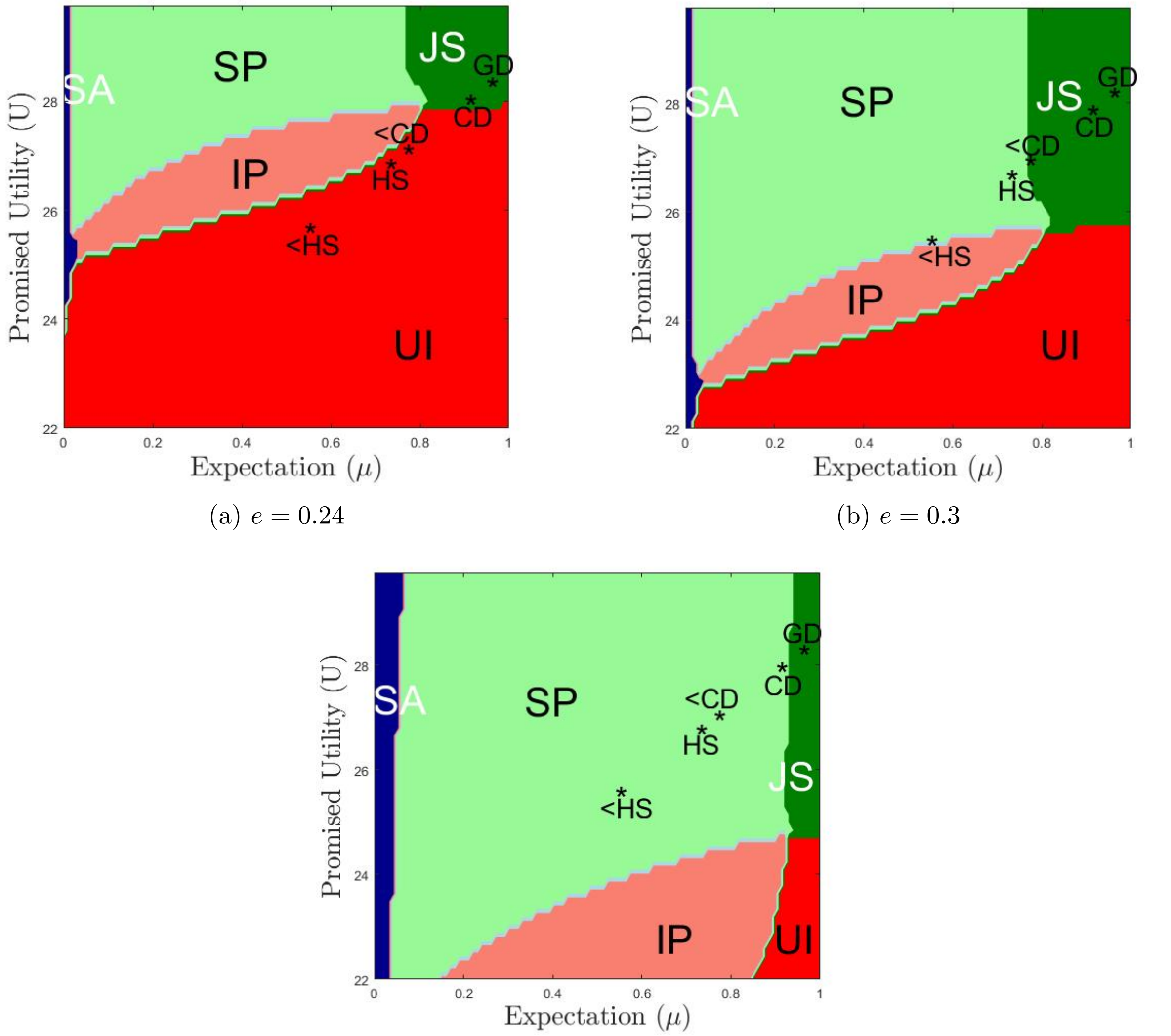
	Less Than HS	HS Diploma	Some College	College	Graduate
Perc. in Program	0.13	0.54	0.17	0.12	0.04
μ_0	0.54	0.72	0.76	0.9	0.95
$c(U_0)$	\$1,275	\$1,436	\$1,474	\$1,617	\$1,671
Per-cap. Welfare Gain	\$14.5	\$29.3	\$26.2	\$3.2	\$0.5

Table 4: Welfare Gains of Profiling per Education Group in Florida (y. 2009)

7.4 Robustness Checks

A relevant dimension on which workers display a large heterogeneity is effort cost. Various studies estimate different costs between men and women (Attanasio et al., 2008; Eckstein and Wolpin, 1989). In addition, they document the existence of a work-effort cost, which is not accounted for in the baseline model of this paper. I therefore conduct a robustness check by allowing for the search-effort cost to vary by ±10% with respect to the baseline value. Second, I assume that the reemployed worker incurs a working cost equal to the search-effort one. Optimal policies in the (μ, U) space for $e = 0.24$, $e = 0.3$ and $e^w = e = 0.27$ are reported in Fig. 7. When the search-effort cost is lower, then search-delegating policies (UI and IP) expand their areas at the expense of assisted-search ones (JS and SP). The opposite occurs when the search-effort cost is larger and all groups of recipients are offered search assistance upon entry. Positive working-effort cost $e^w = 0.27$ produces a comparative disadvantage for active labor-market policies, as reemployment (and effort compensation upon it) is less likely in SA. Therefore, the gain from reallocating workers across SA and JS is larger the higher is the working-effort cost, via relaxation of the Promise-Keeping constraint. For this reason, SP replaces JS for high-end generosities and expectations (see Fig. 7c).

[29]Poe-Yamagata et al. also estimate savings for UI+EUC recipients. However, the authors do not apply any time discount, and this delivers an inflated estimation of net gains. For this reason, I consider only per-capita savings realized on recipients of UI benefits, which come first in chronological order.



(a) $e = 0.24$

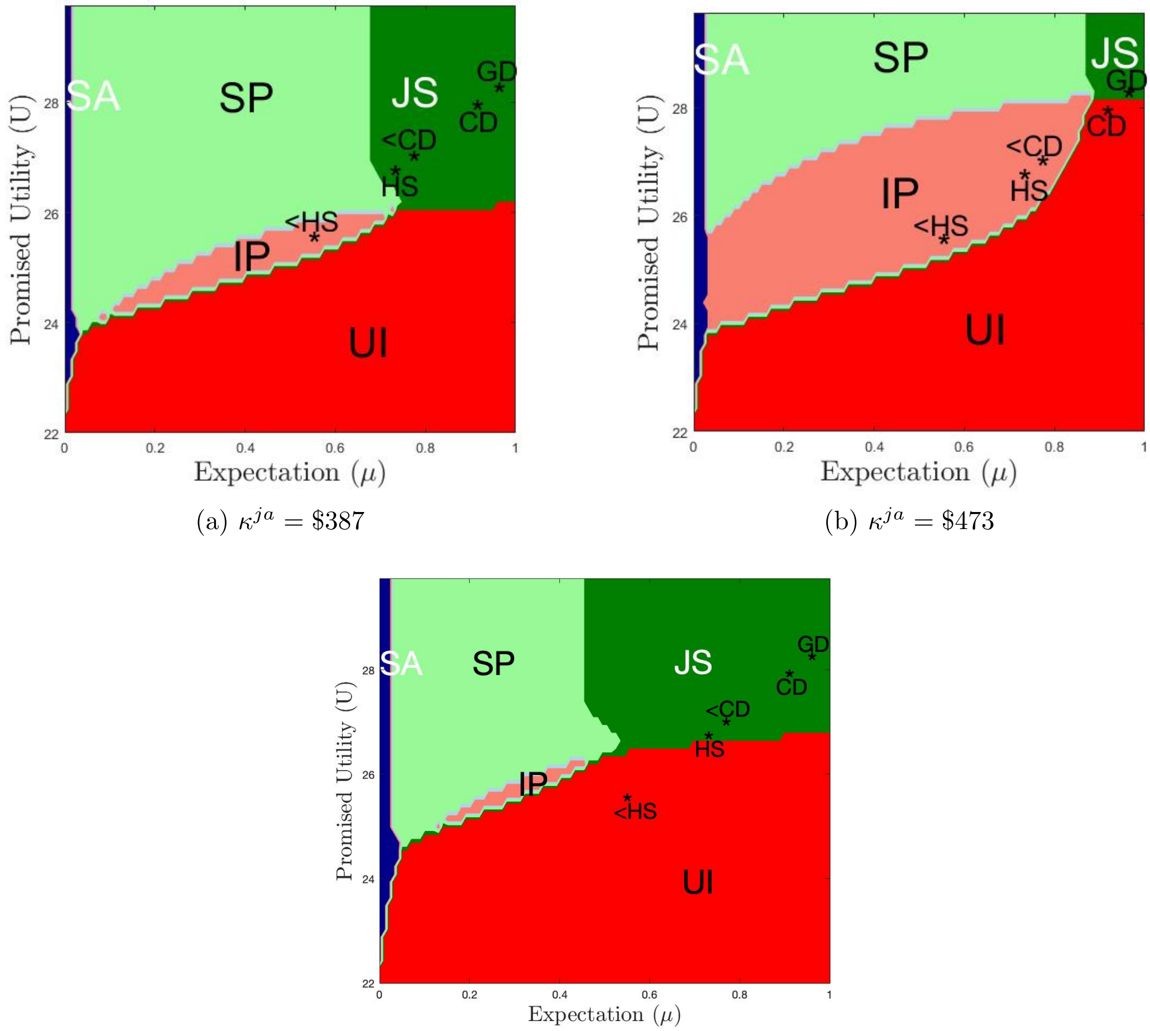
(b) $e = 0.3$

(c) $e^w = e = 0.27$

Figure 7: Optimal Policies over the (μ, U) Space for Different Effort Costs

A second parameter displaying great variability is the cost of assisted search and profiling technologies. Poe-Yamagata et al. (2011) estimate that in Nevada, which merges reemployment services with profiling, κ^{ja} equals \$148. Such a figure is consistent with past estimates of administrative costs of assisted search. For example, Pavoni et al. (2013) compute an average cost of \$150 per person. Fig. 8a and 8b report the state space of policies for $\kappa^{ja} = \$387$ (-10%) and $\kappa^{ja} = \$473$ (+10%). When assisting workers is less costly, JS and SP are optimal also for lower generosities, and the opposite occurs when job-search assistance is more costly and incentivizing workers more convenient also for higher effort-cost compensation.

The cost of profiling varies according to the design of the REA program. Given that all possible levels of accuracy are allowed, I select the largest cost among all four States, i.e. $\kappa^{wp} = \$134$ (Illinois). As a consequence, the areas of SP and IP shrink in favor of the respective welfare policies, JS and UI (see Fig. 8c).



(a) $\kappa^{ja} = \$387$

(b) $\kappa^{ja} = \$473$

(c) $\kappa^{wp} = \$134$

Figure 8: Optimal Policy Space for Different Costs of Assisted Search and Worker Profiling

8 Private Search and Moral Hazard

The government may be unable to observe worker's actions. In particular, worker's search may be private and thus unobservable to it. This may create a misalignment of expectations between the two parties, whenever the agent shirks effort and derives no new information about her permanence in unemployment, while the principal assumes that this was the result of a failed search and revises its expectation accordingly. The next period's contract provides larger incentives, consistently with the failed search hypothesis. Therefore, as the worker retains a more optimistic expectation than the on-equilibrium one, she also expects larger transfers. As a consequence, with private search the worker derives an additional advantage from shirking job search, other than saving on effort. To contrast it, the planner promises larger transfers upfront in case of re-employment. These high-powered incentives enlarge in the prospective duration of private search, following the increase in the number of possible deviations from recommended effort by the agent, and are driven to zero if the next-period contract does not request private search. Indeed, the agent can benefit from deviation only if a dispersion in promised transfers exists in the two alternative scenarios of reemployment and unemployment.

In consequence to the possibility of a covert misalignment of expectations, the planner incurs payment of *learning rents* to induce agent's search. The value of UI and IP thus incorporates such rents. Starting with UI, the problem of the planner when private search ends in the next period is

$$
\begin{aligned}
V_1^{UI}(\mu, U) = & \max_{c^{ui}, U^w, U^u} -c^{ui} + \beta\big[\pi(\mu)W(\mu, U^w) + (1-\pi(\mu))\hat{V}(\mu', U^u)\big] \\
\text{sub:} \quad & \hat{V}(\mu, U) := \max_{i \in \{SA, JS, SP, AP\}} V^i(\mu, U) \\
& U = u(c^{ui}) - e + \beta\big[\pi(\mu)U^w + (1-\pi(\mu))U^u\big] \qquad \text{(PK)} \\
& U \geq u(c^{ui}) + \beta U^u \qquad \text{(IC)}
\end{aligned}
$$

The only difference with the non-contractible effort case is a restriction on the basket of policies to choose among in the next period, in order to be consistent with the current provision of learning rents. The dispersion in continuation utilities is the same as in (6).

Passing to the case of longer UI duration, define $T(\mu, U)$ as the duration of UI, for any worker with initial expectation μ and utility U

$$T(\mu, U) := \inf\big\{n : \, i(\mu_n, U_n) \neq UI\big\} \tag{14}$$

where $\mu_n := \mu^{(n)}(\mu)$ (resp., $U_n := U_n(U)$; resp., $i(\mu_n, U_n)$) is defined as the expectation (resp., continuation utility; resp., policy) after failing the job search for n periods. Next step is to design a contract which is robust to any possible deviation from $t = 0$ (today) to $t = T$ periods ahead. The worker could deviate in the first, second, ... T-th period after being assigned to UI. But she may also decide to shirk multiple times, possibly not successive, before reverting to job search, or even shirk forever after. For this reason, the design of a robust contract is in principle a complicated task. The following holds.

Proposition 10. *Any contract incentivizing search effort for T periods is robust against any possible deviation from the sequence of efforts, whenever it is robust against one-shot deviations* *from that sequence.*[30] *Therefore, IC constraint when implementing UI for T periods reads*

$$U \geq u(c^{ui}) + \beta\big[U^u + \Lambda(T,\mu)\big]$$

with $\Lambda(T,\mu)$ defined by the recursion

$$\begin{cases} \Lambda(1,\mu^{T-1}) = 0 \\ \Lambda(T-j,\mu^j) = \Big(\frac{\pi(\mu^{(j)})}{\pi(\mu^{(j+1)})} - 1\Big)e + \beta\pi(\mu^{(j)})\Big(\frac{1}{\pi(\mu^{(j+1)})} - 1\Big)\Lambda(T-j-1,\mu^{(j+1)}), \quad 0 \leq j \leq T-2 \end{cases} \tag{15}$$

which thus is (i) independent of U^u, (ii) null in $\mu \in \{0,1\}$ and/or $T = 1$, and (iii) increasing in T.

Proof. See Appendix D: Moral Hazard. ■

The dispersion in utilities between re-employment and unemployment now reads

$$U^w - U^u = \frac{e + \beta\Lambda(T,\mu)}{\beta\pi(\mu)}$$

The gap between U^w and U^u, which proxies the cost of incentives, is increasing in Λ. If UI is implemented for $t < T$ periods, then learning rents are lower and the planner incurs a lower contract cost.

Insurance-and-Profiling (IP). When profiling is adopted jointly with private search, the planner's problem reads

$$\begin{aligned} V^{IP}(\mu,U) = \max_{c^{ip},U^w,(U^u_r)_{r=\{p,f\}},\mu_p} & -c^{ip} - \kappa^{wp} + \beta\big[\pi(\mu)W(\mu,U^w) + q(1-\pi(\mu_p))V(\mu'_p,U^u_p) + \\ & + (1-q)(1-\pi_L)V(0,U^u_f)\big] \\ \text{sub:} \quad & U = u(c^{ip}) - e + \beta\big[\pi(\mu)U^w + q(1-\pi(\mu_p))U^u_p + (1-q)(1-\pi_L)U^u_f\big] \\ & U \geq u(c^{ip}) + \beta\big[q(\hat{\Lambda}^{i(\mu_p,U^u_p)}(\mu_p) + U^u_p) + (1-q)(\hat{\Lambda}^{i(0,U^u_f)}(0) + U^u_f)\big] \\ \text{with:} \quad & \hat{\Lambda}^i(\mu) = \begin{cases} \Lambda(t,\mu), & \text{if } i = (UI,t) \\ 0, & \text{otherwise} \end{cases}, \quad \text{(MP)} \end{aligned}$$

Any worker can be profiled at any stage of the unemployment spell, possibly multiple times. If IP is designed to fully reveal the underlying state, the worker is certain to be high-skilled

[30]This result is reminiscent of the way Euler equations are derived. Indeed, Euler equations are conditions imposed on the path of controls (consumption, investment, etc.), which guarantee that the decision maker is never willing to select any different path lying in the feasibility set and differing from the optimal one in one period only. The same property holds for Nash equilibrium strategies in repeated games, which are so if robust to deviations at any single node of the game tree.

after receiving a 'Pass' and low-skilled otherwise. Hence, she does not revise her expectation henceforth, even if she fails to exit unemployment in the next stages of the welfare program. In other words, the policy she is assigned to under either profiling outcome is absorbing. In case profiling is not designed as a perfect signal, instead, the worker who passes it and is referred to any active policy can downward revise her expectation and reenter into IP at a later stage (unless she escapes unemployment in the meantime).

Proposition 11. *When worker's search is private, accuracy of profiling under IP is determined by the need to reduce learning rents, as the 'Pass' posterior is either 1 or solves*

$$V_{\mu}(\mu_p', U_p^u) = \frac{V(\mu_p', U_p^u) - V^{SA}(U_f^u) + V_U^{SA}(U_f^u)(U_f^u - U_p^u)}{\mu_p'} + \left[V_U^{SA}(U_f^u) - W_U(\mu, U^w)\right] \frac{\mu_p \Lambda_{\mu}(t, \mu_p) - \Lambda(t, \mu_p)}{\mu_p'} \tag{16}$$

Proof. See Appendix D: Moral Hazard. ■

The result sheds light on the complementarity between profiling and private search. Indeed, in the case of AP and SP, the government selected the upper posterior more (i.e., 'Pass') by equating the marginal gain of higher informativeness and the marginal cost of lower frequency, which led to the possibility of recommending search also to also a fraction of low-skilled workers. Now, a further component, namely the reduction of learning rents, drives the choice of the upper posterior, in addition to incentive cost reduction.

9 Conclusions

This paper provides an estimate of the welfare gains that can be obtained in programs of unemployment assistance via profiling of recipients. The rationale for embedding profiling into a welfare program stems from the difficulty of inferring recipients' job-finding skills and on-the-job productivity. At the optimum, active labor-market policies and workers' expectations about personal skills and productivity are complementary. Workers who are likely to be low-skilled are thus provided income support only, while those who have moderate or high expectations of being high-skilled are supplied with job-search assistance or search incentives, which come in the form of lower wage taxes or higher wage subsidies. Looking at the dimension of program's generosity, instead, search-incentivizing policies are adopted for low-end generosity, while search-assistance ones are adopted for high-end generosity, due to increasing costs of effort compensation. This causes the dynamic of worker's utility which is implicit in the stream of payments to be decreasing along the spell whenever incentives to worker's search are provided.

The effects of implementing worker profiling within the program divide into gains and losses. The gains from workers' profiling stem from incentive alignment between workers and the government. Indeed, rather than pooling into the same policy and contract both high- and low-skilled workers with equal expectations, profiling allows to refer them to the proper job-search method so to minimize the cost of the program. However, worker profiling entails also a loss for the government on those workers whose expectations are positively biased before being profiled as low-skilled. This loss may be conducive to partial detection of hidden skills aimed at strategic persuasion of (a fraction of) low-skilled workers in the sense of Kamenica and Gentzkow (2011). The deep reason at the base of partial detection of skills is that low-skilled workers are sufficiently productive and the generosity of the program sufficiently low that the government would rather compensate them for searching than putting them at rest. Therefore, some low-skilled workers might better receive a boost of expectations and keep searching for a job, rather than staying inactive, even at the cost of higher incentive payments to high-skilled workers.

The profiling outcome is matched by a fine-tuning of payments, whenever followed by search-incentivizing policies. In particular, an optimal program promises lower transfers to recipients who are profiled as high-skilled and required to search for job, due to agency costs increasing in the level of promised utility. This result, which features actual REA programs, should be accompanied with a decreasing-in-time pattern of unemployment benefits, as opposed to the constant subsidy under SNAP.

Some questions remain unanswered. The main shortcoming of the paper is constituted by the assumption on costs and accuracy of profiling. The actual per-capita cost of profiling depends on the accuracy with which skills are detected. A more accurate detection, indeed, leads to a more expensive profiling process (e.g., longer in-person interviews, more elaborate tasks to perform). In addition, any actual profiling program, as well as any sort of tests aimed at detecting a hidden characteristic, contains a given amount of noise that impedes an exact detection of skills. Assuming (i) the cost of profiling to vary in accordance to the change induced on the initial expectation (i.e., known as *entropy cost* in the information design literature),[31] and/or (ii) the accuracy of information detection to be upper bounded, might lead to different estimates of the value of worker profiling. In this sense, the welfare gain computed in Section 7 can be read as an upper bound on the return from adopting optimal profiling.

A second aspect on which further inspection is required is the absence of any information asymmetry. Both parties are assumed to share the same initial expectation, as claimants truthfully report their personal data to the provider at the beginning of the program. However, if claimants could conceal their personal information, those with high expectations in reemployment would anticipate being requested to search and choose to misreport it,[32] so as to benefit from larger

[31] For a description of costs of information detection that allow the concavification result to survive, see Kamenica and Gentzkow (2014).

[32] In no other case they would find convenient to lie, as all contracts other than incentive-providing ones are independent of expectations.

incentives. If this is the case, the government would need to make search-incentivizing contracts robust to information misreporting and this would further exacerbate the problem of incentive provision. In this sense, as noticed in the case of private worker search (Section 8), profiling might constitute a way for the government to curb the information rents that originate from the agency problem.

References

(1) Ashenfelter O., 1978, ”Estimating the Effect of Training Programs on Earnings”, Review of Economics and Statistics, 60, 47-50;

(2) Atkenson, A., Lucas R. E. J., 1995, ”Efficiency and Equality in a Simple Model of Efficient Unemployment Insurance”, Journal of Economic Theory, 66(1), 64-88;

(3) Bloedel A. W., Segal I., 2018, ”Persuasion with Rational Inattention”, available at SSRN: https://ssrn.com/abstract=3164033;

(4) Boleslavsky R., Kim K., 2021, ”Bayesian persuasion and moral hazard”, *mimeo*;

(5) Card D., Kluve J., Weber A., 2018, ”What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations”, Journal of the European Economic Association, 16(3), 894-931;

(6) Chetty R., 2008, ”Moral Hazard versus Liquidity and Optimal Unemployment Insurance”, Journal of Political Economy, 116(2), 173-234;

(7) Fallick, B. C., 1996, ”A review of the recent empirical literature on displaced workers”, Industrial and Labor Relations Review, 50:1, 5-16;

(8) Fuller D. L., 2014, ”Adverse selection and moral hazard: quantitative implications for unemployment insurance”, Journal of Monetary Economics, 62, 108-122;

(9) Gonzalez F. M., Shi S., 2010, ”An Equilibrium Theory of Learning, Search, and Wages”, Econometrica, 78(2), 509-537;

(10) Habibi A., 2020, ”Motivation and Information Design”, Journal of Economic Behavior and Organization, 169, 1-18;

(11) Hagedorn M., Kaul A., Mennel T., 2010, ”An adverse selection model of optimal unemployment insurance”, Journal of Economic Dynamics and Control, 34(3), 490-502;

(12) Hopenhayn H. A., Nicolini J. P., 1997, ”Optimal Unemployment Insurance”, Journal of Political Economy, 105(2), 412-438;

(13) Kamenica E., Gentzkow M., 2011, ”Bayesian persuasion”, American Economic Review, 101(6), 2590-2615;

(14) Kamenica E., Gentzkow M., 2014, ”Costly Persuasion”, American Economic Review: Papers and Proceedings, 104(5), 457-462;

(15) Keane, M. P. and Wolpin K. I., 1997, "The Career Decisions of Young Men", Journal of Political Economy, 105 (3), 473–522;

(16) Kirby G., Hill H., Pavetti L., Jacobson J., Derr M., Winston P., 2002, "Transitional jobs: stepping stone to unsubsidized employment", Mathematica Policy Research Inc.;

(17) Kluve, J., Schmidt C.M., 2002, "Can training and employment subsidies combat European unemployment?", Economic Policy, 35, 409-448;

(18) Krueger A. B., Mueller A. I., 2010, "Job Search and Unemployment Insurance: New Evidence from Time Use Data", Journal of Public Economics, 94(3-4), 298-307;

(19) Machin S., Manning A., 1999, "The Causes and Consequences of Long-Term Unemployment in Europe", in O. Ashenfelter and D. Card (Eds), Handbook in Labor Economics, Vol. 3C, 3085-3139;

(20) Manoli D., Michaelides M., Patel A., 2018, "Long-Term Effects of Job-Search Assistance: Experimental Evidence Using Administrative Tax Data", *mimeo*;

(21) Meyer B. D., 1990, "Unemployment Insurance and Unemployment Spells", Econometrica, 58(4), 757-782;

(22) Michaelides M., Mueser P., 2020, "The Labor Market Effects Of US Reemployment Policy: Lessons From An Analysis Of Four Programs During The Great Recession", Journal of Labor Economics, 38(4), 1099-1140;

(23) Neal D., 1995, "Industry-Specific Human Capital: Evidence from Displaced Workers", Journal of Labor Economics, 13 (4), 653–677;

(24) Nicholson W. and Needels K., 2011, "The EUC08 Program in Theoretical and Historical Perspective", Mathematica Policy Research, n. 06863.320;

(25) Pavoni N., Setty O., Violante G. L., 2013, "Search and Work in Optimal Welfare Programs", NBER Working Paper No. 18666;

(26) Pavoni N., Setty O., Violante G. L., 2016, "The Design of 'Soft' Welfare-to-Work Programs", Review of Economic Dynamics, 20, 160-180;

(27) Pavoni N., Violante G. L., 2007, "Optimal Welfare-to-Work Programs", Review of Economic Studies, 74(1), 283-318;

(28) Poe-Yamagata E., Benus J., Bill N., Carrington H., Michaelides M., Shen T., 2011, "Impact of the Reemployment and Eligibility Assessemnt (REA) Initiative", IMPAQ International, Task Order No. DOLF091A21507;

(29) Rodina D., 2020, "Information Design and Career Concerns", CRC TR 224 Discussion Paper Series, University of Bonn and University of Mannheim, Germany;

(30) Setty, O., 2019, "Optimal Unemployment Insurance with Monitoring", Quantitative Economics, 10, 693-733;

(31) Shavell S., Weiss L. (1979), "The Optimal Payment of Unemployment Insurance Benefits over Time", Journal of Political Economy, 87(6), 1347-1362;

(32) Shimer R., Werning I., 2008, "Liquidity and Insurance for the Unemployed", American Economic Review, 98(5), 1922-1942;

(33) Sullivan W. F., Jr, Coffey L., Kolovich L., McGlew C. W., Sanford D., Sullivan R., 2007, "Worker Profiling and Reemployment Services. Evaluation of State Worker Profiling Models: Final Report", U.S. Department of Labor, Employment and Training Administration: Office of Workforce Security;

(34) Wang C., Williamson S., 1996, "Unemployment Insurance with Moral Hazard in a Dynamic Economy", Carnegie-Rochester Conference Series on Public Policy, 44, 1-41;

(35) Zapechelnyuk A., 2020, "Optimal Quality Certification", AER: Insights, 2(2), 161–176.

APPENDIX

Appendix A: Properties of SA, JS and UI

Proof of Prop. 1

Proof. Envelope Theorem and first-order conditions imply

$$V_U^{SA}(\mu, U) = -\frac{1}{u'(c_{sa})} = \mathbf{V}_U(\mu, U^u)$$

Now, given that SA is optimal in (μ, U), then $\mathbf{V}_U(\mu, U) = V_U^{SA}(\mu, U) = \mathbf{V}_U(\mu, U^u)$, and concavity of $\mathbf{V}$ in U implies that $U^u = U$. Therefore, the state space (μ, U) is equal in the next period, proving that SA is optimal forever after.

■

Proof of Lemma 1

Proof. The problem of policy $i \in \{SA, JS, UI\}$ reads

$$V^i(\mu, U) = \max_{(z, U^w, U^u) \in \Gamma(\mu, U)} -g(z) - \kappa^i + \beta\big[p^i(\mu)W(\mu, U^w) + (1 - p^i(\mu))\mathbf{V}(\mu^i, U^u)\big]$$

$$\text{sub:} \quad \Gamma^i(\mu, U) = \Big\{(z, U^w, U^u) : \ U = z - e^i + \beta\big[p^i(\mu)U^w + (1 - p^i(\mu))U^u\big], \ U \geq z + \beta U^u\Big\}$$

with $p^{SA}(\mu) = 0, \ p^{JS}(\mu) = p^{UI}(\mu) = \pi(\mu)$ and

$$(e^i, \kappa^i) = \begin{cases} (0, 0) & \text{if } i = SA \\ (0, \kappa^{ja}) & \text{if } i = JS \\ (e, 0) & \text{if } i = UI \end{cases}$$

The following holds.

Lemma 2. *V^i is decreasing in U and increasing in μ. Moreover, if $\mathbf{V}$ is concave in either argument, then so is V^i.*

Proof. To prove concavity of V^i in U/μ, it suffices to show that:

- the objective function is concave in the choice variables and U/μ;
- the graph of the feasibility set is convex.

Simply notice that $g = u^{-1}$ is convex, and that W and V are concave in U^w and U^u, respectively. Moreover, while $\pi(\mu)W(\mu, U)$ is linear in μ, $(1-\pi(\mu))\mathbf{V}(\mu', U)$ is concave in μ if $\mathbf{V}$ is concave in the first argument, as

$$-2(\pi_H - \pi_L)\frac{\partial \mu'}{\partial \mu}V_\mu(\mu', U) + (1-\pi(\mu))\Big[\frac{\partial^2 \mu'}{\partial \mu^2}V_\mu(\mu', U) + \frac{\partial \mu'}{\partial \mu}V_{\mu\mu}(\mu', U)\Big] < 0$$

as $(1-\pi(\mu))\frac{\partial^2 \mu'}{\partial \mu^2} = 2(\pi_H - \pi_L)\frac{\partial \mu'}{\partial \mu}$.
Furthermore, PK constraint is linear in U, z, U^w and U^u, and so is IC constraint, since U^i is linear in U. This means that the graph of Γ^i_μ (i.e., for constant μ) defined as

$$Gr\Gamma^i_\mu = \big\{(z, U^w, U^u, U) : \; U = z - e^i + \beta\big[p^i(\mu)U^w + (1-p^i(\mu))U^u\big], \; U \geq z + \beta U^u\big\}$$

is convex. Same applies to the graph of Γ^i_U (i.e., for constant U), since PK and IC are linear in μ.

To prove (negative) monotonicity in U, one needs to show:

- (negative) monotonicity of the objective function in U;
- (negative) monotonicity of the feasibility set Γ^i_μ in U, i.e.

$$U < \tilde{U} \Longrightarrow \Gamma^i(\mu, \tilde{U}) \subseteq \Gamma^i(\mu, U)$$

The objective function does not directly depend on U, while monotonicity can be shown by rewriting the IC constraint as

$$U^w - U^u \geq \frac{e^i}{\beta p^i(\mu)}$$

which does not depend on U. Therefore, the PK constraint is tightened by an increase of U, which thus leads to a shrinkage of Γ^i_μ.

Proving (positive) monotonicity of V^i in μ is analogous. Indeed, it follows from:

- (positive) monotonicity of the objective function in μ;
- (positive) monotonicity of the feasibility set Γ^i_U in μ, i.e.

$$\mu < \tilde{\mu} \Longrightarrow \Gamma^i(\mu, U) \subseteq \Gamma^i(\tilde{\mu}, U)$$

The objective function is always monotone in μ, as so are W and V in their first argument, μ' is an increasing function of μ and $W(\mu, U^w) \geq V(\mu^i, U^u)$. Monotonicity of $\Gamma^i(., U)$, instead, holds as an increase of μ leads to a relaxation of (IC)[33]. Indeed, (IC) is more slack since

$$\mu < \tilde{\mu} \Longrightarrow U^w - U^u \geq \frac{e^i}{\beta p^i(\mu)} \geq \frac{e^i}{\beta p^i(\tilde{\mu})}$$

■

<u>V^i **and V concave in** U</u>

The proof of concavity of V^i in U follows a recursive argument. Guessing concavity of $\mathbf{V}$ in U, it holds that V^i is concave by Lemma 2 and that, from definition (2) of $\mathbf{V}$,

$$\mathbf{V}_{UU}(\mu, U) = V^i_{UU}(\mu, U)$$

when no randomization over U is conducted. Otherwise, $\mathbf{V}_{UU}(\mu, U) = 0$.

<u>V^i **and V concave in** μ</u>

The proof of concavity in μ follows the same steps, as assuming concavity of $\mathbf{V}$ in μ leads to concavity of V^i. However, the second derivative of $\mathbf{V}$ in μ reads

$$\mathbf{V}_{\mu\mu}(\mu, U) = qV^i_{\mu\mu}(\mu, \overline{U}) + (1-q)V^j_{\mu\mu}(\mu, \underline{U}) < 0, \quad \text{with: } q = \frac{U - \underline{U}}{\overline{U} - \underline{U}}$$

<u>V^{UI} **and** V^{JS} **supermodular**</u>

The derivative of V^{UI} and V^{JS} wrt U reads

$$\begin{aligned}
V^{UI}_U(\mu, U) &= -\frac{1}{u(c_{UI})} = \pi(\mu)W_U(\mu, U^w_{UI}) + (1-\pi(\mu))\mathbf{V}_U(\mu', U^u_{UI}) \\
V^{JS}_U(\mu, U) &= -\frac{1}{u(c_{JS})} = W_U(\mu, U^w_{JS}) = \mathbf{V}_U(\mu', U^u_{JS})
\end{aligned}$$

Thus

$$\begin{aligned}
V^{UI}_{\mu U}(\mu, U) &= (\pi_H - \pi_L)(W_U(\mu, U^w_{UI}) - \mathbf{V}_U(\mu', U^u_{UI})) + \pi(\mu)W_{UU}(\mu, U^w_{UI})\frac{\partial U^w_{UI}}{\partial \mu} + \\
&+ (1-\pi(\mu))\mathbf{V}_{UU}(\mu', U^u_{UI})\frac{\partial U^u_{UI}}{\partial \mu} + (1-\pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu U}(\mu', U^u_{UI}) \\
&= (\pi_H - \pi_L)(W_U(\mu, U^w_{UI}) - \mathbf{V}_U(\mu', U^u_{UI}) + W_{UU}(\mu, U^w_{UI})(U^u_{UI} - U^w_{UI})) + \\
&+ \frac{\partial U^u_{UI}}{\partial \mu}\big(\pi(\mu)W_{UU}(\mu, U^w_{UI}) + (1-\pi(\mu))\mathbf{V}_{UU}(\mu', U^u_{UI})\big) + (1-\pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu U}(\mu', U^u_{UI})
\end{aligned}$$

Convexity of $1/u'$ implies concavity of W_U, which boils down to

$$W_U(\mu, U^w_{UI}) + W_{UU}(\mu, U^w_{UI})(U^u_{UI} - U^w_{UI}) > W_U(\mu, U^u_{UI}) = W_U(\mu', U^u_{UI}) \geq \mathbf{V}_U(\mu', U^u_{UI})$$

[33] (PK) is always relaxed by an increase in μ (recall that $U^w \geq U^u$ at the optimum).

Assume *per contra* that $V^{UI}_{\mu U}(\mu, U) \leq 0$. Then, it must be that $\partial U^u_{UI}/\partial\mu > 0$, which in turn implies that $\partial c_{UI}/\partial\mu < 0$, as $u(c_{UI}) = U - \beta U^u_{UI}$. But this leads to a contradiction as

$$V^{UI}_{\mu U}(\mu, U) = -\frac{\partial}{\partial c_{UI}}\Big(\frac{1}{u'(c_{UI})}\Big)\frac{\partial c_{UI}}{\partial \mu} > 0$$

Passing to JS,

$$V^{JS}_{\mu U}(\mu, U) = W_{UU}(\mu, U^w_{JS})\frac{\partial U^w_{JS}}{\partial \mu} = \mathbf{V}_{\mu U}(\mu', U^u_{JS})\frac{\partial \mu'}{\partial \mu} + \mathbf{V}_{UU}(\mu', U^u_{JS})\frac{\partial U^u_{JS}}{\partial \mu}$$

Per contra, assume that $V^{JS}_{\mu U}(\mu, U) < 0$. Then, it must be that $\partial U^s_{JS}/\partial\mu > 0$, $s \in \{w, u\}$. However, the PK-JS constraint reads

$$U = (1 - \beta + \pi(\mu))U^w_{JS} + \beta(1 - \pi(\mu))U^u_{JS}$$

And so

$$\frac{\partial U^w_{JS}}{\partial \mu} = -\frac{\beta}{1 - \beta + \beta\pi(\mu)}\Big[(\pi_H - \pi_L)(U^w_{JS} - U^u_{JS}) + (1 - \pi(\mu))\frac{\partial U^u_{JS}}{\partial \mu}\Big] < 0$$

where the inequality follows from assumption $U^w_{JS} \geq U^u_{JS}$. Hence, I have reached a contradiction.
$V^{JS}_{\mu\mu}(\mu, U) = 0$ **and** $V^{JS}_U(\mu, U) = V^{SA}_U(U)$ **with no JS → UI transition**
The period before entering SA, the FOC condition is

$$V^{JS}_U(\mu, U) = -\frac{1}{u'(c_{JS})} = -\frac{1}{u'(c_{SA})} = V^{SA}_U(U^u_{JS}) \Longrightarrow U^u_{JS} = (1 - \beta)u(c_{JS}) = U^w_{JS} = U$$

The value of JS wrt μ after imposing $U^u_{JS} = U^w_{JS} = U$ reads

$$V^{JS}(\mu, U) = -u^{-1}((1 - \beta)U) - \kappa^{ja} + \beta\big[\pi(\mu)W(\mu, U) + (1 - \pi(\mu))\mathbf{V}(\mu', U)\big]$$

Therefore, if $\mathbf{V}$ is linear in the first argument, so is V^{JS}, given linearity of $\pi(\mu)W(\mu, U)$ in μ. But then the proof follows from a recursive argument and linearity of V^{SA} in μ.

■

Proof of Prop. 2

Proof. The derivative of the value of each policy i with respect to U is

$$V^i_U(\mu, U) = \mathbf{V}_U(\mu, U) = -\frac{1}{u'(c_i)} \tag{17}$$

which can be obtained by applying the envelope theorem to the problem of each policy.
Third, first-order conditions in UI and JS impose

$$\mathbf{V}_U(\mu', U^u_{UI}) - V^{UI}_U(\mu, U) = \mathbf{V}_U(\mu', U^u_{UI}) + \lambda^{UI} - \chi^{UI} = \frac{\pi(\mu)}{1-\pi(\mu)}\chi^{UI} > 0 \tag{18}$$

$$V^{UI}_U(\mu, U_{UI}) = -(\lambda^{UI} - \chi^{UI}) = W_U(\mu, U^w_{UI}) + \chi^{UI} > W_U(\mu, U^w_{UI}) \tag{19}$$

$$V^{JS}_U(\mu, U) = -\lambda^{JS} = W_U(\mu, U^w_{JS}) = \mathbf{V}_U(\mu', U^u_{JS}) \tag{20}$$

where λ_i (resp., χ_i) is the Lagrange multiplier associated to (PK) (resp., (IC)) constraint. Hence,

$$\mathbf{V}_U(\mu', U^u_{UI}) > V^{UI}_U(\mu, U) > \mathbf{V}_U(\mu', U) \Longrightarrow U^u_{UI} < U \tag{21}$$

where the first inequality holds by FOC and the second by supermodularity of $\mathbf{V}$. Similarly, in JS consumption is constant over time and employment states by (20), which implies that

$$u(c_{JS}) = u(c^w_{JS}) = (1-\beta)U^w_{JS} \Longrightarrow U = \big(1-\beta+\beta\pi(\mu)\big)U^w_{JS} + \beta(1-\pi(\mu))U^u_{JS}$$

and

$$\mathbf{V}_U(\mu, U) = V^{JS}_U(\mu, U) = \mathbf{V}_U(\mu', U^u_{JS}) \leq \mathbf{V}_U(\mu, U^u_{JS}) \tag{22}$$

where the inequality holds since $\mathbf{V}$ is supermodular, the first equality as JS is optimal in (μ, U) and the second equality from FOC (20). Thus, by concavity of $\mathbf{V}$ in U, it holds that $U^u_{JS} \leq U \leq U^w_{JS}$. Supermodularity of JS follows from Lemma 1.

Optimal Policies in the U Space

The proof of the first part of the statement consists of showing that at the crossing point

$$V^{UI}_U(\mu, U) \leq V^{JS}_U(\mu, U) \leq V^{SA}_U(U) = W_U(\mu, U) \tag{23}$$

First, the closed-form expressions of W and V^{SA} deliver

$$W_U(\mu, U) = -\frac{1}{u'(u^{-1}((1-\beta)U))} = V^{SA}_U(U)$$

Then, by (20),

$$V^{JS}_U(\mu, U) = W_U(\mu, U^w_{JS}) \leq W_U(\mu, U) = V^{SA}_U(U)$$

where the inequality follows from $U \leq U^w_{JS}$.

By (21), we know that $U^u_{UI} < U$. Hence,

$$u(c_{UI}) = U - \beta U^u_{UI} > (1-\beta)U \Longrightarrow V^{UI}_U(\mu, U) < -\frac{1}{u'(u^{-1}((1-\beta)U))} = V^{SA}_U(U)$$

If JS refers to SA, then

$$u(c_{JS}) = u(c^u_{JS}) = u(c_{SA}) = (1-\beta)U^u_{JS}$$

So $U^w_{JS} = U^u_{JS} = U$ and

$$V^{JS}_U(\mu, U) = -\frac{1}{u'(u^{-1}((1-\beta)U))} = V^{SA}_U(U)$$

So far, I have shown that if JS is followed by SA, then

$$V^{UI}_U(\mu, U) \leq -\frac{1}{u'(u^{-1}((1-\beta)U))} = V^{SA}_U(U) = V^{JS}_U(\mu, U)$$

What is left to show is that $V^{UI}_U(\mu, U) \leq V^{JS}_U(\mu, U)$, even if JS does not refer to SA in (μ, U). *Per contra*, assume that $V^{UI}_U(\mu, U) > V^{JS}_U(\mu, U)$. First, it must be that the inequality

$$-\frac{1}{u'(u^{-1}((1-\beta)U))} \geq V^{UI}_U(\mu, U) > V^{JS}_U(\mu, U)$$

holds and implies that $(1-\beta)U \leq u(c_{UI}) < u(c_{JS})$. And from (22), it must be that $U^u_{JS} < U < \frac{u(c_{JS})}{1-\beta}$.
Now, by FOCs (18) and (20), it holds that

$$\mathbf{V}_U(\mu', U^u_{UI}) > V^{UI}_U(\mu, U) > V^{JS}_U(\mu, U) = \mathbf{V}_U(\mu', U^u_{JS})$$

By concavity of $\mathbf{V}$ in U, it must hold that $U^u_{UI} < U^u_{JS}$. But this is impossible as

$$u(c_{UI}) + \beta U^u_{UI} = U = u(c_{JS}) + \beta U^u_{JS} + \beta\pi(\mu)\Big[\frac{u(c_{JS})}{1-\beta} - U^u_{JS}\Big] > u(c_{UI}) + \beta U^u_{UI}$$

where the inequality follows from $c_{JS} > c_{UI}$ and $\frac{u(c_{JS})}{1-\beta} > U^u_{JS}$.
Therefore, it has been shown that $V^{UI}_U(\mu, U) \leq V^{JS}_U(\mu, U)$. Hence, V^{UI} dominates V^{JS} for low-end generosity levels and crosses it from above.

Optimal Policies in the μ Space

Passing to the second part of the statement, it is enough to prove that at the crossing point

$$0 = V^{SA}_\mu(U) < V^{JS}_\mu(\mu, U) < V^{UI}_\mu(\mu, U)$$

The derivatives of V^{JS} and V^{UI} wrt to μ

$$\begin{aligned} V^{JS}_{\mu}(\mu, U) =& \beta(\pi_H - \pi_L)\big[W(\mu, U^w_{JS}) - \mathbf{V}(\mu', U^u_{JS}) - \lambda^{JS}(U^u_{JS} - U^w_{JS})\big] + \\ &+ \beta\Big[\pi(\mu) W_{\mu}(\mu, U^w_{JS}) + (1 - \pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu}(\mu', U^u_{JS})\Big] \qquad (24) \\ V^{UI}_{\mu}(\mu, U) =& \beta(\pi_H - \pi_L)\big[W(\mu, U^w_{UI}) - \mathbf{V}(\mu', U^u_{UI}) - \lambda^{UI}(U^u_{UI} - U^w_{UI})\big] + \\ &+ \beta\Big[\pi(\mu) W_{\mu}(\mu, U^w_{UI}) + (1 - \pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu}(\mu', U^u_{UI})\Big] \qquad (25) \end{aligned}$$

Consider JS being implemented in the current period. Using FOC $-\lambda^{JS} = W_U(\mu, U^w_{JS})$, it holds that

$$W(\mu, U^w_{JS}) + W_U(\mu, U^w_{JS})(U^u_{JS} - U^w_{JS}) - \mathbf{V}(\mu', U^u_{JS}) > W(\mu, U^u_{JS}) - \mathbf{V}(\mu', U^u_{JS}) > 0 = V^{SA}_{\mu}(U)$$

where the first inequality follows from concavity of W in U and the second is a necessary condition for optimality of JS.

Consider a program that implements UI with the additional constraint that $U^u_{UI} \geq U^u_{JS}$, and label its value $\hat{V}^{UI}$. Moreover,

$$W_U(\mu, U^w_{JS}) = V^{JS}_U(\mu, U) \geq V^{UI}_U(\mu, U) > W_U(\mu, U^w_{UI}) \Longrightarrow U^w_{JS} < U^w_{UI}$$

where the first inequality follows from the statement proved above and the second one from FOC of UI.[34] Hence, derivatives (24) and (25) can be rewritten

$$\begin{aligned} V^{JS}_{\mu}(\mu, U) =& \beta(\pi_H - \pi_L)\big[W(\mu, U^w_{JS}) - \mathbf{V}(\mu', U^u_{JS}) + W_U(\mu, U^w_{JS})(U^u_{UI} - U^w_{JS}) + \mathbf{V}_U(\mu', U^u_{JS})(U^u_{JS} - U^u_{UI})\big] + \\ &+ \beta\Big[\pi(\mu) W_{\mu}(\mu, U^w_{JS}) + (1 - \pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu}(\mu', U^u_{JS})\Big] \\ \hat{V}^{UI}_{\mu}(\mu, U) =& \beta(\pi_H - \pi_L)\big[W(\mu, U^w_{UI}) - \mathbf{V}(\mu', U^u_{UI}) + W_U(\mu, U^w_{UI})(U^w_{JS} - U^w_{UI}) + W_U(\mu, U^w_{UI})(U^u_{UI} - U^w_{JS})\big] + \\ &+ \beta\Big[\pi(\mu) W_{\mu}(\mu, U^w_{UI}) + (1 - \pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu}(\mu', U^u_{UI})\Big] \end{aligned}$$

In order to prove the result, it is enough to show that

$$\begin{aligned} & W(\mu, U^w_{JS}) + W_U(\mu, U^w_{JS})(U^u_{UI} - U^w_{JS}) - \mathbf{V}(\mu', U^u_{JS}) + \mathbf{V}_U(\mu', U^u_{JS})(U^u_{JS} - U^u_{UI}) < \\ < & W(\mu, U^w_{UI}) + W_U(\mu, U^w_{UI})(U^w_{JS} - U^w_{UI}) + W_U(\mu, U^w_{UI})(U^u_{UI} - U^w_{JS}) - \mathbf{V}(\mu', U^u_{UI}) \end{aligned}$$

and

$$\pi(\mu) W_{\mu}(\mu, U^w_{JS}) + (1 - \pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu}(\mu', U^u_{JS}) < \pi(\mu) W_{\mu}(\mu, U^w_{UI}) + (1 - \pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_{\mu}(\mu', U^u_{UI})$$

[34] The additional constraint preserves the FOC $V^{UI}(\mu, U) > W_U(\mu, U^w_{UI})$.

The first inequality holds since:

- $W(\mu, U^w_{JS}) < W(\mu, U^w_{UI}) + W_U(\mu, U^w_{UI})(U^w_{JS} - U^w_{UI})$, by concavity of W in U;
- $W_U(\mu, U^w_{JS})(U^u_{UI} - U^w_{JS}) < W_U(\mu, U^w_{UI})(U^u_{UI} - U^w_{JS})$, as $U^u_{UI} < U \leq U^w_{JS} < U^w_{UI}$;
- $\mathbf{V}(\mu', U^u_{UI}) \leq \mathbf{V}(\mu', U^u_{JS}) + \mathbf{V}_U(\mu', U^u_{JS})(U^u_{UI} - U^u_{JS})$, by concavity of $\mathbf{V}$ in U.

The second inequality holds since $W_{\mu U} = 0$ and $\mathbf{V}_\mu(\mu', U^u_{JS}) \leq \mathbf{V}_\mu(\mu', U^u_{UI})$, by assumption $U^u_{JS} \leq U^u_{UI}$ and supermodularity of $\mathbf{V}$. Therefore, it has been shown that $\hat{V}^{UI}$ crosses V^{JS} from below in the μ space, and so does V^{UI}, which implies that UI dominates JS for high expectations. ■

Proof of Prop. 3

Unemployment Benefits

Thus, unemployment benefits fall over time during UI and stay constant in JS, as

$$\begin{aligned}
&\mathbf{V}_U(\mu', U^u_{UI}) > V^{UI}_U(\mu, U) \implies c^u_{UI} < c_{UI} \\
&V^{UI}_U(\mu, U_{UI}) > W_U(\mu, U^w_{UI}) \implies c_{UI} < c^w_{UI} \\
&V^{JS}_U(\mu, U) = W_U(\mu, U^w_{JS}) = \mathbf{V}_U(\mu', U^u_{JS}) \implies c_{JS} = c^w_{JS} = c^u_{JS}
\end{aligned}$$

where the implications follow from (17).

Continuation Utility

If JS never refers to UI, then one can start computing backward from the point in time where JS refers to SA. Hence, $V^{JS}_U(\mu, U) = V^{SA}_U(U)$. Therefore $U^w_{JS} = U^u_{JS} = U$ and $\mathbf{V}_{\mu U}(\mu, U) = V^{JS}_{\mu U}(\mu, U) = 0$. The last period before the worker enters SA, the contract satisfies

$$V^{JS}_U(\mu, U) = \mathbf{V}_U(\mu', U^u_{JS}) = \mathbf{V}_U(\mu, U^u_{JS}) \implies U^u_{JS} = U$$

The result is shown by induction argument.

Proof of Prop. 4

Assume that UI is the optimal policy in (μ, U). From the first-order condition on UI, it holds that

$$-g'((1-\beta)U + \eta^{UI}(\mu, U)) = V^{UI}_U(\mu, U) = -g'(U - \beta U^u_{UI})$$

From which it follows that

$$U - \frac{\eta^{UI}(\mu, U)}{\beta} = U^u_{UI}$$

Since by assumption $\eta_U^{UI}(\mu, U) \leq 0$, the left-hand side is increasing in U. Thus,

$$U_{UI}^u = U - \frac{\eta^{UI}(\mu, U)}{\beta} \leq \hat{U}(\mu) - \frac{\eta^{UI}(\mu, \hat{U}(\mu))}{\beta} \leq \hat{U}(\mu')$$

The condition guarantees that anytime UI is adopted in (μ, U), then the next-period state in case of job search failure becomes (μ', U_{UI}^u) where it is still optimal to implement UI (or switch to SA) as $U_{UI}^u \leq \hat{U}(\mu')$. To conclude, the program never switches from UI to JS.

Appendix B: Properties of AP, SP and IP

Proof of Prop. 6

Proof. Define $\mu_{i,j}$ the threshold in the space of expectations where the planner is indifferent between policies i and j.

For $\mu \leq \mu_{js,ui}$, IP is dominated by SP as

$$\begin{aligned}
V^{IP}(\mu, U) = & - c - \kappa^{wp} + \beta[\pi(\mu)W(\mu, U_{ui}^w) + q(1-\pi(\mu_p))\mathbf{V}(\mu_p', U) + (1-q)(1-\pi(\mu_f))\mathbf{V}(\mu_f', U)] \\
\leq & - c - \kappa^{ja} - \kappa^{wp} + \beta\big[\pi(\mu)W(\mu, U) + q(1-\pi(\mu_p))\mathbf{V}(\mu_p', U) + (1-q)(1-\pi(\mu_f))\mathbf{V}(\mu_f', U)] \\
\leq & - c - \kappa^{ja} - \kappa^{wp} + \beta\big[\pi(\mu)W(\mu, U) + \\
& + \max_{\hat{q}, \hat{\mu}_f, \hat{\mu}_p} \big\{\hat{q}(1-\pi(\hat{\mu}_p))\mathbf{V}(\hat{\mu}_p', U) + (1-\hat{q})(1-\pi(\hat{\mu}_f))\mathbf{V}(\hat{\mu}_f', U)\big\}] = V^{SP}(\mu, U)
\end{aligned}$$

where the first inequality follows from $V^{UI}(\mu, U) \leq V^{JS}(\mu, U)$ when $\mu \leq \mu_{js,ui}$. Thus, IP crosses SP in $\mu_{sp,ip} \geq \mu_{js,ui}$.

For $\mu \leq \mu_{sa,js}$, SP is dominated by AP as

$$\begin{aligned}
V^{SP}(\mu, U) = & - c - \kappa^{ja} - \kappa^{wp} + \beta\big[\pi(\mu)W(\mu, U) + q(1-\pi(\mu_p))\mathbf{V}(\mu_p', U) + (1-q)(1-\pi(\mu_f))\mathbf{V}(\mu_f', U)] \\
\leq & - c - \kappa^{wp} + \beta\big[\pi(\mu)\mathbf{V}(U) + q(1-\pi(\mu_p))\mathbf{V}(\mu_p', U) + (1-q)(1-\pi(\mu_f))\mathbf{V}(\mu_f', U)] \\
\leq & - c - \kappa^{wp} + \beta \max_{\hat{q} \in \Delta^*([0,1])} \int_0^1 \hat{q}(x)\mathbf{V}(\hat{\mu}(x), U) = V^{AP}(\mu, U)
\end{aligned}$$

where the first inequality follows from $V^{JS}(\mu, U) \leq V^{SA}(U)$ when $\mu \leq \mu_{sa,js}$, while the second from the fact that the distribution $(q_1, q_2, q_3) = (\pi(\mu), q(1-\pi(\mu_p)), (1-q)(1-\pi(\mu_f)))$ over posteriors $(\mu_1, \mu_2, \mu_3) = ((\mu\pi_H)/\pi(\mu), \mu_p', \mu_f')$ belongs to $\Delta^*([0,1])$.[35] Thus, SP crosses AP in $\mu_{sp,jp} \geq \mu_{sa,js}$. ■

[35] μ_1 solves

$$\pi(\mu)\mu_1 + q(1-\pi(\mu_p))\mu_p' + (1-q)(1-\pi(\mu_f)\mu_f' = \mu$$

Proof of Prop. 7

Proof. Consider the two first-order conditions of the AP problem

$$\mathbf{V}_\mu(\mu_p, U^u_{AP}) - \frac{\mathbf{V}(\mu_p, U^u_{AP}) - \mathbf{V}(0, U^u_{AP})}{\mu_p} = 0$$

$$q\mathbf{V}_U(\mu_p, U^u_{AP}) + (1-q)\mathbf{V}_U(0, U^u_{AP}) + \lambda^{AP} = 0$$

For the pair (μ_p, U^u_{AP}) to be a point of maximum, it must be that the Hessian matrix H of second-order derivatives has positive determinant.

$$H = \begin{bmatrix} q\mathbf{V}_{UU}(\mu_p, U^u_{AP}) + (1-q)\mathbf{V}_{UU}(0, U^u_{AP}) & q\Big(\mathbf{V}_{\mu U}(\mu_p, U^u_{AP}) - \frac{V_U(\mu_p,U^u_{AP}) - V_U(0,U^u_{AP})}{\mu_p}\Big) \\ \mathbf{V}_{\mu U}(\mu_p, U^u_{AP}) - \frac{V_U(\mu_p,U^u_{AP}) - V_U(0,U^u_{AP})}{\mu_p} & \mathbf{V}_{\mu\mu}(\mu_p, U^u_{AP}) \end{bmatrix}$$

Differentiating the two conditions by U yields

$$\mathbf{V}_{\mu\mu}(\mu_p, U^u_{AP})\frac{\partial \mu_p}{\partial U^u_{AP}} + \mathbf{V}_{\mu U}(\mu_p, U^u_{AP}) - \frac{\mathbf{V}_U(\mu_p, U^u_{AP}) - \mathbf{V}_U(0, U^u_{AP})}{\mu_p} = 0$$

$$\Big\{q\Big[\mathbf{V}_{\mu U}(\mu_p, U^u_{AP}) - \frac{\mathbf{V}_U(\mu_p, U^u_{AP}) - \mathbf{V}_U(0, U^u_{AP})}{\mu_p}\Big]\frac{\partial \mu_p}{\partial U^u_{AP}} + q\mathbf{V}_{UU}(\mu_p, U^u_{AP}) + (1-q)\mathbf{V}_{UU}(0, U^u_{AP})\Big\}\frac{\partial U^u_{AP}}{\partial U} + \\ + \lambda^{AP}_U = 0$$

Plugging the expression of $\frac{\partial \mu_p}{\partial U^u_{AP}}$ from the first equation into the second one, the term in curly brackets becomes

$$\Delta := -\frac{q}{\mathbf{V}_{\mu\mu}(\mu_p, U^u_{AP})}\Big[\mathbf{V}_{\mu U}(\mu_p, U^u_{AP}) - \frac{\mathbf{V}_U(\mu_p, U^u_{AP}) - \mathbf{V}_U(0, U^u_{AP})}{\mu_p}\Big]^2 + q\mathbf{V}_{UU}(\mu_p, U^u_{AP}) + (1-q)\mathbf{V}_{UU}(0, U^u_{AP})$$

with $\Delta < 0$, as $det(H) > 0$ and $\mathbf{V}_{\mu\mu}(\mu_p, U^{u,p}_{AP}) < 0$. Therefore the second equation becomes

$$\frac{\partial U^u_{AP}}{\partial U}\Delta + \lambda^{AP}_U = 0$$

which shows that $\frac{\partial U^u_{AP}}{\partial U} > 0$, since $V^{AP}_{UU}(\mu, U) = -\lambda^{AP}_U < 0$. From the first equation, local supermodularity of $\mathbf{V}$ and $\mathbf{V}_U(\mu_p, U) < \mathbf{V}_U(0, U)$ (see Prop. 2) yield $\frac{\partial \mu_p}{\partial U^u_{AP}} > 0$, which deliver the result.

Finally, full accuracy of profiling whenever 'Pass' refers to JS and no incentive is provided along the spell follows from linearity of V^{JS} in μ (see proof of Lemma 2).

■

Appendix C: Statistical Profiling

Proof of Prop. 8 and 9

Proof. $\mu_f = 0$ descends form the linearity of SA and JS in μ. In addition, any worker who receives a 'Pass' (resp., 'Fail') is referred to UI/JS (resp., SA).

Assistance-and-Profiling (AP)

At the optimum,

$$V_U^{AP}(\mu, U) = \mathbf{V}_U(\mu_p, U_{AP}^{u,p}) = \mathbf{V}_U(0, U_{AP}^{u,f}) \tag{26}$$

$$-\frac{1}{u'(c_{AP})} = V_U^{AP}(\mu, U) = \mathbf{V}_U(0, U_{AP}^{u,f}) = -\frac{1}{u'(c_{AP}^{u,f})} \tag{27}$$

which implies that $c_{AP} = c_{AP}^{u,p} = c_{AP}^{u,f}$, and $U_p^u \leq U \leq U_f^u$. Indeed, by (27), it follows

$$U = u(c_{AP}) + \beta(qU_{AP}^{u,p} + (1-q)U_{AP}^{u,f}) = (1-\beta q)U_{AP}^{u,f} + \beta q U_{AP}^{u,p}$$

where the passage follows from $u(c_{AP}) = u(c_{AP}^{u,f}) = (1-\beta)U_{AP}^{u,f}$, and the expression of consumption in SA (see Prop. 1). If referred to JS -which is optimal only for high-end generosities-, then $\mu_p = 1$ given the linearity of JS in μ. Moreover, for U high enough, JS never refers to UI, and so $U_{JS}^w = U = U_{JS}^u$, which in turn implies that $u(c_{JS}) = (1-\beta)U_{JS}^w = (1-\beta)U = u(c_{SA})$ and

$$V_U^{JS}(\mu, U) = -\frac{1}{u'(c_{JS})} = -\frac{1}{u'(c_{SA})} = V_U^{SA}(U)$$

Therefore, if referred to JS/SA forever after, then $U_{AP}^{u,p} = U_{AP}^{u,f} = U_{AP}^u$. So, nothing changes with respect to the case with ND constraint, whenever AP refers workers to SA and JS forever after, that is, for higher generosities.

Assume, instead, AP refers to UI directly, or to JS which later refers to UI. Then $\mathbf{V}_U(\mu, U) <$ $V_U^{SA}(U)$

$$V_U^{SA}(U_{AP}^{u,f}) = V_U(\mu_p, U_{AP}^{u,p}) \Longrightarrow U_{AP}^{u,p} < U < U_{AP}^{u,f}$$

I now show that the 'Pass' posterior μ_p in AP is increasing in U.

$$\mathbf{V}_U(\mu_p, U_{AP}^{u,p}) + \lambda^{AP} = 0$$

$$-\frac{\mathbf{V}(\mu_p, U_{AP}^{u,p}) - \mathbf{V}(0, U_{AP}^{u,f}) + \lambda^{AP}(U_{AP}^{u,p} - U_{AP}^{u,f})}{\mu_p} + \mathbf{V}_\mu(\mu_p, U_{AP}^{u,p}) = 0$$

The Hessian matrix of second-order derivatives reads

$$H = \begin{bmatrix} \mathbf{V}_{UU}(\mu_p, U^{u,p}_{AP}) & \mathbf{V}_{\mu U}(\mu_p, U^{u,p}_{AP}) \\ \mathbf{V}_{\mu U}(\mu_p, U^{u,p}_{AP}) & \mathbf{V}_{\mu\mu}(\mu_p, U^{u,p}_{AP}) \end{bmatrix}$$

$(\mu_p, U^{u,p}_{AP})$ are a point of maximum of the objective function if and only if

$$\mathbf{V}_{UU}(\mu_p, U^{u,p}_{AP}) < 0, \quad det(H) > 0$$

The first condition holds as $\mathbf{V}$ is concave in each argument (see proof of Lemma 2). Differentiating the two FOCs wrt U yields

$$\begin{aligned} \mathbf{V}_{\mu U}(\mu_p, U^{u,p}_{AP})\frac{\partial \mu_p}{\partial U} + \mathbf{V}_{UU}(\mu_p, U^{u,p}_{AP})\frac{\partial U^{u,p}_{AP}}{\partial U} &= -\frac{\partial \lambda^{AP}}{\partial U} \\ \mathbf{V}_{\mu\mu}(\mu_p, U^{u,p}_{AP})\frac{\partial \mu_p}{\partial U} + \mathbf{V}_{\mu U}(\mu_p, U^{u,p}_{AP})\frac{\partial U^{u,p}_{AP}}{\partial U} &= \frac{U^{u,p}_{AP} - U^{u,f}_{AP}}{\mu_p}\frac{\partial \lambda^{AP}}{\partial U} \end{aligned}$$

and solving the system, one obtains

$$\begin{bmatrix} \frac{\partial \mu_p}{\partial U} \\ \frac{\partial U^{u,p}_{AP}}{\partial U} \end{bmatrix} = det(H)^{-1} \begin{bmatrix} \mathbf{V}_{\mu\mu}(\mu_p, U^{u,p}_{AP}) & -\mathbf{V}_{\mu U}(\mu_p, U^{u,p}_{AP}) \\ -\mathbf{V}_{\mu U}(\mu_p, U^{u,p}_{AP}) & \mathbf{V}_{UU}(\mu_p, U^{u,p}_{AP}) \end{bmatrix} \begin{bmatrix} \frac{U^{u,p}_{AP} - U^{u,f}_{AP}}{\mu_p} \\ -1 \end{bmatrix} \frac{\partial \lambda^{AP}}{\partial U}$$

Both derivatives are positive, since $U^{u,p}_{AP} - U^{u,f}_{AP} < 0$ and an increase in U makes it harder for the planner to satisfy (PK) constraint (i.e., $\partial \lambda^{AP}/\partial U > 0$).

Search-assistance-and-Profiling (SP)

At the optimum

$$\begin{aligned} &V^{SP}_U(\mu, U) = W_U(\mu_p, U^{w,p}_{SP}) = W_U(\mu_f, U^{w,f}_{SP}) = \mathbf{V}_U(\mu'_p, U^{u,p}_{SP}) = \mathbf{V}_U(\mu'_f, U^{u,f}_{SP}) \\ \Longrightarrow &c_{SP} = c^w_{SP} = c^{u,p}_{SP} = c^{u,f}_{SP}, \quad U^{u,p}_{SP} \leq U \leq U^{u,f}_{SP} = U^{w,p}_{SP} = U^{w,f}_{SP} \end{aligned}$$

since

$$\frac{u(c_{SP})}{1-\beta} = \frac{u(c^w_{SP})}{1-\beta} = U^w_{SP} = \frac{u(c^{u,f}_{SP})}{1-\beta} = U^{u,f}_{SP}$$

where the last equality follows from referral to SA upon 'Fail'. So

$$U = (1 - \beta + \beta\pi(\mu) + \beta(1-q)(1-\pi(\mu_f)))U^{u,f}_{SP} + \beta q(1-\pi(\mu_p))U^{u,p}_{SP}$$

and the same argument in AP applies, meaning that the continuation utility upon 'Pass' falls if and only if the outcome refers workers directly or indirectly to UI.

Insurance-and-Profiling (IP)

The optimal IP contract satisfies

$$\mathbf{V}_U(\mu_p', U_{IP}^{u,p}) - V_U^{IP}(\mu, U) = \frac{\pi(\mu_p)}{1-\pi(\mu_p)}\chi^{IP} > \frac{\pi_L}{1-\pi_L}\chi^{IP} = \mathbf{V}_U(0, U_{IP}^{u,f}) - V_U^{IP}(\mu, U) \Longrightarrow U_{IP}^{u,p} < U_{IP}^{u,f}$$
$$\mathbf{V}_U(\mu_p', U_{IP}^{u,p}) > \mathbf{V}_U(0, U_{IP}^{u,f}) > V_U^{IP}(\mu, U) = -(\lambda^{IP} - \chi^{IP}) > W_U(\mu, U_{IP}^{w}) \Longrightarrow c_{IP}^{u,p} < c_{IP}^{u,f} < c_{IP} < c_{IP}^{w} \tag{28}$$

Moreover, $U_{IP}^{u,p} < U$, as

$$(1-\beta)U_{IP}^{u,p} \leq u(c_{IP}^{u,p}) < u(c_{IP}) = U - \beta[qU_{IP}^{u,p} + (1-q)U_{IP}^{u,f}] < U - \beta U_{IP}^{u,p}$$

where the first inequality follows from (23), the second one from (28) and the last one from $U_{IP}^{u,p} < U_{IP}^{u,f}$.

Passing to the equation that determines the 'Pass' posterior in IP, the first-order condition of μ_p reads

$$\frac{1}{\mu_p}\big[(1-\pi_L)\mathbf{V}(0, U_{IP}^{u,f}) - (1-\pi(\mu_p))\mathbf{V}(\mu_p', U_{IP}^{u,p})\big] - (\pi_H - \pi_L)\mathbf{V}(\mu_p', U_{IP}^{u,p}) + \frac{(1-\pi_H)(1-\pi_L)}{1-\pi(\mu_p)}\mathbf{V}_\mu(\mu_p', U_{IP}^{u,p}) +$$
$$+ \lambda^{IP}\Big[\frac{1}{\mu_p}\big[(1-\pi_L)U_{IP}^{u,f} - (1-\pi(\mu_p))U_{IP}^{u,p}\big] - (\pi_H - \pi_L)U_{IP}^{u,p}\Big] + \frac{\chi^{IP}}{\mu_p}\big[U_{IP}^{u,p} - U_{IP}^{u,f}\big] = 0$$

Rearranging the terms and using the first order condition on $U_{IP}^{u,f}$

$$(1-\pi_L)\big[\mathbf{V}_U(0, U_{IP}^{u,f}) + \lambda^{IP}\big] = \chi^{IP}$$

it yields

$$\mathbf{V}_\mu(\mu_p', U_{IP}^{u,p}) = \frac{\mathbf{V}(\mu_p', U_{IP}^{u,p}) - \mathbf{V}(0, U_{IP}^{u,f}) + \mathbf{V}_U(0, U_{IP}^{u,f})(U_{IP}^{u,f} - U_{IP}^{u,p})}{\mu_p'}$$

with $U_{IP}^{u,f} > U_{IP}^{u,p}$. ■

Appendix D: Moral Hazard

Proof of Prop. 10

Proof. The first part of the proof is contained in the Technical Appendix. It shows that multiple deviations can be accounted for by single one-shot deviations, that is, deviations from recommended action lasting only one period. Now consider the recursion (15)

$$
\begin{aligned}
\Lambda(1,\mu^{T-1}) &= 0 \\
\Lambda(2,\mu^{T-2}) &= u(c_{T-1}) - e + \beta\pi(\mu^{T-2})U^w_T + \beta(1-\pi(\mu^{T-2}))U^u_T - U^u_{T-2} \\
&= U^u_{T-2} + \beta[\pi(\mu^{T-2}) - \pi(\mu^{T-1})](U^w_T - U^u_T) - U^u_{T-2} = \beta[\pi(\mu^{T-2}) - \pi(\mu^{T-1})]\frac{e}{\beta\pi(\mu^{T-1})} \\
\Lambda(T-j,\mu^j) &= u(c_{j+1}) - e + \beta\pi(\mu^j)U^w_{j+2} + \\
&\quad + \beta(1-\pi(\mu^j))\underbrace{[u(c_{j+2}) - e + \beta\pi(\mu^{j+1})U^w_{j+3} + \beta(1-\pi(\mu^{j+1}))[...]]}_{\Lambda(T-j-1,\mu^{j+1})+U^u_{j+2}} - U^u_{j+1} \\
&= U^u_{j+1} + \beta[\pi(\mu^j) - \pi(\mu^{j+1})](U^w_{j+2} - U^u_{j+2}) + \beta(1-\pi(\mu^j))\Lambda(T-j-1,\mu^{j+1}) - U^u_{j+1} \\
&= \beta[\pi(\mu^j) - \pi(\mu^{j+1})]\frac{\beta\Lambda(T-j-1,\mu^{j+1}) + e}{\beta\pi(\mu^{j+1})} + \beta(1-\pi(\mu^j))\Lambda(T-j-1,\mu^{j+1}) \\
&= \Big(\frac{\pi(\mu^j)}{\pi(\mu^{j+1})} - 1\Big)e + \beta\pi(\mu^j)\Big(\frac{1}{\pi(\mu^{j+1})} - 1\Big)\Lambda(T-j-1,\mu^{j+1}), \quad 0 \le j \le T-1
\end{aligned}
$$

And notice that the constraint $(\hat{IC},t)$, defined as

$$U_s(\mathcal{W},\mu^s,\sigma^s) = u(c_s(\sigma^s)) + \beta\big[U_{s+1}(\mathcal{W},\mu^{s+1},(\sigma^s,u)) + \Lambda(T-s,\mu^s)\big]$$

makes the contract robust against any possible deviation after period t, thanks to the recursive definition of Λ. In particular,

$$(\hat{IC},t) \iff (IC,s),\ \forall s \ge t$$

Hence the whole set of IC constraints can be expressed by

$$U = U_0(\mathcal{W},\mu,\sigma_0) = u(c) + \beta\big[U^u + \Lambda(T,\mu)\big], \quad (\hat{IC},0)$$

Λ is defined by the recursion in (15), and is independent of U^u. In addition, $\Lambda(t+1,\mu) \ge \Lambda(t,\mu)$, with inequality being strict for $\mu \in (0,1)$.[36] Indeed, taking the difference between $\Lambda(t+1,\mu)$ and $\Lambda(t,\mu)$, it holds:

$$
\begin{cases}
\Lambda(2,\mu) - \Lambda(1,\mu) = \Big(\frac{\pi(\mu)}{\pi(\mu')} - 1\Big)e > 0 \\
\Lambda(t+1,\mu) - \Lambda(t,\mu) = \beta\pi(\mu)\Big(\frac{1}{\pi(\mu')} - 1\Big)(\Lambda(t,\mu') - \Lambda(t-1,\mu')) > 0, \quad \forall t \ge 2
\end{cases}
$$

■

Lemma 3. *The value of $(UI,t)_{t\ge1}$ is increasing in μ.*

[36]In $\mu \in \{0,1\}$, no learning occurs and learning rents are null.

Proof. The problem of policy $(UI, t)_{t\geq 1}$ reads

$$\begin{aligned} V_t^{UI}(\mu, U) = \max_{(z, U^w, U^u)\in\Gamma(\mu,U)} & -g(z) + \beta\big[\pi(\mu)W(\mu, U^w) + (1-\pi(\mu))V_{t-1}^{UI}(\mu', U^u)\big] \\ \text{sub:} \quad \Gamma(\mu, U) = & \Big\{(z, U^w, U^u) : \ U = z - e + \beta\big[\pi(\mu)U^w + (1-\pi(\mu))U^u\big], \\ & U \geq z + \beta\big[U^u + \Lambda(t,\mu)\big]\Big\} \end{aligned}$$

V_1^{UI} is monotone increasing in μ (see Lemma 1). By induction, assume that V_{t-1}^{UI} is increasing in μ. Positive monotonicity of V_t^{UI} in μ follows from:

- (positive) monotonicity of the objective function in μ;
- (positive) monotonicity of the feasibility set Γ_U^i in μ, i.e.

$$\mu < \tilde{\mu} \Longrightarrow \Gamma_U^i(\mu) \subseteq \Gamma_U^i(\tilde{\mu})$$

The objective function is always monotone in μ, as so are W and V_{t-1}^{UI} in their first argument, μ' is an increasing function of μ and $W(\mu, U) \geq V_{t-1}^{UI}(\mu', U)$ at the optimum. Monotonicity of Γ_U, instead, holds whenever an increase of μ leads to a relaxation of (IC).[37] Now, if $\Lambda(t, .)$ is constant or decreasing, this always holds. Indeed, (IC) is more slack if $\Lambda(t, .)$ is decreasing as

$$\mu < \tilde{\mu} \Longrightarrow U^w - U^u \geq \frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)} > \frac{e/\beta + \Lambda(t,\tilde{\mu})}{\pi(\tilde{\mu})}$$

To prove that monotonicity holds also when $(\Lambda(t, .))_{t>1}$ is increasing in μ, I prove that the RHS is decreasing in μ.

From the definition of Λ in (15), I can rewrite

$$\frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)} = \frac{e}{\pi(\mu)}\Big(\frac{1}{\beta} - 1\Big) - \beta\Lambda(t-1,\mu') + \beta\Big(\frac{e/\beta + \Lambda(t-1,\mu')}{\pi(\mu')}\Big) \tag{29}$$

Define $f(\mu) := \frac{\pi(\mu)}{\pi(\mu')}$, and notice that it is concave in μ. Indeed:

$$\begin{aligned} f_\mu(\mu) &= (\pi_H - \pi_L)^2 \frac{(1-\mu)^2\pi_L(1-\pi_L) - \mu^2\pi_H(1-\pi_H)}{\big[(1-\pi_H)\pi_H\mu + (1-\pi_L)\pi_L(1-\mu)\big]^2} \\ f_{\mu\mu}(\mu) &= -\frac{2(\pi_H - \pi_L)^2\pi_H\pi_L(1-\pi_H)(1-\pi_L)}{\big[(1-\pi_H)\pi_H\mu + (1-\pi_L)\pi_L(1-\mu)\big]^3} < 0 \end{aligned}$$

Thus, the derivative of $\Lambda(t, \mu)$ by μ reads

$$\Lambda_\mu(t,\mu) = f_\mu(\mu)e + \beta\big[f_\mu(\mu) - (\pi_H - \pi_L)\big]\Lambda(t-1,\mu') + \beta\big[f(\mu) - \pi(\mu)\big]\frac{\partial\mu'}{\partial\mu}\Lambda_\mu(t-1,\mu') \tag{30}$$

[37] (PK) is always relaxed by an increase in μ (recall that $U^w \geq U^u$ in optimum).

Two cases are possible:

1. $f_\mu(\mu) \geq \pi_H - \pi_L$
2. $f_\mu(\mu) < \pi_H - \pi_L$

If the first case applies, then

$$\Lambda_\mu(t,\mu) > 0 \Longrightarrow \Lambda_\mu(t-1,\mu') > 0$$

Assume *per contra* that $\Lambda_\mu(t-1,\mu') < 0$. But then by (strict) concavity of f, $f_\mu(\mu') > \pi_H - \pi_L$. Which, coupled with the expression of the derivative in (30), implies that for the assumption to be true, it must be that $\Lambda_\mu(t-2,\mu'') < 0$, and so on, until

$$f_\mu(\mu^{(t-2)})e = \Lambda_\mu(2,\mu^{(t-2)}) < 0 < \pi_H - \pi_L < f_\mu(\mu^{(t-2)})$$

Therefore, I have reached a contradiction.

Now, I am ready to prove by induction that

$$\Lambda_\mu(t,\mu) > 0 \wedge f_\mu(\mu) \geq \pi_H - \pi_L \Longrightarrow \frac{\partial}{\partial \mu}\Big(\frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)}\Big) < 0$$

Base Step $(t = 2)$

Notice that the result is always true for $t = 2$, as the expression reads

$$\frac{e/\beta + \Lambda(2,\mu)}{\pi(\mu)} = \frac{e}{\pi(\mu)}\Big(\frac{1}{\beta} - 1\Big) + \frac{e}{\pi(\mu')}$$

Induction Step

Assume *per contra* that

$$\Lambda_\mu(t,\mu) > 0 \;\wedge\; f_\mu(\mu) \geq \pi_H - \pi_L \;\wedge\; \frac{\partial}{\partial \mu}\Big(\frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)}\Big) > 0$$

Since the first two addends of (29) have been shown to be decreasing in μ, for it to be true it must be that $\frac{\partial}{\partial \mu'}\Big(\frac{e/\beta+\Lambda(t-1,\mu')}{\pi(\mu')}\Big) > 0$. However,

$$\Lambda(t,\mu) > 0 \wedge f_\mu(\mu) \geq \pi_H - \pi_L \Longrightarrow \Lambda_\mu(t-1,\mu') > 0 \wedge f_\mu(\mu') > \pi_H - \pi_L \Longrightarrow \frac{\partial}{\partial \mu'}\Big(\frac{e/\beta + \Lambda(t-1,\mu')}{\pi(\mu')}\Big) < 0$$

where the second implication follows by induction hypothesis. Hence the contradiction.

What is left to be shown is the following:

$$\Lambda_\mu(t,\mu) > 0 \wedge f_\mu(\mu) < \pi_H - \pi_L \Longrightarrow \frac{\partial}{\partial \mu}\Big(\frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)}\Big) < 0$$

Base Step $(t = 2)$

Same as in the case above, as the thesis always applies.

Induction Step

The derivative by μ has the following expression

$$\frac{\partial}{\partial\mu}\Big(\frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)}\Big) = \frac{1}{\pi(\mu)}\Big[\Lambda_\mu(t,\mu) - (\pi_H - \pi_L)\frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)}\Big]$$

So assume *per contra* that it is positive. Then this means that $\Lambda_\mu(t,\mu) > (\pi_H - \pi_L)\frac{e/\beta+\Lambda(t,\mu)}{\pi(\mu)}$. Moreover, by (29), it either means that $\Lambda_\mu(t-1,\mu') < 0$ or that

$$\frac{\partial}{\partial\mu'}\Big(\frac{e/\beta + \Lambda(t-1,\mu')}{\pi(\mu')}\Big) > 0$$

The first case can not apply, as (30) would imply that

$$(\pi_H - \pi_L)\frac{e}{\beta\pi(\mu)} < (\pi_H - \pi_L)\frac{e/\beta + \Lambda(t,\mu)}{\pi(\mu)} < \Lambda_\mu(t,\mu) < f_\mu(\mu)e < (\pi_H - \pi_L)e$$

which is impossible, as $\frac{1}{\beta\pi(\mu)} > 1$. Therefore, it must be the case that

$$\frac{\partial}{\partial\mu'}\Big(\frac{e/\beta + \Lambda(t-1,\mu')}{\pi(\mu')}\Big) > 0 \Longrightarrow \Lambda_\mu(t-1,\mu') > (\pi_H - \pi_L)\frac{e/\beta + \Lambda(t-1,\mu')}{\pi(\mu')} > 0$$

Now, if $f_\mu(\mu') \geq \pi_H - \pi_L$, I have reached a contradiction, since I have shown above that

$$\Lambda_\mu(t-1,\mu') > 0 \wedge f_\mu(\mu') \geq \pi_H - \pi_L \Longrightarrow \frac{\partial}{\partial\mu'}\Big(\frac{e/\beta + \Lambda(t-1,\mu')}{\pi(\mu')}\Big) < 0$$

If, instead, $f_\mu(\mu') < \pi_H - \pi_L$, then

$$\Lambda_\mu(t-1,\mu') > 0 \wedge f_\mu(\mu') < \pi_H - \pi_L \Longrightarrow \frac{\partial}{\partial\mu'}\Big(\frac{e/\beta + \Lambda(t-1,\mu')}{\pi(\mu')}\Big) < 0$$

by induction hypothesis, and a contradiction is reached in this case, too.

■

Proof of Prop. 11

Define $\Lambda(t,\mu)$ as the learning rents necessary to implement UI for t prospective periods ahead, and notice that, if $\mu_f = 0$, $\Lambda(t,\mu_f) = 0$. Then, from the definition of IP, the first-order condition reads

$$\frac{\partial q}{\partial \mu_p}\big[(1-\pi(\mu_p))V(\mu_p', U_p^u) - (1-\pi_L)V(0, U_f^u)\big] - q(\pi_H - \pi_L)V(\mu_p', U_p^u) +$$
$$+q\frac{(1-\pi_H)(1-\pi_L)}{1-\pi(\mu_p)}V_\mu(\mu_p', U_p^u) + \lambda\Big[\frac{\partial q}{\partial \mu_p}\big[(1-\pi(\mu_p))U_p^u - (1-\pi_L)U_f^u\big] - q(\pi_H - \pi_L)U_p^u\Big] -$$
$$-\chi\Big[\frac{\partial q}{\partial \mu_p}\Lambda(t, \mu_p) + q\Lambda_\mu(t, \mu_p)\Big] = 0$$

Which can be rewritten as

$$V_\mu(\mu_p', U_p^u) = \frac{V(\mu_p', U_p^u) - V(0, U_f^u)}{\mu_p'} + \Big(\lambda - \frac{\chi}{1-\pi_L}\Big)\frac{U_p^u - U_f^u}{\mu_p'} + \frac{\chi}{1-\pi_L}\frac{\mu_p\Lambda_\mu(t, \mu_p) - \Lambda(t, \mu_p)}{\mu_p'}$$

and plugging in $-V_U^{SA}(U_f^u) = -V_U(0, U_f^u) = \lambda - \frac{\chi}{1-\pi_L}$ and $\lambda = -W_U(\mu, U^w)$ delivers the result.

Appendix E: Estimation of hazard rates

In order to infer the hazard rates $\{\pi_H,\, \pi_L\}$, I proceed as follows. First, from the basic monthly Current Population Survey (CPS), I derive the fraction of high- and low-skilled workers for each level of educational attainment θ_i, $i \in \{LessHighSc.,\, HighSc.,\, SomeCollege,\, College,\, Graduate\}$.[38] Then, I compute the hazard rate out of unemployment for each time horizon $(\pi_t)_{t\geq 1}$, from the cross-section of jobless workers who report to have been unemployed for t periods of time, using the following formulas

$$\pi_1 = 1 - Prob(t>1) = 1 - \frac{\#\, jobless\, for\, t > 1}{\# jobless}$$
$$\pi_1 + (1-\pi_1)\pi_2 = 1 - Prob(t>2) = 1 - \frac{\#\, jobless\ for\ t > 2}{\# jobless}$$

....

Third, by looking at the same cross-sections, I compute the share of those with same spell duration (at the time the survey is conducted) who also have attained the same educational level, $\psi_{i,t}$. Lastly, I compute $\{\pi_H,\, \pi_L\}$ that minimize

$$\{\pi_H,\, \pi_L\} = \arg\min_{\hat{\pi}_H, \hat{\pi}_L} \sum_t \big(\sum_i \psi_{i,t}(\theta_i\hat{\pi}_H + (1-\theta_i)\hat{\pi}_L) - \pi_t\big)^2$$

that is,

$$\pi_H = \frac{\sum_t b_t \sum_s \pi_s a_s - \sum_s \pi_s \sum_t a_t b_t}{12\sum_t a_t^2 - (\sum_t a_t)^2}, \quad \pi_L = \frac{(\sum_t \pi_t)(\sum_s a_s^2) - \sum_s \pi_s a_s \sum_t a_t}{12\sum_t a_t^2 - (\sum_t a_t)^2}$$

[38] High-skilled workers are defined as those who earn a wage higher than the mean of ω_H and ω_L, that is, \$2,527.

with $a_t = \sum_i \psi_{it}\theta_i,\ b_t = \sum_i \psi_{it}(1-\theta_i) = 1 - a_t$.[39] The results are reported in Table 5. The hazard rate π_t is quite stable over time, as well as the share of any education level among all jobless people with same duration of unemployment spell, ψ_{it}. The estimated hazard rates are $\pi_H = 0.27$ and $\pi_L = 0.14$.

	Total	< High Sch.	High Sch. D.	< Col.[40]	Col. D.	Grad. D.	
θ_i	39,333	0.54	0.72	0.76	0.9	0.95	
Horizon	Total	ψ_{it}=Pr(Education=i $\mid$ Horizon $\geq t$)					Haz. Rate (π_t)
t=1	3,481	0.11	0.31	0.29	0.28	0.01	0.22
t=2	2,517	0.11	0.32	0.29	0.27	0.01	0.28
t=3	1,742	0.11	0.32	0.29	0.27	0.01	0.31
t=4	1,316	0.11	0.32	0.29	0.28	0.01	0.24
t=5	1,081	0.11	0.32	0.29	0.28	0	0.18
t=6	815	0.11	0.33	0.28	0.27	0	0.25
t=7	586	0.12	0.33	0.28	0.27	0	0.28
t=8	468	0.12	0.33	0.28	0.27	0	0.2
t=9	356	0.11	0.32	0.29	0.27	0	0.24
t=10	274	0.11	0.31	0.28	0.29	0	0.23
t=11	215	0.11	0.33	0.26	0.29	0	0.22
t=12	167	0.11	0.34	0.25	0.3	0	0.22

Table 5: Education-cohort size for any unemployment spell duration.

[39] First-order conditions for π_H and π_L return the minimizers of the convex objective function.

[40] '< Col.' item includes workers who attended college, but have not earned a degree, and workers with an Associate Degree, which is a post-secondary course of study lasting 2 or 3 years.

Technical Appendix

Setting

- $T < \infty$

- $\sigma_t \in \{u, w\}$ describes the worker status, either unemployed or employed. If $\sigma_t = w$, the worker finds reemployment, which is an absorbing state. Hence $p(\sigma_{t+1} = w|\sigma_t = w) = 1$.

- $\sigma^t = \{\sigma_0, ..., \sigma_t\}$ is a public history describing the employment status of the worker

- $c_t(\sigma^t)$ is the transfer function, with $c_t(\sigma^t) \geq 0$ for every σ^t. Let $\mathbf{c}(\alpha \backslash \sigma^t)$ be the stream of transfers downstream of node σ^t

- $a_t(\sigma^t)$ is the effort level, with

$$a_t(\sigma^t) \in \begin{cases} \{0, e\}, & \text{if} \quad \sigma_t = u \\ e, \quad \text{if} \quad \sigma_t = w \end{cases}$$

  The effort is unobservable by the government. Denote by $\mathbf{a}(\alpha \backslash \sigma^t)$ the continuation plan of effort costs downstream of node σ^t, and $\mathbf{a}(\sigma^t)$ its upstream counterpart

- $h \in \{L; H\}$ is the hidden state, which is revealed once W finds reemployment

- $\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))$ -with $\sigma^t = (\sigma^{t-1}, \sigma_t)$- is the expectation held by W during unemployment, expressing the probability about state H. This is clearly a non-contractible variable, as W can hide it from G. $\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))$ impacts the probability of future σ_{t+1}. in particular,

$$p(\sigma_{t+1} = w|\sigma_t = u, \mu_t, a_t(\sigma^t) = e) = \pi(\mu_t), \quad p(\sigma_{t+1} = w|\sigma_t = u, \mu_t, a_t(\sigma^t) = 0) = 0$$

  where I have dropped dependence of μ_t by $(\sigma^t, \mathbf{a}(\sigma^{t-1}))$ to ease notation. Moreover, $\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))$ undergoes an updating process every time W exerts effort in t and remains unemployed in $t + 1$

$$\mu_{t+1}(\sigma^t, \mathbf{a}(\sigma^{t-1}), \sigma_{t+1} = u, a_t(\sigma^t) = e) = \frac{\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))(1 - \pi_H)}{\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1})))(1 - \pi_H) + (1 - \mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1})))(1 - \pi_L)} \tag{31}$$

  Instead, if no effort is exerted, W does not revise expectation[41]

$$\mu_{t+1}(\sigma^t, \mathbf{a}(\sigma^{t-1}), \sigma_{t+1} = u, a_t(\sigma^t) = 0) = \mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1})) \tag{32}$$

[41] Notice that $\mu_{t+1}(\sigma^t, \mathbf{a}(\sigma^{t-1}), \sigma_{t+1} = w, a_t(\sigma^t) = 0)$ is not defined as h is disclosed once $\sigma_{t+1} = w$.

- If $\sigma_s = w$,

$$r_s(\sigma^s, \mathbf{a}(\sigma^{s-1})) = \tilde{\omega}(\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))), \quad \text{if } t = \inf_s\{y_s = w\} - 1$$

Otherwise, if $\sigma_t = u$, $r_s(\sigma^s, \mathbf{a}(\sigma^{s-1})) = 0$.

Worker's Problem in UI

Let $\mathcal{W}(\sigma^t) = (\mathbf{c}, \mathbf{a})(\alpha \backslash \sigma^t) = \big\{c_s(\sigma^s), a_s(\sigma^s)\big\}_{s=t}^T$ denote the contract offered by G to W. W's expected utility reads

$$\begin{aligned}
&U_t(\mathcal{W}, \mathbf{a}(\sigma^{t-1}), \sigma^t) = \mathbf{E}\Big\{\sum_{s=t}^T \beta^{s-t}\big(u(c_s(\sigma^s)) - a_s(\sigma^s)\big)\Big|\mathcal{W}(\sigma^t), \mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))\Big\} + \\
&+ \beta^{T+1-t} \sum_{\sigma^{T+1}} p(\sigma^{T+1}|\sigma^t, \mu_t, \mathbf{a}(\sigma^T)) U_{T+1}(\sigma^{T+1}) \\
&= \sum_{s=t}^T \beta^{s-t} \sum_{\sigma^s} p(\sigma^s|\sigma^t, \mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1})), a_t(\sigma^t))\Big\{u(c_s(\sigma^s)) - a_s(\sigma^s)\Big|\mathcal{W}(\sigma^t)\Big\} \\
&+ \beta^{T+1-t} \sum_{\sigma^{T+1}} p(\sigma^{T+1}|\sigma^t, \mu_t, \mathbf{a}(\sigma^T)) U_{T+1}(\sigma^{T+1}) \\
&= u(c_t(\sigma^t)) - a_t(\sigma^t) + \\
&+ \beta\Big[p(\sigma_{t+1} = w|\sigma_t = u, \mu_t, a_t(\sigma^t)) \sum_{s=t+1}^T \beta^{s-(t+1)} \sum_{h \in \{H,L\}} p(h|\sigma_{t+1} = w, \mu_t)\Big\{u(c_s(\sigma^s)) - e\Big|\mathcal{W}'(\sigma^t, w, h)\Big\} + \\
&+ p(\sigma_{t+1} = u|\sigma_t = u, \mu_t, a_t(\sigma^t)) \sum_{s=t+1}^T \beta^{s-(t+1)} \times \\
&\qquad \times \sum_{\sigma^s} p(\sigma^s|\sigma_{t+1} = u, \mu_{t+1}^u, a_{t+1}(\sigma^t, u))\Big\{u(c_s(\sigma^s)) - a_s(\sigma^s)\Big|\mathcal{W}'(\sigma^t, u)\Big\}\Big] + \\
&+ \beta^{T+1-t} \sum_{\sigma^{t+1}} p(\sigma^{t+1}|\sigma^t, \mu_t, \mathbf{a}(\sigma^t)) \sum_{\sigma^{T+1}} p(\sigma^{T+1}|\sigma^{t+1}, \mu_{t+1}, \mathbf{a}(\sigma^T)) U_{T+1}(\sigma^{T+1}) \\
&= u(c_t(\sigma^t)) - e + \\
&+ \beta \sum_{h \in \{H,L\}} p(h|\mu_t) p(\sigma_{t+1} = w|h)\Big[\sum_{s=t+1}^T \beta^{s-(t+1)}\Big\{u(c_s(\sigma^s)) - e\Big|\mathcal{W}'(\sigma^t, w, h)\Big\} + \beta^{T-t} U_{T+1}(\sigma^t, \mathbf{w}, h)\Big] + \\
&+ \beta(1 - \pi(\mu_t))\Big[\sum_{s=t+1}^T \beta^{s-(t+1)} \sum_{\sigma^s} p(\sigma^s|\sigma_{t+1} = u, \mu_{t+1}^u, a_{t+1}(\sigma^t, u))\Big\{u(c_s(\sigma^s)) - a_s(\sigma^s)\Big|\mathcal{W}'(\sigma^t, u), \mu_{t+1}^u\Big\} + \\
&\qquad + \sum_{\sigma^{T+1}} p(\sigma^{T+1}|\sigma^{t+1}, \mu_{t+1}^u, \mathbf{a}(\sigma^T)) U_{T+1}(\sigma^{T+1})\Big] \\
&= u(c_t(\sigma^t)) - e + \beta\Big\{\pi(\mu_t)\Big[\frac{\mu_t \pi_H}{\pi(\mu_t)} U_{t+1}(\mathcal{W}', \mathbf{a}(\sigma^t), (\sigma^t, w, H)) + \frac{(1-\mu_t)\pi_L}{\pi(\mu_t)} U_{t+1}(\mathcal{W}', \mathbf{a}(\sigma^t), (\sigma^t, w, L))\Big] + \\
&\qquad + (1 - \pi(\mu_t)) U_{t+1}(\mathcal{W}', \mathbf{a}(\sigma^t), (\sigma^t, u))\Big\}
\end{aligned}$$

with $\mu_{t+1}^u = \mu_{t+1}(\sigma^t, \sigma_{t+1} = u, \mathbf{a}(\sigma^{t-1}), a_t(\sigma^t))$[42].

The IC constraint starting from time t reads

$$U_t(\mathcal{W}, \mathbf{a}(\sigma^{t-1}), \sigma^t) \geq U_t((\mathbf{c}, \hat{\mathbf{a}})(\alpha \backslash \sigma^t), \mathbf{a}(\sigma^{t-1}), \sigma^t), \quad \forall\, \hat{\mathbf{a}}(\alpha \backslash \sigma^t) \in A_t(\alpha \backslash \sigma^t) \tag{33}$$

Government's Problem in UI

The problem for G reads

$$V^{UI}(U, \mathbf{a}(\sigma^{t-1}), \sigma^t) = \max_{\mathcal{W}} \mathbf{E}\Big\{ \sum_{s=t}^{T} \beta^{s-t} \big(r_s(\sigma^s, \mathbf{a}(\sigma^{s-1})) - c_s(\sigma^s)\big) \Big| \mathcal{W}(\sigma^t), \mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))\Big\}$$
$$\text{sub:} \quad (33), \quad U_s(\mathcal{W}'(\sigma^s), \mathbf{a}(\sigma^{s-1}), \sigma^s) \geq U, \quad \forall s \geq t$$

Given that expectation is revised upon (failure and) effort exerted only in the last period, it follows a Markovian process, meaning that expectation in $t+1$ can be predicted by expectation μ_t and effort in t, and realization of σ_{t+1}. Thus, I define $x_t = (\mu_t(\sigma^t, x_{t-1}), a_t(\sigma^t))$ and write

$$\mu_{t+1}(\sigma^t, \sigma_{t+1} = u, \mathbf{a}(\sigma^t)) = \mu_{t+1}(\sigma^t, \sigma_{t+1} = u, \mu_t(\sigma^t, x_{t-1}), a_t(\sigma^t)) = \mu_{t+1}(\sigma^t, \sigma_{t+1} = u, x_t)$$

And given that reemployment is absorbing and discloses the state, there exists an isomorphism between all unemployment histories $(\sigma^s, \sigma_{s+1} = u)$ and terminal realizations $\sigma_{s+1} = u$, as no extra information is contained in σ^s which can not be inferred by observing $\sigma_{s+1} = u$. As a result, next expectation μ_{t+1} only depends on current expectation μ_t and effort a_t and future realization of σ_{t+1}:

$$\mu_{t+1}(\sigma^t, \sigma_{t+1} = u, x_t) = \mu_{t+1}(\sigma_{t+1} = u, \mu_t, a_t) = \begin{cases} \mu_t, & \text{if} \quad a_t = 0 \\ \frac{\mu_t(1-\pi_H)}{1-\pi(\mu_t)}, & \text{if} \quad a_t = e \end{cases}$$

Therefore:

$$\begin{aligned} U_t(\mathcal{W}, \mu_t, \sigma^t) &= u(c_t(\sigma^t)) - e + \beta\Big\{\pi(\mu_t)\Big[\frac{\mu_t \pi_H}{\pi(\mu_t)} U_{t+1}(\mathcal{W}', H, \sigma^t, \sigma_{t+1} = w) + \frac{(1-\mu_t)\pi_L}{\pi(\mu_t)} U_{t+1}(\mathcal{W}', L, \sigma^t, \sigma_{t+1} = w \\ &\qquad + (1-\pi(\mu_t)) U_{t+1}(\mathcal{W}', \mu_{t+1}^u, \sigma^t, \sigma_{t+1} = u)\Big\} \\ &= u(c_t(\sigma^t)) - e + \beta\Big\{\mu_t \pi_H U_{t+1}(\mathcal{W}', H, \sigma^t, \sigma_{t+1} = w) + (1-\mu_t)\pi_L U_{t+1}(\mathcal{W}', L, \sigma^t, \sigma_{t+1} = w) + \\ &\qquad + (1-\pi(\mu_t)) U_{t+1}(\mathcal{W}', \mu_{t+1}^u, \sigma^t, \sigma_{t+1} = u)\Big\} \end{aligned}$$

[42] Note that by assumption on absorbing nature of re-employment, $\sigma^s = (\sigma^t, (y_j = w)_{t+1}^s), \quad \forall\, \sigma^s \succsim (\sigma^t, \sigma_{t+1} = w)$.

G's problem can be rewritten as

$$V^{UI}(U,\mu_t,\sigma^t) = \max_{\mathcal{W}\in\Omega(\mu_t,\sigma^t)} \mathbf{E}\Big\{\sum_{s=t}^{T}\beta^{s-t}\big(r_s(\sigma^s,\mu_s)-c_s(\sigma^s)\big)\Big|\mathcal{W}(\sigma^t),\mu_t\Big\}$$

$$= \max_{c_t(\sigma^t),a_t(\sigma^t)\in\Gamma(\mu_t,\sigma^t)} r_t(\sigma^t,\mu_t)-c_t(\sigma^t)+$$

$$+\beta\Big[\pi(\mu_t)\max_{\mathcal{W}'\in\Omega'(\mu_{t+1},\sigma^t,\sigma_{t+1}=w)}\mathbf{E}\Big\{\sum_{s=t+1}^{T}\beta^{s-(t+1)}\big(\overbrace{r_s(\sigma^s,\mu_s)}^{=\tilde{\omega}(\mu_t)}-c_s(\sigma^s)\big)\Big|\mathcal{W}'(\sigma^t,\sigma_{t+1}=w),\mu_{t+1}\Big\}+$$

$$+(1-\pi(\mu_t))\max_{\mathcal{W}'\in\Omega'(\mu_{t+1},\sigma^t,\sigma_{t+1}=u)}\mathbf{E}\Big\{\sum_{s=t+1}^{T}\beta^{s-(t+1)}\big(r_s(\sigma^s,\mu_s)-c_s(\sigma^s)\big)\Big|\mathcal{W}'(\sigma^t,\sigma_{t+1}=u),\mu_{t+1}\Big\}\Big]$$

$$\text{sub:}\quad U_t(\mathcal{W},\mu_t,\sigma^t)\geq U_t((\mathbf{c},\mathbf{a}')(\alpha\backslash\sigma^t),\mu_t,\sigma^t),\quad \forall\,\mathbf{a}'(\alpha\backslash\sigma^t)\in A_t(\alpha\backslash\sigma^t)\quad (IC)$$

$$\mu_t := \mu_t(\sigma^t,\mu_{t-1},a_{t-1}(\sigma^{t-1}))$$

Lemma 4. *G prefers to insure W against the risk of h realization upon reemployment.*

Lemma 5. *Define (IC, s) the constraint that makes contract $\mathcal{W}$ robust to the alternative strategy $\mathbf{a}'(\alpha\backslash\sigma^s) = (0, e\mathbf{1}_{T-s}) \in A_s(\alpha\backslash\sigma^s)$ that shirks in s and sticks to effort from $s+1$ to the final period T*

$$U_s(\mathcal{W},\mu_s,\sigma^s)\geq U_s((\mathbf{c},\mathbf{a}')(\alpha\backslash\sigma^s),\mu_s,\sigma^s) = u(c_s(\sigma^s))+\beta U_{s+1}(\mathcal{W},\mu_s,(\sigma^s,u))$$

If (IC, s)$_{s=t}^{T}$ are all binding under contract $\mathcal{W}$, then contract $\mathcal{W}$ is feasible.

Proof. First, consider that no learning motive or moral hazard problem is present upon reemployment, when state is disclosed, nor there is any chance that any reemployed W falls back into unemployment ($p(y_s = u|y_t = w) = 0,\ \forall s > t$). Hence in order to verify (33), one can only focus on continuation histories $\sigma^s \succsim \sigma^t$ where $\sigma^s = \big(\sigma^t,(y_j)_{j=t+1}^s\big) = (\sigma^t,(u)_{j=t+1}^s)$. For this reason, I therefore adopt the convention that $\sigma^{s+1} = (\sigma^s, u)$. Define also μ_s as the expectation in period s if contract $\mathcal{W}$ is followed.

Notice that continuation utility at time $s > t$ upon reemployment ($y_s = w$) follows

$$U_s(\mathcal{W},h,\sigma^s) = u(c_s(\sigma^s)) - e + \beta U_{s+1}(\mathcal{W},h,(\sigma^s,w))$$

while upon failure ($y_s = u$), it follows

$$U_s(\mathcal{W},\mu_s,\sigma^s) = u(c_s(\sigma^s)) - e + \beta\Big[\mu_s\pi_H U_{s+1}(\mathcal{W},H,(\sigma^s,w)) + (1-\mu_s)\pi_L U_{s+1}(\mathcal{W},L,(\sigma^s,w))+$$
$$+(1-\pi(\mu_s))U_{s+1}(\mathcal{W},\mu^u_{s+1},(\sigma^s,u))\Big]$$
$$\text{with:}\quad \mu^u_{s+1} = \mu_{s+1}(\sigma^s,\sigma_{s+1}=u,\mu_s,e)$$

By Lemma 4, focusing on contracts $\mathcal{W}$ such that

$$
\begin{aligned}
&U_{s+1}(\mathcal{W}, H, (\sigma^s, w)) = U_{s+1}(\mathcal{W}, L, (\sigma^s, w)) \\
\Longrightarrow &\mu\pi_H U_{s+1}(\mathcal{W}, H, (\sigma^s, w)) + (1-\mu)\pi_L U_{s+1}(\mathcal{W}, L, (\sigma^s, w)) = \pi(\mu) U_{s+1}(\mathcal{W}, (\sigma^s, w))
\end{aligned}
$$

is without loss of generality.

Second, the following holds true:

$$
\begin{aligned}
&U_s(\mathcal{W}, \mu_s, \sigma^s) = u(c_s(\sigma^s)) + \beta U_{s+1}(\mathcal{W}, \mu_s, \sigma^{s+1}) \\
\Longrightarrow &U_s(\mathcal{W}, \mu, \sigma^s) \geq u(c_s(\sigma^s)) + \beta U_{s+1}(\mathcal{W}, \mu, \sigma^{s+1}), \quad \forall \mu : \mu > \mu_s
\end{aligned}
\tag{34}
$$

The proof of (34) will be given by induction, joint with the main statement.

<u>Base Step ($t = T$)</u>

UI contract ends in $t = T$, where the only possible deviation is $\hat{\mathbf{a}}(\alpha \backslash \sigma^T) = \hat{a}_T(\sigma^T) = 0$, Thus, for W to be robust to this deviation, it must be that

$$
\begin{aligned}
U_T(\mathcal{W}, \mu_T, \sigma^T) = u(c_T(\sigma^T)) - e + \beta\big[&\mu_T \pi_H U_{T+1}(\sigma^T, w, H) + (1-\mu_T)\pi_L U_{T+1}(\sigma^T, w, L) + \\
&+ (1 - \pi(\mu_T)) U_{T+1}(\sigma^T, u)\big]
\end{aligned}
$$

$$
U_T(\mathcal{W}, \mu_T, \sigma^T) \geq u(c_T(\sigma^T)) + \beta U_{T+1}(\sigma^T, u)
$$

Since (IC, T) is binding by assumption, it holds

$$
U_T(\mathcal{W}, \mu_T, \sigma^T) = u(c_T(\sigma^T)) + \beta U_{T+1}(\sigma^T, u), \quad U_{T+1}(\sigma^T, w) - U_{T+1}(\sigma^T, u) = \frac{e}{\beta \pi(\mu_T)} > 0
$$

and then, for $\mu > \mu_T$,

$$
\begin{aligned}
&U_T(\mathcal{W}, \mu, \sigma^T) - \beta U_{T+1}(\mathcal{W}, \mu, (\sigma^T, u)) = \\
=&u(c_T(\sigma^T)) - e + \beta\pi(\mu)\big[U_{T+1}(\mathcal{W}, (\sigma^T, w)) - U_{T+1}(\mathcal{W}, (\sigma^T, u))\big] > \\
>&u(c_T(\sigma^T)) - e + \beta\pi(\mu_T)\big[U_{T+1}(\mathcal{W}, (\sigma^T, w)) - U_{T+1}(\mathcal{W}, (\sigma^T, u))\big] = \\
=&U_T(\mathcal{W}, \mu_T, \sigma^T) - \beta U_{T+1}(\mathcal{W}, \mu_T, (\sigma^T, u))
\end{aligned}
$$

which proves (34) for $t = T$.

<u>Induction Step ($t \leq T - 1$)</u>

First, notice that

$$\begin{aligned}
&U_t(\mathcal{W},\mu_t,\sigma^t) = u(c_t(\sigma^t)) + \beta U_{t+1}(\mathcal{W},\mu_t,\sigma^{t+1})\\
\Longrightarrow &-e+\beta\pi(\mu_t)\big[U_{t+1}(\mathcal{W},(\sigma^t,w)) - U_t(\mathcal{W},\mu_{t+1},\sigma^{t+1})\big]\\
&= \beta\big[U_{t+1}(\mathcal{W},\mu_t,\sigma^{t+1}) - U_{t+1}(\mathcal{W},\mu_{t+1},\sigma^{t+1})\big]\\
&= \beta\big[-e+\beta\pi(\mu_t)\big(U_{t+2}(\mathcal{W},(\sigma^{t+1},w)) - U_{t+2}(\mathcal{W},\mu_{t+1},\sigma^{t+2}\big)\big]\\
\Longrightarrow &U_{t+1}(\mathcal{W},(\sigma^t,w)) - U_{t+1}(\mathcal{W},\mu_{t+1},\sigma^{t+1}) > \beta\big[U_{t+2}(\mathcal{W},(\sigma^{t+1},w)) - U_{t+2}(\mathcal{W},\mu_{t+1},\sigma^{t+2})\big]
\end{aligned} \tag{35}$$

where the equalities hold since (IC, t) and (IC, $t+1$) are binding. Now, by induction hypothesis, $\forall\mu:\ \mu>\mu_{t+1}$,

$$\begin{aligned}
&U_{t+1}(\mathcal{W},\mu_{t+1},\sigma^{t+1}) = u(c_{t+1}(\sigma^{t+1})) + \beta U_{t+2}(\mathcal{W},\mu_{t+1},\sigma^{t+2})\\
\Longrightarrow &U_{t+1}(\mathcal{W},\mu,\sigma^{t+1}) \geq u(c_{t+1}(\sigma^{t+1})) + \beta U_{t+2}(\mathcal{W},\mu,\sigma^{t+2})
\end{aligned}$$

Therefore, take any $\mu>\mu_t$ and

$$\begin{aligned}
&U_t(\mathcal{W},\mu_t,\sigma^t) - \beta U_{t+1}(\mathcal{W},\mu_t,\sigma^{t+1}) = u(c_t(\sigma^t)) - \beta u(c_{t+1}(\sigma^{t+1})) - e(1-\beta)+\\
&+\beta\pi(\mu_t)\big[U_{t+1}(\mathcal{W},(\sigma^t,w)) - \beta U_{t+2}(\mathcal{W},(\sigma^{t+1},w))\big] + \beta(1-\pi(\mu_t))\big[U_{t+1}(\mathcal{W},\mu_t^u,\sigma^{t+1}) - \beta U_{t+2}(\mathcal{W},\mu_t^u,\sigma^{t+2})\big] <\\
&<u(c_t(\sigma^t)) - \beta u(c_{t+1}(\sigma^{t+1})) - e(1-\beta)+\\
&+\beta\pi(\mu)\big[U_{t+1}(\mathcal{W},(\sigma^t,w)) - \beta U_{t+2}(\mathcal{W},(\sigma^{t+1},w))\big] + \beta(1-\pi(\mu))\big[U_{t+1}(\mathcal{W},\mu_t^u,\sigma^{t+1}) - \beta U_{t+2}(\mathcal{W},\mu_t^u,\sigma^{t+1})\big]\\
&<u(c_t(\sigma^t)) - \beta u(c_{t+1}(\sigma^{t+1})) - e(1-\beta)+\\
&+\beta\pi(\mu)\big[U_{t+1}(\mathcal{W},(\sigma^t,w)) - \beta U_{t+2}(\mathcal{W},(\sigma^{t+1},w))\big] + \beta(1-\pi(\mu))\big[U_{t+1}(\mathcal{W},\mu^u,\sigma^{t+1}) - \beta U_{t+2}(\mathcal{W},\mu^u,\sigma^{t+2})\big]\\
&=U_t(\mathcal{W},\mu,\sigma^t) - \beta U_{t+1}(\mathcal{W},\mu,\sigma^{t+1})
\end{aligned}$$

where the first inequality follows from (35) above, as $\pi(\mu)>\pi(\mu_t)$, while the second inequality follows from induction hypothesis. I can thus conclude that (34) holds also for t.

I now pass to the proof of the main part of the proposition, that is, that binding IC constraints is a sufficient condition to account for all possible deviations occurring from t onward. By induction hypothesis, $\mathcal{W}$ satisfies all $(IC,\ s)_{s=t}^T$ with equality, and that guarantees robustness to all possible deviations over histories in $\alpha\backslash\sigma^{t+1}$, i.e.

$$U_{t+1}(\mathcal{W},\mu_{t+1},\sigma^{t+1}) \geq U_{t+1}((\mathbf{c},\mathbf{a}')(\alpha\backslash\sigma^{t+1}),\mu_{t+1},\sigma^{t+1}),\quad \forall\ \mathbf{a}'(\alpha\backslash\sigma^{t+1}) \in A_{t+1}(\alpha\backslash\sigma^{t+1})$$

What it is to show is that $\mathcal{W}$ is robust also to all possible deviations in $\alpha\backslash\sigma^t$, i.e.

$$U_t(\mathcal{W},\mu_t,\sigma^t) \geq U_t((\mathbf{c},\mathbf{a}')(\alpha\backslash\sigma^t),\mu_t,\sigma^t), \quad \forall\, \mathbf{a}'(\alpha\backslash\sigma^t) \in A_t(\alpha\backslash\sigma^t)$$

First of all, notice that $A_t(\alpha\backslash\sigma^t) = \{0,e\} \times A_{t+1}(\alpha\backslash\sigma^{t+1})$ can be decomposed into:

- all effort histories with positive effort in t, i.e. $A_e = A_t(\alpha\backslash\sigma^t) \cap \{a_t(\sigma^t) = e\}$;
- all effort histories with zero effort in t, i.e. $A_0 = A_t(\alpha\backslash\sigma^t) \cap \{a_t(\sigma^t) = 0\}$;

Second, assumption on robustness to any $\mathbf{a}'(\alpha\backslash\sigma^{t+1}) \in A_{t+1}(\alpha\backslash\sigma^{t+1})$ guarantees robustness of $\mathcal{W}$ to the first set of deviations A_e, since $\mu_{t+1} = \frac{\mu_t(1-\pi_H)}{1-\pi(\mu_t)} = \mu_t^u$. Indeed, pick any $\mathbf{a}'(\alpha\backslash\sigma^t) \in A_e$. Then, it follows

$$\begin{aligned} U_t(\mathcal{W},\mu_t,\sigma^t) &= u(c_t(\sigma^t)) - e + \beta\big[\pi(\mu^t)U_{t+1}(\mathcal{W},\mu_t^w,(\sigma^t,w)) + (1-\pi(\mu^t))U_{t+1}(\mathcal{W},\mu_t^u,\sigma^{t+1})\big] \\ \geq& u(c_t(\sigma^t)) - e + \beta\big[\pi(\mu^t)U_{t+1}(\mathcal{W},\mu_t^w,(\sigma^t,w)) + (1-\pi(\mu^t))U_{t+1}(\mathcal{W}',\mu_t^u,\sigma^{t+1})\big] = U_t(\mathcal{W}',\mu_t,\sigma^t) \end{aligned}$$

where the inequality follows from robustness to $\mathbf{a}'(\alpha\backslash\sigma^{t+1}) \in A_{t+1}(\alpha\backslash\sigma^{t+1})$.
What is left to show is robustness of $\mathcal{W}$ to A_0. By assumption, (IC, t) and (IC, $t+1$) are binding, which means that

$$U_t(\mathcal{W},\mu_t,\sigma^t) = U_t((\mathbf{c},\tilde{\mathbf{a}})(\alpha\backslash\sigma^t),\mu_t,\sigma^t) = U_t((\mathbf{c},\hat{\mathbf{a}})(\alpha\backslash\sigma^t),\mu_t,\sigma^t) \tag{36}$$

with $\hat{\mathbf{a}}(\alpha\backslash\sigma^t) = (0, e\mathbf{1}_k)$, $\tilde{\mathbf{a}}(\alpha\backslash\sigma^t) = (e,0,e\mathbf{1}_{k-1})$. Define $\tilde{\mathcal{W}} = (\mathbf{c},\tilde{\mathbf{a}})(\alpha\backslash\sigma^t)$ and $\hat{\mathcal{W}} = (\mathbf{c},\hat{\mathbf{a}})(\alpha\backslash\sigma^t)$. Thus, by construction

$$U_{t+2}(\tilde{\mathcal{W}},\mu_t^u,\sigma^{t+2}) = U_{t+2}(\hat{\mathcal{W}},\mu_t^u,\sigma^{t+2}) \tag{37}$$

Indeed, both alternative strategies prescribe to set effort cost to 0 either at stage t or $t+1$ (but not both), and therefore the expectation at node $\sigma^{t+2} = (\sigma^t,u,u)$ is equal to μ_t^u under both strategies. Moreover, they prescribe positive effort forever after, until the last period T.
Pick any $\mathbf{a}'(\alpha\backslash\sigma^t) \in A_0$. There are two possibilities: $a'_{t+1}(\sigma^{t+1}) = e$ or $a'_{t+1}(\sigma^{t+1}) = 0$. If the first case applies, consider the alternative deviation strategy $\mathbf{a}''(\alpha\backslash\sigma^t) \in A_e$ so constructed:

$$a''_t(\sigma^t) = e, \quad a''_{t+1}(\sigma^{t+1}) = 0, \quad \mathbf{a}''(\alpha\backslash\sigma^{t+2}) = \mathbf{a}'(\alpha\backslash\sigma^{t+2})$$

Likewise, define $\mathcal{W}' = (\mathbf{c},\mathbf{a}')(\alpha\backslash\sigma^t)$ and $\mathcal{W}'' = (\mathbf{c},\mathbf{a}'')(\alpha\backslash\sigma^t)$. Hence, by construction,

$$U_{t+2}(\mathcal{W}',\mu_t^u,\sigma^{t+2}) = U_{t+2}(\mathcal{W}'',\mu_t^u,\sigma^{t+2}) \tag{38}$$

for the same reason as in (37), and

$$U_t(\mathcal{W}', \mu_t, \sigma^t) = U_t(\hat{\mathcal{W}}, \mu_t, \sigma^t) + \beta^2(1 - \pi(\mu_t))\big[U_{t+2}(\mathcal{W}', \mu_t^u, \sigma^{t+2}) - U_{t+2}(\hat{\mathcal{W}}, \mu_t^u, \sigma^{t+2})\big]$$
$$U_t(\mathcal{W}'', \mu_t, \sigma^t) = U_t(\tilde{\mathcal{W}}, \mu_t, \sigma^t) + \beta^2(1 - \pi(\mu_t))\big[U_{t+2}(\mathcal{W}'', \mu_t^u, \sigma^{t+2}) - U_{t+2}(\tilde{\mathcal{W}}, \mu_t^u, \sigma^{t+2})\big]$$

which follows from the fact that $\mathcal{W}'$ is identical to $\hat{\mathcal{W}}$ in periods t and $t+1$, and the same holds true for $\mathcal{W}''$ and $\tilde{\mathcal{W}}$.

One can easily see that the RHS of the two equations are equal, by (36), (37) and (38), which causes also the LHS to be equal

$$U_t(\mathcal{W}', \mu_t, \sigma^t) = U_t(\mathcal{W}'', \mu_t, \sigma^t)$$

But then, given that $\mathcal{W}$ is robust to any alternative strategy in A_e,

$$\mathbf{a}''(\alpha \backslash \sigma^t) \in A_e \Longrightarrow U_t(\mathcal{W}, \mu_t, \sigma^t) \geq U_t(\mathcal{W}'', \mu_t, \sigma^t) = U_t(\mathcal{W}', \mu_t, \sigma^t)$$

proving that $\mathcal{W}$ is robust to $\mathbf{a}'(\alpha \backslash \sigma^t)$, too.

Now, consider the case where $a'(\sigma^{t+1}) = 0$ and the strategies $\hat{\mathbf{a}}$ and $\tilde{\mathbf{a}}$ defined as above, and also $\ddot{\mathbf{a}}(\alpha \backslash \sigma^t) = (0, 0, e\mathbf{1}_{k-1})$. I first show that

$$U_t(\mathcal{W}, \mu_t, \sigma^t) \geq U_t(\ddot{\mathcal{W}}, \mu_t, \sigma^t)$$

under the assumption of (IC, t) being binding

$$U_t(\mathcal{W}, \mu_t, \sigma^t) = U_t(\hat{\mathcal{W}}, \mu_t, \sigma^t) = u(c_t(\sigma^t)) + \beta U_{t+1}(\hat{\mathcal{W}}, \mu_t, \sigma^{t+1})$$

which boils down to prove that

$$U_{t+1}(\hat{\mathcal{W}}, \mu_t, \sigma^{t+1}) \geq U_{t+1}(\ddot{\mathcal{W}}, \mu_t, \sigma^{t+1}) = u(c_{t+1}(\sigma^{t+1})) + \beta U_{t+2}(\ddot{\mathcal{W}}, \mu_t, \sigma^{t+2})$$
$$\Longrightarrow U_{t+1}(\mathcal{W}, \mu_t, \sigma^{t+1}) \geq u(c_{t+1}(\sigma^{t+1})) + \beta U_{t+2}(\mathcal{W}, \mu_t, \sigma^{t+2}) \tag{39}$$

where the first inequality follows from the fact that both strategies prescribe no effort in t, and the second inequality follows from the fact that $\hat{\mathcal{W}} = \mathcal{W}$ (resp., $\ddot{\mathcal{W}} = \mathcal{W}$) over $\alpha \backslash t+1$ (resp., $\alpha \backslash t+2$). By assumption, (IC, $t+1$) is binding

$$U_{t+1}(\mathcal{W}, \mu_t^u, \sigma^{t+1}) = u(c_{t+1}) + \beta U_{t+2}(\mathcal{W}, \mu_t^u, \sigma^{t+2})$$

which, jointly with (34) and since $\mu_t^u < \mu_t$, causes (39). Now, $\mathbf{a}'$ and $\ddot{\mathbf{a}}$ prescribe the same action in periods t and $t+1$. Therefore, in order to prove that $\mathcal{W}$ is robust against $\mathbf{a}'(\alpha\backslash\sigma^t)$, it is enough to show that

$$U_{t+2}(\mathcal{W},\mu_t,\sigma^{t+2}) = U_{t+2}(\ddot{\mathcal{W}},\mu_t,\sigma^{t+2}) \geq U_{t+2}(\mathcal{W}',\mu_t,\sigma^{t+2}) \tag{40}$$

Now, there are two possibilities, $a'_{t+2}(\sigma^{t+2})$ can either be 0 or e. If the first case occurs, in order to prove (40) it is enough to show

$$U_{t+3}(\mathcal{W},\mu_t,\sigma^{t+3}) \geq U_{t+3}(\mathcal{W}',\mu_t,\sigma^{t+3}) \tag{41}$$

Indeed, (IC, $t+2$) binding and (34) jointly cause

$$U_{t+2}(\mathcal{W},\mu_t,\sigma^{t+2}) \geq u(c_{t+2}(\sigma^{t+2})) + \beta U_{t+3}(\mathcal{W},\mu_t,\sigma^{t+3})$$

On the other hand, if $a'_{t+2}(\sigma^{t+2}) = e$, then

$$U_{t+2}(\breve{\mathcal{W}},\mu_t,\sigma^{t+2}) = u(c_{t+2}(\sigma^{t+2})) - e + \beta\big[\pi(\mu_t)U_{t+3}(\mathcal{W},(\sigma^{t+2},w)) + (1-\pi(\mu_t))U_{t+3}(\breve{\mathcal{W}},\mu_t^u,\sigma^{t+3})\big],$$
$$\breve{\mathcal{W}} = \{\mathcal{W},\mathcal{W}'\}$$

But then proving (40) boils down to show (41). I have just established the following implication

$$U_{j+1}(\mathcal{W},\mu'_{j+1},\sigma^{j+1}) \geq U_{j+1}(\mathcal{W}',\mu'_{j+1},\sigma^{j+1}) \Longrightarrow U_j(\mathcal{W},\mu'_j,\sigma^j) \geq U_j(\mathcal{W}',\mu'_j,\sigma^j), \quad \forall j:\, t \leq j \leq T$$

where μ'_j is the expectation in period j if strategy $\mathbf{a}'$ is applied. But then the proof is complete, as

$$U_{T+1}(\mathcal{W},\mu'_{T+1},\sigma^{T+1}) = U_{T+1}(\mathcal{W},\sigma^{T+1}) = U_{T+1}(\mathcal{W}',\mu'_{T+1},\sigma^{T+1}) \Longrightarrow U_t(\mathcal{W},\mu'_t,\sigma^j) \geq U_t(\mathcal{W}',\mu'_t,\sigma^t)$$

■

Lemma 5 proves to be useful in light of the following result.

Lemma 6. *In optimum, all (IC, s)$_{s=0}^T$ constraints are binding.*

Proof. By contradiction, assume that $\mathcal{W} = (\mathbf{c},\mathbf{a})(\alpha\backslash\sigma^0)$ is optimum and that (IC, t) is slack

$$U_t(\mathcal{W},\mu_t,\sigma^t) > u(c_t(\sigma^t)) + \beta U_{t+1}(\mathcal{W},\mu_t,(\sigma^t,u))$$

Then there exists $\varepsilon > 0$ such that

$$\begin{cases} c'_{t+1}(\sigma^t, w) = c_{t+1}(\sigma^t, w) - \varepsilon \\ U_t(\mathcal{W}', \mu_t, \sigma^t) = u(c_t(\sigma^t)) + \beta U_{t+1}(\mathcal{W}', \mu_t, \sigma^{t+1}) \end{cases}$$

where $\mathcal{W}' = (\mathbf{c}', \mathbf{a})(\alpha \backslash \sigma^0)$ is defined as

$$c'_s(\sigma^s) = c_s(\sigma^s),\ \forall \sigma^s \neq (\sigma^t, w), \quad c'_{t+1}(\sigma^t, w) = c_{t+1}(\sigma^t, w) - \varepsilon$$

Now, G's payoff is larger under $\mathcal{W}'$ than under $\mathcal{W}$, as payment to W in history (σ^t, w) is lower in the former case. Moreover, by Lemma 5, $\mathcal{W}'$ is also feasible, since it satisfies all (IC, $s)_{s=0}^T$ constraints with equality. But this contradicts that $\mathcal{W}$ is optimum. ■

Thus, robustness against all one-shot deviations from the prescribed effort sequence constitutes a necessary condition for a contract to be optimum (by Lemma 6) and sufficient one for it to be robust against any multiple deviation (by Lemma 5). Therefore, focusing on the set of contracts with such characteristic is without loss of generality.