SERGIO CAPPELLINI

University of Padova

OPTIMAL UNEMPLOYMENT INSURANCE WITH WORKER PROFILING

November 2022

Marco Fanno Working Papers – 294

Department of Economics and Management "Marco Fanno"
University of Padova
Via del Santo 33, 35123 Padova

Optimal Unemployment Insurance with Worker Profiling*

[Click here for the latest version]

Sergio Cappellini†

November 29, 2022

ABSTRACT

I study the design of an optimal profiling policy within welfare-to-work programs, by embedding dynamic learning about worker's ability into the principal-agent framework of Pavoni and Violante (2007). In optimal profiling, a fraction of low-skilled workers is persuaded to be highly skilled and referred to delegated search together with actual high-skilled workers (positive type-II error), whenever the government prefers to have over-optimist low-skilled workers searching for jobs (at small incentive costs) rather than referring them to passive labor-market policies. On the contrary, no high-skilled worker is ever profiled as low skilled and referred to any passive policy (zero type-I error). To ease agency costs, workers who are profiled on a statistical basis and referred to delegated search suffer a reduction in the generosity of transfers. In Florida, the optimal profiling program would generate around 540 millions USD savings each year.

JEL classification: D82, I38, 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, Alessia Russo, Nenad Kos, Antonella Trigari, and the 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.

†Postdoctoral Fellow, University of Padua, e-mail: sergio.cappellini@unipd.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, as it allows to tailor assistance to their needs and abilities. Profiling of welfare claimants is present in most OECD countries and is usually employed as a tool to support and improve the design of existing ALMPs.[2]

In the US, Worker Profiling and Reemployment Services (WPRS) and Reemployment and Eligibility Assessment (REA) are the two main Federal-funded programs dedicated to profiling of welfare claimants.[3] 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, they improve 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, WPRS and REA allow policymakers to design 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

[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,

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

US profiling programs are therefore ancillary to welfare policies, which deal with income support and provision of search incentives and assistance. Welfare support in the US is funded partly by the Federal government and partly by single States, while the organization and design is mainly delegated 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 not, and (v) what transfers they should be receive.

This paper develops a framework suitable to rigorously address each of these policy-relevant questions related to the design of a profiling program within a welfare-to-work context. An optimal unemployment insurance program minimizes the cost/maximizes the revenues of a risk-neutral public welfare provider (hereafter, 'the government'), while guaranteeing a certain standard of generosity of assistance to risk-averse recipients (hereafter, 'the workers'), that is proxied by their continuation utility. Workers and government have imperfect -but symmetric- knowledge about the ability of any worker to find a job. Abilities and skills are used as synonyms and refer to the reemployment perspectives and working productivity of each worker. Any failed attempt to find a job leads expectations about skills to be revised downward, as new information about worker's reemployment perspectives is acquired.

A welfare program can be framed as a dynamic agency problem, whose state is formed by expectations and promised utility, and where the government and the worker play the role of the principal and agent, respectively. This approach allows for a recursive formulation of the problem, as in Shavell and Weiss (1979), Hopenhayn and Nicolini (1997), where revision of expectations replaces human capital depreciation (Pavoni and Violante, 2007). According to the literature on dynamic UI (Pavoni et al., 2013), any policy instrument is a combination of (i) job search intensity, (ii) unemployment benefits, (iii) wage taxes (or subsidies) contingent on reemployment, (iv) worker profiling, and (v) job-search assistance, this latter encompassing all programs where employment offices assist/replace workers in the job search. Workers differ in their initial expectations and thus referred to different programs, which arise as a sequence of policies over time, until the worker exits unemployment.

Any profiling policy can be thought of as a test conducted over an unknown quantity -worker's skills- expectations about which are the sole information available beforehand. As in WPRS and REA programs, the test is used to determine whether a worker is better given job-search assistance, and whether there exist overpayments to eliminate. Test accuracy can be read in terms of type-I and type-II errors. The paper finds out that type-I error must always be minimized, i.e.

---

https://www.dol.gov/agencies/eta/american-job-centers/RESEA). Hence, the two versions of the program have similar targets.

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

the test must avoid to profile a high-skilled worker as low skilled. On the contrary, for low values of generosity, the optimal welfare-to-work program must profile a share of low-skilled workers as high skilled so to raise the expected return of their job search and lower the cost of search incentives for the government. This finding derives from the shape of incentive costs and a well established result in the context of Bayesian persuasion (Kamenica and Gentzkow, 2011) that introduces asymmetries between type I and type II errors in order to align incentives between principal and agent. The paper also states sufficient conditions for the cost of search incentives to increase in the generosity of payments. For this reason, the government lowers the payments to jobseekers who are profiled as high skilled, in the attempt to contrast the rise in agency cost due to positive type-II error and ease incentive provision. This finding rationalizes a feature of the REA program, that spots workers who are likely to exit unemployment on their own, with the aim to lower their unemployment benefits.

The second objective of this paper is to construct a benchmark to evaluate existing programs, as well as an estimate of the welfare gains of profiling (in the form of budget savings). To this aim, the paper solves for the optimal program with and without profiling, and uses the difference of returns in the two cases as a proxy of the value of profiling. The welfare gains for the State of Florida are estimated to be in the order of millions of dollars, a result which is comparable to the estimate contained in Poe-Yamagata et al. (2011). The analysis can be applied to all existing welfare programs, based on the generosity implicit in their transfers and the reemployment expectations of their recipients. While previous performance assessments of WPRS and REA focus on specific margins and targets (see Section 2), the advantage of this paper's approach, instead, lies in the absence of any arbitrary choice, 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 test-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-to-work program. The paper provides an analysis of the gains and losses of profiling, in conjunction with others labor-market policies, when workers' skills are imperfectly observed by the government and workers themselves.

Profiling consists of information detection about any welfare recipient over dimensions that are relevant for government's targets. Berger et al. (2000) outline the trade-off between equity and efficiency in the design of statistical profiling programs. Depending on the ultimate goal of the welfare provider, profiling implements different allocation mechanisms and targets different characteristics of recipients (for ex., likelihood to exhaust UI benefits). Attempts have been made to estimate returns of existing profiling programs. Berger et al. (2000) find evidence that Kentucky WPRS is successful in targeting long-term unemployed, but not so much in increasing their reemployment rates, while Black et al. (2007) estimate a positive return in terms of shorter spells, lower UI benefits and higher earnings. Other works wonder about the optimal design of profiling programs. Sullivan et al. (2007) rank 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). I find that such a choice has a rationale. 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. Poe-Yamagata et al. (2011) 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 estimates of cost savings have the same order of magnitude (millions of US Dollars, see Section 7.4) of this paper's estimates.

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. As human capital of jobless workers depreciates over time (for ex., because lack of training harms on-the-job productivity), workers are efficiently reassigned to different policies during the unemployment spell. Likewise, in this paper any transition to a different policy follows the deterioration of expected reemployment perspectives. Yet, such a deterioration stems from an endogenous learning process which brings agents to revise their

[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).

initial expectations. 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. It has long been acknowledged that jobseekers learn about occupations in a gradual time-consuming way (Miller, 1984; Neal, 1999; Gibbons and Waldman, 1999; Gibbons et al., 2005; Papageorgiou, 2014; Groes et al., 2015) and revise job expectations accordingly (Kudlyak et al., 2014; Belot et al., 2019). Similarly, in this paper's framework the duration of unemployment spells has a discouragement effect on jobseekers.

Expectation revision also ensues external interventions of the welfare provider, like job counselling and worker profiling. The effects of job counselling on reemployment rates are found to be stronger on long-term unemployment (Altmann et al., 2018), yet little in the aggregate (Belot et al., 2019).[8] In the spirit of the literature on job counselling, this paper studies profiling not just as a mechanism to allocate reemployment services, but also as a means to persuade jobless recipients to take some action by inducing a change in 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 about a payoff-relevant, yet unknown, state from a principal/sender to an agent/receiver. Strategic signalling may lead to partial information disclosure, with the aim of making the agent choose an action which is most favorable to the principal, yet not to the agent itself. The optimal signal has a geometric characterization, as its return coincides with the concave closure of the pre-signal payoff function. Previous works (Bergemann and Morris, 2018; Kolotilin, 2018; and Galperti and Perego, 2018) have highlighted the limitations of this method and proposed a new formulation of the problem that can be solved with linear programming.[9] The first attempt to model profiling in the context of unemployment insurance has been made by Cappellini (2020). The paper outlines the trade-off in profiling between information detection and principal-agent incentive alignment and shows that it may lead to partial detection of skills, aimed at persuading the worker to search at lower incentive costs. This paper builds on that result, by highlighting a second channel of gains from profiling that originates from reduction of effort-compensation costs on workers referred to delegated search upon profiling. The peculiarity of this paper's framework is the 'hybrid' nature of the problem, that requires the principal to deal with both information and incentive design. Consequently, the well-known concavification result à la Kamenica and Gentzkow only holds when utility promised to the agent is kept constant over profiling outcomes, and so the

[8]Belot et al. (2019) find that job counselling broadens the job search to other occupations and increases the number of job interviews, yet fails to increase the job-finding rate.

[9]In particular, Galperti and Perego have proved the existence of a dual of the original problem, where the 'shadow price' of the probability of each state is declining in a measure of agent's persuasion.

problem can be framed as one of pure information design (like in Cappellini, 2020). Boleslavsky and Kim (2021) extend the concavification result to a setting with three players (sender, agent and receiver) and incentive provision. The sender commits to a signal about a hidden state, with the twofold objective of convincing the agent to exert private effort and so affecting the state distribution with some probability, and persuading the receiver to take the sender's preferred action. Rodina (2020) considers a similar setting where the agent effort is not private. Similarly this paper, both works find that incentive provision determines a change in the concavification result by Kamenica and Gentzkow. However, in my paper the agent/receiver's effort affects the employment state, not the hidden one, and a combination of detected information and payments provides search incentives. 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, so far no work has studied the relationship between information design and incentive provision in the context of unemployment insurance.

3 Economic Environment

The paper extends the economic framework of Pavoni and Violante (2007) to the case of unobservable worker's reemployment chances and labor productivity. To this purpose, worker's expectations replace her human capital and the revision of expectations that ensues from failed job search replaces human capital depreciation.

Players' Interaction. A risk-neutral government (principal, it) and a risk-averse worker (agent, she) are infinitely-lived. Time is discrete and their discount factor is $\beta \in (0, 1)$. The worker can be employed or not, and the government observes her employment state. In period 0, (i) the worker is unemployed, (ii) her ability is unknown and both players hold common expectations about it, and (iii) the government offers her an insurance contract contingent on any possible future employment state and information about her ability. At the beginning of any time $t > 0$, a shock realizes that assigns a job to the worker, with a probability that depends on the search activity and the expectations on worker's ability at $t-1$. If the worker finds a job, she exists the program. Otherwise, the government assists her in period t. Therefore, the insurance contract maximizes the net present value of revenues -i.e., minus the cost- of future transfers and wage taxes upon reemployment V, conditional on current expectations μ and promising to her (expected) utility U.

Ability and Job Search. Worker's ability can be high ($h = H$), or low ($h = L$). Workers with high (resp., low) ability 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 $\pi_L \in (0, \pi_H)$. Job search is public, but non-contractible, and makes any worker incur effort cost e. Worker's utility v derives from consumption c and separable effort a, and is 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 worker's ability ($\omega_H > \omega_L$). 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*.[10] 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 skills and employability. The evaluation of any claimant is often 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 ability with some precision, and returns a publicly observable outcome, at cost κ^{wp}.[11] Profiling can be thought of as a test that returns a binary outcome -'Pass' (p) or 'Fail' (f)- with predetermined probabilities, conditional on ability h. The commitment power on government's side brings both parties to revise prior expectation μ to μ_f and μ_p, upon observing the test outcome.[12] Since expectations are correct on average, the distribution of revised expectations has mean equal to the prior, a property known as Martingale Property (MP, henceforth).

$$q\mu_p + (1 - q)\mu_f = \mu, \quad \text{(MP)}$$

where q is the probability that a 'Pass' is returned. This Property can be interpreted as a restriction that guarantees the credibility of profiling. Indeed, considering all workers who share

[10] 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.

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

[12] Without loss of generality, 'Pass' is assumed to mainly signal high skills, for which reason μ_p is larger than μ_f.

the same initial expectation μ, some must be induced to revise their expectation down to μ_f upon 'Fail', for someone else to revise their expectation up to μ_p upon 'Pass'. The timing of profiling is as follows. In the current period, the selected worker undergoes test. The profiling outcome is disclosed in the next period, after the new employment state realizes. The government is only allowed to base future unemployment benefits on the new information derived from profiling. Hence, if the worker exists unemployment in the meantime, profiling is useless as wage taxes do not depend on it.

Policies. Any policy arises as the composition of (i) recommended search effort, (ii) consumption, and (iii) wage tax/subsidy upon reemployment, and (iv) assisted-search and/or profiling technology (or neither of them). Combinations of search effort levels and technologies give 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 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' ($i = JS$). The instruments that do not envisage profiling are labelled *welfare policies*. Profiling without any search gives rise to 'Assistance and Profiling' ($i = AP$), whereas 'Insurance and Profiling' ($i = IP$) arises when the profiling technology is adopted together with worker's search. Finally, 'Search-Assistance and Profiling' ($i = SP$) originates if both technologies are jointly adopted. The instruments that envisage profiling are labelled *profiling policies*.

Welfare Policies (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 Policies			
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

It is useful to further discuss some of the assumptions made while defining the economic environment. The hazard rate π being increasing in expectations determines a complementarity between search and expected skills. In addition, separable search cost makes the dispersion of utilities (and payments, thereof) between successful and unsuccessful job search -a feature of delegated-search policies necessary to make payments incentive compatible- decrease in expectations in a hyperbolic fashion. The concavity of returns of delegated search in the μ space has consequences on the design of profiling policies (see Section 5).

4 Welfare Policies (No Profiling)

The planner's problem can be written recursively by keeping track of worker's expected ability 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, contingent on reemployment ($s = w$) or not ($s = u$), if policy i implements the job search. Current expectation μ and program's generosity U jointly determine the choice of the policy instrument. The government chooses the optimal policy $i(\mu, U)$ by solving

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

At the end of each period, the planner randomizes over worker's utility and policy instruments, under the constraint that the promised utility to be delivered in expectation must be equal to the promised one at the beginning of the current period. To this end, the operator $\mathbf{V}$ is defined as the net present value of expected returns, before knowing the outcome of the lottery over the space of worker's utilities.

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

One can show that the upper envelope V displays a tendency toward between-policy convexity in the U space, that makes this randomization welfare-improving (see Pavoni and Violante, 2007).

Before passing to the definition of the welfare policies SA, JS and UI, I describe the value of wage tax (or subsidies, if negative) that the government raises (or pay) in case the worker is reemployed.

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

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

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 to smooth worker's consumption over time ($U^w = U$).[13] 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

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

It is convenient to define also the expected wage tax/subsidy, prior to skills revelation.

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

I can now define welfare policies.

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

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

The planner transfers c^{sa} in the current period, without requiring the worker to exert any effort. In the next period, it pledges continuation utility equal to U^{sa}, and implements the lottery over utilities, before referring the worker to the next policy. The continuation value of the program thus equals $\mathbf{V}$. 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.

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. ■

This result is equivalent to Pavoni and Violante (2007). As SA implements no job search or profiling, no new information about worker's ability is acquired, and no expectation revision

[13] At the optimum, U^w solves

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

occurs. On the other hand, consumption smoothing delivers $U^{sa} = U$, as in the case of wage taxes. This means that expectations and promised utility in SA remain constant over time, which in turn implies that, if SA is optimal at time t, it will be so also at time $t+1$. The absorbing nature of SA is admittedly quite extreme to adopt 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 workers with low expected ability. The value of SA is independent of μ and has a closed-form expression

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

Job-Search assistance (JS). When resorting to assisted search, the government looks for employment on worker's behalf at cost κ^{ja}. With probability $\pi(\mu)$, the search is successful and the government extracts (expected) taxes on worker's wage, that depend on the level of promised utility upon reemployment U^w. If the search is unsuccessful -which occurs with probability $(1-\pi(\mu))$-, the government pledges to supply continuation utility equal to U^u and refers the worker to the optimal program $\mathbf{V}$. In this case, both the government and the worker downward revise initial expectation μ. The revised expectation that ability is high follows a Bayesian updating rule

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

The revised probability is lower than initial one, with equality holding only in the extreme case where ability is known already ($\mu \in \{0,1\}$). 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 $\pi(\mu) \equiv \mu\pi_H + (1-\mu)\pi_L$ of them manages to find a new employment, which implies that the high-skilled ones who remain unemployed in the next 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 low ability.[14]

[14] If the failed attempts to exit unemployment are t, 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 their ability, no update ever occurs;
- $\lim_{t\to\infty} \mu^{(t)} = 0$, if $\mu^{(0)} < 1$.

The value of JS reads

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

The government insures worker's consumption against the risk related to successful job search and smooths it over time and future employment state.[15]

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 requires her current promised utility U to be larger than the off-the-equilibrium utility she can enjoy by saving on effort and foregoing the chance of reemployment. This additional requirement boils down to an Incentive Compatibility constraint (hereafter, IC).

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

Promise Keeping in UI takes into account the effort cost e exerted by the job-seeker agent

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

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. Incentive costs are increasing in the cost of effort and decreasing in the level of patience (β) and expectations (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

[15] The other source of risk is related to ability realization upon reemployment, which is insured since the curvature of W in the U space is independent of h, i.e.

$$W_U(L, U) = W_U(H, U)$$

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

ensuing from it.

The following proposition sheds light on the components of V.

Proposition 2 (**Slope of** V **over** μ **and** U)**.** *The slopes of the value functions with respect to* U *satisfy*

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

The slopes of the value functions with respect to μ *satisfy*

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

where the last inequality of (8) holds (at least) at the crossing point.

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

The value of welfare policies equals the difference between search returns and contract costs. Search returns coincide with the expected labor productivity upon reemployment, and are thus linear in expectations.[16] A higher expectation μ increases worker's job-finding probability and expected wage tax, which makes the marginal value of μ positive in active policies (and null in SA). On the other hand, contract costs divide into technology, baseline and agency costs. Technology cost is present only when the government assists the worker in the job search (hence, only in JS) and does not depend on expectations, nor promised utility. Baseline costs are defined as the (constant) annuity consumption that delivers U, that is, $c^{sa} = u^{-1}((1-\beta)U)$. Agency costs are present only when workers supply positive effort (hence, only in UI) and equal the sum of effort compensation and incentive costs. Transfers must increase in U to compensate for the disutility of effort, as the marginal utility of risk-averse workers becomes smaller. In addition, a positive third derivative of u^{-1} guarantees incentive costs also to be increasing in U.[17] In sum, the agency costs of delegated search are higher when program's generosity is high. And this accounts for the slopes of V over U in (7). In addition, incentive costs in UI decline, as the dispersion in reemployment and unemployment transfers -proportional to the utility gap on the left-hand side of inequality (6)- shrinks, and this is why the slope of V along μ is steeper in UI than in JS, as reported in (8).

Fig. 1 reports the state space of optimal policy instruments. The blue area corresponds to combinations of μ and U for which SA is optimal. Not surprisingly, workers are referred to SA when their expected skills are low (and so low is μ). Fixing U and spanning the space of μ from left to right, SA is replaced by either JS and UI, as reemployment becomes more likely and

[16] $\pi(\mu)W(\mu, .)$ is a linear function of μ.

[17] See Pavoni and Violante (2007) for more insights about this result.

remunerative for the planner. Finally, UI is optimal for lower U, as incentivizing effort is less costly in less generous programs.

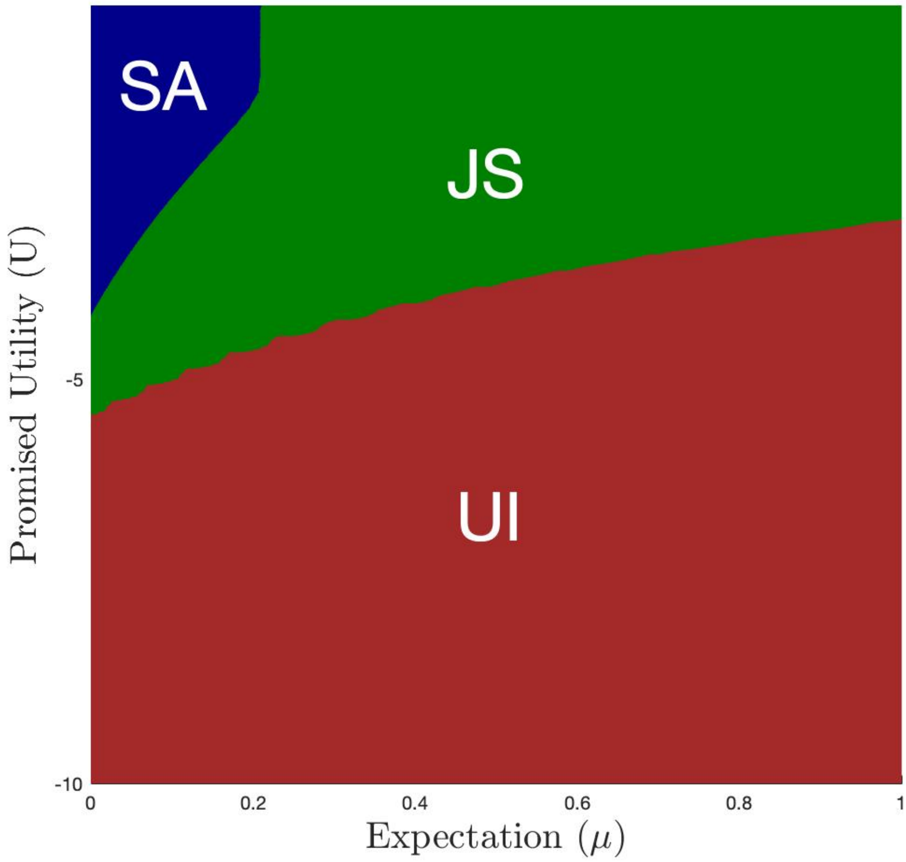


Figure 1: Optimal Welfare Policies over the (μ, U) State Space

The implications of Prop. 2 on the shape of upper envelope V are deep. The upper envelope inherits from each policy value V^i a 'local' tendency toward *within-policy supermodularity* -i.e., $V_{\mu U}(\mu, U) \geq 0$- (shown in the proof of Prop. 2). However, the 'global' shape of V is *between-policy submodular*, as a consequence of (7) and (8). The slope of V with respect to U is higher when SA is optimal, for low-end μ, and lower when UI or JS is optimal, for high-end μ. Prop. 7 in Section 5 sheds light on how the shape of V determines the relationship between profiling accuracy and program's generosity.

4.1 Optimal Welfare-to-Work Programs

The main advantage of the framework introduced in this Section lies in the possibility to characterize the optimal pattern of transfers and policies of any welfare-to-work program, starting from the initial expectation μ_0 and the promised utility U_0 to any new recipient in the program.

[17]The parametrization of Sections 4, 5, and 6 is: $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$ with $\sigma = 3$, $\beta = 0.9$, $e = 0.7$, $\kappa^{ja} = 8$, $\kappa^{wp} = 0.5$, $\omega_H = 10$, $\omega_L = 3$, $\pi_H = 0.27$, $\pi_L = 0.14$.

Prop. 3 and 4 describe the main features of optimal transfers and possible policy transitions over time.

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 contract costs of delegated search increase in the level of worker's utility, search incentives are a mix of reward upon reemployment and punishment upon job-search failure. Hence, promised utility decreases during UI ($U^u_{UI} < U$). Furthermore, whenever assisted search is followed by delegated search, worker's utility starts decreasing while the worker is still enrolled in JS ($U^u_{UI} < U^u_{JS} < U$). This finding sheds light on the possible policy patterns that can arise as a function of worker's initial expectation and program's generosity.

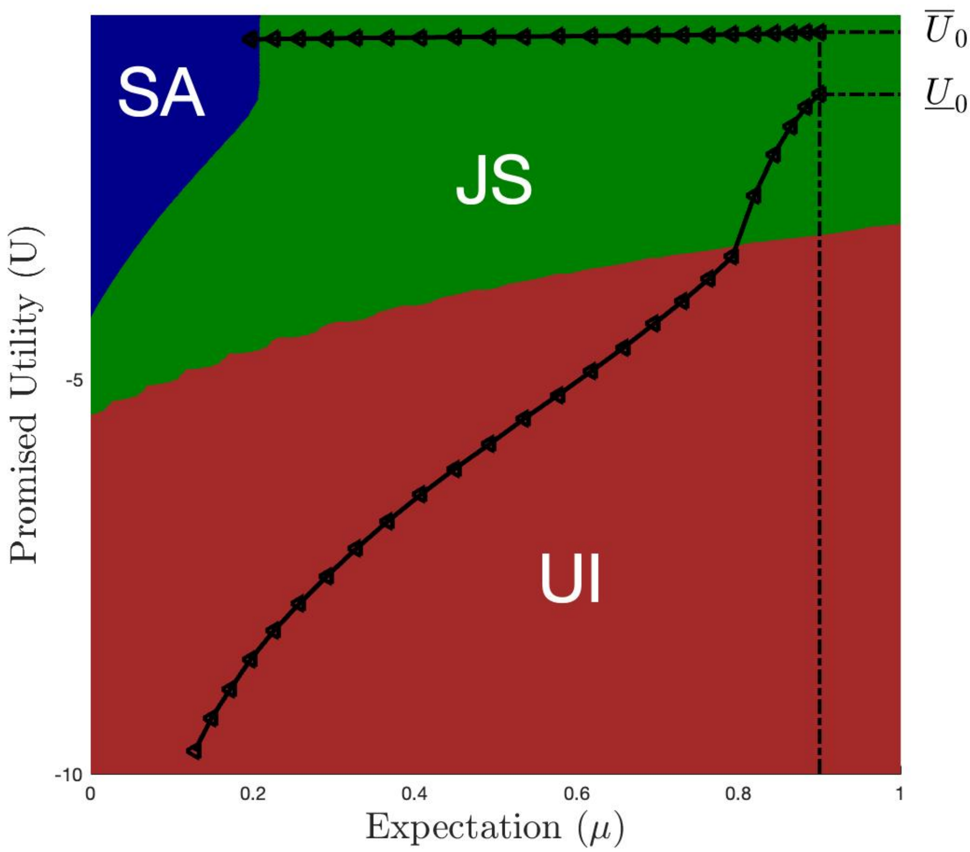


Figure 2: Optimal Policy Transitions in the (μ, U) State Space.

In Fig. 2, the state space is used as a phase diagram to track down the evolution of μ and U, as well as policy transitions, over the unemployment spell. Two instances of optimal policy sequences are reported, for same initial expectation ($\mu_0 = 0.9$) and different generosity levels. When generosity is higher ($c(\overline{U}_0) = 290$ in consumption-equivalent terms),[18] the worker never enters UI and only moves from JS to SA. Thus, she is required to exert no effort at any time and granted both consumption and utility insurance along the spell. So, she eventually exits unemployment or enters SA with the same utility level and consumption as she entered the program. When, instead, generosity is lower ($c(\underline{U}_0) = 190$), the worker is relocated from JS to UI at some point of the spell. For this reason, her utility decreases over time until she finds a job and exits unemployment. What Fig. 2 shows is that, at the optimum, more generous programs are mainly focused on income support and assisted search, and less on search incentives.

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 drop in the contract costs of delegated search, and makes UI more appealing with respect to JS *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

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

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

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

Then, the optimal program never switches from UI to JS.

[18] The expression of consumption equivalent of U is

$$c(U) = u^{-1}((1-\beta)U)$$

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. 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 relatively small variation of U (of the same 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 Test-Based Profiling

5.1 Profiling Policies and Persuasion Mechanism

Profiling publicly discloses worker's ability, up to a level of accuracy chosen by the government. For this purpose, profiling programs implement different strategies to infer worker's skills. Some assign the profiled claimant a given task and assess ability based on her performance. An example of this type of profiling is adopted by Missouri, where claimants selected for REA are required to fill in a self-eligibility assessment (Annex I). In this case, profiling takes the form of a test with two mutually exclusive outcomes, 'Pass' ($r = p$) and 'Fail' ($r = f$). Conditional on the test outcome, the worker revises expectations and is referred to a different policy instrument. In test-based profiling programs, test accuracy is determined by the minimum requirement to pass it. Assume that each profiled worker is assigned a score, and is returned a 'Pass' whenever the score is above a given threshold. A low threshold allows for a fraction of low-skilled workers to obtain the 'Pass' together with high-skilled ones, and this occurs whenever type-II error is positive. Given the nature of this profiling methodology, any profiled worker could pretend to be low skilled by intentionally failing the test. It is thus necessary that the continuation utility upon 'Pass' is non smaller than the continuation utility upon 'Fail'. Such a requirement, labelled No Discrimination constraint (ND, hereafter), imposes a restriction on the contract offered under either profiling outcome.[19]

$$U_p^u \geq U_f^u \quad \text{(ND)}$$

[19] In principle, the worker's utility for intentionally failing the test has a complex expression, as the worker does not learn about her skills if she intentionally fails and transfers must be evaluated with the initial expectation μ. The expression would be complicated by write down, if expectations determined the generosity of payments upon 'Fail'. However, a consequence of Prop. 5 is that, since SA is an absorbing policy and no job search is conducted ever after, the generosity of payments does not depend on the expected job-finding probability of reemployment, nor on current and future expectations thereof. Therefore, (ND) constraint admits a closed-form expression.

I can now define profiling policies. They can be read as the natural counterpart of welfare policies, with the addition of the profiling technology. The planner's problem that defines profiling policies has a compact formulation. What test-based profiling does is to effectively randomize over the value of the continuation program in the μ space. The problem for $i \in \{AP, SP, IP\}$ thus reads

$$V^i(\mu, U) = \max_{(c, U^w, U^u_f, U^u_p)} -c - \kappa^i + \beta\big[p^i(\mu)W(\mu, U^w) + q(1 - p^i(\mu_p))\mathbf{V}(\mu^i_p, U^u_p) + (1 - q)(1 - p^i(\mu_f))\mathbf{V}(\mu^i_f, U^u_f)\big] \tag{10}$$

$$\begin{aligned} \text{subject to:} \quad & U = u(c) - e^i + \beta\big[p^i(\mu)U^w + q(1 - p^i(\mu_p))U^u_p + (1 - q)(1 - p^i(\mu_f))U^u_f\big] \quad \text{(PK)} \\ & U \geq u(c) + \beta\big[qU^u_p + (1 - q)U^u_f\big] \quad \text{(IC- only in IP), (ND) and (MP)} \end{aligned}$$

with $p^{AP} = 0,\ p^{SP} = p^{IP} = \pi$ and $\mu^{AP} = \mu,\ \mu^{SP} = \mu^{IP} = \mu'$.

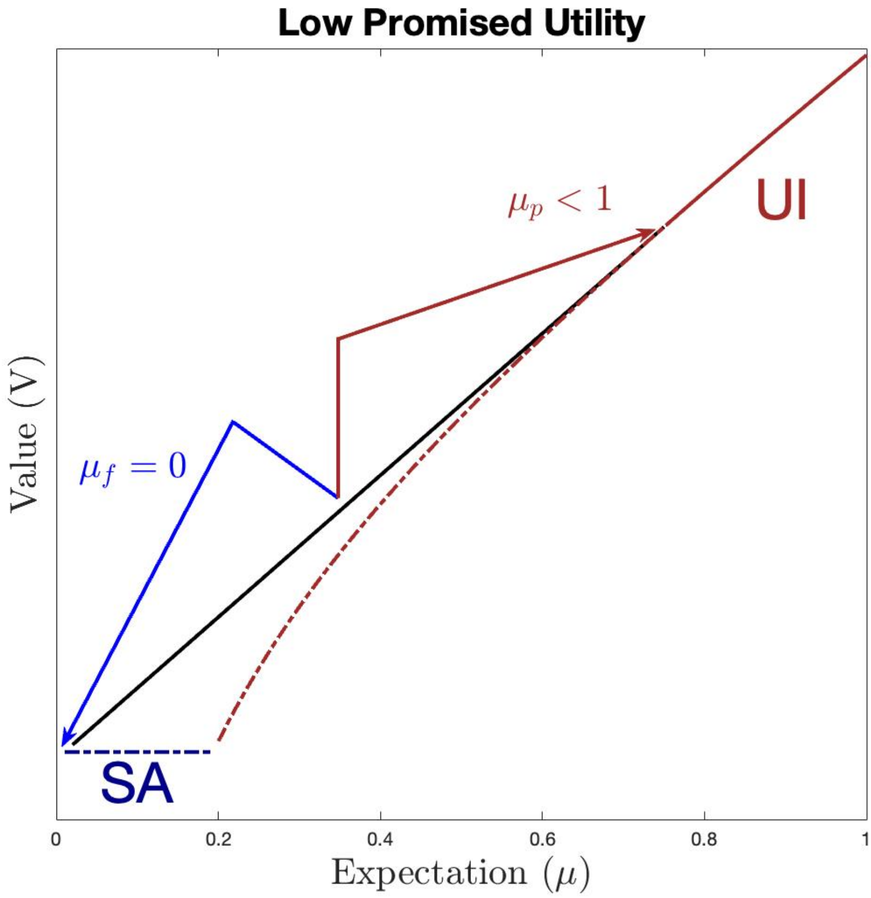


Figure 3: Value of SA and UI Policies over Expectations.

Fig. 3 reports the section of V^{SA} (blue line) and V^{UI} (brown line) in the μ space for constant U.[20] As anticipated by the utility dispersion in (6), Section 4, UI contract costs have a convex

[20]Promised utility is chosen low so that UI is optimal for high-end μ's. To keep things as simple as possible, the value of JS is not reported in the figure.

hyperbolic shape in the μ space. Concavity of V^{UI} in μ follows from convexity of incentive costs and the linearity of returns. The purple line represents the return of the optimal profiling in the Assistance-and-Profiling (AP) policy, where the government finds it optimal not to profile worker's ability with full precision, and leaves those receiving 'Pass' outcome more optimistic, yet doubtful about their actual ability. The return of profiling connects to a well established result of Bayesian persuasion, in which the optimal test delivers the concave closure of the V section (Kamenica and Gentzkow, 2011). The figure shows that both the posterior expectation upon 'Fail' and the type-I error are null, meaning that no high-skilled worker is profiled as low skilled.

Proposition 5 (**No Type-I Error**). *Define type-I error as the occurrence where a high-skilled worker is profiled as low skilled. Therefore, at the optimum the probability of type-I error is null ($\mu_f = 0$), that is, only low-skilled workers are given a 'Fail' and referred to Social Assistance henceforth.*

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

Intuitively, the planner uses profiling to align incentives with agents. Those agents whose incentives are less costly to provide are the high skilled. For this reason, it makes no sense to depress expectations of high-skilled agents by 'Fail'-ing them.

Passing to type-II error, the planner persuades a fraction of low-skilled agents to be high skilled and have them search at a rather low incentive cost, by pooling them in the same ('Pass') outcome with high-skilled ones. This would lead to a posterior upon 'Pass' μ_p smaller than 1, as 'Pass' does not precisely 'detect' high ability. What profiling is effectively doing is to make the high-skilled workers in the pool a bit more pessimistic compared to the perfect signal ($\mu_p = 1$), while proportionally increasing the optimism of the low-skilled ones, whose perfect signal $\mu_f = 0$ is replaced by μ_p. Therefore, increasing type-II error (that is, lowering μ_p) produces a gain from 'relocating' the marginal low-skilled worker to delegated search, which initially outweighs the loss from causing higher incentives costs on the pool of high-skilled profiles. As μ_p falls/the probability of type-II error increases, the marginal incentive cost increases due to its convex shape up to the point where garbling profiles does not pay off any further.

Proposition 6 (**Optimal Type-II Error**). *The type-II error is pinned down by the 'Pass' posterior μ_p, that in test-based profiling corresponds to the minimum between 1 and the solution of*

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

with $p^i = \pi$, $\mu_p^i = \mu_p'$ if $i = SP$, IP and $p^i = 0$, $\mu_p^i = \mu_p$ if $i = AP$.

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

A positive type-II error implies that the planner prefers not to disclose information to any low-skilled agent who expects to be high skilled (for ex., whose prior expectation μ larger than μ_p in Fig. 3). In other words, the planner finds that any overconfident low-skilled agent is mistaken in the direction that is most favourable to it, and that any new information about her actual abilities would only misalign incentives and make job search more expensive to delegate.

5.2 Optimal Profiling Policies in the (μ, U) Space

I am now ready to locate the optimal profiling policies in each region of the (μ, U) state space. First, no profiling policy is optimal for very high or very low expectations, as the administrative expenses (κ^{wp}) outweigh the prospective gains from relocating workers across different policy instruments. Indeed, under the assumption that expectations are correct on average, workers with very high or very low expectations are likely to be assigned to a suitable policy instrument already, and prospective benefits from reallocation are quite small. Second, there exists a 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 optimal welfare policy is the one that adopts the same job-search method.

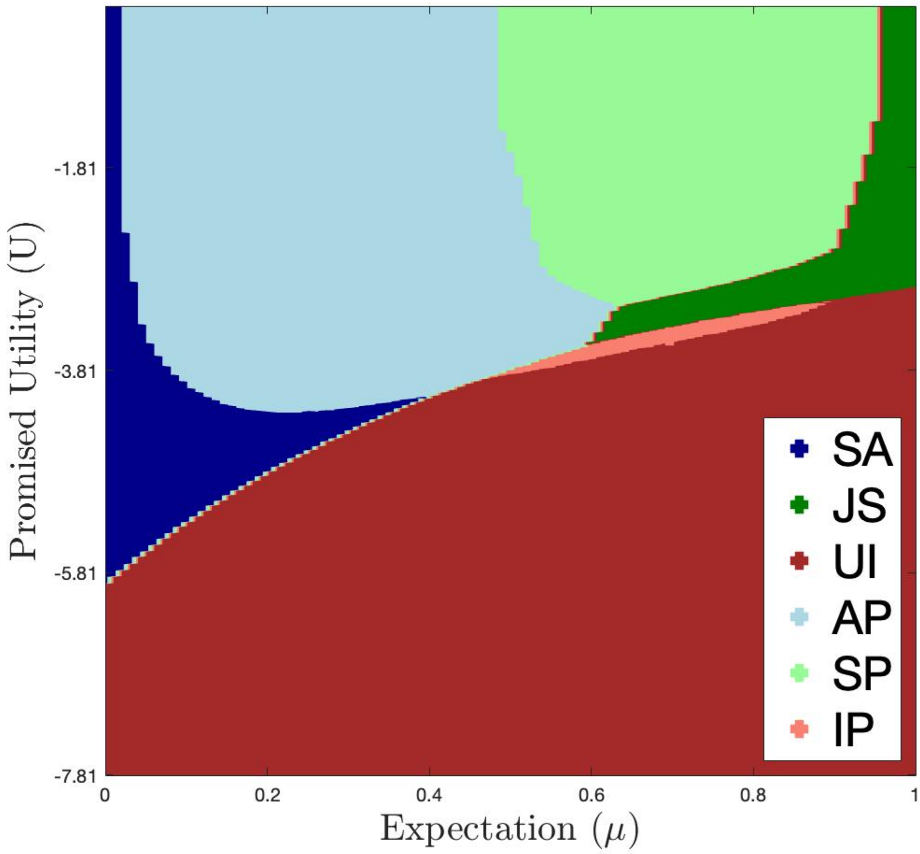


Figure 4: Optimal Policies in the Space of Expectation and Generosity.

Fig. 4 reports the optimal policies in the (μ, U) space. The choice of the optimal profiling policy mirrors the complementarity of search effort with expectations and its substitutability with promised utility. In particular, AP does not implement search and is thus optimal for lower-intermediate expectations. SP and IP, on the contrary, are active policies, and are thus optimal for upper-intermediate expectations. As generosity rises (moving vertically from bottom to top of Fig. 4), higher costs of search-effort compensation and incentive provision (Prop. 2) lower the return of delegated search (UI and IP), which is replaced by assisted search (JS and SP) for high-end expectations, and by no search (SA and AP) for low-end expectations.

An important aspect of a profiling policy is the optimal probability of type-II error. Indeed, the value of information to the planner may be negative when workers' prior expectations are so high (i.e., $\mu > \mu_p$) that incentives between principal and agent can not be aligned any further (Prop. 6). The following result establishes a monotone relationship between profiling accuracy and program's generosity.

Proposition 7 (Type-II Error and Generosity). *For fixed expectation μ, the probability of type-II error is decreasing in generosity U. $(\partial\mu_p/\partial U \geq 0)$.*

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 within-policy supermodular and between-policy submodular. Looking at the determinants of μ_p, within-policy supermodularity produces an increase in the left-hand side of (11) in response to a rise in U, and a drop in the right-hand side due to between-policy submodularity. To reestablish equality between the two sides of the equation, μ_p must increase.

For high-end generosities, the optimal program never resorts to delegated search, making the job-search return linear in μ, and worker's utility does not fall over time (Prop. 2 and 3). In such a case, skills are fully revealed ($\mu_f = 0,\ \mu_p = 1$) and worker's utility is kept constant over time ($U^u_{SP} = U$). The gains from profiling exclusively originate from the efficient match between workers and policy instruments, as high-skilled workers are referred to Reemployment Services and low-skilled ones to passive income support. This type of profiling is reminiscent of the WPRS program, where information on workers' skills is used as a criterion to allocate services. On the contrary, for low-end generosities, the government delegates the job search to workers who pass the test. In any such case, information about skills is used also to fine-tune unemployment benefits, that decline over time (Prop. 3), as it is the case in the REA program. Savings of this second type are more sizable, when profiling is based on the statistical evidence coming from past unemployment spells rather than test outcomes, and no opportunistic behavior is possible thereof. Indeed, in statistical profiling workers are referred to 'more generous' or 'less generous' programs based on their profile, in order to ease the cost of incentive provision on those who are classified as high skilled (see Section 6).

6 Statistical Profiling

The presence of (ND) constraint in test-based profiling prevents 'punishment' of workers who are referred to delegated search upon it. However, this restriction does not apply whenever skills are 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, based on a large number of past observations. Given that the outcome of statistical profiling does not rely on worker's commitment to it, (ND) constraint is not present 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.

Proposition 8 (**Optimal Statistical Profiling**)**.** *The 'Pass' posterior μ_p is either 1 (i.e., no* *type-II error) or solves*

$$\mathbf{V}_\mu(\mu_p^i, U_p^u) = \frac{\mathbf{V}(\mu_p^i, U_p^u) - \mathbf{V}(0, U_f^u) + \mathbf{V}_U(0, U_f^u)(U_f^u - U_p^u)}{\mu_p^i} \tag{12}$$

with $\mu_p^{AP} = \mu_p$ and $\mu_p^{SP} = \mu_p^{IP} = \mu_p'$. As in case of test-based profiling, the frequency of type-II error is decreasing in generosity (i.e., $\partial\mu_p^i/\partial U > 0$).[21]

Proof. See Appendix C: Statistical Profiling. ■

The government sets different continuation utilities according to the profiling outcome. The cost of incentive provision and effort compensation is increasing in generosity (Prop. 2), which makes the marginal cost of U larger in UI than in SA. Hence, the government finds it optimal to lower the net discounted value of future payments upon 'Pass' (i.e., $U_p^u < U_f^u$). The result matches a characteristic of the actual REA program, where any worker who is found high skilled is referred to SNAP transfers, the minimum form of income support, until she is reemployed. The rationale of this rule is that generous transfers to high-skilled workers a disincentive to search for people who are likely to easily find reemployment.

The possibility to randomize over continuation utilities modifies the optimal probability of 'Pass'. Condition 12 pins down a new balance between incentive cost reduction in UI, the likelihood of being referred to it, and the new channel arising from the relaxation of the (IC) constraint.[22] Lowering the probability of type-II error, indeed, also increases the frequency of 'Fail', conditional on which the planner pledges a larger utility. Hence, expected promised utility for the agent is larger if the type-II error is less likely and the 'Pass' outcome is more informative *ceteris paribus* -i.e., μ_p is higher-, which allows the planner to further lower promised payments in order to restore binding (PK) constraint.

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

- *If the policy sequence after 'Pass' includes UI, utility upon 'Pass' (resp., 'Fail') is lower (resp., higher) than current utility.*

$$U_i^{u,p} < U < U_i^{u,f}$$

[21] For $i = AP, SP$, statistical profiling at the optimum delivers the concave closure of $\mathbf{V}$ over the two dimensions. Indeed, condition (12) can be rewritten

$$\begin{bmatrix}\mathbf{V}_\mu(\mu_p^i, U_p^u) & \mathbf{V}_U(\mu_p^i, U_p^u)\end{bmatrix}\begin{bmatrix}\frac{\mu_p^i}{||\mathbf{h}||} & \frac{U_p^u - U_f^u}{||\mathbf{h}||}\end{bmatrix}^T = \frac{\mathbf{V}(\mu_p^i, U_p^u) - \mathbf{V}(0, U_f^u)}{||\mathbf{h}||}$$

with $||\mathbf{h}|| = \sqrt{{\mu_p^i}^2 + (U_p^u - U_f^u)^2}$.

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

Otherwise, continuation utilities are constant over time and profiling outcome.

$$U_i^{u,p} = U = U_i^{u,f}$$

- *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'.*

$$c_{IP}^{u,p} < c_{IP}^{u,f} < c_{IP}$$

- *In IP (resp., SP), the net wage upon reemployment is larger than (resp., equal to) current unemployment benefits.*

$$c_{IP} < c_{IP}^{w}, \quad c_{SP} = c_{SP}^{w}$$

Proof. See Appendix C: Statistical Profiling. ∎

Fig. 5 plots the patterns of policy transitions, unemployment benefits, current expectation and promised utility, for a worker who enters the program with $\mu_0 = 0.5$ and $c(U_0) = \$113$. The worker is referred to delegated search and profiled right upon entering the program. Profiling does not fully disclose ability upon 'Pass' (positive type-II error). Thus, if the worker is low-skilled, chances are that she is incentivized to actively search for a job and suffers a reduction in the generosity of transfers, which keep falling along the spell.

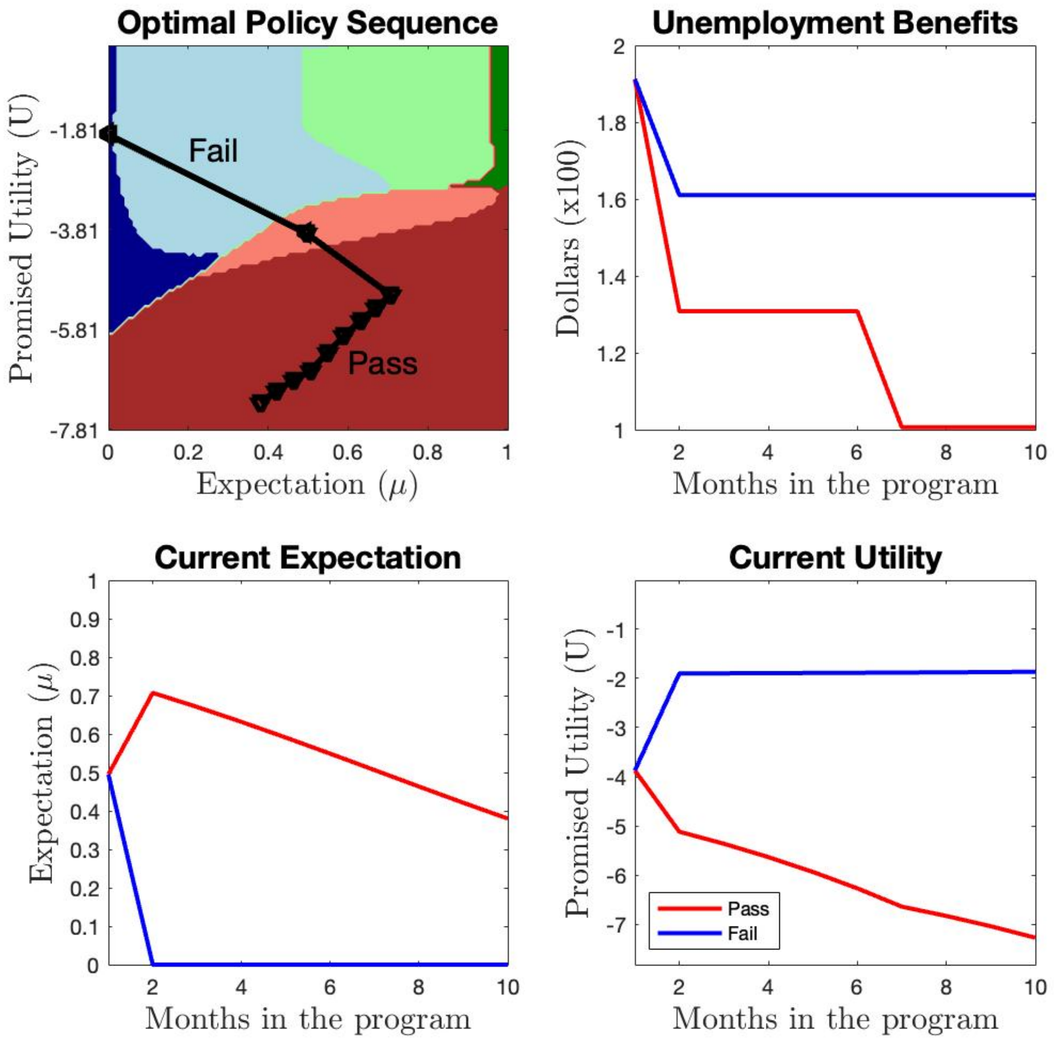


Figure 5: Consumption pattern upon profiling in a program with decreasing utility and $\mu_0 = 0.5$ and $c(U_0) = 100 \times u^{-1}((1-\beta)U_0) = \$113$.[23]

7 Quantitative Analysis

The main objectives of this paper are to design optimal welfare programs with profiling, and to estimate welfare gains they produce. To this aim, this section estimates the return of optimal profiling in the State of Florida, by comparing it to an alternative scenario where only policies SA, JS and UI are implemented. The estimated welfare gains of profiling equal the difference in returns between the two cases.

7.1 Program Description

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 (see

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

Section 1). In the US, for example, there are two 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 for, whose goal is to reduce fraud and misallocation of funds, 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 collected via profiling. With REA, efficiency gains realize by means of transfer reduction, whereas WPRS generates savings by matching each worker with the most efficient job-search 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 (Manoli et al., 2018). Moreover, neither program allows workers enrolled in a 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%.[24] Each State adopts a different selection process to refer claimants to either program. In particular, Florida randomly assigns eligible claimants to REA, PREP or a control group, where no reemployment or profiling service is supplied (Annex II).[25] REA participants are required to undergo an in-person eligibility review at a public employment office, while PREP participants are referred to group orientation where they receive information and are supplied job-search services,

7.2 Parameterization

Turning to to the choice of parameters (reported in Table 2), the values to choose 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}) and assisted search (κ^{ja}). 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 $\overline{e}^m = 0.67$ and $\overline{e}^w = 0.97$ given the

[24]The relative weight assigned to each State depends on the number of participants to the program. 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.

[25]In Florida, the WPRS program is called Priority RE-employment Planning (PREP).

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	$\{c_H^w, c_L^w\}$	{$2,498, $1,128}	Poe-Yamagata et al. (2011)
Wage Tax	$\{\tau_H, \tau_L\}$	{$178, -$224}	EIC, FICA
Worker Profiling			
Administrative Cost	κ^{wp}	$50	Poe-Yamagata et al. (2011)
Assisted Search			
Administrative Cost	κ^{ja}	$430	Balducchi and O'Leary (2018)
REA programs (FL, ID, IL, NV)	c^i,		
Generosity (consumption equivalent)	$i = FL, ID, IL, NV$	[$1,519,$1,697]	Nicholson and Needels (2011)

Table 2: Choice of Parameters Value

logarithmic specification[26]. And given that the percentage of male participants within the four 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 about three quarters before displacement and steadily declines later on. The decline is consistent with the phenomenon known as *Ashenfelter's dip*, suggesting that wages fall in the pre-layoff period (Ashenfelter, 1978). To prevent this effect from distorting estimates it would require considering the wage some quarters prior to layoff. However, the paper does not consider skill depreciation along the unemployed spell, which is instead well documented by the empirical literature (Keane and Wolpin, 1997; Neal, 1995) and harms both labor productivity and reemployment perspectives. 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.[27] 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, reported by Poe-Yamagata et al. (2011). Given that there are two unknowns and three States, I compute

[26] Logarithmic utility allows for separation of consumption utility from working disutility in a natural way, according 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.

[27] 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 $\{c_H^w, c_L^w\}$.

$\{c_H^w, c_L^w\}$ as the pair that minimizes the loss function

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

with φ_i being the fraction of welfare recipients in country i. The computation delivers monthly (net) 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 from the net amounts. In the 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 half levied on employers and half on employees, and amounted to 15.3% of Adjusted Gross Income (AGI) in 2009. Moreover, taxpayers are entitled to an Earned Income Credit (EIC), if their AGI is lower than a certain amount, depending on their marital status and number of children. 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 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, which boils down to a gross income of $\$10,844$. High-skilled recipients, instead, have a gross income of $\$32,112$ and a net one of $\$29,976$.[28] 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 method-of-moments, the probability of reemployment after t periods is computed as the fraction of workers who exit unemployment in t. 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 of no search to be equal to half the 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 no-search option, indeed, is higher with respect to the case with $\hat{\pi} = 0$, and

[28] 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 $\$2,139$.

the incentive constraint now requires satisfying a more stringent 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 (for ex., $\kappa^{ja} = \$500$ in Pavoni et al. (2016)), as well as cost estimates of programs that perform different activities, like 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 (2007) estimate a monthly cost of monitoring of \$478 per claimant.

The generosity of any program depends on the amount and duration of flow endowments. Poe-Yamagata et al. (2011) collects data about average weekly benefits 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 the 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 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), which replaced the Food Stamps Program in 2008. This constitutes a typical purely income-support measure of last resort, consisting of a constant allowance for the purchase of food, with no additional eligibility requirement 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 for the whole prospective duration of these programs, i.e. 73 weeks, and that exhaustees who are still unemployed at the end of UI+EUC+EB enter into SNAP. The average total payment is \$7,930 under EUC and \$3,844 under EB (Nicholson and Needels, 2011), hence amounting to a monthly endowment of $c^{EUC/EB} = \$645$, while a family of four people is receiving a \$501 monthly benefit from SNAP.[29] The program's generosity for each of the four States is computed backward from the moment the welfare recipient enters into SNAP or finds a job, up until the first month of regular UI benefits. Worker's utility of reemployment in case she is high-(resp., low-)skilled amounts to[30]

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

while in SNAP with no search it is equal to

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

Condition $U_L^w > U_{e=0}^{SNAP}(0) + \frac{e}{\beta(\pi_L - \hat{\pi})}$ implies that any worker always finds it convenient to search also under SNAP, no matter what their expected skills are. Hence, the expected generosity of SNAP can be rewritten as a function of 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 month 25, the following recursion is implemented

$$U_{j,t}^{WEL}(\mu) = u(c_t^j) - e + \beta[\mu\pi_H U_H^w + (1-\mu)\pi_L U_L^w + (1-\pi(\mu))U_{j,t+1}^{WEL}(\mu')], \quad 1 \le t \le 25,$$
$$j = \{FL, ID, IL, NV\},\ i = \{< HS, HS, < CD, CD, GD\}$$

with $U_{j,26}^{WEL} = U^{SNAP}$ and j indexing the four US States. The initial probability of being high-skilled, μ_i^0, equals the share of high-skilled individuals with same educational attainment, ϑ_i, and $U_{j,1}^{WEL}(\vartheta_i)$ represents the continuation utility of the regular assistance program for workers who are never profiled over the spell. If instead they do undergo profiling after t periods of benefits, their promised utility reads

$$U_{j,t}^{REA}(\mu) = u(c_t^j) - e + \beta\big[\mu\big(\pi_H U_H^w + (1-\pi_H)U^{SNAP}(1)\big) + (1-\mu)\big(\pi_L U_L^w + (1-\pi_L)U_{j,t+1}^{WEL}(0)\big)\big]$$

The generosity of the program is the weighted average between the promised utilities in the two alternative scenarios. According to Poe-Yamagata et al. (2011), welfare claimants are randomly assigned to REA after some time spent in the program, that ranges from 2-3 (Nevada) up to 4-6

[29] See SNAP Data Tables at: https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap.
[30] All monetary amounts are normalized so that 1 consumption unit corresponds to \$100.

weeks (Florida), and are scheduled for REA 2 weeks after the assignment. Hence, the expression for the program's generosity reads

$$U^{PROF}_{j,t}(\mu) = u(c^j_t) - e + \beta[\mu\pi_H U^w_H + (1-\mu)\pi_L U^w_L + (1-\pi(\mu))U^{PROF}_{j,t+1}(\mu')], \quad 1 \le t \le \bar{t}_j$$
$$U^{PROF}_{j,\bar{t}_j}(\mu) = u(c^j_{\bar{t}_j}) - e + \beta\big[\mu\pi_H U^w_H + (1-\mu)\pi_L U^w_L + (1-\pi(\mu))\big(\nu_j U^{REA}_{j,\bar{t}_j+1}(\mu') + (1-\nu_j)U^{WEL}_{j,\bar{t}_j+1}(\mu')\big)\big]$$

where $\bar{t}_j \in \{1,2\}$ is the actual timing of REA implementation and ν_j is the fraction of REA-eligible claimants who benefit from the program in State j.[31] The generosity levels of each program and educational attainment, expressed in consumption-equivalent terms, 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 $U^w_H > U^w_L$. Among the four States, Nevada (resp., Florida) is the most (resp., least) generous one for all levels of education.

States	Less Than HS	HS Diploma	Some College	College	Graduate
Florida	\$ 1,288	\$ 1,447	\$ 1,485	\$ 1,623	\$ 1,675
Idaho	\$ 1,458	\$ 1,628	\$ 1,668	\$ 1,816	\$ 1,872
Illinois	\$ 1,418	\$ 1,562	\$ 1,596	\$ 1,718	\$ 1,763
Nevada	\$ 1,476	\$ 1,647	\$ 1,687	\$ 1,835	\$ 1,890

Table 3: Program generosity for any State and educational level (consumption equivalent).[32]

7.3 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 likely to exhaust their UI benefits. 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 is statistical, as it allows 'punishment' of high-skilled recipients.

Fig. 6 reports the optimal policies in the state-space of programs' generosity and expectation, and locates Florida's welfare claimants on it according to their initial (μ_0, U_0).[33] Recipients with less than a high-school degree should better search and at the same time undergo profiling right after entering the program. Those with a higher educational attainment should be initially assisted in the search and profiled as they enter the program (high-school graduates) or after

[31] Poe-Yamagata et al. (2011) report that in 2009 ν_j ranged from 4% in Idaho to 99% in Illinois (see Exhibit 5, p. 16).

[32] The expression of consumption equivalent of U is

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

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

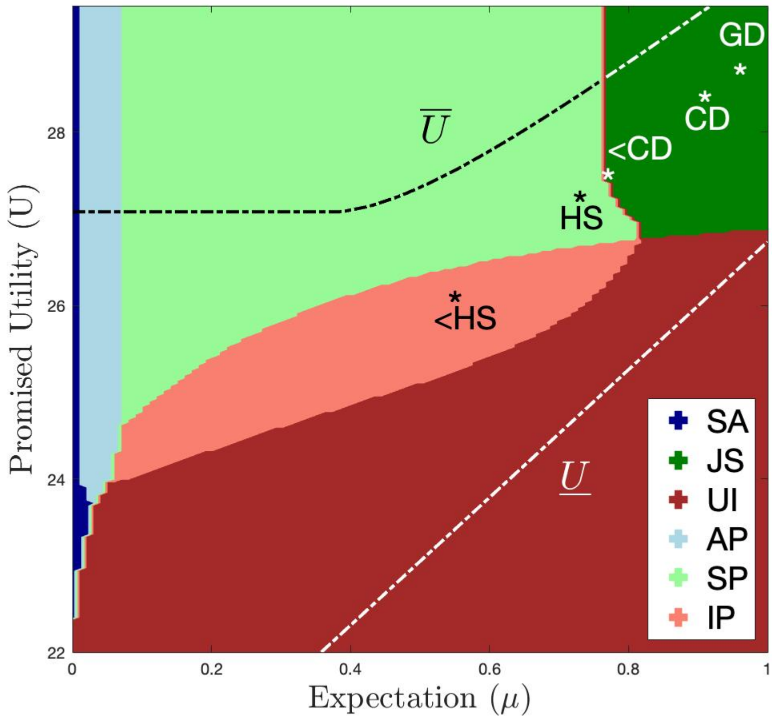


Figure 6: Optimal Policies for Florida's REA Recipients.
Note: <HS=Less Than High School, HS=High School Diploma, <CD=Some College, CD=College Degree, GD=Graduate Degree

some failed attempts of the government to find them a job (all those who have attended college). Contrary to the common idea, 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 thus with higher implicit utility U_0, which in turn makes graduates' effort too expensive to compensate for the government. The figure displays also upper and lower bounds for the initial location of claimants in the spate space. The two lines, indeed, represent the initial generosity of programs that pay perpetual monthly instalments of \$1,500 and \$500 to claimants with initial expectation displayed on the x axis.[34] The upper bound becomes flat for low-end expectations, as claimants refuse to search when their reemployment perspectives are low enough.

7.4 Welfare Gains

A relevant aspect for policymakers is the return of optimal profiling. In order to estimate savings, I compare two distinct programs, one with only welfare policies SA, UI and JS ($\mathcal{W}$) and the other

[34]The value of $\overline{U}(\mu)$ is defined as

$$\overline{U}(\mu) = \max\left\{\frac{u(\bar{c})}{1-\beta}, u(\bar{c}) - e + \beta\big[\mu\pi_H U_H^w + (1-\mu)\pi_L U_L^w + (1-\pi(\mu))\overline{U}(\mu')\big]\right\}$$

with $\bar{c} = \$1,500$. An analogous definition holds for $\underline{U}$ with $\underline{c} = \$500$.

featuring all six policies ($\mathcal{P}$). Fig. 7 reports the optimal patterns of promised utility, unemployment benefits and wage taxes/subsidies in $\mathcal{W}$ and $\mathcal{P}$ programs for Florida's jobseekers with a high school diploma (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 $c(U_0) = \$1,447$, respectively. As shown in Fig. 6, in $\mathcal{P}$ high-school diplomates are profiled under SP and referred to either UI or SA, while in $\mathcal{W}$ they never exit UI. The pattern of promised utility and unemployment benefits is thus falling over time for all workers in $\mathcal{W}$ and only for high-skilled ones in $\mathcal{P}$ (Fig. 7). However, the decline of benefits and utility is faster in W, due to higher incentive costs and lower search returns. The reemployment tax displays a monotone increasing pattern in $\mathcal{P}$ upon profiling (in the weak sense for low-skilled workers) due to declining 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 over time.

Table 4 reports the per-capita welfare gain of profiling for each educational group, computed as a difference between the return of the welfare program with ($V_{\mathcal{P}}$) and without ($V_{\mathcal{W}}$) profiling. Estimates increase in the educational level of the recipient and range from \$1,026 (less than high-school diploma) to \$1,305 (graduate). These figures are up to three times as large as the net per-capita savings estimated in Poe-Yamagata et al. (2011), that amounts to \$395.[35] This is evidence that the current program design has large room for improvement. Given the relative size of each educational level and the number of REA-eligible workers in 2009 (459,021), the annual aggregate welfare gain of Florida would have been \$540.4 millions.[36] The expected total number of claimants to be profiled is 288,180, amounting to 62.8% of all welfare recipients, for an overall (discounted) cost of \$136,660. Therefore, I can conclude that every dollar spent on REA in Florida generates positive net savings equal to \$3,954.

	Less Than HS	HS Diploma	Some College	College	Graduate
Perc. in Program	0.13	0.54	0.17	0.12	0.04
$\pi(\mu_0)$	0.21	0.23	0.24	0.26	0.27
$c(U_0)$	\$ 1,288	\$1,447	\$1,485	\$1,623	\$1,675
Per-cap. Welfare Gain	\$1,026	\$1,128	\$1,140	\$1,250	\$1,305

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

[35]The reader should recall that Poe-Yamagata et al. (2011) estimate the return of the current program, not of the optimal one.

[36]Poe-Yamagata et al. (2011) report a number of 64,263 REA recipients in Florida in 2009, accounting for 14% of overall REA-eligible recipients (Exhibit 5, p. 16).

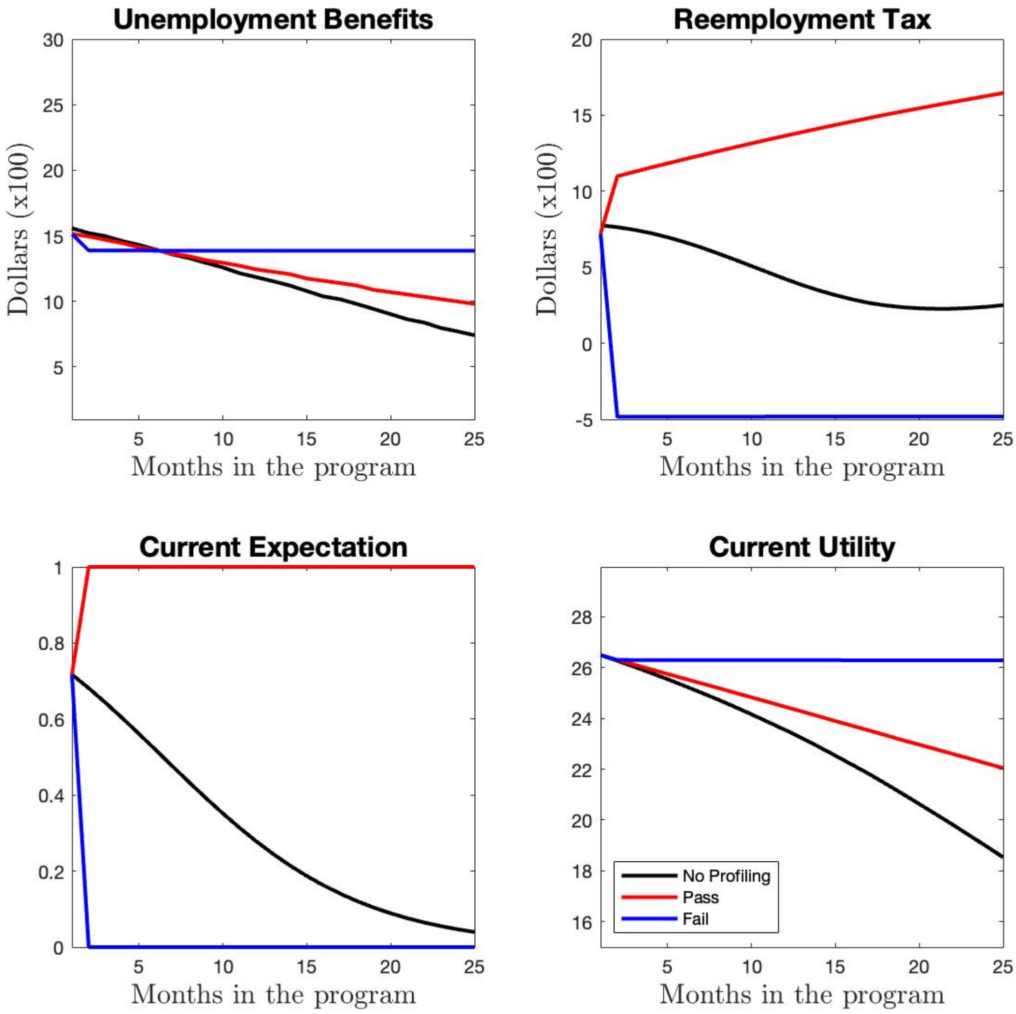


Figure 7: Optimal REA program of Florida for recipients with a high-school diploma over a 25-month horizon (UI+EUC+EB). Initial expectation and generosity are $\mu_0 = 0.72$ and $c(U_0) = \$1,447$, respectively.

7.5 Robustness Checks

A relevant dimension in which workers display a large heterogeneity is the search-effort cost. Various studies report men and women to have different effort costs (Attanasio et al., 2008; Eckstein and Wolpin, 1989). In addition, they document a positive work-effort cost, which is instead assumed zero in the baseline version of the model. I therefore conduct a robustness check by allowing for the search-effort cost to vary by $\pm 10\%$ with respect to the baseline value. Second, I assume that any reemployed worker incurs a working cost equal to the search-effort one. The optimal policies in the (μ, U) space for $e = 0.24$, $e = 0.3$ and $e^w = e = 0.27$ are reported in Fig. 8. When the search-effort cost is lower, then search-delegating policies (UI and IP) expand their areas at the expense of search-assisting 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 search-delegating policies, as it adds extra risk on consumption upon search, and more in general for active labor-market policies, where reemployment is more likely to occur and thus the working effort more likely to be compensated by the government.

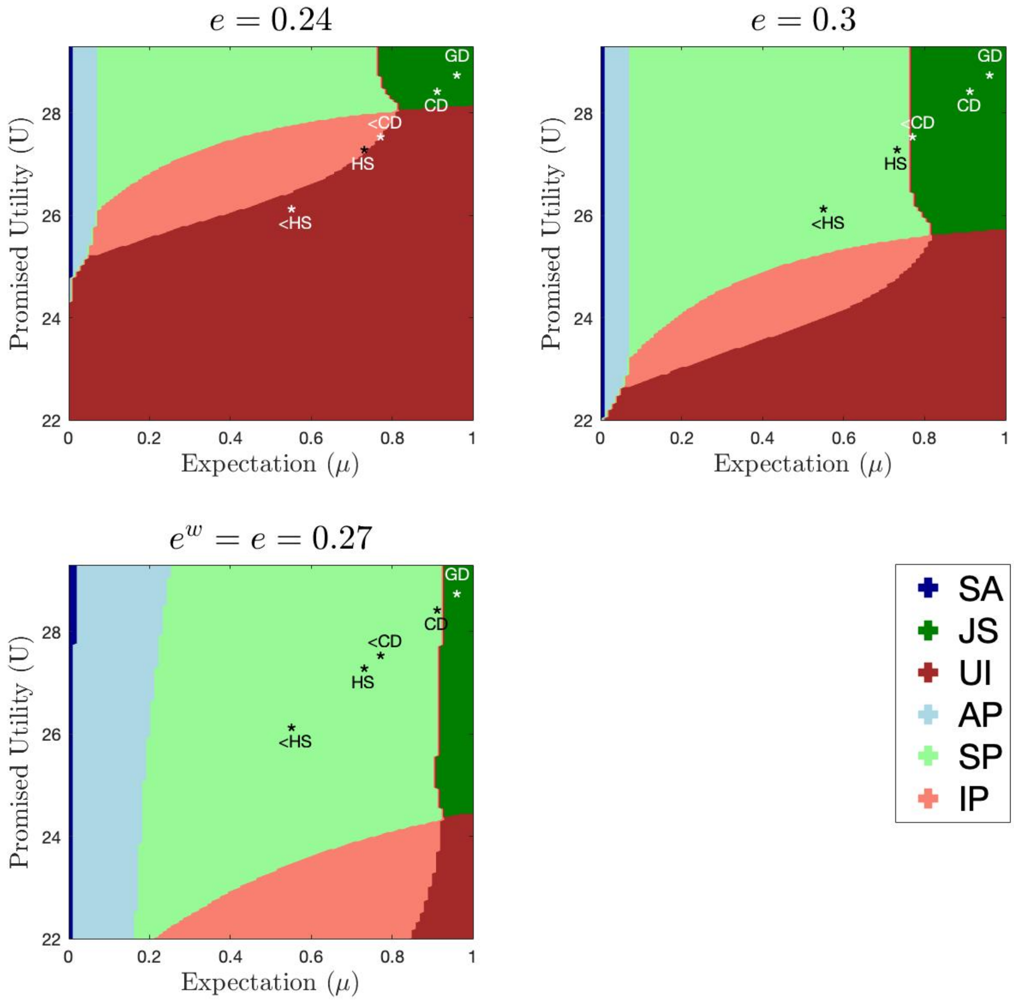


Figure 8: Optimal Policies Space for Different Effort Costs

A second parameter displaying great variability is the cost of assisted search and profiling technologies. Poe-Yamagata et al. (2011) estimate that in Nevada, where reemployment and profiling services are adopted together, 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. 9 (panel a-b) reports the state space of policies for κ^{ja} = \$387 (-10%) and κ^{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 is more convenient *ceteris paribus*. The cost of profiling varies according to the accuracy of the REA program. Given that all possible accuracy levels are allowed, I select the largest cost among all four States, i.e. $\kappa^{wp} = \$134$ (Illinois). As a consequence, the areas of AP, SP and IP shrink in favor of the corresponding welfare policies (see Fig. 9-c).

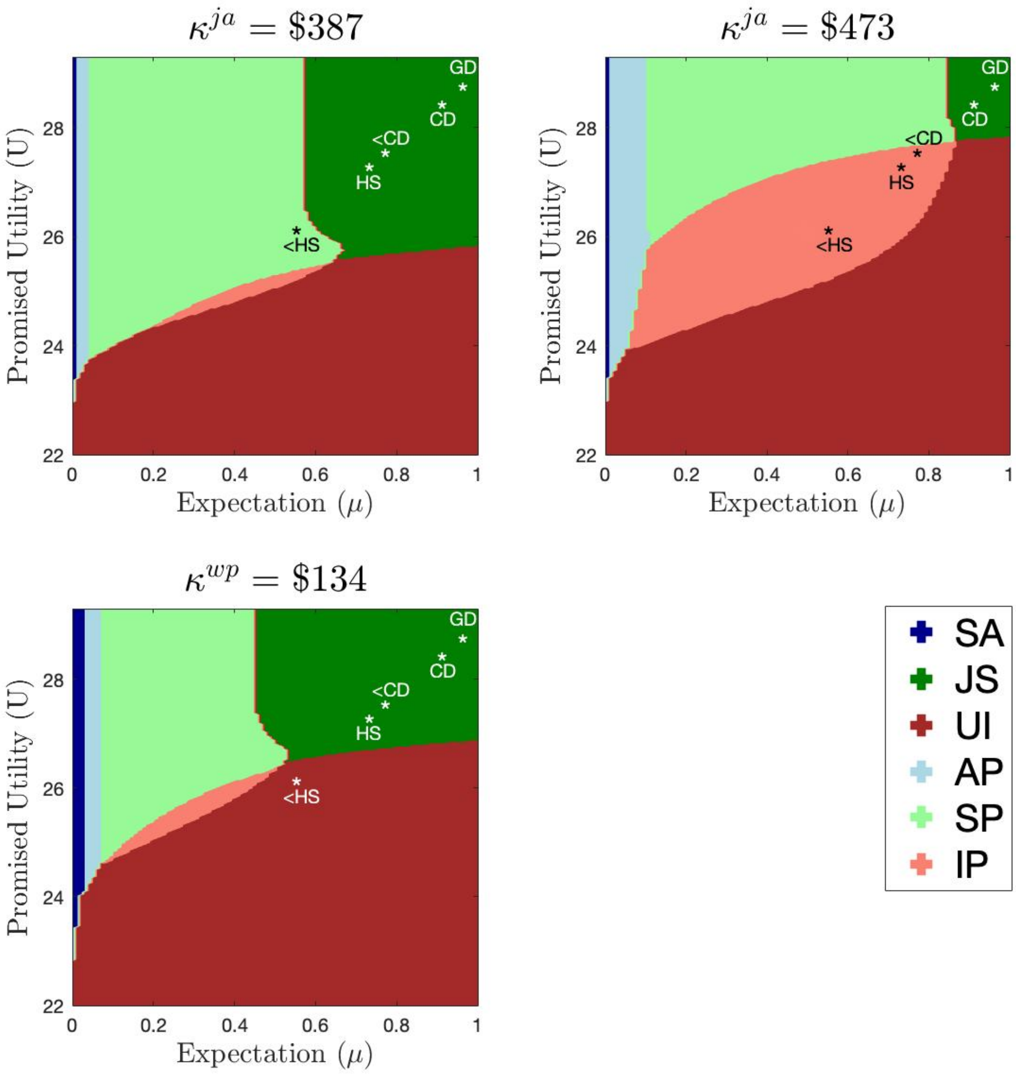


Figure 9: Optimal Policy Space for Different Costs of Assisted Search and Worker Profiling

8 Private Search and Moral Hazard

The government may be unable to observe worker's actions. In particular, worker's search may be private. This may create a misalignment of expectations between the two parties of the agency, whenever the agent shirks effort and thus derives no new information about her skills, while the principal assumes permanence in unemployment to be the result of a failed search and revises expectations 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 effort, other than sparing its cost. To contrast it, the planner promises larger transfers upfront in case of re-employment. Such high-powered incentives are increasing in the prospective duration of private search, following the high number of possible deviations from recommended effort by the agent, and are driven to zero if the next-period policy does not delegate search. Indeed, the agent can benefit from deviation only if a dispersion in promised transfers exists in the two alternative scenarios of reemployment and persistent unemployment.

In response to the possibility of a covert misalignment of expectations, the planner promises *learning rents* to make the agent search. The value of UI and IP thus incorporates such rents. The next step consists of designing 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 deviate multiple -possibly not successive- times, before reverting to search, or even shirk forever after. For this reason, the design of a robust contract could be a complicated task. However, the following result greatly simplifies the problem, by rewriting the (IC) constraint in a compact way.

Proposition 10. *Any contract incentivizing search effort for T periods is robust against any possible deviation, whenever it is robust against one-shot deviations from the sequence of efforts.*[37] *Therefore, the (IC) constraint when implementing UI for T periods reads*

$$U \geq u(c^{ui}) + \beta\big[U^u + \Lambda(T, \mu)\big] \tag{13}$$

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

$$\Lambda(1, \mu^{T-1}) = 0 \tag{14}$$

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

which 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 problem of the planner when private search ends in T periods reads

$$\begin{aligned} V_T^{UI}(\mu, U) &= \max_{c^{ui}, U^w, U^u} -c^{ui} + \beta\big[\pi(\mu)W(\mu, U^w) + (1-\pi(\mu))V_{T-1}^{UI}(\mu', U^u)\big] \\ \text{sub:} \quad V_0^{UI}(\mu, U) &:= \max_{i \in \{SA, JS, SP, AP\}} V^i(\mu, U), \quad \text{(PK)} + (13) \end{aligned}$$

[37]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 feasible path that differs from the optimal one just in a given period. 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.

The only difference with the case of observable effort (other than the (IC) constraint) is constituted by the 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 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.

Passing to IP, 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 (zero type-I and type-II errors), the worker is certain to be high-skilled 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 $\mu = 1$, the net return of profiling is negative, as no information has to be detected. In other words, the worker whose ability is profiled with full precision does not undergo profiling a second time. In case of positive type-II error, instead, workers who are referred to UI can downward revise her expectation and reenter into IP at any later stage (unless they escape 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*

$$\begin{aligned} V_\mu(\mu'_p, U^u_p) = & \frac{V(\mu'_p, U^u_p) - V^{SA}(U^u_f) + V^{SA}_U(U^u_f)(U^u_f - U^u_p)}{\mu'_p} + \\ & + \left[V^{SA}_U(U^u_f) - W_U(\mu, U^w)\right] \frac{\mu_p \Lambda_\mu(t, \mu_p) - \Lambda(t, \mu_p)}{\mu'_p} \end{aligned} \tag{15}$$

Proof. See Appendix D: Moral Hazard. ■

The result sheds light on the complementarity between profiling and private search. Indeed, in AP and SP, the posterior upon 'Pass' is chosen so to equate the marginal gain of persuading low-skilled workers to the marginal cost of reducing profiling accuracy and incurring larger incentive costs in active recipients. In IP, instead, an additional determinant of the optimal type-II error is the reduction of learning rents.

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' skills and employability. At the optimum, active labor-market policies and workers' expectations about personal skills are complementary. Workers who are expected 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 in the form of lower wage taxes or higher wage subsidies. Looking at program generosity, instead, delegated-search policies are adopted for low-end generosities, while assisted-search policies are adopted for high-end generosities, due to agency costs being increasing in program's generosity. During delegated search the pattern of payments decreases over time to ease the agency cost, and makes the program less and less generous along the unemployment spell. Thus, it may occur that assisted search is initially the preferred policy, when the program is more generous, and that it is replaced by delegated search at a later stage of the unemployment spell.

Worker profiling enables the government to align incentives with workers. The implementation of worker profiling divides into gains and losses. On the one hand, profiling allows to match welfare recipients with the most suitable job-search method, according to their actual skills. On the other hand, profiling entails an operative cost -due to personnel and administrative expenses- and a loss due to the detection of low skills on those recipients who have a positive bias on their reemployment expectations at the time they are profiled. The prospective incentive misalignment between government and recipients may be so large to lead to the adoption of a noisy profiling strategy, no matter how little the operative cost is. This type of strategy foregoes profiling the recipients with high-end expectations, whose incentive cost of search delegation is already pretty low, and classifies as high skilled also a fraction of low-skilled workers with lower initial expectations so to persuade them to search at low incentive costs. The rationale at the base of noisy profiling (i.e., positive type-II error) is that low-skilled workers are sufficiently employable and the generosity of the program sufficiently low that, absent incentive costs, the government would compensate their search effort rather than assign them to a passive labor-market policy. Therefore, some of them might better become overconfident about their reemployment chances and be supplied search incentives, even at the cost of higher incentive costs on high-skilled workers.

While noisy profiling is purely driven by a cost-efficiency argument, the government may find it convenient also for other reasons. First, a fully accurate profiling could be discriminatory against ethnic minorities, which are likely to be less employable and, for this reason, referred to passive policies. Thus, the government may rather prefer to classify (at least) a fraction of minority groups as likely to be reemployed and provide them with search incentives. Second, noisy profiling may alleviate the unemployment stigma effect on recipients who have been jobless for longer periods,[38] by signalling higher job-finding chances to prospective employers. These beneficial effects provide further reasons for efficiently 'cheating' on profiled workers about their actual reemployment chances.

Profiling is also used to fine-tune the generosity of downstream transfers. 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 programs like the Reemployment and Eligibility Assessment in the US, should be accompanied by a decreasing pattern of unemployment benefits, as opposed to the constant subsidy under SNAP. On the contrary, workers classified as low skilled should be granted a constant unemployment benefit.

Some questions remain unanswered. First, the paper abstracts from the trade-off between administrative cost of profiling and its accuracy. Nonetheless, in reality per-capita costs of profiling depend on the accuracy with which skills are detected. A more accurate detection leads to a more expensive profiling process (e.g., longer in-person interviews, more elaborate tasks to perform, etc.). In addition, any actual profiling program, like any sort of tests aimed to detect a hidden characteristic, contains a given amount of unavoidable noise that impedes full accuracy. Assuming (i) the cost of profiling to vary in accordance to the change induced on the initial expectation (i.e., known in the information design literature as *entropy cost*),[39] and (ii) information detection to have limited accuracy, might lead to different estimates of the value of worker profiling. In this sense, the welfare gain computed in Section 7 is to 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 have private information on personal reemployment expectations, those assigned to delegated search may underreport expectations, so as to benefit from larger incentives.[40] If this is the case, the government would need to make search-incentivizing policies robust to lying on expectations and this would further exacerbate the cost of incentive provision. In this sense, as noticed in the case of private worker search (Section 8), profiling may constitute a way for the government to curb the information rents that originate in the agency. Third, many welfare programs currently feature hybrid policies devoted to worker profiling and training. Training can improve worker's human capital and reemployment chances, thereof. Therefore, the government needs to design a profiling program that provides workers

[38] A stigma effect on the long-term unemployed has been documented to cause negative duration dependence in the escape rate from unemployment (see Heckman and Borjas, 1980, and Krug et al., 2019).

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

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

with incentives to train, as well as to detect hidden skills. Boleslavsky and Kim (2021) constitute an useful starting point to understand the interplay between incentives to train and incentives to undergo profiling.

References

(1) Altmann, S., Falk A., Jger S. et al., 2018, “Learning about Job Search: a Field Experiment with Job Seekers in Germany", Journal of Public Economics, 164, 33-49;

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

(3) 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;

(4) Attanasio, O., Low H., Sánchez-Marcos V., 2008, “Explaining Changes in Female Labor Supply in a Life-Cycle Model", American Economic Review, 98(4), 1517-1552;

(5) Belot, M., Kircher, P., Muller, P., 2019, “Providing Advice to Jobseekers at Low Cost: an Experimental Study on Online Advice", Review of Economic Studies, 86, 1411-1447;

(6) Bergemann, D., Morris S., 2018, “Information Design: A Unified Perspective", Journal of Economic Literature, 57(1), 44-95;

(7) Berger, M. C., Black D., Smith J. A., 2000, “Evaluating profiling as a means of allocating government services" in Lechner, M. and Pfeiffer, F. (eds), *Econometric Evaluation of Labour Market Policies* (Heidelberg: Physica), 59-84;

(8) Black, D. A., Galdo J., Smith J. A., 2007, “Evaluating the Worker Profiling and Reemployment Services System Using a Regression Discontinuity Approach", American Economic Review, 97(2), 104-107;

(9) Boleslavsky, R., Kim K., 2021, “Bayesian Persuasion and Moral Hazard", *mimeo*;

(10) Cappellini, S., 2020, “Worker Profiling and Reemployment Services", *mimeo*;

(11) 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;

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

(13) Eckstein, Z., Wolpin K. I., 1989, “Dynamic Labour Force Participation of Married Women and Endogenous Work Experience", Review of Economic Studies, 56 (3), 375–390;

(14) Fallick, B. C., 1996, “A Review of the Recent Empirical Literature on Displaced Workers", Industrial and Labor Relations Review, 50:1, 5-16;

(15) Fuller, D. L., 2014, “Adverse Selection and Moral Hazard: Quantitative Implications for Unemployment Insurance", Journal of Monetary Economics, 62, 108-122;

(16) Galperti, S., Perego J., 2018, “A Dual Perspective on Information Design", Available at SSRN: https://ssrn.com/abstract=3297406;

(17) Gibbons, R., Waldman, M., 1999, “A Theory of Wage and Promotion Dynamics Inside Firms", Quarterly Journal of Economics, 114, 1321-1358;

(18) Gibbons, R., Katz L. F., Lemieux T. et al., 2005, “Comparative Advantage, Learning, and Sectoral Wage Determination", Journal of Labor Economics, 23, 681-724;

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

(20) Groes, F., Kircher P., Manovskii I., 2015, “The U-shapes of Occupational Mobility", Review of Economic Studies, 82, 659-692;

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

(22) 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;

(23) Heckman, J. J., Borjas, G., 1980, “Does Unemployment Cause Future Unemployment? Definitions, Questions, and Answers from a Continuous Time Model of Heterogeneity and State Dependence", Economica, 47, 247-283;

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

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

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

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

(28) Kirby, G., Hill H., Pavetti L., Jacobson J., Derr M., Winston P., 2002, “Transitional Jobs: Stepping Stone to Unsubsidized Employment", Mathematica Policy Research Inc.;

(29) Kluve, J., Schmidt C.M., 2002, “Can Training and Employment Subsidies Combat European Unemployment?", Economic Policy, 35, 409-448;

(30) Kolotilin, A., 2018, “Optimal Information Disclosure: a Linear Programming Approach", Theoretical Economics, 13, 607–635;

(31) 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;

(32) Krug, G., Drasch K., Jungbauer-Gans M., 2019, “The social stigma of unemployment: consequences of stigma consciousness on job search attitudes, behaviour and success", Journal of Labour Marker Research, 53(11);

(33) Kudlyak, M., Lkhagvasuren D., Sysuyev R., 2014, “Systematic Job Search: New Evidence from Individual Job Application Data", FRB Richmond Working Paper No. 12-03R;

(34) 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;

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

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

(37) 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;

(38) Miller, R. A., 1984, “Job Matching and Occupational Choice", Journal of Political Economy, 92, 1086-1120;

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

(40) Neal, D., 1999, “The Complexity of Job Mobility among Young Men", Journal of Labor Economics, 17, 237-261;

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

(42) Papageorgiou, T., 2014, “Learning Your Comparative Advantage", Review of Economic Studies, 8, 1263-1295;

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

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

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

(46) 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;

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

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

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

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

(51) 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;

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

(53) 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. The proof follows the same steps as in Pavoni et al. (2016).

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.

■

Lemma 1 (Slopes of the value functions with respect to μ and U). *The value V^i of welfare policy i is increasing in expectations, and decreasing in promised utility. In addition, V^i is concave in U and μ. $\mathbf{V}$ inherits the sign of first- and second-order derivatives of V^i.*

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

$$\begin{aligned}
V^i(\mu, U) &= \max_{(z, U^w, U^u) \in \Gamma(\mu, U)} -g(z) - \kappa^i + \beta\big[p^i(\mu)W(\mu, U^w) + (1 - p^i(\mu))\mathbf{V}(\mu^i, U^u)\big] \\
\text{sub:} \quad & \Gamma^i(\mu, U) = \Big\{(z, U^w, U^u) : \ U = z - e^i + \beta\big[p^i(\mu)U^w + (1 - p^i(\mu))U^u\big], \ U \geq z + \beta U^u\Big\}
\end{aligned}$$

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

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

$\mathbf{V}$ is a contraction. Thus, the proof of the Lemma follows a guess-and-verify approach. Assume that $\mathbf{V}$ is concave both U and in μ. To prove concavity of V^i, 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 $\mathbf{V}$ are concave in U^w and U^u, respectively. Moreover, while $p^i(\mu)W(\mu, U)$ is linear in μ, $(1 - p^i(\mu))\mathbf{V}(\mu', U)$ is concave in μ if $\mathbf{V}$ is concave in the first argument, as

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

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 μ, z, U^w and U^u, and so is the IC constraint. This means that the graph of Γ^i_U (i.e., for constant U) defined as

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

is convex. Therefore, I have shown that V^i is concave in μ. Concavity of V^i in U follows as well, as the graph of Γ^i_μ is convex in U, due to the linearity of PK and IC 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 $\Gamma^i(\mu,.)$ in U, i.e.

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

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

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

which does not depend on U. Therefore, the PK constraint is tightened by an increase of U, which thus leads to a shrinkage of $\Gamma^i(\mu,.)$. And so $V^i_U(\mu,U)\leq 0$.

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 $\mathbf{V}$ in their first argument, μ' is an increasing function of μ and $W(\mu,U^w)\geq\mathbf{V}(\mu^i,U^u)$. Monotonicity of $\Gamma^i(.,U)$, instead, holds as an increase of μ leads to a relaxation of (IC)[41]. Indeed, (IC) is more slack since

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

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

And so $V^i_\mu(\mu, U) \geq 0$.

<u>$\mathbf{V}$ is derivable and $\mathbf{V}_U < 0, \mathbf{V}_{UU} < 0, \mathbf{V}_{\mu\mu} < 0 < \mathbf{V}_\mu$</u>

Assume V defined in (1) is not derivable in U. Then, it must be that U lies at the conjunction between policies i and j. Define i (resp., j) the optimal policy in a left (resp., right) neighborhood of U. Therefore, it must be that

$$V_U(\mu, U^-) = V^i_U(\mu, U^-) \leq V^j_U(\mu, U^+) = V_U(\mu, U^+) \leq 0 \tag{16}$$

Indeed, if the opposite was true, $V^i(\mu, .)$ would strictly dominate $V^j(\mu, .)$ in a right neighborhood of U, which would lead to a contradiction as $V^j(\mu, U^+) = V(\mu, U^+) \geq V^i(\mu, U^+)$ by definition of V.

By (16), it must be the case that there exist $U_*, U^* :\ U_* < U < U^*$ such that

$$\mathbf{V}(\mu, U) = \rho V(\mu, U_*) + (1-\rho)V(\mu, U^*) > V(\mu, U), \quad \text{with } \rho U_* + (1-\rho)U^* = U$$

Recall definition (2) and let λ be the Lagrange multiplier of constraint

$$U = \int_0^1 \rho(x)U(x)dx$$

From the Envelope theorem and the first-order condition, it follows that

$$\mathbf{V}_U(\mu, U) = -\lambda = V_U(\mu, U(x)), \quad \forall x :\ \rho(x) > 0$$

which, in the case under analysis, implies that $\mathbf{V}_U(\mu, U) = V_U(\mu, U_*) = V_U(\mu, U^*)$. Thus, $\mathbf{V}$ is derivable over the whole U space.

Concavity of $\mathbf{V}$ in U follows from concavity of V^i in the second argument, linearity of the objective function (2) in the choice variables $\rho(x)$, and linearity of the Promise Keeping constraint in $U(x)$, $\rho(x)$ and U.

Similarly, concavity of $\mathbf{V}$ in μ follows from concavity of V^i in both arguments, linearity of the objective function (2) in the choice variables $\rho(x)$, and linearity of the Promise Keeping constraint in $U(x)$ and $\rho(x)$. ■

Proof of Prop. 2

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

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

by the envelope theorem and first-order condition of c_i.

As $\mathbf{V}$ is the fixed-point of a contraction, the proof follows a guess-and-verify approach. I guess $\mathbf{V}$ is 'locally' supermodular, i.e.

$$\mathbf{V}_{\mu U}(\mu, U) \geq 0$$

Optimal Policies in the U Space

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

$$V_U^{UI}(\mu, U) \leq V_U^{JS}(\mu, U) \leq V_U^{SA}(U) = W_U(\mu, U) \tag{18}$$

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

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

I now prove $V_U^{JS}(\mu, U) \leq V_U^{SA}(U)$. By the envelope theorem and first-order conditions,

$$-\frac{1}{u'(c_{JS})} = V_U^{JS}(\mu, U) = W_U(\mu, U_{JS}^w) = -\frac{1}{u'(g((1-\beta)U_{JS}^w))}$$

which implies that (PK) constraint can be rewritten as

$$U = u(c_{JS}) + \beta\big[\pi(\mu)U_{JS}^w + (1-\pi(\mu))U_{JS}^u\big] = \big(1 - \beta + \beta\pi(\mu)\big)U_{JS}^w + \beta(1-\pi(\mu))U_{JS}^u \tag{19}$$

Moreover, by the envelope theorem and first-order conditions,

$$\mathbf{V}_U(\mu, U) = V_U^{JS}(\mu, U) = \mathbf{V}_U(\mu', U_{JS}^u) \leq \mathbf{V}_U(\mu, U_{JS}^u) \tag{20}$$

where the first equality as JS is optimal in (μ, U) and the inequality holds since $\mathbf{V}$ is supermodular. Thus, by concavity of $\mathbf{V}$ in U and (19), it holds that

$$U_{JS}^u \leq U \leq U_{JS}^w = \frac{u(c_{JS})}{1-\beta} \tag{21}$$

Therefore, I have shown that $V_U^{JS}(\mu, U) = -\frac{1}{u'(c_{JS})} \leq -\frac{1}{u'(g((1-\beta)U))} = V_U^{SA}(U)$.

What is left to show is that $V_U^{UI}(\mu, U) \leq V_U^{JS}(\mu, U)$. *Per contra*, assume that $V_U^{UI}(\mu, U) > V_U^{JS}(\mu, U)$. First, the inequality

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

implies that $(1-\beta)U \leq u(c_{UI}) < u(c_{JS})$. And from (21), it must be that $U_{JS}^u < U < \frac{u(c_{JS})}{1-\beta} =$

U_{JS}^w. Merging envelope and first-order conditions of UI yields

$$\mathbf{V}_U(\mu', U_{UI}^u) - V_U^{UI}(\mu, U) = \mathbf{V}_U(\mu', U_{UI}^u) + \lambda^{UI} - \chi^{UI} = \frac{\pi(\mu)}{1-\pi(\mu)}\chi^{UI} > 0 \tag{22}$$

Now, by (20) and (22)

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

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

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

where the inequality follows from $c_{JS} > c_{UI}$ and $\frac{u(c_{JS})}{1-\beta} > U_{JS}^u$.

Therefore, it has been shown that $V_U^{UI}(\mu, U) \leq V_U^{JS}(\mu, U)$.

Optimal Policies in the μ Space

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

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

By (3), the derivative of V^{SA} is null.

The derivative of V^{JS} with respect to μ reads

$$\begin{aligned} V_\mu^{JS}(\mu, U) =& \beta(\pi_H - \pi_L)\big[W(\mu, U_{JS}^w) - \mathbf{V}(\mu', U_{JS}^u) - \lambda^{JS}(U_{JS}^u - U_{JS}^w)\big] + \\ &+ \beta\Big[\pi(\mu) W_\mu(\mu, U_{JS}^w) + (1-\pi(\mu))\frac{\partial \mu'}{\partial \mu}\mathbf{V}_\mu(\mu', U_{JS}^u)\Big] \end{aligned} \tag{23}$$

Consider JS being implemented in the current period. Using the first-order condition $-\lambda^{JS} = W_U(\mu, U_{JS}^w)$, it holds that

$$\begin{aligned} &W(\mu, U_{JS}^w) + W_U(\mu, U_{JS}^w)(U_{JS}^u - U_{JS}^w) - \mathbf{V}(\mu', U_{JS}^u) > W(\mu, U_{JS}^u) - \mathbf{V}(\mu', U_{JS}^u) > \\ &> W(\mu', U_{JS}^u) - \mathbf{V}(\mu', U_{JS}^u) > 0 \end{aligned}$$

where the first inequality follows from concavity of W in U and the second from monotonicity of W in μ. In addition, $\mathbf{V}_\mu(\mu', U_{JS}^u) \geq 0$ by Lemma 1, which proves that $V_\mu^{JS}(\mu, U) > 0$.

Consider a policy that implements UI with the additional constraint that $U_{UI}^u \geq U_{JS}^u$, and label its value $\hat{V}^{UI}(\mu, U, U^u_{JS})$. Its derivative with respect to μ reads

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

where the second equality follows from the first-order conditions $-\lambda^{UI}=W_U(\mu,U^w_{UI})$.

I now prove two facts, that are used in the remainder of the proof. The first fact is that $U^w_{JS}<U^w_{UI}$, since

$$W_U(\mu,U^w_{JS})=V^{JS}_U(\mu,U)\geq V^{UI}_U(\mu,U)=-(\lambda^{UI}-\chi^{UI})>-\lambda^{UI}=W_U(\mu,U^w_{UI}) \tag{25}$$

where the first inequality follows from (18). The second fact is that $U^u_{UI}<U$, since

$$-\frac{1}{u'(c_{UI})}=V^{UI}_U(\mu,U)<V^{SA}_U(U)=-\frac{1}{u'(c_{SA})}\Longrightarrow(1-\beta)U=u(c_{SA})<u(c_{UI})=U-\beta U^u_{UI} \tag{26}$$

where the first inequality follows from (18), and the second from (3) and the (IC) constraint in the definition of V^{UI}.

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

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

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 $U^u_{UI}<U\leq U^w_{JS}<U^w_{UI}$ by (21)-(25)-(26);
- $\mathbf{V}(\mu',U^u_{UI})\leq\mathbf{V}(\mu',U^u_{JS})+\mathbf{V}_U(\mu',U^u_{JS})(U^u_{UI}-U^u_{JS})$, by concavity of $\mathbf{V}$ in U.

The second inequality holds since $W_{\mu U}(\mu,U)=0$ and $\mathbf{V}_\mu(\mu',U^u_{JS})\leq\mathbf{V}_\mu(\mu',U^u_{UI})$, by assumption

$U^u_{JS} \le U^u_{UI}$ 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.

V^{UI} and V^{JS} are supermodular

In order to check that the guess about supermodularity of $\mathbf{V}$ makes sense, one must verify that each V^i is supermodular too.

First, from (3), $V^{SA}_{\mu U}(U) = 0$.

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

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

Assume *per contra* that $V^{UI}_{\mu U}(\mu, U) \le 0$. 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 (21). Hence, I have reached a contradiction.

■

Proof of Prop. 3

Unemployment Benefits

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

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

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

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

The result is shown by induction argument.

Proof of Prop. 4

Assume that UI is the optimal policy in (μ, U). By Prop. 2 and definition of $\hat{U}$, it follows that $U \leq \hat{U}(\mu)$.

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

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

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

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

where the second inequality holds by assumption (9). 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 never optimal to switch from UI to JS, as $U_{UI}^u \le \hat{U}(\mu')$.

Appendix B: Properties of AP, SP and IP

Proof of Prop. 5

Proof. Consider the definition of the general policy i as in (10) and define λ the Lagrange multiplier of Promise Keeping constraint, χ the Lagrange multiplier of Incentive Compatibility constraint, ξ the Lagrange multiplier of No Discrimination constraint, while the Martingale Property constraint is rewritten as

$$q = \frac{\mu_p - \mu_f}{\mu - \mu_f}$$

The first-order derivative with respect to μ_f reads

$$\begin{aligned}
&-\beta p^{i\prime}(\mu_f)(1-q)\big[\mathbf{V}(\mu_f^i, U_f^u) + \lambda U_f^u\big] + \beta(1-q)(1-p^i(\mu_f))\mathbf{V}_\mu(\mu_f^i, U_f^u)\frac{\partial \mu_f^i}{\partial \mu_f} + \\
&+\frac{\partial q}{\partial \mu_f}\beta\big[(1-p^i(\mu_p))(\mathbf{V}(\mu_p^i, U_p^u) + \lambda U_p^u) - (1-p^i(\mu_f))(\mathbf{V}(\mu_f^i, U_f^u) + \lambda U_f^u) - \chi(U_p^u - U_f^u)\big] \\
=& -A\Big\{p^{i\prime}(\mu_f)\big[\mathbf{V}(\mu_f^i, U_f^u) + \lambda U_f^u\big] - (1-p^i(\mu_f))\mathbf{V}_\mu(\mu_f^i, U_f^u)\frac{\partial \mu_f^i}{\partial \mu_f} + \\
&+\frac{1}{\mu_p - \mu_f}\big[(1-p^i(\mu_p))(\mathbf{V}(\mu_p^i, U_p^u) + \lambda U_p^u) - (1-p^i(\mu_f))(\mathbf{V}(\mu_f^i, U_f^u) + \lambda U_f^u) - \chi(U_p^u - U_f^u)\big]\Big\} \\
=& -A\Big\{-(1-p^i(\mu_f))\mathbf{V}_\mu(\mu_f^i, U_f^u)\frac{\partial \mu_f^i}{\partial \mu_f} + \frac{1-p^i(\mu_p)}{\mu_p - \mu_f}\big[\mathbf{V}(\mu_p^i, U_p^u) - \mathbf{V}(\mu_f^i, U_f^u) + \lambda(U_p^u - U_f^u)\big] - \frac{\chi(U_p^u - U_f^u)}{\mu_p - \mu_f}\Big\} \\
<& -A\Big\{(1-p^i(\mu_f))\frac{\partial \mu_f^i}{\partial \mu_f}(\mathbf{V}_\mu(\mu_p^i, U_f^u) - \mathbf{V}_\mu(\mu_f^i, U_f^u)) + \frac{1-p^i(\mu_p)}{\mu_p - \mu_f}\big[\mathbf{V}(\mu_p^i, U_p^u) - \mathbf{V}(\mu_p^i, U_f^u) + \lambda(U_p^u - U_f^u)\big] - \\
&\qquad - \frac{\chi(U_p^u - U_f^u)}{\mu_p - \mu_f}\Big\} \\
<& -A\Big\{(1-p^i(\mu_f))\frac{\partial \mu_f^i}{\partial \mu_f}(\mathbf{V}_\mu(\mu_p^i, U_f^u) - \mathbf{V}_\mu(\mu_f^i, U_f^u)) + \frac{U_p^u - U_f^u}{\mu_p - \mu_f}\Big[(1-p^i(\mu_p))\big(\mathbf{V}(\mu_p^i, U_p^u) + \lambda\big) - \chi\Big]\Big\} \\
=& -A(1-p^i(\mu_f))\frac{\partial \mu_f^i}{\partial \mu_f}\big(\mathbf{V}_\mu(\mu_p^i, U_f^u) - \mathbf{V}_\mu(\mu_f^i, U_f^u)\big) < 0
\end{aligned}$$

with $A = \beta(1-q)$.

The first inequality follows from concavity of $\mathbf{V}$ in the first argument

$$\mathbf{V}(\mu^i_f, U^u_f) < \mathbf{V}(\mu^i_p, U^u_f)+(\mu^i_f-\mu^i_p)\mathbf{V}_\mu(\mu^i_p, U^u_f) = \mathbf{V}(\mu^i_p, U^u_f)+\frac{\partial \mu^i_f}{\partial \mu_f}\frac{1-p^i(\mu_f)}{1-p^i(\mu_p)}(\mu_f-\mu_p)\mathbf{V}_\mu(\mu^i_p, U^u_f)$$

The second inequality follows from concavity of $\mathbf{V}$ in the second argument

$$\mathbf{V}(\mu^i_p, U^u_f) < \mathbf{V}(\mu^i_p, U^u_p) + \mathbf{V}_U(\mu^i_p, U^u_p)(U^u_f - U^u_p)$$

The second-to-last passage follows from the slack constraint $\xi(U^u_p - U^u_f) = 0$ and first-order condition on U^u_p,

$$\beta q\{(1-p^i(\mu_p))\big[V_U(\mu^i_p, U^u_p) + \lambda\big] - \chi\} + \xi = 0$$

And the third (and last) inequality follows as (i) profiling outcomes refer workers to different policies,[42] (ii) $\mu^i_f < \mu^i_p$, and (iii) condition (8) of Prop. 2.
Therefore, it must be that $\mu_f = 0$. Referral to SA follows from the fact that SA is optimal (if any) at low-end expectations. ■

Proof of Prop. 6

Consider the definition of the general policy i as in (10), with $\mu_f = 0$ and assume that $\mu_p < 1$. The first-order condition for μ_p reads

$$\begin{aligned}
&-\beta p^{i\prime}(\mu_p)q\big[\mathbf{V}(\mu^i_p, U^u_p) + \lambda U^u_p\big] + \beta q(1-p^i(\mu_p))\mathbf{V}_\mu(\mu^i_p, U^u_p)\frac{\partial \mu^i_p}{\partial \mu_p}+\\
&+\frac{\partial q}{\partial \mu_p}\beta\big[(1-p^i(\mu_p))(\mathbf{V}(\mu^i_p, U^u_p) + \lambda U^u_p) - (1-p^i(0))(\mathbf{V}(0, U^u_f) + \lambda U^u_f) - \chi(U^u_p - U^u_f)\big] =\\
&= -\frac{A}{\mu_p}\Big\{(1-p^i(0))\big[\mathbf{V}(\mu^i_p, U^u_p) - \mathbf{V}(0, U^u_f) + \lambda(U^u_p - U^u_f) - \mu^i_p\mathbf{V}_\mu(\mu^i_p, U^u_p)\big] - \chi(U^u_p - U^u_f)\Big\}\\
&= -\frac{A}{\mu_p}(1-p^i(0))\Big\{\mathbf{V}(\mu^i_p, U^u_p) - \mathbf{V}(0, U^u_f) - (U^u_p - U^u_f)\mathbf{V}_U(0, U^u_f) - \mu^i_p\mathbf{V}_\mu(\mu^i_p, U^u_p)\Big\} = 0
\end{aligned}$$

with $A = \beta q$. The first passage holds as $p^i(0) + \mu_p p^{i\prime}(\mu_p) = p^i(\mu_p)$ and

$$\mu_p(1-p^i(\mu_p))\frac{\partial \mu^i_p}{\mu_p} = \mu_p\frac{(1-p^i(0))(1-p^i(1))}{(1-p^i(\mu_p))} = \mu^i_p(1-p^i(0))$$

The second passage holds as the first-order condition of U^u_f reads

$$\beta(1-q)\big\{(1-p^i(0))[\mathbf{V}_U(0, U^u_f) + \lambda] - \chi\big\} - \xi = 0$$

[42]Since each policy is concave in the first argument, referring workers to the same policy, irrespective of the profiling outcome, would produce a net loss.

Hence,

$$(U_p^u - U_f^u)[(1 - p^i(0))\lambda - \chi] = -(U_p^u - U_f^u)(1 - p^i(0))\mathbf{V}_U(0, U_f^u)$$

Lastly, I will show that $U_p^u = U_f^u$. Assume *per contra* that $U_p^u > U_f^u$. Then, the first-order conditions of U_p^u and U_f^u lead to

$$(1 - p^i(\mu_p))[\mathbf{V}_U(\mu_p^i, U_p^u) + \lambda] = (1 - p^i(0))[\mathbf{V}_U(0, U_f^u) + \lambda] > (1 - p^i(\mu_p))[\mathbf{V}_U(0, U_f^u) + \lambda] > 0$$
$$\Longrightarrow \mathbf{V}_U(\mu_p^i, U_p^u) > \mathbf{V}_U(0, U_f^u) = V^{SA}(U_f^u) \geq \mathbf{V}_U(\mu_p^i, U_f^u)$$

where the second inequality follows from (7) in Section 4. And concavity of $\mathbf{V}$ in U leads to $U_p^u < U_f^u$, which contradicts the assumption.

Therefore, at the optimum, μ_p equals 1 or solves condition (11).

Proof of Prop. 7

Proof. Consider the two first-order (and envelope) conditions of the AP problem

$$\mathbf{V}_\mu(\mu_p, U_{AP}^u) - \frac{\mathbf{V}(\mu_p, U_{AP}^u) - \mathbf{V}(0, U_{AP}^u)}{\mu_p} = 0$$
$$q\mathbf{V}_U(\mu_p, U_{AP}^u) + (1 - q)\mathbf{V}_U(0, U_{AP}^u) - V_U^{AP}(\mu, U) = 0$$

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

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

Differentiating the two conditions by U yields

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

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

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

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

$$\frac{\partial U_{AP}^u}{\partial U}\Delta - V_{UU}^{AP}(\mu, U) = 0$$

which shows that $\frac{\partial U_{AP}^u}{\partial U} > 0$, since $V_{UU}^{AP}(\mu, U) < 0$. From the first equation, local supermodularity of $\mathbf{V}$ and $\mathbf{V}_U(\mu_p, U) < \mathbf{V}_U(0, U)$ (see Prop. 2) yields $\frac{\partial \mu_p}{\partial U_{AP}^u} > 0$, which delivers the result.
The proof of μ_p being monotone increasing in U also under SP and IP follows accordingly.

■

Appendix C: Statistical Profiling

Proof of Prop. 8 and 9

Proof. The expression (12) follows from the first-order condition in Prop. 6.

Assistance-and-Profiling (AP)

At the optimum,

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

which implies that $c_{AP} = c_{AP}^{u,p} = c_{AP}^{u,f}$, and $U_{AP}^{u,p} \leq U \leq U_{AP}^{u,f}$. 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, by Prop. 2,

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

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

$$\begin{aligned}
&\mathbf{V}_U(\mu_p, U^{u,p}_{AP}) + \lambda^{AP} = 0 \\
&-\frac{\mathbf{V}(\mu_p, U^{u,p}_{AP}) - \mathbf{V}(0, U^{u,f}_{AP}) + \lambda^{AP}(U^{u,p}_{AP} - U^{u,f}_{AP})}{\mu_p} + \mathbf{V}_\mu(\mu_p, U^{u,p}_{AP}) = 0
\end{aligned}$$

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 **??**). 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

$$\begin{aligned}
&V^{SP}_U(\mu, U) = W_U(\mu_p, U^{w,p}_{SP}) = W_U(0, U^{w,f}_{SP}) = \mathbf{V}_U(\mu'_p, U^{u,p}_{SP}) = \mathbf{V}_U(0, U^{u,f}_{SP}) \\
\Longrightarrow &c_{SP} = c^w_{SP} = c^{u,p}_{SP} = c^{u,f}_{SP}, \quad U^{u,p}_{SP} \leq U \leq U^{u,f}_{SP} = U^{w,p}_{SP} = U^{w,f}_{SP}
\end{aligned}$$

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

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

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

$$\begin{aligned}
&\mathbf{V}_U(\mu'_p, U^{u,p}_{IP}) - V^{IP}_U(\mu, U) = \frac{\pi(\mu_p)}{1-\pi(\mu_p)}\chi^{IP} > \frac{\pi_L}{1-\pi_L}\chi^{IP} = \mathbf{V}_U(0, U^{u,f}_{IP}) - V^{IP}_U(\mu, U) \Longrightarrow U^{u,p}_{IP} < U^{u,f}_{IP} \\
&\mathbf{V}_U(\mu'_p, U^{u,p}_{IP}) > \mathbf{V}_U(0, U^{u,f}_{IP}) > V^{IP}_U(\mu, U) = -(\lambda^{IP} - \chi^{IP}) > W_U(\mu, U^w_{IP}) \Longrightarrow c^{u,p}_{IP} < c^{u,f}_{IP} < c_{IP} < c^w_{IP}
\end{aligned} \tag{28}$$

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

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

where the first inequality follows from (18), the second one from (28) and the last one from $U^{u,p}_{IP} < U^{u,f}_{IP}$.

■

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 (14)

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

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 (14), and is independent of U^u. In addition, $\Lambda(t+1,\mu) \ge \Lambda(t,\mu)$, with inequality being strict for $\mu \in (0,1)$.[43] 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 2. $(V^{UI}_t)_{t\ge1}$ *is increasing in* μ.

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

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

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

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).[44] 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 (14), 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

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

[44] (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

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

<u>Base Step $(t = 2)$</u>

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

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

<u>Induction Step</u>

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

When profiling is adopted jointly with private search, the planner's problem reads

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

Then, the first-order condition reads

$$
\begin{aligned}
& \frac{\partial q}{\partial \mu_p}\big[(1-\pi(\mu_p))V(\mu'_p,U^u_p)-(1-\pi_L)V(0,U^u_f)\big]-q(\pi_H-\pi_L)V(\mu'_p,U^u_p)+ \\
& +q\frac{(1-\pi_H)(1-\pi_L)}{1-\pi(\mu_p)}V_\mu(\mu'_p,U^u_p)+\lambda\Big[\frac{\partial q}{\partial \mu_p}\big[(1-\pi(\mu_p))U^u_p-(1-\pi_L)U^u_f\big]-q(\pi_H-\pi_L)U^u_p\Big]- \\
& -\chi\Big[\frac{\partial q}{\partial \mu_p}\Lambda(t,\mu_p)+q\Lambda_\mu(t,\mu_p)\Big]=0
\end{aligned}
$$

Which can be rewritten as

$$
V_\mu(\mu'_p,U^u_p)=\frac{V(\mu'_p,U^u_p)-V(0,U^u_f)}{\mu'_p}+\Big(\lambda-\frac{\chi}{1-\pi_L}\Big)\frac{U^u_p-U^u_f}{\mu'_p}+\frac{\chi}{1-\pi_L}\frac{\mu_p\Lambda_\mu(t,\mu_p)-\Lambda(t,\mu_p)}{\mu'_p}
$$

and plugging in $-V^{SA}_U(U^u_f)=-V_U(0,U^u_f)=\lambda-\frac{\chi}{1-\pi_L}$ 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. From the basic monthly Current Population Survey (CPS), I first derive the fraction of high- and low-skilled workers for each level of educational attainment ϑ_i, $i \in$\{*LessHighSc.*, *HighSc.*, *SomeCollege*, *College*, *Graduate*\}, with high-skilled workers being those who earn a wage higher than the mean. 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

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

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, ψ_{it}. Lastly, I compute $\{\pi_H, \pi_L\}$ that minimize

$$\{\pi_H, \pi_L\} = \arg\min_{\hat{\pi}_H, \hat{\pi}_L} \sum_t \Big(\sum_i \psi_{it}(\vartheta_i \hat{\pi}_H + (1-\vartheta_i)\hat{\pi}_L) - \pi_t \Big)^2$$

that is,

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

with $a_t = \sum_i \psi_{it}\vartheta_i$, $b_t = \sum_i \psi_{it}(1-\vartheta_i) = 1 - a_t$.[45] 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.[46]	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 $\vert$ 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.

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

[46] '< 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.

Annex I Missouri Self-Eligibility Assessment Form

DIVISION OF EMPLOYMENT SECURITY

P.O. Box 3915
Jefferson City, MO 65102-3915
www.labor.mo.gov/DES

Date Mailed:

Reemployment Services and Eligibility Assessment

You have been selected to participate in a Reemployment Services and Eligibility Assessment (RESEA) program designed to help you return to work. As a condition of continued eligibility for unemployment benefits, you must participate in an interview at the Missouri Job Center shown below. If you have already returned to work, contact the Missouri Job Center at the phone number listed below.

In the interview, you will receive individual assistance with your reemployment plan, reemployment services and finding training that may be available.

You are required to bring photo identification, your completed Work Search Record, and your completed Self-Evaluation Questionnaire (enclosed) to the interview. For your convenience, a Work Search Record is located on the reverse of the Self-Evaluation Questionnaire. If you have a resume, bring the most recent version with you.

The interview may last up to 90 minutes. You can significantly reduce the time you spend at the Job Center if you pre-register with MoJobs at https://app-jobs.mo.gov prior to your appointment.

Participation in this program is mandatory and failure to attend any scheduled RESEA appointment will result in an interruption of your unemployment benefits. If you are unable to attend this interview for any reason, you must call the Missouri Job Center prior to the scheduled appointment date and time and explain why you cannot attend.

Report on at to:

MISSOURI JOB CENTER

Local Phone

***IMPORTANT:** If needed, call 573-751-9040 for assistance in the translation and understanding of the information in this document.*
***¡IMPORTANTE!:** Si es necesario, llame al 573-751-9040 para asistencia en la traducción y entendimiento de la información en este documento.*
Missouri Division of Employment Security is an equal opportunity employer/program. Auxiliary aids and services are available upon request to individuals with disabilities. TDD/TTY: 800-735-2966 Relay Missouri: 711

MISSOURI DEPARTMENT OF LABOR & INDUSTRIAL RELATIONS

MODES-4632 (04-18)
Benefits

DIVISION OF EMPLOYMENT SECURITY

MISSOURI DEPARTMENT OF LABOR AND INDUSTRIAL RELATIONS

ELIGIBILITY ASSESSMENT

1. Have you worked since you filed for unemployment insurance benefits? This includes full-time work, part-time work, or temporary work. ☐ Yes ☐ No If Yes, provide dates of employment.
   Beginning Employment Date: ____________________ Ending Employment Date: _________________
2. Please provide your rate of pay on your last job.
   Hourly wage: $__________ or Salary: $_______________ ☐ Weekly ☐ Monthly
3. How much experience did you have on that job? (check one)
   ☐ Less than 6 months ☐ 6 months – 1 year ☐ 1 year – 3 years ☐ 3 years – 5 years ☐ 5+ years
4. Are you looking for: ☐ Full-time work ☐ Part-time work ☐ Both
5. What type of work are you seeking?
   ☐ Construction ☐ Retail ☐ Office Services ☐ Management
   ☐ Manufacturing ☐ Transportation ☐ Health Care ☐ Other _________________________
6. What days are you available for work? (check all that apply)
   ☐ Sunday ☐ Monday ☐ Tuesday ☐ Wednesday ☐ Thursday ☐ Friday ☐ Saturday
7. What hours are you available for work?
   From: _______________ ☐ a.m. ☐ p.m. To: _______________ ☐ a.m. ☐ p.m.
8. What is the lowest pay you will accept for work?
   Hourly wage: $__________ or Salary: $_______________ ☐ Weekly ☐ Monthly
9. What type of transportation do you have to get to a job? (check one)
   ☐ Private Vehicle ☐ Public Transportation ☐ Family/Friend ☐ None ☐ Other _______________
10. How many miles are you willing to travel to a job (one way)? (check one)
    ☐ 0-5 miles ☐ 5-10 miles ☐ 10-20 miles ☐ 20-30 miles ☐ More ____________________
11. Do you attend or plan to attend school or training? ☐ Yes ☐ No
    If currently attending school or training, provide name of educational or training institution: ____________
12. Are you self-employed? ☐ Yes ☐ No
    If Yes, please provide the number of hours worked per week. __________ hours worked per week.
13. Do you have limitations that may keep you from performing the type of work that you are seeking?
    ☐ Yes ☐ No If Yes, please explain. ____
    ____
    ____
    ____
14. Do you have dependents who require care during work hours? ☐ Yes ☐ No
    If Yes, will you be able to make arrangements for the dependents if you are offered work? ☐ Yes ☐ No

Name ____ Date ______________

MISSOURI DEPARTMENT OF LABOR AND INDUSTRIAL RELATIONS

WORK SEARCH RECORD

Name	Social Security Number XXX-XX-

Keep a list of all the employers and labor unions you contact each week while claiming unemployment benefits. Make at least as many contacts each week as you were instructed when you first filed. **You must bring your completed Work Search Record with you when you report for your Reemployment Services and Eligibility Assessment interview.** You can get additional copies of this form by visiting: labor.mo.gov/sites/default/files/pubs_forms/4633-AI.pdf or you may use your own sheet.

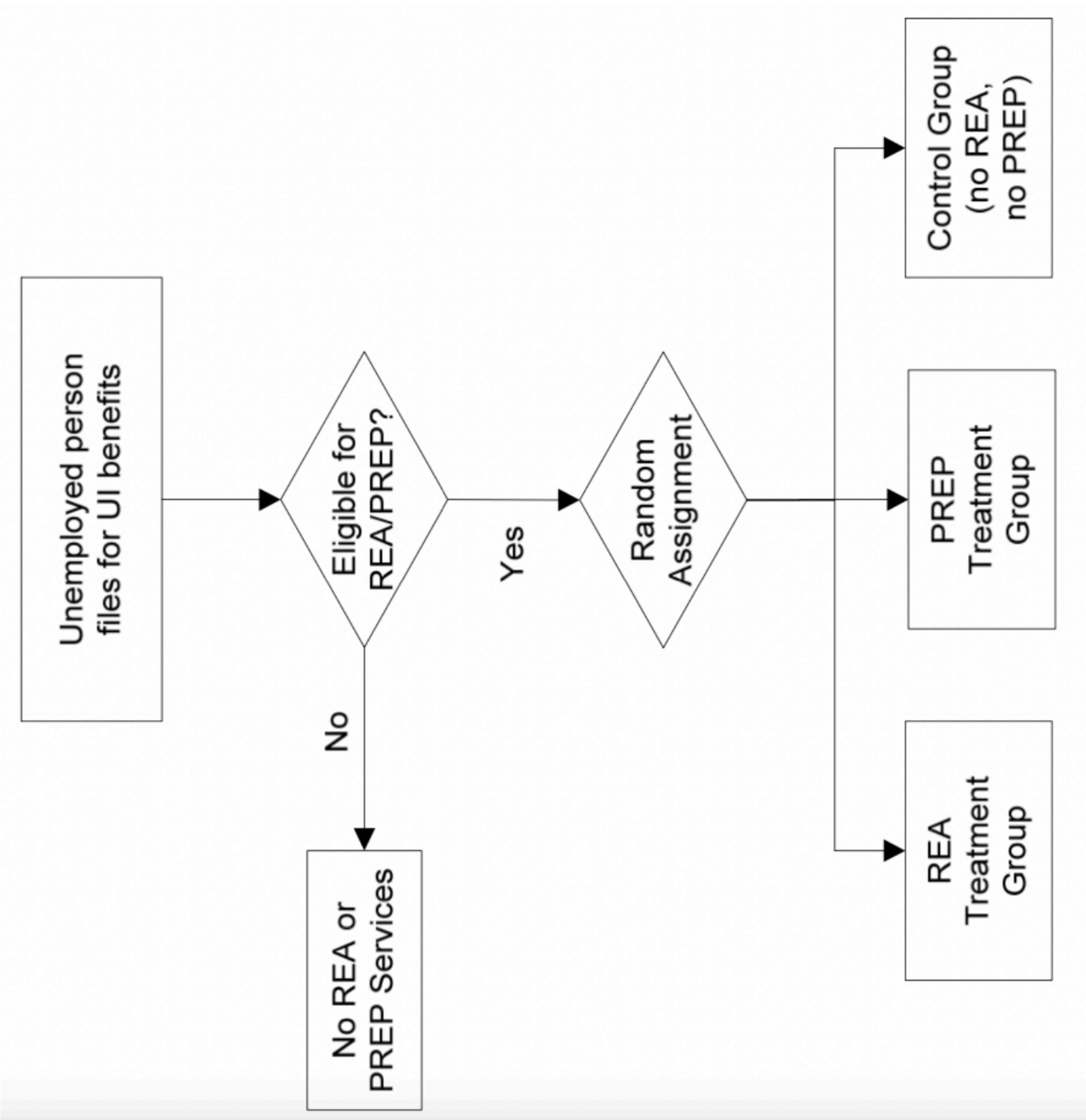
Date of Contact	Employer's Name Address, and Phone Number	Method of Contact*	Name/Title of Person Contacted	Position Applied For	Was Application Taken?	Result of Contact
1-25-16	*ABC Company - 829 Juniper Kansas City, MO 64111 816-555-1221*	*T*	*Eric Dean, Manager*	*Warehouse*	*Yes*	*Check back in Feb.*

*T-Telephone P-In Person R-Sent Resume I-Internet

IMPORTANT: *If needed, call 573-751-9040 for assistance in the translation and understanding of the information in this document.*
¡IMPORTANTE!: *Si es necesario, llame al 573-751-9040 para asistencia en la traducción y entendimiento de la información en este documento.*

Missouri Division of Employment Security is an equal opportunity employer/program. Auxiliary aids and services are available upon request to individuals with disabilities. TDD/TTY: 800-735-2966 Relay Missouri: 711

Annex II Florida REA Selection Process



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

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

  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 the worker finds reemployment
- $\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))$ -with $\sigma^t = (\sigma^{t-1}, \sigma_t)$- is the expectation held by the worker during unemployment, expressing the probability about state H. This is clearly a non-contractible variable, as the worker can hide it from the government. $\mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))$ 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 the worker 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, the worker does not revise expectation[47]

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

[47] 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 $\sigma_{t+1} = w$.

- $r_s(\sigma^s)$ is the return for the government in period s upon history σ^s. If $\sigma_s = w$,

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

Otherwise, if $\sigma_t = u$, $r_s(\sigma^s) = 0$.

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 the government to the worker. Worker'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^u_{t+1}, 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^u_{t+1}, a_{t+1}(\sigma^t, u))\Big\{u(c_s(\sigma^s)) - a_s(\sigma^s)\Big|\mathcal{W}'(\sigma^t, u), \mu^u_{t+1}\Big\} + \\
&\qquad + \sum_{\sigma^{T+1}} p(\sigma^{T+1}|\sigma^{t+1}, \mu^u_{t+1}, \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 $\mu^u_{t+1} = \mu_{t+1}(\sigma^t, \sigma_{t+1} = u, \mathbf{a}(\sigma^{t-1}), a_t(\sigma^t))$[48].

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 the government reads

$$V^{UI}(U, \mathbf{a}(\sigma^{t-1}), \sigma^t) = \max_{\mathcal{W}} \mathbf{E}\Big\{ \sum_{s=t}^{T} \beta^{s-t}\big(r_s(\sigma^s) - c_s(\sigma^s)\big) \Big| \mathcal{W}(\sigma^t), \mu_t(\sigma^t, \mathbf{a}(\sigma^{t-1}))\Big\}$$

$$\text{sub:} \quad (33), \quad U_s(\mathcal{W}'(\sigma^s), \mathbf{a}(\sigma^{s-1}), \sigma^s) \geq U, \quad \forall s \geq t$$

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:

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

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

Government's problem can be rewritten as

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

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

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

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

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

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

Lemma 3. *The government prefers to insure the worker against the risk of h realization upon reemployment.*

Lemma 4. *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 3, 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.

Base Step $(t = T)$

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

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

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 $t = T$.

Induction Step $(t \leq T-1)$

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

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

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 4 proves to be useful in light of the following result.

Lemma 5. *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, government's payoff is larger under $\mathcal{W}'$ than under $\mathcal{W}$, as payment to the worker in history (σ^t, w) is lower in the former case. Moreover, by Lemma 4, $\mathcal{W}'$ 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 5) and sufficient one for it to be robust against any multiple deviation (by Lemma 4). Therefore, focusing on the set of contracts with such characteristic is without loss of generality.