Discussion Paper Series

IZA DP No. 18934

September 2026

At-Risk or Responsive? Targeting in Unemployment Policy

Sebastian Y. D. Ernst
LSE

Andreas I. Mueller
University of Zurich
and IZA@LISER

Johannes Spinnewijn
LSE

Joost Bollens
VDAB

IZA@LISER NETWORK
ISSN: 2365-9793

At-Risk or Responsive? Targeting in Unemployment Policy*

Abstract

This paper evaluates the use of algorithmic risk profiling to direct job search support toward individuals most at risk of long-term unemployment. Using three regression discontinuity designs and rich administrative data from the Flemish Public Employment Service (VDAB) in Belgium, we address two key dimensions: whom and when to target. First, job search counseling raises employment on average, but treatment effects decline sharply with predicted risk. This reveals a fundamental trade-off between targeting those most at risk and those most responsive to intervention. Second, we find no evidence that treatment effects depreciate at the individual level, but strong dynamic selection: as spells lengthen, surviving job seekers have both higher baseline risk and lower treatment responsiveness, with opposing implications for the returns to treatment. We develop and calibrate a conceptual framework for optimal targeting that highlights the key role of the ratio of predicted treatment effects to predicted job-finding probabilities. Targeting rules combining both objects substantially outperform the status quo, yielding welfare gains nearly three times as large as those from random assignment, while targeting on risk scores alone performs worse than random assignment.

JEL classification

J64, J65, H43, C21

Keywords

unemployment, targeting, treatment effect heterogeneity, regression discontinuity, public employment services, machine learning

Corresponding author

Andreas I. Mueller

andreas.mueller@econ.uzh.ch

* We thank Michele Belot, Guillaume Bied, Lorenzo Casaburi, Bart Cockx, Tijl De Bie, François Gérard, Maarten Goos, Anders Humlum, Xavier Jaravel, Philipp Kircher, Pat Kline, Rafael Lalive, Rob Metcalfe, Emilie Rademaekers, Johannes Schmieder and participants at seminars and presentations at the Central Planning Bureau of the Netherlands, Cornell University, Ghent University, IFAU in Uppsala, LMU Munich, the London School of Economics, the NBER SI joint Labor/Public Finance, the Tinbergen Institute in Rotterdam, the University of Essex, the University of Zurich, the VDAB and at the workshops on job search at Copenhagen University and Université Libre de Bruxelles. We are grateful to the VDAB for its support throughout the project, and especially to Stijn Van de Velde for his help with the data and institutional details. Spinnewijn acknowledges funding from the European Research Council through Advanced Grant 101200402.

1 Introduction

A central challenge for governments operating under limited resources is how to allocate interventions where they create the greatest social value. Public employment services must decide which unemployed workers receive intensive job search assistance [Berger et al., 2001; Black et al., 2003]. Health systems must decide which patients should receive treatment [Obermeyer et al., 2019; Yadlowsky et al., 2025], while poverty alleviation programs determine which households receive transfers [Haushofer et al., 2025]. Increasingly, these decisions are guided by predictive algorithms that estimate which individuals are most at risk of poor outcomes. Yet, prediction and targeting are fundamentally different problems [Kleinberg et al., 2015; Athey, 2017]. Intuitively, a planner allocating scarce resources should care not only about who is most at risk, but also about who stands to benefit most from the intervention. Whether these two objects coincide is ultimately an empirical question. Nevertheless, many policy applications implicitly assume that they do. Public employment services across OECD countries increasingly prioritize job seekers predicted to remain unemployed longest, thereby treating measures of risk as measures of treatment value [Desiere et al., 2019].

This paper studies the optimal targeting of unemployment policy and shows that the distinction between prediction and targeting is first-order in this context. We consider the allocation of job search counseling by the Flemish Public Employment Service (VDAB), which uses a machine-learning model to prioritize workers predicted to be at greatest risk of long-term unemployment. The setting provides a unique opportunity to study how scarce interventions should be allocated when both baseline risk and treatment responsiveness are heterogeneous. We make three contributions. First, we develop a conceptual framework that characterizes the trade-offs of targeting job search counseling across both workers and unemployment durations. Second, we provide quasi-experimental evidence on how treatment effects vary across these two dimensions, exploiting three complementary regression discontinuity (RD) designs that help identifying heterogeneity by long-term unemployment risk, and the evolution of treatment effects over the unemployment spell. Third, we combine the conceptual framework with the empirical estimates to evaluate alternative targeting policies, showing how employment services should decide whom to treat and when to intervene.

We consider a stylized model of unemployment, where job search counseling increases the probability that workers transition from unemployment into employment. The framework accommodates observed and unobserved heterogeneity in both job-finding probabilities and treatment effects, as well as duration dependence in job finding and treatment efficacy. Our results generalize to other settings where policymakers seek to accelerate transitions out of adverse states (e.g., illness, poverty, or homelessness). In particular, we show that the return to treatment depends jointly on its effect on the transition probability and on the continuation value of remaining in the adverse state. In fact, these two objects combine in a remarkably simple sufficient statistic. To a first-order approximation, the return to treatment is proportional to the treatment effect τ divided by the underlying transition probability π,

$$R \approx \frac{\tau}{\pi}.$$

Intuitively, a given increase in the probability of finding employment is more valuable for workers who would otherwise remain unemployed for longer. This simple characterization provides the organizing principle for the remainder of the paper.

First and foremost, it provides a unified way of thinking about both *whom* to target and *when* to intervene. Across individuals, it highlights the fundamental trade-off between risk and responsiveness. Workers with lower job-finding probabilities have more to gain from leaving unemployment, but they need not be those who respond most to job search counseling. The optimal allocation therefore depends on the balance between these two forces rather than either object in isolation.

The same logic applies across unemployment durations. Interestingly, the return can change dynamically over the unemployment spell, because treatment effects and job-finding probabilities may change as unemployment lengthens, but also because the composition of the unemployed changes through dynamic selection. The optimal timing of intervention therefore depends jointly on duration dependence in treatment effects and in job-finding probabilities. At the same time, delaying treatment risks losing the opportunity to reach high-return job seekers who exit unemployment in the meantime, creating a harvesting motive for intervening earlier.

Second, it provides estimable sufficient statistics for allocating scarce interventions across both individuals and time. To bring the framework to the data, we study the allocation of job search counseling by the VDAB. Our data covers the universe of registered unemployment spells between 2018 and 2022, linked to rich information on labor market histories, risk scores, caseworker interactions, and employment outcomes. Since 2018 the VDAB has relied on a machine-learning model using more than 400 granular predictors to estimate each unemployed worker's *risk score*, defined as the probability of finding employment within six months. Job seekers whose predicted job-finding probability falls below a threshold of 0.65 are prioritized for counseling: after five weeks of unemployment, they are proactively contacted by the service line and considered for assignment to a caseworker. Those above the threshold, however, are not actively contacted until three months later. The VDAB's system exemplifies the international trend toward algorithmic profiling in active labor market policy [Desiere et al., 2019].

The combination of institutional detail, data richness, and quasi-experimental variation provides an unusually clean setting in which to study how risk-based targeting performs in practice. In particular, the institutional setting generates complementary sources of quasi-experimental variation that map directly into the two allocation problems studied in the conceptual framework. First, a regression discontinuity around the algorithmic risk-score threshold identifies the average effects of job search counseling for workers near the margin of treatment. Second, a discontinuity generated by the automatic reassignment of recently unemployed workers to their previous caseworker allows us to estimate treatment effects across the distribution of predicted risk scores, providing evidence on whom to target. Third, the risk-score discontinuity reappears with the opposite sign later in the spell, when the VDAB reaches out to workers who have remained unemployed without previously being assigned to a caseworker. This design helps identify how the value of intervention evolves over the unemployment spell.

Our empirical analysis delivers three main sets of findings.

First, being prioritized for job search counseling substantially improves employment outcomes on average. Prioritized workers find employment sooner, increasing days worked by about 9 percent over the following three months. If anything, they also transition into more stable employment relationships, suggesting that the intervention improves job matches rather than simply accelerating acceptance of lower-quality jobs. The threshold shifts a bundle of interactions with the employment service: it substantially increases contact with service-line operators (by 65.8pp) and increases both the prospect and realization of subsequent caseworker assignment (by 9.5pp). The resulting differences in downstream interventions are concentrated in job-search assistance and monitoring rather than formal training. While it is difficult to attribute the employment effect to any single component of counseling, the sizable overall response makes the allocation of this intervention bundle quantitatively important.

Second, treatment effects vary sharply across workers. Contrary to the logic underlying current targeting practice, we find a strong negative relationship between long-term unemployment risk and treatment responsiveness. We find a precise zero effect of job search counseling for workers with the highest risk of long-term unemployment — the very workers prioritized by the existing targeting algorithm. By contrast, for workers with better employment prospects, we find large and statistically significant effects. The planner therefore faces a first-order trade-off between targeting workers who are most at risk and those for whom treatment generates the largest returns.

Beyond this comparison across risk-score groups, we find substantial additional heterogeneity in treatment effects, estimated using an instrumental forest [Athey et al., 2019]. Importantly, the estimated treatment effects based on the reassignment to caseworkers also predict heterogeneity in our baseline RD, where both the population and the treatment environment differ, although this cross-design comparison is statistically imprecise. We further show that the pattern is not explained by differential assignment to programs, which mitigates concerns that the heterogeneity is driven by lock-in effects or differences in the type of support received.

To quantify the importance of this heterogeneity, we can decompose the variation in returns to treatment using the sufficient statistic, $R \approx \tau/\pi$. Based on a Shapley decomposition of the variance in the targeting index τ/π, we find that 82% of the variation is attributable to heterogeneity in treatment effects. Moreover, part of the variation arising from differences in risk scores is offset by its negative covariance with treatment effects. As a result, targeting exclusively on risk scores misses considerable variation in the returns to treatment.

Third, we find no evidence that treatment effects vary with the timing of treatment among comparable workers. We exploit the reversal of the risk-score discontinuity when the VDAB contacts remaining unemployed workers later in the spell, allowing us to compare counseling effects for workers around the same initial risk-score threshold but treated at different unemployment durations. The estimated effects are similar across the first and second treatment rounds, although both estimates are imprecise. We subject this comparison to several validation checks, including tests for differential survival into the second round and for discontinuities in predicted treatment responsiveness among untreated survivors. These checks support the interpretation that the similar estimates reflect limited duration dependence in treatment effects for comparable workers.

This does not imply, however, that returns to treatment are constant over the unemployment spell. Workers with stronger employment prospects leave unemployment earlier, so the remaining pool becomes increasingly selected toward workers with lower job-finding probabilities and lower responsiveness to counseling. Through the lens of our sufficient statistic $R \approx \tau/\pi$, these forces have opposite implications for the timing of intervention. Declining treatment effects reduce the return to counseling, while declining baseline job-finding probabilities increase the continuation value of treatment for those who remain unemployed. Quantitatively, we find that average treatment effects decline by roughly 51% over the first year, while job-finding probabilities decline by roughly 57%. These two observable forces are of comparable magnitude, but the decline in job finding clearly dominates once we condition on observables — treatment effects show no decline for comparable workers, while their job-finding prospects continue to deteriorate — implying that the return to treating the average surviving job seeker rises over the spell.

In the last part of the paper we use the empirical estimates to calibrate our conceptual framework and evaluate alternative targeting policies. The welfare gains from optimal targeting are nearly three times those from random assignment, reflecting substantial heterogeneity in both risk and responsiveness across job seekers. A simple rule based on the ratio of predicted treatment effect to predicted job-finding probability performs nearly as well as the fully optimal policy. By contrast, the current practice of targeting exclusively on risk scores performs worse than random assignment. The shortfall is quantitatively large: in our calibrated model, risk-score targeting yields only about 70% of the welfare gains of random assignment.

Extending the analysis to the timing dimension confirms that the average return to treating surviving job seekers rises over the unemployment spell. Yet when optimizing jointly across job seekers and durations, the harvesting motive favors earlier intervention, since fewer job seekers with high returns remain available for treatment. The calibrated optimal policy therefore treats most targeted job seekers early in the spell, even though average returns among those who remain unemployed continue to rise.

Taken together, the results imply that employment services should not treat risk predictions as targeting rules. Risk scores are valuable inputs into allocation decisions, but optimal targeting also requires information on treatment responsiveness and on how returns evolve over the unemployment spell. In our setting, this points toward earlier intervention and substantially greater weight on predicted responsiveness, rather than relying primarily on predictions of long-term unemployment risk.

A final component of targeting we briefly investigate in our calibrated model concerns *how* targets should be estimated. The VDAB's algorithm, like most risk-profiling tools, is trained on realized employment outcomes, which reflect a mixture of baseline job-finding prospects and treatment effects for treated job seekers. This conflation creates adverse selection into treatment: individuals with high treatment effects will have higher predicted job-finding rates, pushing them above the threshold and out of the treatment pool. The appropriate risk measure for targeting should instead predict job finding in the absence of treatment, not in its presence. Our finding that treatment effects are increasing in predicted job-finding probability may partly reflect this algorithmic bias. Correcting risk scores for the effects of past treatment weakens the relationship between treatment effects and predicted job finding by roughly a fifth. However, it does not overturn it, nor does the bias invalidate our analysis of current practice: the VDAB assigns counseling based on the observed risk scores, so our quasi-experimental estimates and our evaluation of score-based targeting refer to the scores actually used for assignment.

Related Literature. Our paper contributes to several strands of literature. First, it connects two large literatures on unemployment policy that have largely developed separately. One literature studies heterogeneity in individual job-finding prospects and how this heterogeneity shapes the dynamic composition of the unemployed over the spell [e.g., Ahn and Hamilton, 2020; Mueller et al., 2021; Ahn, 2023; He and Kircher, 2023; Alvarez et al., 2024; Gregory et al., 2025; Mueller and Spinnewijn, 2025; Lalive et al., 2026]. While we also find substantial and predictable heterogeneity in job-finding prospects, for targeting purposes, predicting who is likely to remain unemployed is not sufficient: the planner must also know who responds to intervention. A second literature estimates the effects of active labor market programs, documenting substantial heterogeneity across programs, populations, and unemployment durations [e.g., Crépon and van den Berg, 2016; Card et al., 2018; van den Berg and Vikström, 2022; Cockx et al., 2023; Le Barbanchon et al., 2024; Humlum et al., 2025].[1] Our contribution is to bring these two strands together: we extend the analysis of heterogeneity and dynamics from job-finding prospects to treatment effects and show how job-finding prospects and treatment responsiveness should jointly determine the allocation of scarce employment services.

Second, our paper contributes to work on profiling, algorithmic assistance, and digital tools in public employment services. Statistical profiling has long been used to identify unemployment insurance claimants at risk of long spells and target them for reemployment services, but has received limited attention in the economics literature [e.g., Berger et al., 2001; Black et al., 2003; Desiere et al., 2019]. However, several recent papers study how predictive algorithms and recommender systems can improve job search and matching, including by recommending vacancies, occupations, or search strategies to job seekers [e.g., Altmann et al., 2022; Bied et al., 2023; Le Barbanchon et al., 2023; Behaghel et al., 2024; Belot et al., 2026; Bächli et al., 2025]. We differ from this literature in focusing on the allocation of job search counseling and the heterogeneity in treatment responsiveness. Our results show that even when algorithms predict well, using those predictions as targeting rules can perform poorly if predicted risk is negatively related to treatment responsiveness.

Third, our paper relates to a broader literature on targeting interventions when social need and treatment impact need not coincide. The closest paper is Haushofer et al. [2025], who study the trade-off between targeting deprivation and targeting impact in the context of cash transfers in Kenya. They show that targeting solely on deprivation is unattractive unless redistributive preferences are sufficiently strong, a conclusion that mirrors our finding that targeting solely on unemployment risk can be dominated by rules that also account for responsiveness. More broadly, our analysis builds on the empirical welfare approach to treatment assignment pioneered by Manski [2004] and developed in subsequent work on welfare-based policy choice and policy learning [e.g., Kitagawa and Tetenov, 2018; Athey and Wager, 2021; Kasy and Sautmann, 2021]. A related literature emphasizes the distinction between prediction and causal decision-making, documenting both the promise of algorithmic predictions for high-stakes decisions and their limitations when the prediction target differs from the object relevant for policy [e.g., Kleinberg et al., 2015; Athey, 2017; Kleinberg et al.,

[1] For example, across programs, existing evidence suggests that job search counseling often has larger effects for workers closer to employment, while training and information-based interventions may display different patterns [e.g., Koning, 2009; Liu et al., 2014; Altmann et al., 2018; Belot et al., 2026].

2018; Obermeyer et al., 2019; Athey et al., 2025]. We contribute to this literature by developing a simple sufficient-statistic framework for interventions that accelerate transitions out of adverse states. We characterize the targeting and timing trade-offs that arise when the planner must decide not only whom to treat but also when to intervene, and quantify these trade-offs empirically, in the spirit of the sufficient-statistics literature in unemployment insurance [e.g., Kolsrud et al., 2018; Spinnewijn, 2020; Gerard and Gonzaga, 2021; Ferey, 2022].

The remainder of the paper is organized as follows. Section 2 develops the conceptual framework and characterizes optimal targeting. Section 3 describes the institutional setting and data. Section 4 estimates the average effects of job search counseling. Section 5 documents heterogeneity in treatment effects across the risk distribution. Section 6 examines duration dependence in treatment effects. Section 7 presents the quantitative evaluation of alternative targeting policies. Section 8 concludes.

2 Conceptual Framework

This section presents a conceptual framework that clarifies both *who* should be targeted for treatment and *when* during an unemployment spell treatment is most valuable. The framework encompasses heterogeneity across unemployed workers and dynamic changes over the unemployment spell, but remains sufficiently tractable to allow for a simple characterization of the individual returns to treatment as a function of predicted treatment effects and baseline job-finding prospects. By linking these objects, the analysis yields transparent sufficient statistics for optimal targeting, both highlighting the key trade-offs that govern the allocation of scarce program resources across workers and durations and guiding our empirical analysis in order to quantify these trade-offs.

2.1 Setup

We consider an unemployment agency that decides whether and when to treat a cohort of unemployed workers. The workers may differ in their response to treatment τ_d as well as their baseline job finding probability π_d at duration d. Both the responsiveness to treatment and the job finding probability can differ in ways that are observable and unobservable to the agency and can change over the unemployment spell.

Job Finding and Survival. We assume that the job-finding probability at unemployment duration d takes the following form:

$$\begin{aligned} P_d^T(\pi,\tau,\varepsilon_\pi,\varepsilon_\tau) &= \pi_d + \tau_d \times T_d(\pi,\tau) && (1)\\ &\equiv (\pi+\varepsilon_\pi)\,\theta_{\pi,d} + (\tau+\varepsilon_\tau)\,\theta_{\tau,d} \times T_d(\pi,\tau). && (2) \end{aligned}$$

An individual's type is represented by $(\pi,\tau,\varepsilon_\pi,\varepsilon_\tau)$, where π denotes an individual's baseline observable job-finding type, and ε_π captures unobserved heterogeneity in baseline job-finding prospects. Similarly, τ denotes the individual's predictable treatment effect, while ε_τ represents unobserved heterogeneity in treatment responsiveness. We assume that the dynamics of the baseline job finding and treatment effects are the same across workers. The term $\theta_{\pi,d}$ governs how baseline job-finding probabilities evolve with unemployment duration, while $\theta_{\tau,d}$ captures how treatment effects depreciate (or appreciate) over the spell. We assume that observed and unobserved components of the baseline job-finding probabilities and treatment effects are independent, although the observed components as well as the unobserved components across the baseline and treatment types may be correlated. Formally, types are distributed according to the joint density $f(\pi, \tau, \varepsilon_\pi, \varepsilon_\tau) = g(\pi, \tau)h(\varepsilon_\pi, \varepsilon_\tau)$.

The indicator $T_d(\pi, \tau) \in \{0, 1\}$ specifies whether an individual of observed type (π, τ) is treated at duration d. We restrict targeting rules to depend only on observable characteristics, so that treatment cannot be conditioned on the unobserved components. A specific targeting policy T specifies a vector of treatment dummies $\{T_d(\pi, \tau)\}$ for each type (π, τ) at any duration d. We define the survival probability of individual of type $(\pi, \tau, \varepsilon_\pi, \varepsilon_\tau)$ as:

$$S_d^T(\pi, \tau, \varepsilon_\pi, \varepsilon_\tau) = \prod_{\delta=0}^{d-1}(1 - P_\delta^T(\pi, \tau, \varepsilon_\pi, \varepsilon_\tau)).$$

The superscript T on P and S denotes that the job-finding probability is conditional on the specific targeting policy.

Note that, throughout, we use unemployment risk and job-finding prospects interchangeably: the risk of remaining unemployed is the complement of the probability of finding a job, so the workers most at risk of long-term unemployment are those with the lowest job-finding probability π — and, empirically, the lowest VDAB risk score P, the predicted probability of finding employment within six months.

Utility losses from unemployment and cost of treatment. While our empirical focus is on job-finding probabilities and treatment effects, we also introduce utility losses from unemployment and costs of treatment in order to meaningfully evaluate welfare and characterize optimal policy. We assume that unemployed individuals incur a per-period utility loss u_d. In analogy to the job-finding probability and treatment effect, we can allow this utility loss to vary across job seekers and over the unemployment spell, i.e., $u_d = (u + \varepsilon_u)\,\theta_{u,d}$, where the unobserved component ε_u may be distributed jointly with $(\varepsilon_\pi, \varepsilon_\tau)$. In principle, targeted policies can condition on u and we can characterize the returns to targeting observable types (π, τ, u). However, given our empirical focus, we will often leave this dimension implicit to keep the notation light. We further assume that providing treatment entails a constant per-period cost c, and that there is no discounting over time.

As a result, the expected utility cost of a spell of unemployment for a worker of type $(\pi, \tau, \varepsilon_\pi, \varepsilon_\tau)$ is:

$$L^T(\pi, \tau, \varepsilon_\pi, \varepsilon_\tau) = \sum_{\delta=0} u_\delta S_\delta^T(\pi, \tau, \varepsilon_\pi, \varepsilon_\tau),$$

which is equal to u times the expected duration of unemployment in the case where $u_d = u$. The expected budgetary cost of treatment is:

$$B^T(\pi,\tau,\varepsilon_\pi,\varepsilon_\tau) = \sum_{\delta=0} cT_\delta(\pi,\tau)S_\delta^T(\pi,\tau,\varepsilon_\pi,\varepsilon_\tau).$$

Integrating over unobserved heterogeneity $(\varepsilon_\pi, \varepsilon_\tau)$ with density h gives type-level objects $L^T(\pi,\tau)$ and $B^T(\pi,\tau)$ in the natural way. We can then further aggregate across the population to write the expected utility loss L^T and budgetary cost of an unemployment spell B^T averaged across all job seekers. We adopt the short form notation $X^T = X$ in the absence of treatment.

While the welfare ingredients are stylized, the model can encompass or be extended with heterogeneity and dynamics in search, preferences, household resources and other features that have been central in related work [e.g., Kolsrud et al., 2018; Landais and Spinnewijn, 2021]. In principle, the treatment effect τ can capture search responses to the agency's interventions, while the catch-all variable u can capture differences in preferences, consumption, and social welfare weights, but also the fiscal costs of unemployment benefits. The latter could also be added directly to the budgetary cost.

2.2 Optimal Targeting and Timing

We focus on the agency's allocation problem: how to distribute treatment across a subset of unemployed individuals and across durations given an exogenously determined budget. The agency's objective is to minimize the expected welfare losses associated with unemployment as follows:

$$\begin{aligned} \min_{T} \quad & \int\int L^T(\pi,\tau)g(\pi,\tau)d\pi d\tau \\ & s.t.\ \bar{B} \geq \int\int B^T(\pi,\tau)g(\pi,\tau)d\pi d\tau. \end{aligned} \tag{3}$$

Since we consider the problem of an unemployment agency operating under a fixed budget $\bar{B}$, we abstract from issues related to distortionary taxation and the optimal overall size of the program. Moreover, by focusing on at most a single intervention per worker, providing treatment does not generate budgetary savings on other interventions that job seekers would otherwise receive over the spell. This allows us to characterize the optimal policy and prove the following proposition:

Proposition 1. ***Optimal targeting across individuals and durations.*** *Consider the agency's problem of allocating treatment across types and durations, where each type (π,τ) can be treated for at most one period, and assume that the distribution function $g(\pi,\tau)$ is continuous. Define the return to targeting type (π,τ) with utility-loss level u at duration d as the reduction in continuation losses per unit of treatment cost,*

$$R_d^T(\pi,\tau,u) \ \equiv\ -\frac{L_d^T(\pi,\tau,u) - L_d(\pi,u)}{c}. \tag{4}$$

Then there exists a threshold return $\bar{R} \geq 0$, determined by the budget constraint, such that the *optimal policy:*

(i) treats all types whose return at their optimal targeting duration exceeds the threshold,

$$T_{d^\star}(\pi,\tau,u) = 1 \iff R^T_{d^\star}(\pi,\tau,u) \geq \bar{R};$$

(ii) targets each treated type at the duration that maximizes its survival-weighted excess return,

$$d^\star(\pi,\tau,u) = \arg\max_d \; S_d(\pi,\tau)\big[R^T_d(\pi,\tau,u) - \bar{R}\big].$$

In the absence of unobserved heterogeneity and duration dependence ($\sigma_{\varepsilon_\pi} = \sigma_{\varepsilon_\tau} = \sigma_{\varepsilon_u} = 0$ and $\theta_{\pi,d} = \theta_{\tau,d} = \theta_{u,d} = 1$ for all d), the return simplifies to

$$R^T_d(\pi,\tau,u) \;=\; \frac{u}{c}\cdot\frac{\tau}{\pi} \qquad \textit{for all } d,$$

so that the optimal policy treats all types with $u\,\tau/\pi$ above a threshold at entry, $d^\star = 0$.

Proof. See Appendix A.

Proposition 1 separates the policy problem into two economic margins. Part (i) describes the targeting problem as a threshold rule across individuals: the agency ranks types by their return per unit of treatment cost and treats those whose return, evaluated at their optimal targeting duration, exceeds the common threshold $\bar{R}$, which measures the shadow value of the budget. Within a given duration, this is the familiar logic of ranking policies by their returns. Part (ii) governs the timing of treatment and adds an option-value consideration: treating a type at duration d only reaches the fraction S_d of that type still unemployed, so postponing treatment sacrifices the opportunity to treat those who exit in the meantime. The agency therefore compares excess returns $R^T_d - \bar{R}$ across durations, weighted by survival. This implies that delaying treatment is optimal only when returns rise sufficiently over the spell to compensate for the shrinking pool of treatable job seekers.

The return to treatment is central for both margins of the policy problem. A one-period treatment at duration d raises the job-finding probability of the treated by τ_d, so the return to targeting can be written as

$$R^T_d(\pi,\tau,u) = \frac{\tau_d}{c} L_{d+1}(\pi,u),$$

where $L_{d+1}(\pi,u) = \sum_{\delta=d+1}^{\infty} u_\delta\, S_{\delta|d+1}(\pi)$ denotes the continuation loss from unemployment from period $d+1$ onward and $S_{\delta|d+1}(\pi)$ is the survival probability to period $\delta \geq d+1$, conditional on remaining unemployed until $d+1$. The return is thus the treatment effect relative to its cost, times the avoided unemployment loss. Hence, this return is highest when both treatment effects and continuation losses are large.

The special case in Proposition 1 illustrates the relevant forces most clearly. Without any unobserved heterogeneity and depreciation, the continuation loss simplifies to u/π, giving the simplification in the proposition. The avoided unemployment loss is larger for *at-risk* job seekers with low job-finding probabilities — so the optimal policy ranks job seekers by the index $u\,\tau/\pi$, which reduces to τ/π when utility losses are identical, trading off risk against responsiveness. The timing implication is equally stark: because the return is the same at every duration while survival declines, delaying treatment shrinks the group that can be reached without raising its return, and all targeting optimally occurs at entry. Away from this benchmark, both forces in part (ii) of the proposition are active: returns may rise or fall over the spell, and the agency weighs this against the declining survival rate.

The next proposition characterizes how the returns vary across job seekers and over the unemployment spell. While approximate, these results continue to accommodate unobserved heterogeneity and dynamics in the model:

Proposition 2. ***Approximate returns to targeting.*** *Suppose baseline job-finding probabilities, treatment effects, and utility losses evolve geometrically at the individual level,* $\pi_d = (\pi + \varepsilon_\pi)\,\theta_\pi^d$, $\tau_d = (\tau + \varepsilon_\tau)\,\theta_\tau^d$, *and* $u_d = (u + \varepsilon_u)\,\theta_u^d$. *For type* (π, τ, u), *let* $\mathbb{E}_d[\pi_d]$, $\mathbb{E}_d[\tau_d]$, *and* $\mathbb{E}_d[u_d]$ *denote the expectation over* $(\varepsilon_\pi, \varepsilon_\tau, \varepsilon_u)$ *among untreated survivors at duration* d.[2] *One-period targeting returns* R_d^T *(defined in equation 4) are approximately proportional to average treatment effects and average utility losses, and inversely proportional to average job-finding probabilities, among the surviving unemployed:*

(i) across types. For two types (π, τ, u) *and* (π', τ', u') *at the same duration* d*:*

$$\frac{R_d^T(\pi', \tau', u') - R_d^T(\pi, \tau, u)}{R_d^T(\pi, \tau, u)} \approx \frac{\mathbb{E}_d[\tau'_d] - \mathbb{E}_d[\tau_d]}{\mathbb{E}_d[\tau_d]} - \frac{\mathbb{E}_d[\pi'_d] - \mathbb{E}_d[\pi_d]}{\mathbb{E}_d[\pi_d]} + \frac{\mathbb{E}_d[u'_d] - \mathbb{E}_d[u_d]}{\mathbb{E}_d[u_d]}.$$

(ii) across durations. For the same type (π, τ, u) *across consecutive durations* d *and* $d+1$*:*

$$\frac{R_{d+1}^T - R_d^T}{R_d^T} \approx \frac{\mathbb{E}_{d+1}[\tau_{d+1}] - \mathbb{E}_d[\tau_d]}{\mathbb{E}_d[\tau_d]} - \frac{\mathbb{E}_{d+1}[\pi_{d+1}] - \mathbb{E}_d[\pi_d]}{\mathbb{E}_d[\pi_d]} + \frac{\mathbb{E}_{d+1}[u_{d+1}] - \mathbb{E}_d[u_d]}{\mathbb{E}_d[u_d]}.$$

Part (i) holds in a neighborhood of $\theta_\pi = \theta_u = 1$. *Part (ii) holds in a neighborhood of* $\theta_\tau = \theta_\pi = \theta_u = 1$.[3] *Without unobserved heterogeneity, parts (i) and (ii) hold with individual characteristics in place of survivor averages.*

Proof. See Appendix A.

The key empirical value of Proposition 2 is that it maps the returns to targeting into empirical objects defined among the surviving unemployed at each duration. These are the average treatment effect $\mathbb{E}_d[\tau_d]$, the average job-finding probability $\mathbb{E}_d[\pi_d]$, and — when utility losses vary unobservably — the average utility loss $\mathbb{E}_d[u_d]$ among job seekers of a given observable type who are still unemployed at duration d.

Part (i) shows that the cross-sectional logic of Proposition 1 carries over to these survivor averages: returns are approximately proportional to average treatment effects and to the average utility

[2] Continuation losses are interpreted as truncated at a finite maximum spell length: for $\theta_\pi < 1$, individual job-finding probabilities vanish geometrically over the spell and a positive fraction of job seekers would never exit unemployment, so that without discounting the untruncated continuation loss would be infinite. We further assume that $\pi + \varepsilon_\pi$ is bounded away from zero and one, and that $u + \varepsilon_u$ is positive and bounded.

[3] The approximation is exact in θ_τ when part (ii) is written in ratio form; the displayed difference form drops a cross term that is second order only for θ_τ near one (see the proof).

losses from unemployment, and inversely proportional to average job-finding probabilities. Unobserved heterogeneity within observable types matters for the returns to targeting only through the survivor averages themselves, up to terms that are second order in its dispersion.

Part (ii) characterizes how the returns evolve over the unemployment spell: returns rise or fall according to whether the observed treatment effects among survivors decline more or less quickly than their observed job-finding probabilities, net of the observed change in utility losses among survivors. Moreover, building on Mueller and Spinnewijn [2025], we can show that the observed decline in job finding compounds two distinct forces through an exact identity, $\mathbb{E}_{d+1}[\pi_{d+1}]/\mathbb{E}_d[\pi_d] = \theta_\pi\big[1 - \mathrm{Var}_d(\pi_d)/(\mathbb{E}_d[\pi_d](1 - \mathbb{E}_d[\pi_d]))\big]$: true duration dependence, captured by θ_π, and dynamic selection, driven by the variance $\mathrm{Var}_d(\pi_d)$ in job-finding probabilities at duration d. Job seekers with high job-finding probabilities are more likely to exit unemployment, leaving behind an increasingly negatively selected pool. Importantly, the two forces matter for the returns to targeting only through their combined — and directly estimable — effect on the survivor averages, so the agency need not disentangle them. If observed treatment effects fall over the spell while observed job finding remains stable, returns are highest early in the spell. Conversely, if treatment effects are stable among survivors but observed job finding deteriorates, returns increase over the spell. While the latter may appear counterintuitive at first, it reflects that continuation losses are much larger late in the spell: early on, many job seekers would exit unemployment even without treatment, which reduces the value of intervening at short durations.

Combining Proposition 2 with Proposition 1 makes clear, however, that increasing returns among survivors do not by themselves imply that treatment should be delayed. Waiting also reduces the probability that a worker remains unemployed and can still be reached. If returns are constant over duration, treatment is therefore optimally front-loaded. More generally, later treatment can be optimal only if returns rise sufficiently to compensate for this survival loss.

Discussion. The propositions provide a clear roadmap for our empirical analysis, while abstracting from several issues. We probe these issues directly in the empirical analysis and assess their potential quantitative importance through extensions and sensitivity exercises in Section 7.

First, Proposition 1 restricts each worker to a single intervention. With repeated interventions, the same marginal-return logic applies, but planned future interventions affect the returns to intervening earlier by changing both the continuation losses that an earlier intervention can avoid and the expected budgetary costs of subsequent treatment. Under a fixed budget, repeated treatment also allows resources to be concentrated on fewer, higher-return workers. In the quantitative evaluation, we relax this restriction by allowing treatment in two consecutive periods, while continuing to abstract from persistence in treatment effects and potentially decreasing returns to repeated treatment.

Second, the characterization treats the intervention as uniform. In practice, the unemployment agency assigns workers to a variety of programs, ranging from job search assistance to training [e.g., Lechner and Smith, 2007; Bolhaar et al., 2020]. With multiple interventions, the planner would instead compare the returns across worker-program pairs, accounting also for differences in program costs. Our counterfactuals therefore characterize improved targeting of the current intervention bundle: they optimize whom and when to target while holding the downstream assignment of programs at its observed pattern. Heterogeneity in estimated treatment effects could in principle partly reflect differences in this program assignment. Empirically, however, we find little evidence that the types of interventions assigned by caseworkers vary systematically across risk groups.

Third, the simple sufficient-statistic characterization in Proposition 2 relies on approximations. We therefore compare it with the full model-based return in the quantitative evaluation and find that the sufficient-statistic approximation performs nearly as well.

Fourth, our characterization treats baseline job-finding probabilities and treatment effects as known functions of observable characteristics. In practice, both objects must be estimated from finite data, and their relative precision and data requirements may affect the performance of feasible targeting rules. We abstract from the full statistical decision problem, which is the focus of a broader literature on treatment choice, policy learning, and experimental design for policy choice [e.g., Athey and Wager, 2021; Kasy and Sautmann, 2021]. Our quantitative evaluation nevertheless probes the sensitivity of the targeting gains to more limited treatment-effect heterogeneity.

Fifth, even accurately estimated targets may not correspond to the targets required by the theory. In addition to treatment effects, the relevant object is job finding in the absence of treatment, whereas a prediction model trained on realized outcomes under an existing targeting regime may partly incorporate the effects of past treatment. This issue arises directly in our context. We show that this policy-induced 'algorithmic bias' contributes to the positive relationship between predicted job finding and treatment responsiveness, and quantify how it affects the performance of targeting based on predicted job finding alone.

Finally, our welfare analysis relies on a reduced-form specification of the utility costs of unemployment u, as our empirical estimation focuses primarily on baseline job-finding prospects π and treatment effects τ. In practice, heterogeneity or dynamics in u that countervail the patterns we document for π and τ could attenuate our policy conclusions. In particular, introducing curvature into the social welfare function would assign higher welfare weights to long-term unemployed individuals with low baseline job-finding probabilities and lower weights to more responsive individuals with higher job-finding probabilities when treated. This creates a force toward targeting need rather than responsiveness, similar to the trade-off emphasized by Haushofer et al. [2025]. We assess the potential importance of these considerations in the quantitative evaluation by allowing utility losses to vary dynamically and introducing curvature in the social welfare function.

3 Context and Data

This section describes the data and institutional details relevant for our empirical analysis. In Belgium, unemployed individuals must register with the regional Public Employment Service in order to receive unemployment benefits. The VDAB is the Public Employment Service for the Flemish Region, the Dutch-speaking part of Belgium and home to over 6.8 million inhabitants. Core aspects of the VDAB's mission include helping all working-age residents find sustainable employment, providing jobseekers with employment services and training and eliminating the mismatch between supply and demand in the labor market [VDAB, 2024]. We exploit quasi-experimental variation in the assignment of job seekers to job search counseling, combined with rich administrative data from the VDAB on unemployment spells, risk scores, interactions with the employment service, and subsequent employment outcomes.

3.1 Institutional Details

Caseworkers and Service Line Operators Caseworkers are the primary providers of individualized support to job seekers at the VDAB. They are based in regional offices throughout the Flemish Region. Job seekers who have been assigned to a caseworker must meet with their caseworker regularly (possibly over the phone). Caseworkers provide job search counseling services to job seekers, monitor their job search activity and oversee their assignment to orientation and training programs. Job seekers who fail to show up for meetings or actively search for work face the threat of unemployment benefit sanctions. Assignment to a caseworker is typically mediated through the 'service line', the VDAB's call center. Operators working at the service line provide first-line advice to job seekers and determine which job seekers should be assigned to a caseworker at a regional office.

Assignment to caseworkers Our empirical analysis leverages three sources of variation in the assignment to caseworkers. We describe the assignment procedure here in detail and provide a schematic overview in Figure 1.[4] Upon registration with the VDAB, job seekers receive an introductory email describing what they need to do in order to receive unemployment benefits. While job seekers are not required to call the service line in the first four weeks of unemployment, some do so voluntarily. Some of these early callers are then assigned to a caseworker.[5] After four weeks of unemployment, all job seekers below the age of 60 who have not yet been assigned to a caseworker receive an email instructing them to call the service line. These calls involve an in-depth half-hour interview with an operator who must then decide whether a job seeker is *self-reliant* or should be assigned to a caseworker. Job seekers classified as self-reliant are left to continue their search for work independently for the time being. Those assigned to a caseworker receive an appointment at a nearby regional office. Time until the first appointment depends on backlog, possibly taking up to eight weeks at the busiest offices.

After five weeks of unemployment, the VDAB calls those job seekers who have not yet called and who are considered to be at risk of long-term unemployment. In order to determine a job seeker's risk of long-term unemployment, the VDAB developed a machine learning prediction model that predicts job seekers' probability of employment within 6 months. Those with a probability below 65% are called and either assigned to a caseworker or determined to be self-reliant. We exploit this variation in our *'Baseline RD'*. Job seekers who do not respond to calls from the VDAB face unemployment benefit sanctions. After seven weeks of unemployment, job seekers with a score above 0.65 who have not called the VDAB are automatically classified as self-reliant. While in principle available to service line operators, risk scores are not further used in the assignment decision, nor

[4] The described process was in place from October 2018 until May 2023. Prior to October 2018, the VDAB did not use predicted risk to assist with assignment. From May 2023, each step of the caseworker assignment process was brought forward by three weeks.

[5] Although uncommon, job seekers can, in theory, also show up to a regional VDAB office in person and be assigned to a caseworker at any time if capacity permits.

in a job seeker's later interactions with the VDAB. Appendix B.1 provides statistics on caseworker assignment up until this point.

Eleven weeks after being classified as self-reliant, job seekers who are still unemployed receive another email instructing them to call the service line. Those who do not call within the next week are called by the service line. Risk scores are not used for these follow-up interviews. Instead, all job seekers must speak to a mediator. Job seekers who had a score above 0.65 during the first outreach campaign are therefore more likely to be called by the service line at around 19 weeks of unemployment. We exploit this variation in a second RD, which we refer to as the *'Dynamics RD'* as it allows us to estimate treatment effects later in the unemployment spell and compare them relative to the Baseline RD. In this second round of interviews, mediators again either assign job seekers to a caseworker or classify them as self-reliant. The process is repeated one last time eleven weeks later in a third round of interviews, during which all remaining job seekers are assigned to a caseworker.

Finally, job seekers who find employment, but who become unemployed again within 6 months are automatically re-assigned to their previous caseworker, if they had one. Otherwise, they again follow the procedure described above. This creates a third source of variation that we exploit in what we will refer to as the *'Heterogeneity RD'*, as it allows us to estimate heterogeneity in treatment effects, also by job seekers' risk score.

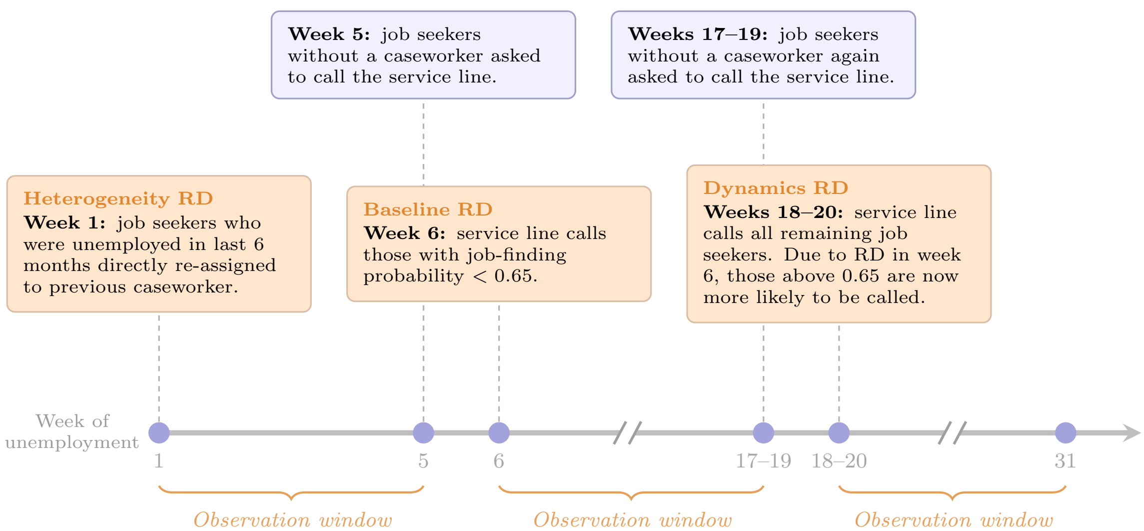


Figure 1: Assignment to Caseworkers

The risk score The VDAB's prediction model, called *Kans op Werk* or "Employment Prospects," predicts the probability that a job seeker begins a period of employment lasting at least 28 consecutive days within the next 6 months. The random forest model uses more than 400 input variables, with a particular focus on labor market history and the current unemployment spell. Other variables include socio-demographic characteristics, job search preferences (e.g., occupation and geography) and job search behavior (e.g., interactions with the VDAB website). The model is retrained monthly and predictions are updated daily. Prior to October 2020, risk scores were typically generated and sent to the service line once per week (usually on a Friday). In October 2020, however, the "INOS" system was introduced. Under this system, the risk scores used by the service line when deciding who to call are updated automatically on a daily basis. See Ernst et al. [2024] for a more detailed description of the prediction model and the process of assigning job seekers to counseling.

3.2 Data

We use administrative data from the Flemish Public Employment Service (VDAB) covering the universe of registered unemployment spells starting between 1 September 2018 and 30 November 2022 in Flanders, Belgium.[6] The full sample includes over 780,000 spells. After restricting to individuals who remain unemployed for at least five weeks, who do not call the service line when invited to, who are part of the outreach campaign (e.g., below the age of 60) and who are near the risk score threshold, the estimation sample used for the baseline RD includes around 79,000 spells.[7] For the dynamics RD, we apply the same sample restrictions, but exclude job seekers who exit unemployment before reaching the 20th week. This leaves a sample of approximately 38,000 spells. For the heterogeneity RD, we restrict to individuals who had contact with a caseworker in their previous unemployment spell and who ended their previous spell around 6 months before the start of their current spell. The remaining sample consists of approximately 131,000 unemployment spells. In Appendix B.2, we provide summary statistics for the full sample as well as all three estimation samples. In Appendix B.3, we provide statistics on transitions out of unemployment by duration.

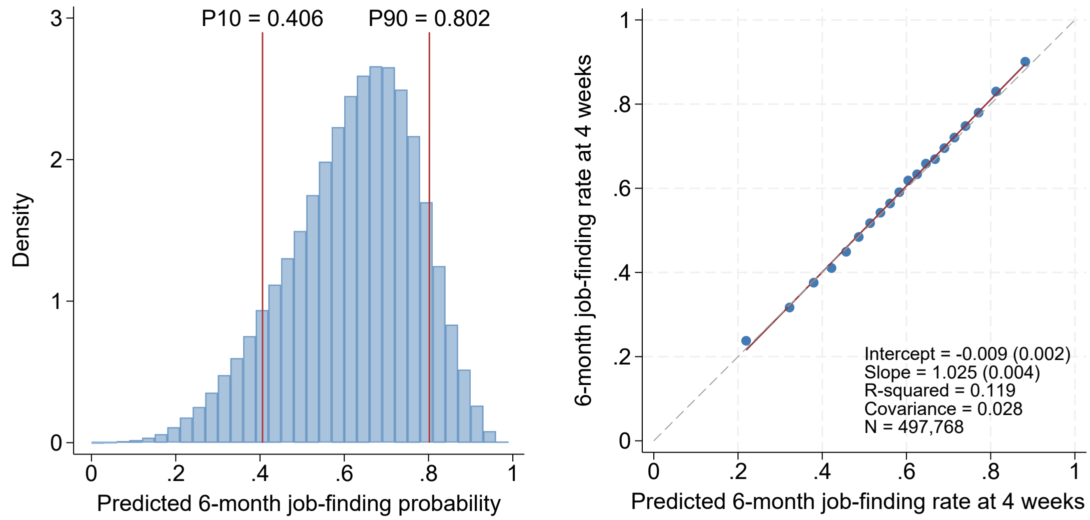
The data includes a rich set of variables on individuals' socio-demographic characteristics, labor market histories and outcomes, the VDAB's risk scores, interactions with the VDAB (e.g., contact with the service line, caseworkers, participation in programs, benefit sanctions), job search behavior (e.g., logins on the VDAB website, self-reported job applications), occupational experience and occupational preferences.

Risk scores. We have data on the 6-month job-finding probabilities predicted by the VDAB for job seekers in our sample at different durations. Panel (a) of Figure 2 illustrates the dispersion in these risk scores, plotting the distribution of predicted job-finding probabilities for all job seekers in the second week of the spell. The 90th percentile faces an 80 percent chance of finding a job in the next 6 months, while for the 10th percentile this is only 40 percent. Panel (b) of Figure 2 illustrates the accuracy of the prediction model at 4 weeks of unemployment, showing that the average 6-month job-finding rate closely tracks the predicted 6-month job-finding rate, even at the top and the bottom of the distribution.[8] Our empirical strategy relies on the risk scores that were provided to the service

[6] To avoid the peak of the COVID-19 crisis, we also show robustness to excluding observations during the pandemic in Appendix C.4.

[7] Risk score predictions from 3 October 2020 onward were handled by the INOS system. For the pre-INOS sample, we only have access to one prediction per job seeker, meaning that for job seekers who were called by the service line in two or more separate spells between September 2018 and September 2020, only one spell is included in the estimation sample. We check robustness to just using the pre-INOS or INOS sample in Appendix C.4. Additionally, we only observe contact with the service line and caseworkers until 5 January 2023. We therefore restrict the sample to spells beginning at least 17 weeks earlier for the baseline RD, 4 weeks earlier for the heterogeneity RD and 31 weeks for the dynamics RD.

[8] In Appendix B.4, we also show binned scatter plots for predictions made at four other durations.



(a) Distribution of Predictions

(b) Binned Scatter Plot: Predictions vs. Outcomes

Figure 2: The VDAB Risk Score

Note: Panel (a) shows a histogram of predicted 6-month job-finding probabilities generated on the 8th day of unemployment. Panel (b) shows a binned scatter plot comparing the predicted 6-month job-finding rates generated on the 29th day of unemployment to the empirically observed 6-month job-finding rates for the same spells.

line and used to profile job seekers.

Employment. The VDAB defines employment as a period of at least 28 days in work. The VDAB also classifies individuals as employed if they are engaged in temporary work or in subsidized work-and-training programs lasting 1–6 months.[9] Many job seekers transition from such engagement back into unemployment. For our baseline analysis, we therefore count these engagements as employment if they are directly followed by a transition into employment. Otherwise, they are counted as unemployment. We check the robustness of our results to always counting temporary and subsidized work as employment as well as never counting them as employment. While we only have job seekers with unemployment spells beginning between 1 September 2018 and 30 November 2022, we have the employment outcomes for all of these job seekers until 31 July 2023 and detailed information on their unemployment spells prior to 1 September 2018.

Interactions with the VDAB. We have detailed day-by-day data on job seekers' contact with caseworkers, including both in-person contact and phone calls. From 3 October 2020, following the introduction of the INOS system, we also have detailed data on interactions with the service line, including days on which calls were made, whether the job seeker or the service line made the call, whether a call was made to the job seeker but not answered and whether the service line classified a

[9]The VDAB classifies an individual as engaged in temporary work if they have done at least 10 days of temporary work in the past 28 days. Temporary work involves fixed short-term contracts and is very common in Belgium. It is used by firms to replace workers on leave, handle temporary increases in workloads and trial new employees.

job seeker as self-reliant or assigned them to a caseworker. For all analysis of interactions with the service line, we therefore restrict our sample to the period covered by this dataset.[10]

We also have data on all programs that job seekers participate in. This includes, for example, job search orientation, training, job application requirements and external employment services. Programs vary greatly in terms of intensity, lasting from one day to many months, as shown in Appendix B.5.[11]

4 Baseline Design

We first turn to the discontinuity at a risk score of 0.65 in week 6 of the unemployment spell, which plays a central role in the VDAB's targeting strategy for assigning job seekers to caseworkers. This discontinuity provides the cleanest and most powerful source of quasi-experimental variation for estimating the effects of job search counseling in the Flemish context. It also serves as a benchmark for our alternative RD designs, which we use below to assess heterogeneity in treatment effects by risk score and dynamics in treatment effects over the unemployment spell.

4.1 Empirical Strategy

After five weeks of unemployment, job seekers with a predicted 6-month job-finding rate below 0.65 are prioritized for contact by the service line. During the resulting interview, the service line operator decides whether the job seeker should be assigned to a caseworker. The threshold therefore generates a discontinuous increase both in the probability of interacting with the service line and in the probability of subsequent assignment to a caseworker.

We employ an RD design, estimating the effect of having a score below 0.65 using the following baseline specification:

$$y_i = \alpha + \beta \mathbb{1}\{P_i < 0.65\} + \gamma_1 (P_i - 0.65) + \gamma_2 (P_i - 0.65) \mathbb{1}\{P_i < 0.65\} + \boldsymbol{X}_i' \boldsymbol{\delta} + \epsilon_i, \tag{5}$$

where y_i is the outcome variable for unemployment spell i, P_i is the predicted 6-month job-finding probability used by the service line when determining whether to contact the job seeker and $\boldsymbol{X}_i$ is a vector of spell characteristics. The sample is restricted to job seekers who were invited to call the service line after four weeks of unemployment but did not call within a week and were therefore added to the service line's call list.

The service line begins calling job seekers below the cutoff after five weeks of unemployment. Because the next round of service-line interviews begins approximately three months later, we focus on weeks 6–17 of the unemployment spell. This observation window captures the effects of the initial targeting decision while avoiding contamination from later interventions. Our baseline outcome is the number of days worked per week during this 12-week period. This measure captures both the

[10] While we have some data on service line contact prior to 3 October 2020, these records are incomplete.

[11] See Cockx et al. [2023] for more details on VDAB's training programs and their estimated impacts. Our data also includes information on investigations and sanctions by the VDAB. Investigations are initiated for job seekers who do not show up to meetings or fulfill tasks. Sanctions can include partial or full suspensions of unemployment benefits.

probability of finding employment and the speed with which employment is obtained. We consider two treatment measures: whether the job seeker had contact with the service line and whether the job seeker had contact with a caseworker during weeks 6–17 of unemployment.

We define the running variable as the first risk score the service line had on record in weeks 6–7 of unemployment.[12] For our baseline specification, we include observations with risk scores within a bandwidth of 0.15 on either side of the cutoff, allow for differential linear trends and apply a uniform kernel. We also add a rich set of controls, which include quarter-year dummies, gender, education, age, arrondissement, and some prior unemployment outcomes.

As risk scores are generated automatically by an algorithm, strategic manipulation of the running variable is unlikely to be a concern. Nevertheless, discontinuities could arise mechanically if the prediction algorithm assigns scores in a non-smooth manner around the threshold. Appendix C.1 shows no evidence of bunching in the distribution of risk scores at the cutoff. We further verify in Appendix C.2 that predetermined covariates are balanced around the threshold. While we find some significant differences, these differences seem driven by the bandwidth choice and linear specification. The reported variables are also included in our set of controls.

Panel (a) of Figure 3 plots the contact rate with the service line by risk score around the risk score threshold. We observe slightly lower compliance in the immediate neighborhood of the threshold due to the frequent updating of risk scores, which leads to some individuals crossing the threshold.[13] To address this, we estimate a donut RD, excluding job seekers within 0.01 of the 0.65 threshold.

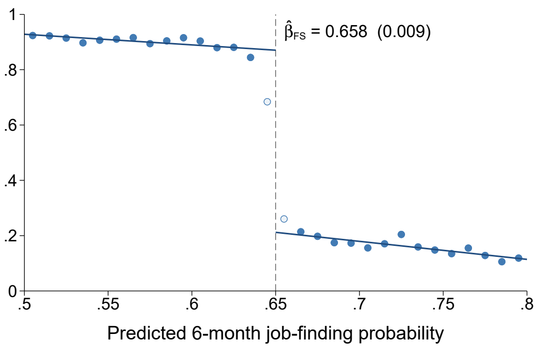
4.2 Results

The discontinuity identifies the effect of being prioritized for outreach by the service line. Since service line contact also affects subsequent assignment to caseworkers, the RD estimates the effect of a bundled intervention consisting of service-line counseling and, for a subset of job seekers, caseworker support.

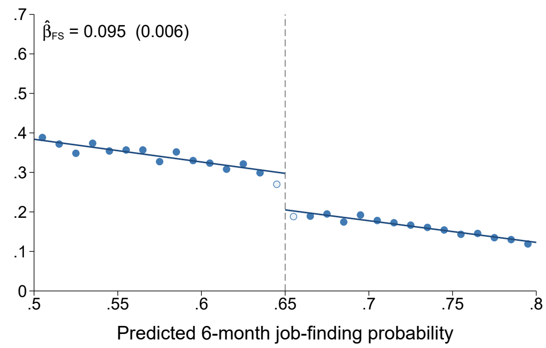
Panels (a) and (b) of Figure 3 plot the proportions of job seekers around the cutoff who had contact with the service line and with a caseworker, respectively. Both panels reveal clear discontinuities at the 0.65 threshold, which are particularly strong outside the donut. The probability of contact with the service line increases by 66 percentage points (s.e. 0.9), from a baseline of approximately 20 percent just above the threshold. This translates into a 9.5 percentage point increase (s.e. 0.6) in the probability of contact with a caseworker, relative to a baseline of just under 20 percent. The relationship between the risk score and both contact measures is close to linear on either side of the threshold, contributing to the robustness of the estimated first stages. While we aggregate contacts over the full observation window of weeks 6–17, we explore the timing of these interactions in detail in Section 6.

[12] Prior to October 2020, we use risk scores recorded by the service line between days 32 and 49 of unemployment, as many predictions were generated a few days before week 6.

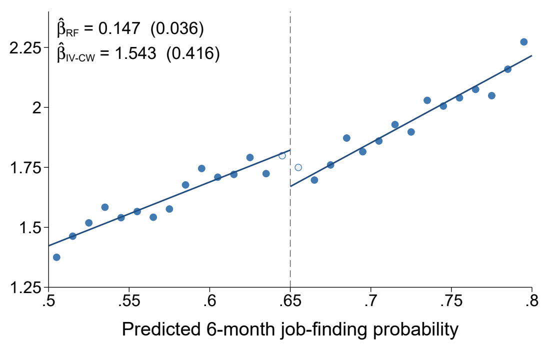
[13] We see that the drop in compliance is larger just below the threshold than just above. This is due to those above the threshold being automatically classified as self-reliant after 7 weeks of unemployment. Subsequent drops in the risk score therefore have no effect. It can take time, however, for those below the threshold to be reached, meaning subsequent increases in the risk score can lead to them being classified as self-reliant.



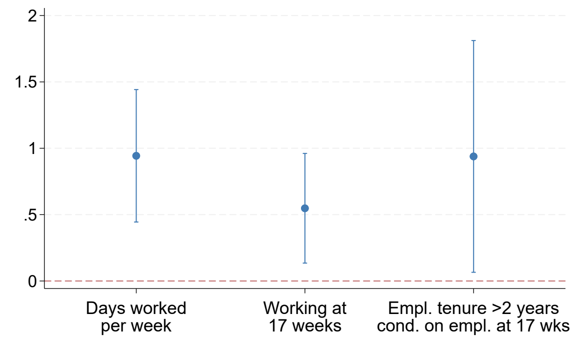
(a) Contact with the Service Line



(b) Contact with a Caseworker



(c) Days Worked per Week



(d) Caseworker IV Estimates (Relative to Control)

Figure 3: Risk Score Regression Discontinuity

Note: Panels (a) and (b) plot the proportion of spells in which job seekers had contact with the service line or a caseworker in weeks 6–17, respectively. Panel (c) plots days worked per week in weeks 6–17, after residualizing on a set of controls and adding the mean. The controls are quarter-year, age, gender, education, arrondissement, average previous unemployment duration, indicators for whether the last job and current unemployment spell began on the first day of the month, and the day of the week on which the last job and the current unemployment spell began. All reported RD estimates control for this same set of controls. Light blue markers indicate bins in the donut region, $[0.64, 0.66]$. Panel (d) plots IV estimates of the effect of caseworkers on the baseline outcome as well as two additional employment outcomes relative to the right-sided RD threshold intercept. The second outcome is a binary indicator of employment at week 17, conditional on being unemployed after 5 weeks. The third is entry into an employment spell lasting at least two years conditional on entering employment in weeks 6–17. For all panels, the sample is restricted to job seekers for whom the service line receives a risk score after 5 weeks of unemployment, thereby excluding those who had already exited unemployment or been assigned to a caseworker. The baseline estimation sample includes 78,675 spells beginning between 1 September 2018 and 9 September 2022. The sample for panel (a) is restricted to only include spells in the service line's INOS database, which began collecting data on 3 October 2020. Standard errors, reported in parentheses, are clustered at the individual level and calculated using the nearest-neighbor variance estimator. We report 95% confidence intervals.

Panel (c) of Figure 3 plots the number of days worked per week during weeks 6–17 of unemployment. As expected, employment outcomes improve steadily with the predicted job-finding probability. More importantly, we observe a clear discontinuity at the 0.65 threshold, with individuals just below the cutoff working more days on average than those just above it. The reduced-form estimate implies that being prioritized for outreach by the service line increases days worked per week by 0.15 (s.e. 0.04). Relative to a baseline of approximately 1.64 days worked per week, this corresponds to an increase of roughly 9 percent.

Combining the reduced-form estimate with the first-stage caseworker contact estimate yields a fuzzy RD estimate [Hahn et al., 2001], reported as $\hat{\beta}_{IV-CW}$. We estimate that caseworker contact corresponds to an increase of 1.54 days worked per week (s.e. 0.42), or a 94 percent increase relative to baseline. This estimate should not be interpreted as the causal effect of caseworker contact alone. Service-line interactions may themselves affect job finding and anticipation of future meetings may alter job-search behavior before actual caseworker contact occurs. Nevertheless, $\hat{\beta}_{IV-CW}$ provides a useful benchmark for comparison with the alternative RD designs introduced below.

The RD estimates are visually compelling and robust to alternative specifications. Appendix C.3 reports estimates using a wide range of bandwidths, using a triangular kernel and including observations within the donut hole.

Additional Outcomes. The positive employment effects are not unique to our baseline measure of days worked per week. Panel (d) of Figure 3 shows that job seekers just below the threshold are also more likely to be employed after 17 weeks, the end of our observation window. An important concern is whether the increase in employment reflects job seekers accepting lower-quality jobs in order to leave unemployment more quickly [e.g., Osikominu, 2013; van den Berg and Vikström, 2014]. To examine this possibility, we consider the duration of subsequent employment spells. If anything, job seekers just below the threshold appear to enter longer-lasting employment relationships than those just above, as reported in panel (d) of Figure 3. These findings are consistent with job search counseling improving not only the speed with which job seekers find employment, but also the quality or stability of the resulting job matches.[14]

We also examine intermediate outcomes that may help shed light on the mechanisms underlying the employment effects. In Appendix C.6, we show that job seekers below the threshold are substantially more likely to log into the VDAB job-search platform as well as report job applications. These patterns are consistent with the intervention increasing search activity, either through counseling provided by the service line or through the monitoring and guidance associated with subsequent caseworker contact (e.g., Schiprowski et al. [2024]). In Appendix C.7, we consider participation in the various programs offered by the VDAB, ranging from short informational interventions to long-term formal education. We find that those just below the cutoff are approximately 6 percentage points more likely to participate in one of these interventions in weeks 6–17 of unemployment. The

[14] The corresponding RD plot is provided in Appendix C.6. Because employment at week 17 is itself affected by treatment, the comparison of subsequent employment duration conditional on employment need not identify a causal effect on match quality. If workers who are more responsive to counseling are also more likely to enter stable employment, treatment will change the composition of workers employed at week 17 and can mechanically generate greater subsequent employment stability.

largest differences at the threshold are in job application requirements and sanctions investigations, with smaller increases in job search information and assistance. Taken together, these patterns suggest that the targeted intervention at this stage of the unemployment spell is primarily focused on job search monitoring, deterrence, and assistance. By contrast, the increases in training and outsourced employment services are each less than 1 percentage point. This makes it unlikely that the short-run employment effects are confounded by participation in more intensive training programs, which may initially reduce job finding through lock-in and generate benefits only at longer horizons (e.g., Lechner et al. [2011]; Cockx et al. [2023]).

The results in this section establish that job search counseling can have substantial effects on employment outcomes. The key question for targeting, however, is not whether the intervention works on average, but for whom and at what point in the unemployment spell it is most effective. We turn to these questions next.

5 Heterogeneity in Effects

While the estimated effects in our baseline RD are substantial, they may mask important heterogeneity in responsiveness to job search counseling. Understanding such heterogeneity is central to evaluating the VDAB's targeting strategy, which assigns treatment based on predicted job-finding prospects. However, the baseline RD does not allow us to estimate treatment effects separately by risk score, as the risk score is the running variable. We therefore introduce a second research design that exploits a different source of quasi-random variation and allows us to study how responsiveness varies across the risk distribution.

5.1 Empirical Strategy

A separate discontinuity arises when job seekers re-enter unemployment. As discussed before, individuals who become unemployed again within six months of a previous spell are automatically reassigned to their former caseworker if they had one during the earlier spell. Individuals returning after more than six months are not. This reassignment rule creates a discontinuity in caseworker contact around the six-month threshold that we use to estimate employment effects. We refer to this empirical strategy as the "heterogeneity RD."

Unlike our baseline RD, this design allows us to estimate treatment effects separately by predicted risk score. We split job seekers into three groups based on their predicted probability of finding a job within six months, generated on day 8 of unemployment: high-risk ($P_i < 0.6$), medium-risk ($P_i \in [0.6, 0.7)$), and low-risk ($P_i \geq 0.7$).[15] The middle group is chosen to correspond most closely to the population around the 0.65 threshold in the baseline RD.

[15] We use the prediction generated on day 8 of unemployment, as earlier predictions are considered less reliable. Predictions on day 8 are also available for individuals who find work before reaching 8 days of unemployment. The predictions do not take into account that these individuals are already in work and therefore differ little from those who find work shortly after day 8.

For each risk group, we estimate the following local linear RD specification:

$$y_i = \alpha + \beta \mathbb{1}\{d_i < 183\} + \gamma_1(d_i - 183) + \gamma_2(d_i - 183)\mathbb{1}\{d_i < 183\} + \boldsymbol{X}_i'\boldsymbol{\delta} + \epsilon_i, \tag{6}$$

where y_i is the outcome variable for unemployment spell i, d_i is the duration in days between the beginning of the current unemployment spell and the end of the previous unemployment spell and $\boldsymbol{X}_i$ is a vector of spell characteristics. The sample is restricted to job seekers who had contact with a caseworker in their previous unemployment spell.

We focus on employment outcomes during the first four weeks of unemployment, as reassignment occurs immediately whereas the standard VDAB intervention begins only after four weeks. We estimate local linear specifications within a bandwidth of twenty weeks around the threshold.

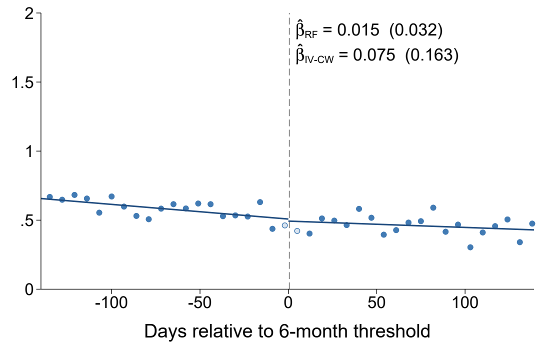
A concern is that the composition of job seekers may change discontinuously around the six-month threshold. Appendix D.1 documents clear bunching at three, six, nine, and twelve months, suggesting that these durations serve as reference points for employers and employees. We therefore use a data-driven procedure to select the donut radius, excluding the region around the six-month threshold in which the observed density of the running variable deviates from an estimated counterfactual density by more than 20%, as further explained in Appendix D.1. We also test for balance in observable characteristics on either side of the cutoff in Appendix D.2. We find some statistically significant differences, but these are not stable across bandwidth choices. All reported covariates are included as controls in our specifications. Finally, because reassignment only applies to job seekers who previously interacted with a caseworker, we can construct placebo estimates for otherwise similar job seekers who are not eligible for reassignment.

We also note that the heterogeneity RD differs from the baseline RD both in sample and treatment. The baseline RD focuses on individuals around the risk-score threshold who do not contact the VDAB when instructed to do so, whereas the heterogeneity RD focuses on job seekers who recently experienced a previous unemployment spell. In addition, treatment in the baseline RD occurs somewhat later in the spell and combines contact with service-line mediators and caseworkers, whereas the heterogeneity RD captures reassignment to a previous caseworker. We therefore view the heterogeneity RD as a complementary source of evidence on treatment effect heterogeneity rather than providing directly comparable estimates.

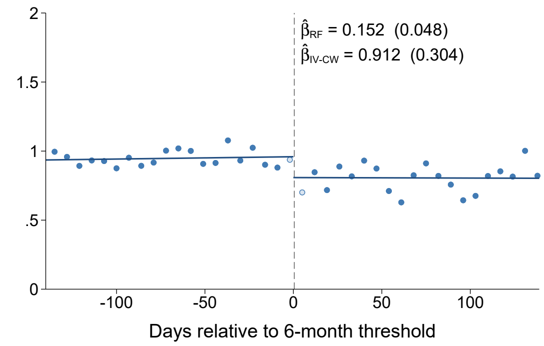
5.2 Results

We define the first stage as contact with a caseworker during the first four weeks of unemployment. Appendix D.3 shows strong first stages for all three risk groups. The discontinuity increases caseworker contact by roughly 16–20 percentage points, with the largest increase among job seekers with the lowest predicted job-finding probabilities.

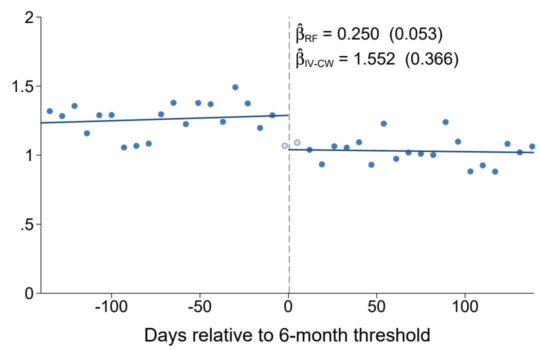
Panels (a)-(c) of Figure 4 plot employment outcomes around the six-month reassignment threshold separately by risk group. A striking pattern emerges. For job seekers with the lowest predicted job-finding probabilities – those most at risk of long-term unemployment and most likely to be targeted under the VDAB's standard approach – employment outcomes evolve smoothly through the threshold. The estimated effect on days worked is essentially zero (0.015; s.e. 0.032).



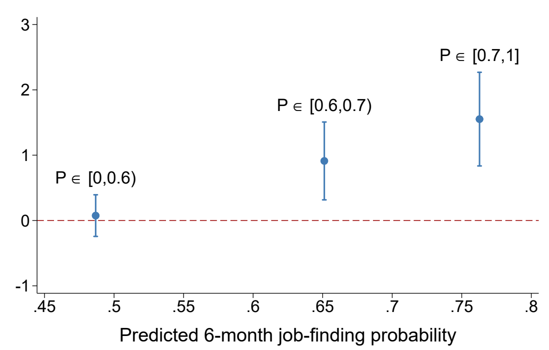
(a) Days Worked per Week ($P \in [0, 0.6)$)



(b) Days Worked per Week ($P \in [0.6, 0.7)$)



(c) Days Worked per Week ($P \in [0.7, 1]$)



(d) Effect on Days Worked per Week (IV Estimates)

Figure 4: Heterogeneity in Treatment Effects

Note: The sample is restricted to job seekers who had contact with a caseworker in their previous spell. Job seekers are then split into three groups by the probability of finding a job within 6 months, P, which was predicted for them after 1 week of unemployment. Panels (a) to (c) plot days worked per week in weeks 1–4 for each risk group. The outcome is residualized on a set of controls and added to the mean. The controls are quarter-year, gender, education, age, arrondissement, average previous unemployment duration, indicators for whether the last job or current unemployment spell began on the first of the month, and the day of the week on which the last job and the current unemployment spell began. All reported RD estimates control for this same set of controls. Light blue markers indicate bins with at least one observation inside the donut hole, the region $[-2, 5]$. Observations in the donut hole are excluded from the estimation sample but are included when computing bin means. We reject the null hypotheses that the IV estimate from panel (a) is equal to the estimates from panels (b) and (c) with p-values of 0.016 and below 0.001, respectively. Panel (d) plots, for each group, the IV RD estimate against the mean of P. We report 95% confidence intervals. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

In contrast, we find large and statistically significant employment effects for the two groups with higher predicted job-finding probabilities. The estimated effect equals 0.152 (s.e. 0.048) for job seekers with $P_i \in [0.6, 0.7)$ and rises further to 0.250 (s.e. 0.053) for those with $P_i \in [0.7, 1]$. We reject equality of treatment effects between the lowest and highest risk groups with a p-value below 0.001.

Panel (d) reports the corresponding fuzzy RD estimates, $\hat{\beta}_{IV-CW}$, using the discontinuity as an instrument for caseworker contact. The IV estimates reinforce the monotonic relationship between predicted job-finding prospects and responsiveness to counseling. For job seekers around the 0.65 threshold, the estimated effect is close to the baseline RD estimate. More importantly, treatment effects decline sharply as predicted risk of long-term unemployment increases.

Appendix D.4 shows that our results are robust to alternative bandwidths, kernels, and donut specifications. For each of the risk groups, the estimates hardly change across specifications and thus also their ranking remains robust. Moreover, while the heterogeneity RD is restricted to a four-week outcome window, Appendix D.5 shows that the same monotonic pattern persists when outcomes are measured over sixteen weeks. We show further robustness to estimating treatment effects for more finely defined risk groups and alternative employment outcomes in Appendices D.6 and D.7, respectively. We also explore effects on employment tenure. One may be concerned that job search counseling could push high-risk individuals into less stable employment. We do not, however, find strong evidence for either a positive or negative effect on employment tenure for any of the three risk groups.

As a final robustness check, we consider a placebo design using job seekers who did not have a caseworker in their previous spell and are therefore not subject to reassignment at the six-month threshold. We report the placebo RD results in Appendix D.8. Pooling all risk groups, the placebo estimate is small and statistically insignificant. Estimates by risk group are noisier and more sensitive to the specification. In our baseline specification, the placebo estimates are insignificant for the highest- and lowest-risk groups, while the estimate for job seekers with $P_i \in [0.6, 0.7)$ is positive and statistically significant, but not robust across specifications.

Taken together, these results reveal a strong and overall robust, negative relationship between unemployment risk and responsiveness to counseling. The job seekers whom the VDAB most actively targets exhibit little or no employment response, whereas those at substantially lower risk respond strongly. The trade-off between targeting individuals who are most at risk and targeting those who are most responsive therefore appears to be first-order in this setting. Before interpreting these results as evidence of genuine heterogeneity in responsiveness, we consider two alternative explanations.

Heterogeneity in Treatments. A natural alternative explanation for the observed heterogeneity is that different job seekers receive different interventions. In particular, high-risk job seekers may be more likely to be referred to intensive training programs. Such programs often involve lock-in effects, reducing job search and employment in the short run while potentially improving longer-run outcomes. If so, the weaker short-run employment effects for high-risk job seekers would reflect differences in treatment content rather than differences in responsiveness to counseling.

We find, however, little evidence that caseworkers systematically assign different interventions across risk groups. Appendix D.9 examines a broad range of tasks and programs and reveals broadly similar treatment patterns. All risk groups face a significant increase in the requirement to apply for jobs, while we find no significant effect on training participation. If anything, the training estimate is most negative for job seekers with $P_i \in [0.6, 0.7)$, going against the interpretation that high-risk job seekers are more likely to be locked into training. We thus find little support for the view that the observed heterogeneity is driven by differences in treatment content.[16]

Compared with the patterns observed in the baseline design (see Appendix C.7), caseworkers assign relatively few additional interventions to repeat job seekers during weeks 1–4. In general, we prefer to interpret the heterogeneity RD estimates not as the effect of training or other intensive services, but rather as the effect of a treatment consisting primarily of one or a few meetings with a caseworker, increased job-application requirements, and the prospect of more intensive treatment in the future. Our findings suggest that the high-risk job seekers targeted by the VDAB are the least responsive to this treatment.

Algorithmic Bias. Since treatment effects are estimated to be heterogeneous, another concern is that they mechanically correlate with predicted risk. The reason is that the risk score prediction algorithm may be trained on outcomes that themselves reflect prior treatment. In particular, the VDAB algorithm predicts job-finding probabilities using historical data generated under an existing targeting regime. If treatment affects job finding, then the algorithm does not simply learn who would find work in the absence of intervention; it also learns who benefited from intervention in the past.

The relevance of this concern becomes clear when considering the relationship between predicted job-finding probabilities P_i and treatment effects τ_i. Put simply, suppose $P_i = \pi_i + \tau_i T_i$ such that:

$$Cov(P_i, \tau_i) = Cov(\pi_i, \tau_i) + Cov(\tau_i T_i, \tau_i).$$

In the presence of treatment T_i, the algorithm learns a combination of baseline job-finding prospects and treatment responsiveness rather than the former alone. Under random assignment, we obtain $Cov(\tau_i T_i, \tau_i) = Pr(T_i = 1)Var(\tau_i)$. In this case, even if untreated job-finding prospects π_i are unrelated to treatment effects, observed job-finding probabilities P_i will be positively correlated with treatment effects.

This observation has important implications for targeting. Ideally, targeting would distinguish between untreated job-finding prospects π_i and treatment effects τ_i. However, when treatment is assigned on the basis of predicted outcomes P_i instead, individuals with low treatment effects τ_i may be adversely selected into treatment. An important question is whether this mechanism is quantitatively important in our setting. Since we do directly observe π_i, we will assess the overall variation in treatment effects in the next section and then return to this issue using the calibrated model in Section 7.

[16] Differential program assignment across caseworkers may nevertheless be important more generally. Humlum et al. [2025], for example, exploit differences in caseworkers' propensities to assign job seekers to different training programs in Denmark and find sizable effects of classroom training.

5.3 Other Observable Heterogeneity

The heterogeneity by risk score reveals an important trade-off between targeting job seekers who are most at risk and those who are most responsive to intervention. However, risk scores represent only one dimension of observable heterogeneity. More generally, efficient targeting requires identifying which observable characteristics predict treatment effects and whether these predictions improve upon targeting based on risk scores alone. We therefore adopt a data-driven approach and estimate conditional average treatment effects (CATEs) using an instrumental forest algorithm following Athey et al. [2019]. This allows us both to predict treatment effects from observables and to identify the characteristics most important for explaining treatment effect heterogeneity.

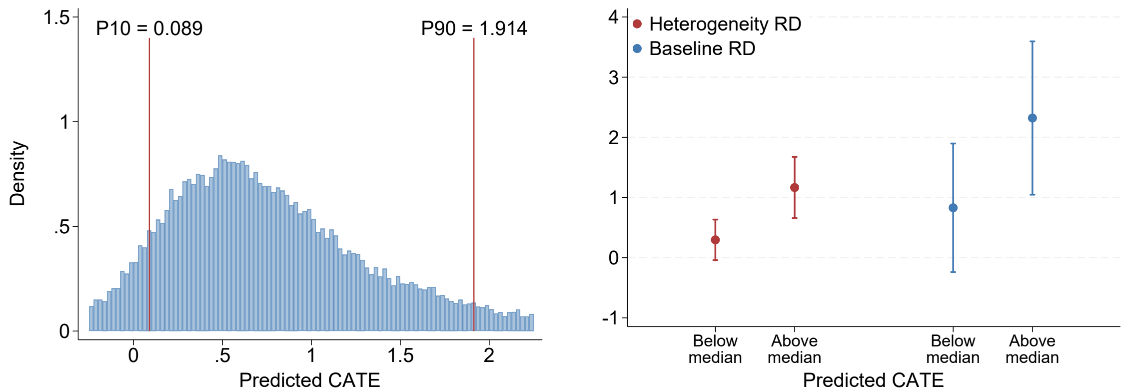
We estimate the instrumental forest using the same treatment variation underlying the heterogeneity RD. Since the algorithm does not accommodate the local-linear structure of the RD, we restrict attention to observations within ten weeks of the six-month cutoff and use caseworker contact as the treatment variable and assignment below the cutoff as the instrument.[17] Following Basu et al. [2018] and Athey and Wager [2019], we train an initial causal forest on a large set of features and then a second causal forest on the features that were split on the most. All reported CATEs are 'out-of-bag' predictions, so that each observation's predicted treatment effect is generated using trees that were not trained on that observation.

Panel (a) of Figure 5 plots the distribution of predicted CATEs. The distribution reveals substantial predictable heterogeneity in treatment effects. Indeed, the relative dispersion in predicted treatment effects is considerably larger than the relative dispersion in risk scores. By Proposition 2, this distinction is important as optimal targeting depends on proportional differences: a job seeker with a higher job-finding probability can still have a higher return to treatment if their treatment effect is proportionally larger.

Panel (b) of Figure 5 assesses both the internal and external validity of the predicted CATEs. For internal validity, we estimate the RD separately for observations with below and above median CATEs. We find a statistically significant difference in the estimated treatment effects, confirming that the CATE estimates capture meaningful treatment effect heterogeneity. Note that, if anything, the CATE estimates underestimate the heterogeneity in RD estimates by risk score. As shown in Appendix D.10, this attenuation is primarily explained by the inability of the instrumental forest procedure to accommodate linear trends in the running variable.

The stronger test is external validity. The CATEs are estimated using the heterogeneity RD, which captures the short-run effects of reassigning recently unemployed job seekers to a previous caseworker. The baseline RD instead captures a different intervention, occurring later in unemployment, for a different population. It is therefore far from obvious that the same characteristics should predict treatment effects across the two settings. Nevertheless, we find that CATEs estimated from the heterogeneity RD also predict treatment effects in the baseline RD. The estimated effect for job seekers with above-median CATEs is more than twice as large as for those with below-median CATEs, although the difference between the two effects is only marginally significant ($p = 0.08$). This is a particularly demanding test: because the baseline RD sample is local to the $P = 0.65$

[17] See Britto et al. [2022] for a similar implementation of causal forest estimation in an RD context.



(a) Distribution of CATEs

(b) Internal and External Validity

Figure 5: Predicted Conditional Average Treatment Effects (CATEs)

Note: Panel (a) plots the distribution of estimated CATEs for the sample on which the instrumental forest was trained (note, each CATE is still an out-of-bag prediction). The red estimates in panel (b) are IV estimates of the effect of a caseworker on days worked per week estimated using the heterogeneity RD for samples with below and above median CATEs. We reject the null hypothesis that effects are equal with a p-value of 0.005. In blue, we report estimates of the effect of a caseworker on days worked per week from the baseline RD by CATE. The p-value for rejecting the null hypothesis that these effects are equal is 0.08. The same median is used for both sets of estimates. We display 95% confidence intervals for the estimated effects.

threshold, the risk score — as we show below, the most important predictor of treatment effects — contributes virtually no variation in predicted CATEs within this sample, so the predictive power of the CATEs derives almost entirely from other characteristics. Given the differences in treatment, timing, and population across the two designs, we thus view this result as suggestive evidence that the heterogeneity uncovered by the heterogeneity RD reflects stable differences in responsiveness rather than idiosyncrasies of a particular treatment environment.

In addition to quantifying the heterogeneity, the instrumental forest estimation allows us to analyze the observables that drive the heterogeneity. In Appendix D.11, we report the "variable importance" for different observables, which is the number of times the forest split on a variable weighted by the depth at which the split occurred. Interestingly, the risk score is the single most important predictor of treatment effects, confirming the earlier evidence that responsiveness is strongly related to predicted unemployment risk. However, the forest also identifies demographic characteristics, unemployment history, and occupation preferences as important predictors. The fact that these variables continue to matter even after conditioning on the risk score suggests that substantial heterogeneity remains within risk-score groups. In other words, the risk score captures an important but incomplete dimension of treatment effect heterogeneity.

To illustrate this further, Appendix D.12 compares individuals with high vs. low long-term unemployment risk and high vs. low predicted treatment effects. While some characteristics contribute to the average trade-off between the two, some characteristics are associated with unemployment risk and treatment responsiveness in the same direction. For example, job seekers aged 45+ as well as those of non-Belgian origin appear both more likely to experience long-term unemployment and more responsive to intervention. Such characteristics strengthen the case for targeting, as they simultaneously increase need and expected returns from treatment. Other characteristics primarily affect responsiveness while having little impact on unemployment risk. For instance, female job seekers exhibit lower predicted treatment effects despite having similar risk scores. A targeting strategy based solely on unemployment risk would therefore overlook dimensions of heterogeneity that are highly relevant for efficient treatment assignment. We return to these implications when comparing alternative targeting strategies in the calibrated model in Section 7 below.

6 Duration Dependence in Effects

The conceptual framework highlights that the timing of treatment depends on potentially opposing forces underlying its return. On the one hand, treatment is more valuable early if treatment effects decline with unemployment duration. On the other hand, treatment becomes more valuable later in the spell if job-finding prospects deteriorate and the consequences of continued unemployment become more severe. Whether employment services should intervene early or late therefore depends on the relative dynamics of treatment effects and job-finding probabilities. As Proposition 1 highlights, however, delaying treatment also shrinks the pool of job seekers who can still be reached, so intervening later is optimal only if returns rise sufficiently to compensate for this survival loss — a trade-off we return to in Section 7.

In this section, we estimate both objects. We first introduce an alternative RD design that allows us to compare treatment effects for similar job seekers treated at different points in the unemployment spell. We then quantify how dynamic selection changes both treatment effects and job-finding probabilities among the surviving unemployed population.

6.1 Empirical Design

To study the dynamics of treatment effects, we exploit a second round of interventions that begins around week 19 of the unemployment spell. As discussed before, all job seekers who are still unemployed at that point and have not previously been assigned to a caseworker are invited again to contact the service line. In week 20, the service line contacts those who did not respond. This second round generates a useful reversal in treatment assignment relative to the baseline RD.

In the baseline RD, we used the risk score threshold to estimate treatment effects for individuals with a predicted six-month job-finding probability just below 0.65. Because job seekers below the threshold are more likely to be treated during this first round, those above the threshold are more likely to remain untreated and therefore more likely to receive treatment during the second round, when the risk-score threshold is no longer used.

We exploit this reversal by estimating the same specification as in the baseline RD,

$$y_i = \alpha + \beta \mathbb{1}\{P_i < 0.65\} + \gamma_1(P_i - 0.65) + \gamma_2(P_i - 0.65)\mathbb{1}\{P_i < 0.65\} + \boldsymbol{X}_i'\boldsymbol{\delta} + \epsilon_i, \qquad (7)$$

but with treatment assignment reversed around the threshold. The running variable remains the same as in the baseline RD: the predicted six-month job-finding probability initially used by the service line in weeks 6–7. We refer to this empirical strategy as the "dynamics RD." Combined with the baseline RD, it allows us to compare treatment effects for job seekers with similar predicted job-finding prospects, but at different points in the unemployment spell.

6.2 Results

Figure 6 plots the estimates for β at different durations d. Panels (a) and (b) first verify the reversal in treatment assignment that underlies the dynamics RD. Job seekers below the risk-score threshold are substantially more likely to interact with the service line or be assigned to a caseworker during the first treatment round, but substantially less likely to do so during the second round.[18] The differences in contact probabilities are largest in the first few weeks after each treatment round begins and then rapidly disappear, although caseworker contact lags service line contact by up to a few weeks and the difference in contact rates decreases more gradually. The reversal in assignment generates quasi-experimental variation in caseworker contact among surviving job seekers around the same risk-score threshold, but at different unemployment durations.

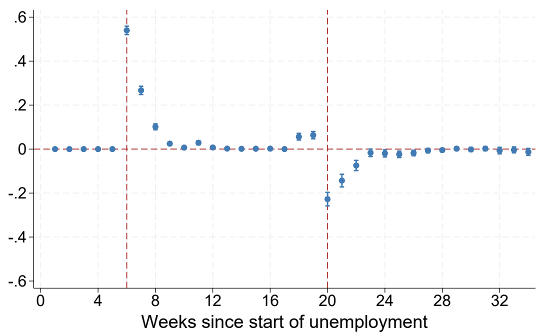
Panel (c) of Figure 6 examines the corresponding employment effects. Employment responds rapidly following both treatment rounds, with effects concentrated in the weeks immediately after the start of the contact round. While the weekly estimates are somewhat noisy, the timing and shape of the employment response do not differ substantially across the two rounds. If anything, the effects appear slightly more concentrated towards the start of the round in the second one.

Panel (d) of Figure 6 summarizes these dynamics using IV estimates of the effect of caseworker contact on days worked per week over a twelve-week horizon. We estimate a treatment effect of 1.54 (s.e. 0.42) days worked per week in the first treatment round and 1.48 (s.e. 0.61) in the second. Appendix E.1 shows the corresponding RD plot for the dynamics RD. While the confidence intervals remain substantial, the estimates provide little evidence that treatment effects would actually decline as unemployment lengthens.

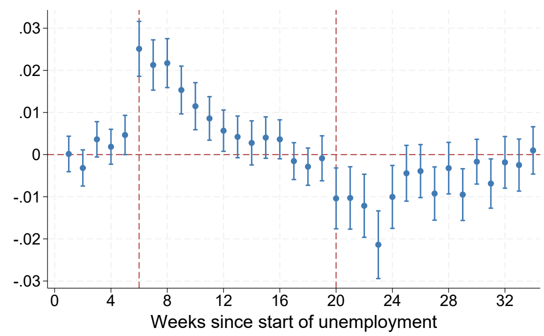
Appendix E.2 further gauges robustness of this conclusion to the bandwidth and weighting choices. We also report estimates for different horizons in Appendix E.3. The results are comparable overall and indicate no important increases or decreases in estimated treatment effects over the spell.

Interpretation From the perspective of the conceptual framework, the comparison between the baseline and dynamics RD captures the relevant notion of duration dependence conditional on the observable risk score. Both designs identify treatment effects for job seekers with predicted job-finding probabilities around the same threshold, but at different points in the unemployment spell. To the extent that the risk score captures all relevant heterogeneity in job-finding prospects, the comparison isolates how treatment effectiveness evolves with unemployment duration for comparable job seekers. The similarity of the estimated effects then suggests little depreciation in treatment effectiveness over the first months of unemployment. However, interpreting the comparison between

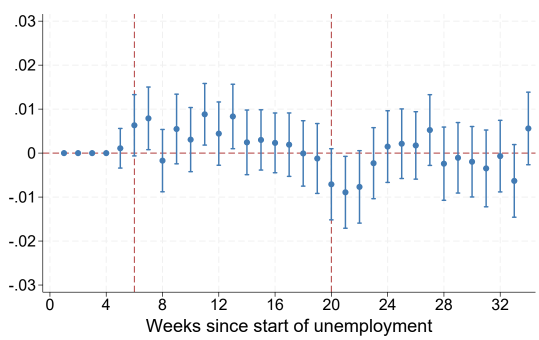
[18] In weeks 18 and 19 we observe slightly more contact with the service line below the threshold than above. This is because job seekers are invited to call the service line 11 weeks after being classified as self-reliant (and get called one week later). Those below the threshold were more likely to be classified as self-reliant in weeks 6 and 7, so are more likely to be called in weeks 18 and 19. Most of those above the threshold were automatically classified as self-reliant after 7 full weeks of unemployment and therefore get called at the start of the 20th week.



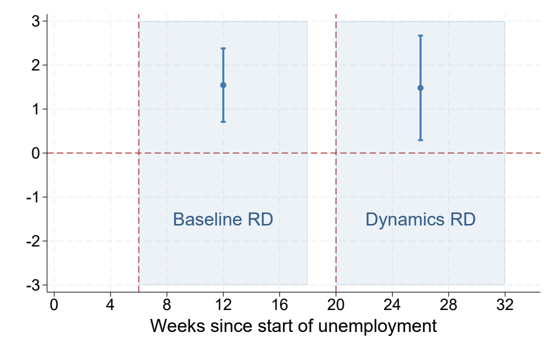
(a) Effect on Contact with Service Line (Weekly)



(b) Effect on First Contact with Caseworker (Weekly)



(c) Effect on Entry into Employment (Weekly)



(d) IV Effect on Days Worked per Week (Over 12 Weeks)

Figure 6: Dynamics

Note: All estimates use the same specifications used in the baseline RD. Panel (a) plots RD estimates of the effect of being below $P = 0.65$ on contact with the service line each week, conditional on being unemployed at the beginning of that week. Panel (b) plots the estimated effect on having the first caseworker meeting of the unemployment spell each week. Panel (c) plots the estimated effect on employment each week. Panel (d) plots the IV estimate of the effect of meeting a caseworker for the first time in the 12-week period on days worked per week in the same period. We display 95% confidence intervals for the estimated effects.

the baseline and dynamics RD as evidence on *true* duration dependence in treatment effects requires ruling out dynamic selection on unobservables.

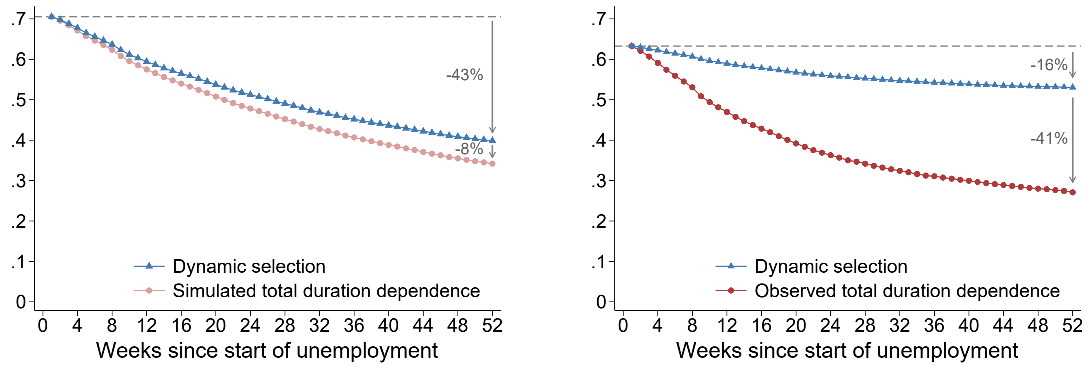
While both designs focus on job seekers with similar predicted job-finding prospects, surviving job seekers in the second treatment round may still differ from those in the first round along dimensions not captured by the risk score. If treatment effects were correlated with such unobserved characteristics, differences between the two estimates could reflect changes in composition rather than genuine duration dependence. In any case, the estimates provide little indication of either substantial duration dependence or strong selection on unobservables. At minimum, neither force appears quantitatively dominant in our setting.

Separately, the institutional structure of the dynamics RD raises two concerns for interpreting the second-round estimates themselves. Both arise because treatment during the first round affects the composition of job seekers who remain eligible for treatment during the second round.

A first concern is differential survival into the second treatment round. Since treatment increases job finding in the first round, job seekers below the threshold are less likely to remain unemployed until week 20. If treatment effects are positively correlated with unobserved job-finding prospects, the surviving population above the threshold could therefore exhibit higher job-finding rates even in the absence of the second-round intervention. We view this concern as limited. The employment effects in Figure 6c emerge precisely at week 20 when the second treatment round begins, while estimates in weeks 18 and 19 are close to zero. This timing is difficult to reconcile with a purely compositional explanation.

A second concern is that surviving untreated job seekers below and above the threshold may differ systematically because mediators exercise discretion when assigning caseworkers during the first treatment round. In particular, some job seekers below the threshold may be classified as self-reliant and thus remain untreated despite being contacted. If these individuals differ in their responsiveness to treatment, the composition of untreated survivors could differ across sides of the threshold. To assess this possibility, we use the CATEs estimated with the instrumental forest in Section 5 and test whether these predicted treatment effects are discontinuous at the threshold among job seekers who survive untreated until week 20 (Appendix E.4). We find no meaningful discontinuities, either in the predicted CATEs or in their main observable determinants. This suggests that differential selection into the pool of untreated survivors is unlikely to explain our findings.

Finally, Appendix E.5 examines program participation around the threshold in weeks 20–31. We find no significant differences in program participation, although these null differences are harder to interpret as they may reflect some convergence in program participation following the first treatment round. Participation levels in interventions such as training and job search orientation near the cutoff are comparable to those in weeks 6–17 (see Appendix C.7). As before, the employment effects in the dynamics RD may instead reflect the impact of service-line and caseworker contacts rather than program participation itself.



(a) Treatment Effects

(b) Job Finding

Figure 7: Duration Dependence in Treatment Effects and Job Finding

Note: Panel (a) simulates the evolution of the average treatment effect of caseworkers on days worked per week at different unemployment durations due solely to dynamic selection. This is done by splitting the sample into predicted risk quintiles and estimating treatment effects and observed survival rates for each quintile. Treatment effects for each quintile are estimated using the heterogeneity RD. Total duration dependence is then estimated by calculating true duration dependence using the baseline and dynamics RD estimates from Figure 6d and assuming it to be homogeneous within the sample and affecting treatment effects at a constant geometric rate with duration. Panel (b) plots the empirical 6-month job-finding rate for the surviving population at each duration (observed total duration dependence) against 6-month job-finding rate predicted by VDAB on day 8 of unemployment for the surviving population at each duration (a lower bound on dynamic selection). All predicted rates are multiplied by a constant so that the predicted and empirical rates are equal at a duration of 1 week.

6.3 Dynamic Selection

The previous subsection provided little evidence that treatment effects decline substantially with unemployment duration for comparable job seekers. This does not imply, however, that the returns to treatment remain constant over the unemployment spell. As unemployment lengthens, the composition of the unemployed population changes. Job seekers with lower job-finding prospects become increasingly overrepresented among those who remain unemployed and, given the heterogeneity documented in Section 5, these individuals also tend to exhibit lower treatment responsiveness. Dynamic selection therefore affects both job-finding probabilities and treatment effects, making it a central determinant of the optimal timing of interventions.

Panel (a) of Figure 7 simulates the dynamic selection in treatment effects by splitting the sample into predicted risk quintiles and simulating the average treatment effect of the surviving population at different unemployment durations.[19] The dynamic selection of job seekers with low predicted job finding into longer unemployment spells causes the average treatment effect to fall from 0.71 for job seekers one week into unemployment to 0.40 for those who have been unemployed for a year. Thus, even in the absence of true duration dependence, average treatment effects decline substantially over the unemployment spell because the remaining unemployed population becomes increasingly negatively selected. We can also simulate total duration dependence by taking the difference between

[19]This assumes treatment effects are the same within predicted risk quintile, but heterogeneous across quintiles. Quintile treatment effects are estimated using the heterogeneity RD.

the baseline RD and the dynamics RD estimates reported in panel (d) of Figure 6 as representative of true duration dependence in treatment effects. Assuming true duration dependence to evolve at a constant geometric rate, we estimate an overall drop in average treatment effects to 0.34 after one year of unemployment, a decline of 51%.

Panel (b) of Figure 7 then investigates duration dependence in job-finding rates. Empirically, we find that 6-month job-finding rates fall from 0.63 at one week of unemployment to 0.27 at one year of unemployment, a reduction of 57%. The decline in average predicted job-finding, with the rates predicted at the start of the spell but averaged across job seekers who survive later into the spell, can give us a lower bound estimate of the proportion of duration dependence in job-finding explained by dynamic selection rather than true duration dependence, following Mueller and Spinnewijn [2025]. In comparison to their findings in the Swedish context, we find that the decline in job-finding caused by dynamic selection is somewhat smaller, while the overall decline in job-finding is much larger. This suggests that, unlike treatment effects, job-finding rates exhibit substantial duration dependence beyond what can be explained by dynamic selection alone.

Taken together, the results imply a decline in average treatment effects of roughly 51% over the first year of unemployment, compared to a decline in job-finding probabilities of roughly 57%. Through the lens of Proposition 2, these two forces have opposing implications for timing. The decline in treatment responsiveness favors earlier intervention, whereas the decline in job-finding probabilities increases the value of helping those who remain unemployed. On balance, the two observable forces are of comparable magnitude, suggesting no major advantage of targeting earlier rather than later in the unemployment spell.[20]

The comparison above characterizes how returns to treatment evolve among job seekers who remain unemployed at different durations. This does not by itself determine the optimal timing of treatment. As Proposition 1 highlights, delaying treatment may raise returns among those who remain unemployed, but comes at the cost of losing the opportunity to treat high-return job seekers who exit in the meantime. This harvesting motive creates a force toward earlier intervention. In the next section, we quantify this trade-off within the calibrated model and characterize optimal targeting jointly across job seekers and unemployment durations.

7 Quantitative Evaluation

The preceding sections estimated the key ingredients for targeting job-search counseling: baseline job-finding prospects, treatment responsiveness, and how both evolve over the unemployment spell. This section combines these ingredients in a calibrated version of the framework to quantify the welfare gains from alternative targeting policies. The exercise serves two main purposes. First, we compare targeting policies to evaluate how much is lost by using unemployment risk alone as a targeting rule, and how closely the simple sufficient statistic $R \approx \tau/\pi$ approximates the full model-based return to treatment. Second, we quantify the timing trade-off, distinguishing between changes in the return to treating job seekers who remain unemployed and changes in the share of job seekers

[20] The strong decline in job-finding probabilities that is not accounted for by observables (about 41%) still points toward increasing returns to targeting later in the spell, conditional on risk scores.

Table 1: Calibrated Parameter Values in Baseline

Object	Description	Value	Source/Targeted Moments
$G(\pi)$	Distribution of observable job finding	—	Empirical distribution of risk scores
σ_{ε_π}	Unobserved heterogeneity in job finding	0.031	Mueller and Spinnewijn [2025]
θ_π	Depreciation of job finding	0.925	12-month decline in job finding
α_τ	Intercept in $\tau = \alpha_\tau + \beta_\tau P + e_\tau$	-0.032	RD estimates by risk score
β_τ	Slope in $\tau = \alpha_\tau + \beta_\tau P + e_\tau$	0.080	RD estimates by risk score
σ_{e_τ}	Orthogonal heterogeneity in treatment effects	0.025	Variance of CATEs
θ_τ	Depreciation of treatment effects	1.000	Estimates of risk-score RDs
$\bar{B}$	Budget constraint	0.530	Share of unemployed targeted by VDAB

who survive to each duration.

7.1 Calibration

We calibrate the model to key moments in the data, as summarized in Table 1. For the job finding we start from the empirical distribution of predicted 6-month job-finding probabilities P from the VDAB, drawing a random sample of 200,000 individuals at the start of the spell. We calibrate the model at the monthly frequency and thus map each risk score into a monthly job-finding probability. We refer to this as the baseline job-finding probability, denoted by π.[21] We allow for additional unobserved heterogeneity in job finding, σ_{ε_π}, which is calibrated using estimates on the relative contribution of observed and unobserved heterogeneity to job-finding risk in Sweden from Mueller and Spinnewijn [2025]. We assume geometric depreciation, $\theta_{\pi,d} = (\theta_\pi)^d$, where θ_π is chosen so that the implied decline in job-finding from both the depreciation and the dynamic selection based on observed and unobserved heterogeneity matches the observed decline in job-finding rates over the first 12 months in the Flemish data. We assume no further depreciation in job finding after 12 months.[22]

Turning to the treatment effects, we assume that observable heterogeneity is linear in the VDAB prediction:

$$\tau = \alpha_\tau + \beta_\tau P + e_\tau,$$

where α_τ is a constant, β_τ captures the relation between treatment effects and risk scores, and e_τ represents heterogeneity orthogonal to risk scores. We assume that treatment increases the job finding probability for one month only and use the estimated reduced-form effects on the probability of starting work within four weeks from the heterogeneity RD. We calibrate α_τ and β_τ by taking the RD estimate for each decile of the risk score distribution and fitting a linear approximation to

[21] We recover π from the VDAB risk score P using a numerical solver. In the absence of depreciation and unobserved heterogeneity, this mapping would be simply given by $\pi = 1 - (1 - P)^{1/6}$. The VDAB predictions are based on realized job-finding outcomes under the existing policy environment, while the model object is baseline job finding in the absence of treatment. In the baseline calibration, we abstract from this distinction, but then explicitly relax it in sensitivity analysis.

[22] We also bound the monthly job-finding probability from below at 1 percent. The bound does not bind at the start of the spell, as the lowest predicted 6-month job-finding probability in our sample is 4 percent, but limits the decline in job finding at long durations for job seekers with the lowest risk scores. Without such a bound, continuation losses, which are inversely proportional to job-finding probabilities, would become arbitrarily large as hazards approach zero, so that welfare results could be driven by a small set of extreme unemployment durations.

these estimates (see Figure F2 in the Appendix). To discipline the orthogonal component, we target the variance of CATEs shown in Figure 5a.[23] In the baseline specification, we assume no additional unobserved heterogeneity in treatment effects, $\sigma_{\varepsilon_\tau} = 0$ and no depreciation in treatment effects, consistent with the comparison of RD estimates at 5 and 19 weeks. We assess robustness to both assumptions.

For evaluating welfare effects we lack direct empirical targets for the utility cost of unemployment u and the treatment cost c. Since our focus is on relative welfare comparisons, we normalize both to one and omit these parameters from the table. We set $\theta_{u,d} = (\theta_u)^d = 1$ in the baseline, while exploring sensitivity to this assumption. We also impose a budget constraint consistent with current VDAB policy, allowing treatment of 53 percent of the sample.[24] Our quantitative analysis is not intended as a formal policy-learning exercise. In particular, we do not directly estimate optimal rankings under the budget constraint, which can differ from rankings based on estimated conditional means [Gu and Koenker, 2023]. Nor do we conduct formal inference on the value of the policy that is estimated to be optimal; see Whitehouse et al. [2026] for recent methods addressing this problem. Instead, we assess the robustness of our quantitative conclusions across a range of alternative specifications and calibrations.

7.2 Optimal Targeting and Timing

We first evaluate targeting rules for newly unemployed workers, fixing treatment at the start of the spell, $d = 0$. This exercise isolates the cross-sectional targeting problem: whom should the agency treat among the entering unemployed, abstracting from the timing dimension considered below. We compare three classes of policies, each subject to the same budget constraint:

1. **Optimal Targeting (T_0^{OPT})**: Constrained optimal targeting of individuals with the highest return, $R^{T_0}(\pi, \tau)$.
2. **Index Targeting (T_0^{IDX})**: Targeting based on a geometric weighted average of π and τ,[25]

$$Q(\pi, \tau, \omega) \quad = \quad \frac{\tau^{1-\omega}}{\pi^{\omega}}. \tag{8}$$

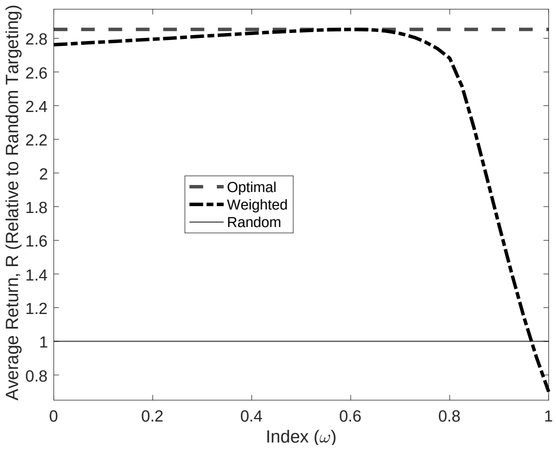
3. **Random Targeting (T_0^{RDM})**: Targeting a random sub-sample of unemployed workers.

The index targeting allows us to evaluate simple rules, ranging from targeting on π ($\omega = 1$) to τ ($\omega = 0$) only, and also includes our sufficient-statistic approximation when targeting on τ/π ($\omega = 0.5$). Panel (a) of Figure 8 reports the returns of different targeting policies relative to the random benchmark, which yields the average return to intervening. Several results stand out.

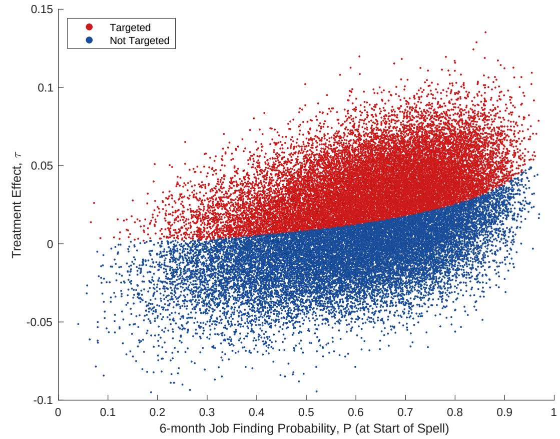
[23] More precisely, the CATEs reported in Figure 5a are based on days worked. In the calibration, however, we target the dispersion of CATEs for the probability of starting work within four weeks. See also Appendix Figure D10 for the heterogeneity RD estimates by risk group for this outcome.

[24] We assume that both ε_π and ε_τ are uniformly distributed with mean zero. Appendix Figure F1 shows the distribution of π's and τ's as well as the duration dependence in job finding, which matches closely the empirically observed duration dependence.

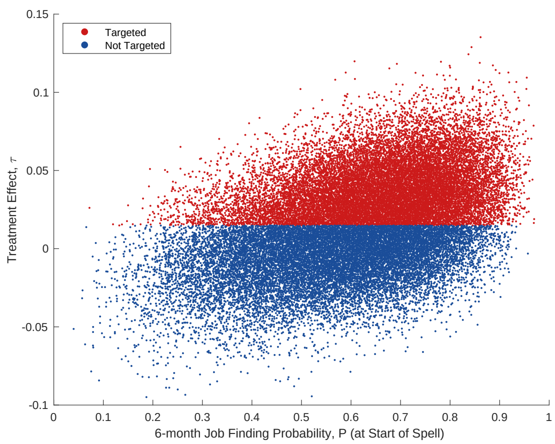
[25] To be precise, for the purposes of computing the index, we assume that the treatment effects are non-negative.



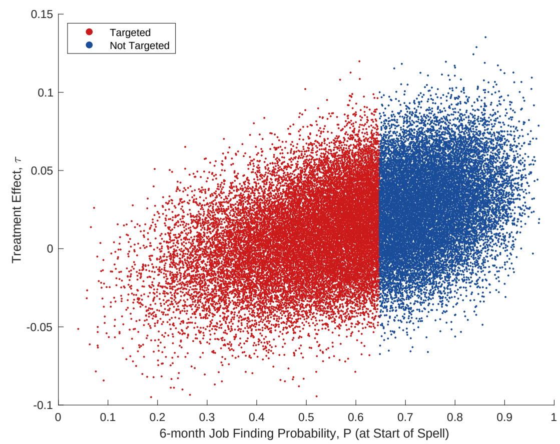
(a) Average Return Relative to Random Targeting



(b) Targeted Types With Optimal Policy



(c) Targeted Types With $\omega = 0$



(d) Targeted Types With $\omega = 1$

Figure 8: Welfare Gains of Targeting and Targeted Types under Alternative and Optimal Policies

Note: Panel (a) shows welfare gains from targeting relative to the benchmark of random assignment. Panel (b) displays the worker types targeted under the optimal policy. Panels (c) and (d) illustrate the targeted types under sub-optimal policies that focus exclusively on treatment effects ($\omega = 0$) or exclusively on job-finding probabilities ($\omega = 1$), respectively. All policies are subject to the same budget constraint.

First, the welfare gain under the optimal policy is nearly three times as large as under the benchmark. This reflects substantial heterogeneity in both observed risk scores and treatment effects across individuals. A Shapley decomposition of the variance of the targeting index τ/π attributes 82 percent to heterogeneity in treatment effects and the remaining 18 percent to heterogeneity in baseline job-finding probabilities (see Appendix Table F1). The corresponding panel (b) in Figure 8 illustrates the types targeted under the optimal policy, clearly highlighting the underlying trade-off between targeting individuals that are responsive vs. at risk.

Second, the index rule can perform very well and realize basically all gains from the optimally targeted policy. The index τ/π is highly correlated with the exact return to targeting (correlation 0.88), supporting the sufficient-statistic approximation. The simplest rule based on this index provides an excellent approximation, but the index attains its maximum at $\omega = 0.60$. Recall that the optimal weight is exactly $\omega = 0.50$ in a stationary environment without duration dependence in job finding.

Third, a policy that targets exclusively on job-finding probabilities ($\omega = 1$) performs poorly and even *worse* than the benchmark of random targeting. This is because individuals at highest risk of long-term unemployment exhibit negligible treatment effects. Conversely, a policy that targets exclusively on treatment effects ($\omega = 0$) performs nearly as well as the optimal policy, as it avoids selecting individuals with low returns to treatment. Panels (c) and (d) display the corresponding targeted types for $\omega = 0$ and $\omega = 1$, respectively, and show substantially less overlap between the optimal policy and the latter compared to the former. Taken together, these results show that targeting based solely on risk scores falls well short of the optimum, as it fails to balance long-term unemployment risk against treatment effectiveness.

We next turn to the timing dimension. The preceding comparison fixes treatment at the start of the spell. In practice, however, the agency may also decide whether to intervene immediately or wait until job seekers have remained unemployed for some time. This introduces a distinct trade-off. Conditional on still being unemployed, the return to treatment may change with duration because baseline job-finding probabilities and treatment effects evolve over the spell. At the same time, delaying treatment reduces the mass of job seekers who remain available for intervention, since some workers exit unemployment before treatment is delivered.

We distinguish between three timing exercises. First, we evaluate the return to treating a fixed share of the surviving unemployed at each duration. This comparison is conditional on survival: at each duration, the agency ranks only those job seekers who remain unemployed. Figure 9 reports the corresponding welfare gains relative to random targeting at the start of the spell. The return to treating surviving job seekers rises over the spell, as also illustrated by the returns to random targeting at different durations d. This reflects the steep decline in baseline job-finding probabilities in the Flemish data, which raises the value of accelerating exit from unemployment, despite adverse selection in treatment responsiveness. This result naturally depends on our baseline assumption that treatment effects do not depreciate at the individual level.[26]

Second, we consider a different exercise in which the agency must spend the same overall budget, but is constrained to treat workers at a common duration d. This comparison incorporates the

[26] For comparison, Appendix Figure F3 shows how the return evolves for a given type.

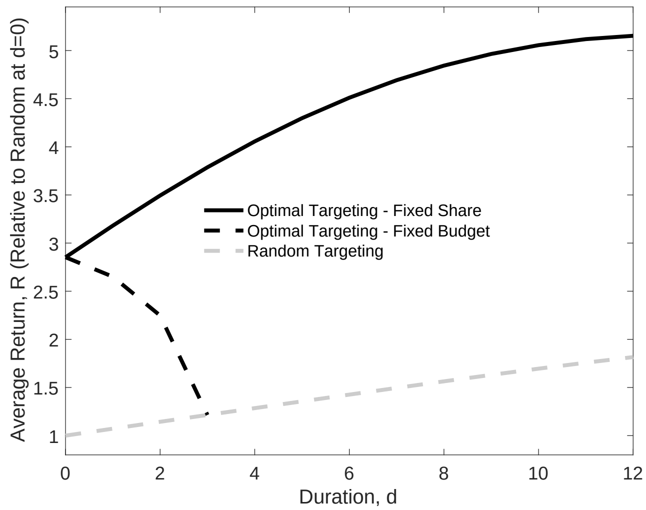


Figure 9: Welfare Gains of Targeting Policies at d, Relative to Random Targeting at $d = 0$

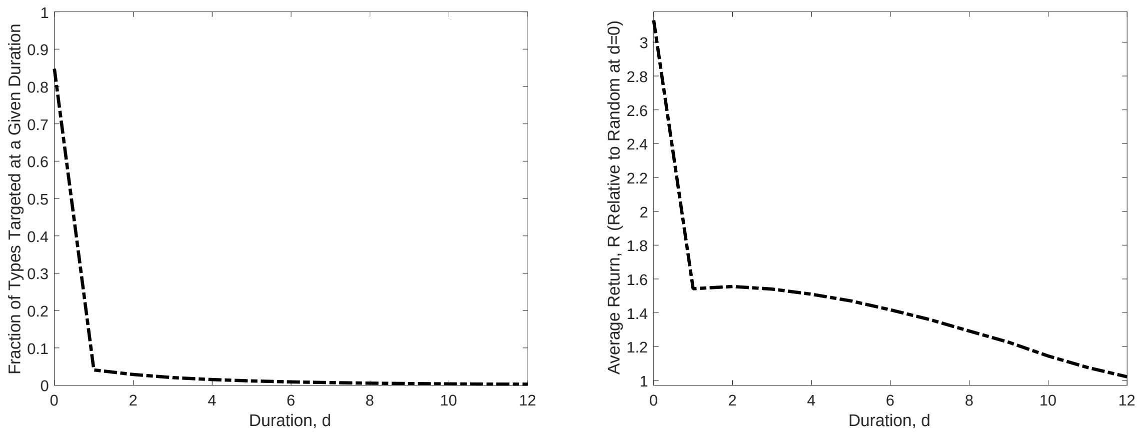
Note: The figure reports the welfare gains from applying three targeting rules: random targeting among survivors, optimal targeting of a fixed share of survivors, and optimal targeting under a fixed budget at unemployment duration d, relative to random targeting at $d = 0$. Welfare gains for the fixed-budget rule are greyed out after month 3, when the optimal policy can no longer exhaust the full budget because too few workers remain unemployed.

shrinking size of the survivor pool. In contrast to the conditional-survivor exercise, welfare gains decline sharply over the first months of the spell. Early in the spell, the unemployed pool is large and heterogeneous, allowing the agency to select many high-return workers. As duration increases, many of these workers have already exited unemployment, and the remaining pool is smaller and less favorable for selective targeting. After month 3, fewer than half of the initial unemployed population remains unemployed, so the agency can no longer exhaust the full budget even by treating all survivors and realizes the average return to treating surviving job seekers.

Finally, we allow the planner to choose jointly whom to treat and when to treat them, subject to the same overall budget and the constraint that each individual can be treated at most once. The optimal policy combines the two forces above. Later treatment can be attractive because conditional returns among survivors are high. But reserving treatment for later also risks losing high-return individuals who exit unemployment before treatment is delivered. The relevant object is therefore not only the conditional return $R_d(\pi, \tau)$, but the survival-weighted value of assigning treatment to a type at duration d, as characterized in Proposition 1. Quantitatively, the optimal policy assigns a much larger share of treatments at the start of the spell and reserves later treatment only for a few selected workers who remain unemployed, as shown in Panel (a) of Figure 10. Panel (b) shows that the workers treated early have higher average returns than those treated later.[27]

Taken together, the calibration clarifies the policy message. Risk scores are informative because

[27] This is not inconsistent with the rising return to treating the average survivor in Figure 9. The global optimum first uses early treatment slots for the highest-return types. Later treatment is then allocated to the best remaining survivor-types for whom waiting is valuable, but these need not have higher returns than the individuals already treated at the start of the spell.



(a) Fraction Targeted at Duration d (b) Average Return of Targeted Types

Figure 10: Optimal Targeting Across Durations

Notes: The figure describes the globally optimal policy, which allocates treatment across job seekers and unemployment durations subject to the aggregate budget constraint. Panel (a) plots the fraction of all targeted job seekers who are treated at each duration d. Panel (b) plots the average return of the types targeted at duration d under this policy, relative to the return of random targeting at $d = 0$.

they predict baseline job-finding probabilities, but they are not satisfactory targets. High-return policies combine information on risk and responsiveness. The timing of treatment depends on two distinct objects: the return to treating job seekers who remain unemployed, and the share of each type that survives to later durations. This explains why returns among survivors can rise with unemployment duration, while the globally optimal policy still treats most high-return job seekers early.

7.3 Algorithmic Bias

We finally use the calibrated model to assess whether the estimated relationship between predicted job-finding prospects and treatment effects could itself reflect the way the risk score is constructed. Because the VDAB score is trained on realized employment outcomes under the existing policy environment, it may partly incorporate the effects of past treatment. As discussed in Section 5, this can mechanically raise predicted job-finding probabilities for groups with high treatment effects and vice versa, generating a positive relationship between prediction and responsiveness.

To quantify this bias, we construct a counterfactual risk score that predicts job finding in the absence of treatment and compare its relationship with treatment effects to that of the observed score. This exercise assumes a specific assignment to treatment underlying the observed scores, as detailed in Appendix F.2. We find that the bias is significant: the regression coefficient of treatment effects on the risk score — the moment targeted in the calibration — falls from 0.53 for the observed score to 0.43 for the treatment-free score. Thus, the observed relationship is amplified by the algorithmic feedback, but not generated by it. This result is robust to different underlying treatment assignments.

The positive relationship between treatment effects and predicted job-finding prospects thus remains strong once the bias is accounted for, so our conclusions are not an artifact of this feedback effect. Note also that the bias itself does not compromise our evaluation of current practice: the VDAB assigns counseling based on the observed risk scores, so both our quasi-experimental estimates and our assessment of score-based targeting apply to the scores actually used for assignment, whether or not these embed the effects of past treatment. Two further remarks are in order. First, the sensitivity analysis in Table 2 below covers the polar case in which the correlation between treatment effects and baseline job-finding prospects is set to zero (column 5) — the case that would apply if the observed relationship were entirely an artifact of algorithmic bias. Removing the correlation naturally improves the performance of risk-based targeting, which then no longer selects against treatment responsiveness, but it leaves the main conclusions intact: risk-only targeting still falls well short of policies that incorporate treatment effects. Second, while an ideal risk score would predict job finding in the absence of treatment, the VDAB does not adjust its risk scores for the effects of past treatment. Such an adjustment would require estimating the prediction algorithm on untreated individuals while controlling for selection into treatment, which is not straightforward in practice.

7.4 Sensitivity Analysis

Table 2 presents a sensitivity analysis with respect to the main calibration parameters. Panel A reports the welfare gains of targeting policies at the start of the unemployment spell. The gains from optimal targeting remain large across all specifications, indicating that the high return to targeting is robust to alternative calibrations. This includes changes in the heterogeneity and/or the dynamics in job-finding probabilities and treatment effects (columns 1-3). The relative welfare gains under the optimal policy are somewhat reduced, but remain sizable, when using caseworker IV-estimates rather than reduced-form treatment effects (column 4) or when the correlation between risk scores and treatment effects is set to zero (column 5). A related concern is that the variance of the estimated CATEs, which disciplines the orthogonal heterogeneity in treatment effects, partly reflects estimation error rather than true heterogeneity in responsiveness. Halving this variance (column 6) reduces the welfare gains from optimal targeting somewhat, but they remain more than twice those of random assignment (2.11 versus 2.85 in the baseline). The gains, however, become even more pronounced when utility losses increase over the spell (column 7) or when the budget is constrained such that only the top 30 percent of individuals can be treated (column 9).

Similarly, the simple index rule targeting τ/π realizes most of the targeting gains across the different specifications and, in fact, all of them in the specification with no depreciation in job finding (column 1). Moreover, targeting solely based on high treatment effects ($\omega = 0$) comes close to the optimal rule across all calibrations, whereas targeting solely based on low risk scores ($\omega = 1$) delivers only about 70 percent of what random targeting does in the baseline calibration. Using the IV-based rather than reduced-form treatment effects reduces the level of the welfare gains somewhat but leaves the comparison across targeting policies unaffected (column 4). Only when the correlation between risk scores and treatment effects is set to zero does targeting only on risk scores perform better and, in fact, outperform the benchmark of random targeting (column 5). Instead, when the budget is restricted to the top 30 percent of individuals (column 9), this policy generates only 23 percent of what random targeting delivers.

Finally, column (10) assesses robustness to redistributive preferences. Our baseline criterion is linear in unemployment losses, so that a given employment gain is valued equally regardless of who receives it. A planner with redistributive preferences would instead assign higher welfare weights to job seekers facing larger expected losses [Haushofer et al., 2025]. Column (10) therefore evaluates all policies and re-solves the optimal ones under log welfare weights that increase with expected unemployment losses. The expected unemployment losses are anchored so that the average unemployed job seeker carries twice the marginal welfare weight of an employed job seeker, consistent with evidence from Landais and Spinnewijn [2021] (see Appendix F.4 for details). The ranking of policies is unchanged: the sufficient-statistic rule still captures almost all of the gains from optimal targeting, risk-score targeting improves — the weights now reward reaching high-loss job seekers — but continues to perform worse than random assignment, and the optimally targeted population overlaps 94 percent with that under the linear criterion. Appendix Table F3 repeats the full sensitivity analysis under the weighted criterion; the main notable difference arises when risk scores and treatment effects are uncorrelated (column 5), in which case risk-score targeting moves closer to the optimal policy.

Panels B and C turn to the timing of interventions. Panel B compares the welfare gains of intervening at three months to those of intervening at entry. When the agency targets a fixed share of the surviving unemployed, the welfare gains relative to the cost of treatment are 33 percent higher at three months than at entry in the baseline calibration (first row), and random targeting exhibits a similar, albeit weaker, increase (third row). The cross-calibration patterns confirm the mechanisms behind these rising returns: the advantage of targeting later vanishes in the stationary environment (column 1) and when treatment effects decline over the spell at the same rate as job finding (column 3) — indeed, in both scenarios random targeting performs *worse* at three months than at entry — while it is amplified when the utility loss from unemployment rises over the spell (column 7) and dampened when continuation losses are truncated after two years (column 8). While the latter assumption is counterfactual, it can be interpreted as an environment in which the policymaker cares only about UI benefit payments, which expire after a fixed duration.

The picture reverses when the agency must exhaust a fixed budget at a single duration (second row): the welfare gains at three months drop to less than half of those at entry in the baseline calibration, because too few workers remain unemployed for the policy to stay selective. The exception is the constrained budget of column (9), where only 30 percent of the initial cohort can be treated: waiting until three months is then almost costless, although we verified that the gains decline more strongly when waiting beyond three months.

Panel C considers the optimal policy when treatment can be targeted across durations. In the baseline calibration, the gains of the global optimum are only marginally larger than those of optimal targeting at entry (2.88 versus 2.85 in Panel A). Indeed, 85 percent of targeted individuals are treated

Table 2: Welfare Gains of Targeting, Across Alternative Calibration Scenarios

	Baseline	(1) $\sigma_{\varepsilon_\pi} = 0$ $\theta_\pi = 1$	(2) $\sigma_{\varepsilon_\tau} > 0$	(3) $\theta_\tau < 1$	(4) $\tau(IV)$	(5) $corr(\pi, \tau)$ $= 0$	(6) Lower σ_{e_τ}	(7) $\theta_u > 1$	(8) At $d = 25$: no contin- uation loss	(9) Target 30%	(10) Welfare weights
Panel A. Targeting at $d = 0$											
Optimal Targeting vs. Random	2.85	2.68	2.61	2.85	1.89	2.17	2.11	2.96	2.46	3.84	6.55
Sufficient Statistics Targeting (τ/π) vs. Random	2.84	2.68	2.58	2.84	1.88	2.16	2.09	2.95	2.46	3.79	6.35
Targeting High Treatment Effects (τ) vs. Random	2.76	2.64	2.49	2.76	1.78	2.11	1.98	2.85	2.43	3.32	5.79
Targeting Low Risk Scores (π) vs. Random	0.70	0.56	0.82	0.70	0.75	1.46	0.61	0.73	0.65	0.23	0.89
Panel B. Targeting at $d > 0$											
Optimal (Fixed Share) at $d = 3$ vs. $d = 0$	1.33	1.01	1.34	1.05	1.29	1.44	1.29	1.42	1.17	1.34	1.55
Optimal (Fixed Budget) at $d = 3$ vs. $d = 0$	0.43	0.59	0.50	0.32	0.65	0.67	0.54	0.45	0.44	0.94	0.25
Random at $d = 3$ vs. $d = 0$	1.21	0.86	1.31	0.92	1.22	1.46	1.14	1.33	1.08	1.21	1.64
Panel C. Targeting Across Durations											
Global Optimal Targeting vs. Random at $d = 0$	2.88	2.68	2.70	2.85	1.94	2.18	2.16	3.01	2.47	4.00	6.58
Fraction Targeted at $d = 0$	0.85	1.00	0.81	0.98	0.72	0.86	0.70	0.80	0.89	0.54	0.85

Notes: This table reports the welfare gains of targeting policies under alternative calibration scenarios. Panel A reports the welfare gains of policies applied at $d = 0$, relative to random targeting at $d = 0$. Panel B reports the ratio of the welfare gains from applying a policy at $d = 3$ to those from applying the same policy at $d = 0$, holding fixed either the share of survivors targeted (fixed share) or the total budget (fixed budget). Panel C reports the welfare gains of the optimal policy across durations, relative to random targeting at $d = 0$, and the fraction of targeted individuals who are treated at $d = 0$. The first column shows the baseline calibration. Column (1) combines $\sigma_{\varepsilon_\pi} = 0$ with $\theta_\pi = 1$. Column (2) allows for unobserved heterogeneity in treatment effects by setting $\sigma_{\varepsilon_\tau} = 0.79\sigma_\tau$, mirroring the relative importance of unobserved and observed heterogeneity for job finding probabilities. Column (3) assumes treatment effects decline at the individual level at the same rate as job finding. Column (4) uses instrumental-variables estimates of treatment effects, τ (IV), instead of reduced-form estimates. Column (5) imposes zero correlation between observed types, $corr(\pi, \tau) = 0$. Column (6) imposes that the variance of the orthogonal heterogeneity in treatment effects is half of the one reported in Table 1. Column (7) assumes that the utility cost of unemployment increases by 50 percent at month 12. Column (8) assumes no continuation loss after period 25 in the model. Column (9) restricts the budget so that only the top 30 percent of individuals can be treated. Column (10) evaluates all policies — re-solving for the optimal ones — under concave (log) social welfare weights that increase with expected unemployment losses, anchored so that the average unemployed job seeker receives twice the marginal welfare weight of an employed job seeker, with weights capped at ten times the employed weight (see Appendix F.4); welfare gains are expressed relative to random targeting under the same weighted criterion.

immediately: the option value of spreading treatment over the spell is small. Consistent with the timing results in Panel B, this option value disappears in the stationary environment of column (1), where all targeted individuals are treated at entry,[28] and is negligible when adding unobserved heterogeneity or depreciation in the treatment effects (columns 2 and 3).

Timing flexibility becomes considerably more valuable, however, when the budget is tight: with only the top 30 percent treated (column 9), the gains rise from 3.84 to 4.00 and just around half of the targeted individuals are treated at entry. Intuitively, a small budget allows the agency to remain selective throughout the spell, and dynamic selection makes waiting attractive: as the spell progresses, survival itself reveals the workers with the poorest job-finding prospects and thus the highest returns to treatment. For similar reasons, targeting later also becomes somewhat more attractive when the utility loss from unemployment rises over the spell (column 7).

Another margin of flexibility concerns the duration of treatment itself, which we have so far restricted to a single period per job seeker. Appendix F.3 relaxes this restriction and shows that allowing the agency to target the same individuals in two consecutive periods further increases the welfare gains from optimal targeting under a fixed budget, as the highest-return job seekers can now be reached more than once. This exercise is similar in nature to restricting the budget to the top 30 percent of individuals (column 9 of Table 2): holding total resources fixed, treating individuals repeatedly means treating fewer of them, so that treatment is concentrated on a more select group of job seekers.

Taken together, the sensitivity analysis confirms the robustness of our main findings. Across all calibration scenarios, targeting rules that combine predicted treatment effects with predicted job-finding probabilities generate welfare gains around two to three times as large as those of random assignment, with the simple index τ/π remaining close to optimal throughout. The timing results are similarly robust: although the returns to treating a given worker rise over the unemployment spell through dynamic selection and duration dependence in job finding, a budget-constrained agency optimally treats the large majority of targeted job seekers at the start of the spell in nearly all scenarios.

8 Conclusion

This paper has examined the optimal targeting of job search counseling in the context of the Flemish Public Employment Service. We have shown that the common practice of directing active labor market interventions toward the most at-risk job seekers (those with the lowest predicted probability of finding employment) is unlikely to be welfare-maximizing. The individuals most at risk are not the same as the individuals most responsive to intervention, and conflating the two leads to a substantial misallocation of scarce program resources.

Our three regression discontinuity designs, exploiting discontinuities in the VDAB's caseworker assignment process at the start of an unemployment spell, at five weeks, and at nineteen weeks, yield two key empirical findings. First, treatment effects are large and positive for low-risk job seekers but effectively zero for those classified as most at risk. Second, there is no evidence of individual-level

[28]This mirrors the special case in Proposition 1, in which all targeting optimally occurs at entry.

depreciation in treatment effects over the unemployment spell, but strong dynamic selection causes the pool of surviving job seekers to be progressively less responsive even as it becomes more at-risk.

These empirical findings motivate our conceptual framework, which characterizes the optimal targeting problem of an unemployment agency that minimizes expected welfare losses subject to a fixed budget. The key insight is that the return to treating a given job seeker depends on both the magnitude of their treatment effect and the continuation loss from unemployment they would otherwise face, with the latter being larger for higher-risk individuals. The optimal targeting index thus involves the ratio of predicted treatment effects to baseline job-finding probabilities, placing positive weight on both responsiveness and risk.

The quantitative evaluation demonstrates that incorporating treatment effect heterogeneity into the targeting rule could nearly triple the welfare gains relative to random assignment. Even a simple rule based on treatment effects alone substantially outperforms the status quo. By contrast, targeting exclusively on risk scores can perform worse than random targeting, because long-term unemployment risk and treatment effects are strongly negatively correlated: those with the lowest predicted job-finding probability are precisely the individuals with the smallest returns to intervention, so that risk-score targeting systematically directs resources away from those who would benefit most. The framework also sheds light on the optimal timing of intervention. Although the return to treating the average surviving job seeker rises over the unemployment spell, delaying treatment means losing the opportunity to reach high-return job seekers who exit in the meantime. This harvesting motive favors earlier intervention, and in our calibrated model most targeted job seekers are optimally treated early in the spell. Employment agencies should therefore invest in identifying responsive individuals early in the spell, when the scope for welfare-improving targeting is greatest.

More broadly, our results highlight a general principle that extends beyond unemployment policy: in settings where individual responsiveness to treatment is heterogeneous and negatively correlated with baseline need, targeting based on risk alone can be counterproductive. This consideration may be particularly important for more intensive and costly active labor market programs, where treatment effects are likely to be highly heterogeneous and the gains from improved targeting correspondingly large. Understanding how best to predict and exploit this heterogeneity remains an important question for future research.

References

Ahn, Hie Joo, "The Role of Observed and Unobserved Heterogeneity in the Duration of Unemployment," *Journal of Applied Econometrics*, 2023, *38* (1), 3–23.

_ **and James D. Hamilton**, "Heterogeneity and Unemployment Dynamics," *Journal of Business & Economic Statistics*, 2020, *38* (3), 554–569.

Altmann, Steffen, Anita Marie Glenny, Robert Mahlstedt, and Alexander Sebald, "The Direct and Indirect Effects of Online Job Search Advice," Technical Report Discussion Paper No. 15830, IZA – Institute of Labor Economics 2022.

— , **Armin Falk, Simon Jäger, and Florian Zimmermann**, "Learning about Job Search: A Field Experiment with Job Seekers in Germany," *Journal of Public Economics*, 2018, *164*, 33–49.

Alvarez, Fernando, Katarína Borovičková, and Robert Shimer, "Decomposing Duration Dependence in a Stopping Time Model," *Review of Economic Studies*, November 2024, *91* (6), 3151–3189.

Athey, Susan, "Beyond Prediction: Using Big Data for Policy Problems," *Science*, 2017, *355* (6324), 483–485.

— **and Stefan Wager**, "Estimating Treatment Effects with Causal Forests: An Application," *Observational Studies*, 2019, *5* (2), 37–51.

— **and** — , "Policy Learning with Observational Data," *Econometrica*, 2021, *89* (1), 133–161.

— , **Julie Tibshirani, and Stefan Wager**, "Generalized Random Forests," *The Annals of Statistics*, 2019, *47* (2), pp. 1148–1178.

— , **Niall Keleher, and Jann Spiess**, "Machine Learning Who to Nudge: Causal vs. Predictive Targeting in a Field Experiment on Student Financial Aid Renewal," *Journal of Econometrics*, 2025, *249*, 105945.

Basu, Sumanta, Karl Kumbier, James B. Brown, and Bin Yu, "Iterative random forests to discover predictive and stable high-order interactions," *Proceedings of the National Academy of Sciences*, 2018, *115* (8), 1943–1948.

Behaghel, Luc, Sofia Dromundo, Marc Gurgand, Yagan Hazard, and Thomas Zuber, "The Potential of Recommender Systems for Directing Job Search: A Large-Scale Experiment," IZA Discussion Paper 16781, IZA Institute of Labor Economics, Bonn 2024.

Belot, Michèle, Philipp Kircher, and Paul Muller, "Do the Long-Term Unemployed Benefit from Automated Occupational Advice during Online Job Search?," *The Economic Journal*, January 2026, *136* (673), 184–206.

Berger, Mark C., Dan A. Black, and Jeffrey A. Smith, "Evaluating Profiling as a Means of Allocating Government Services," in Michael Lechner and Friedhelm Pfeiffer, eds., *Econometric Evaluation of Labour Market Policies*, Vol. 13 of *ZEW Economic Studies*, Heidelberg: Physica-Verlag, 2001, pp. 59–84.

Bied, Guillaume, Solal Nathan, Elia Perennes, Morgane Hoffmann, Philippe Caillou, Bruno Crépon, Christophe Gaillac, and Michèle Sebag, "Toward Job Recommendation for All," in Edith Elkind, ed., *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, International Joint Conferences on Artificial Intelligence Organization 8 2023, pp. 5906–5914. AI for Good.

Black, Dan A., Jeffrey A. Smith, Mark C. Berger, and Brett J. Noel, "Is the Threat of Reemployment Services More Effective Than the Services Themselves? Evidence from Random Assignment in the UI System," *American Economic Review*, 2003, *93* (4), 1313–1327.

Bolhaar, Jonneke, Nadine Ketel, and Bas van der Klaauw, "Caseworker's discretion and the effectiveness of welfare-to-work programs," *Journal of Public Economics*, 2020, *183*, 104080.

Britto, Diogo G. C., Paolo Pinotti, and Breno Sampaio, "The Effect of Job Loss and Unemployment Insurance on Crime in Brazil," *Econometrica*, 2022, *90* (4), 1393–1423.

Bächli, Mirjam, Rafael Lalive, and Michele Pellizzari, "Helping Jobseekers with Recommendations Based on Skill Profiles or Past Experience: Evidence from a Randomized Intervention," IZA Discussion Paper 17713, Institute of Labor Economics (IZA), Bonn Feb 2025.

Card, David, Jochen Kluve, and Andrea Weber, "What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, June 2018, *16* (3), 894–931.

Cockx, Bart, Michael Lechner, and Joost Bollens, "Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium," *Labour Economics*, 2023, *80*, 102306.

Crépon, Bruno and Gerard van den Berg, "Active Labor Market Policies," *Annual Review of Economics*, 2016, *8* (1), 521–546.

Desiere, Sam, Katrin Langenbucher, and Ludo Struyven, "Statistical Profiling in Public Employment Services: An International Comparison," Technical Report 224, OECD Social, Employment and Migration Working Papers, Paris 2019.

Ernst, Sebastian, Andreas I. Mueller, and Johannes Spinnewijn, "Risk Scores for Long-Term Unemployment and the Assignment to Job Search Counseling," *AEA Papers and Proceedings*, May 2024, *114*, 572–576.

Ferey, Antoine, "Redistribution and Unemployment Insurance," Rationality and Competition Discussion Paper Series 345, CRC TRR 190 Rationality and Competition Nov 2022.

Gerard, François and Gustavo Gonzaga, "Informal Labor and the Efficiency Cost of Social Programs: Evidence from Unemployment Insurance in Brazil," *American Economic Journal: Economic Policy*, August 2021, *13* (3), 167–206.

Gregory, Victoria, Guido Menzio, and David Wiczer, "The Alpha Beta Gamma of the Labor Market," *Journal of Monetary Economics*, 2025, *150*, 103695. Labor Market Dynamics: Drivers and Consequences.

Gu, Jiaying and Roger Koenker, "Invidious Comparisons: Ranking and Selection as Compound Decisions," *Econometrica*, 2023, *91* (1), 1–41.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Haushofer, Johannes, Paul Niehaus, Carlos Paramo, Edward Miguel, and Michael Walker, "Targeting Impact versus Deprivation," *American Economic Review*, 2025, *115* (6), 1936–1974.

He, Qiwei and Philipp Kircher, "Updating about Yourself by Learning about the Market: The Dynamics of Beliefs and Expectations in Job Search," Working Paper 31940, National Bureau of Economic Research December 2023.

Humlum, Anders, Jakob R. Munch, and Mette Rasmussen, "What Works for the Unemployed? Evidence from Quasi-Random Caseworker Assignments," NBER Working Paper 33807, National Bureau of Economic Research 2025.

Kasy, Maximilian and Anja Sautmann, "Adaptive Treatment Assignment in Experiments for Policy Choice," *Econometrica*, 2021, *89* (1), 113–132.

Kitagawa, Toru and Aleksey Tetenov, "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 2018, *86* (2), 591–616.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions," *Quarterly Journal of Economics*, 2018, *133* (1), 237–293.

_ , Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer, "Prediction Policy Problems," *American Economic Review*, 2015, *105* (5), 491–495.

Kolsrud, Jonas, Camille Landais, Peter Nilsson, and Johannes Spinnewijn, "The Optimal Timing of Unemployment Benefits: Theory and Evidence from Sweden," *American Economic Review*, April 2018, *108* (4-5), 985–1033.

Koning, Pierre, "The Effectiveness of Public Employment Service Workers in the Netherlands," *Empirical Economics*, October 2009, *37*, 393–409.

Lalive, Rafael, Aderonke Osikominu, Jeremy Zuchuat, Lorenzo Pesaresi, and Josef Zweimüller, "Duration Dependence in Finding a Job: Applications, Interviews, and Job Offers," Working Paper 2026.

Landais, Camille and Johannes Spinnewijn, "The Value of Unemployment Insurance," *Review of Economic Studies*, 2021, *88* (6), 3041–3085.

Le Barbanchon, Thomas, Johannes F. Schmieder, and Andrea Weber, "Job Search, Unemployment Insurance, and Active Labor Market Policies," in Christian Dustmann and Thomas Lemieux, eds., *Handbook of Labor Economics*, Vol. 5, Elsevier, 2024, pp. 435–580.

_ , Lena Hensvik, and Roland Rathelot, "How Can AI Improve Search and Matching? Evidence from 59 Million Personalized Job Recommendations," 2023. SSRN Working Paper No. 4604814.

Lechner, Michael and Jeffrey Smith, "What is the value added by caseworkers?," *Labour Economics*, 2007, *14* (2), 135–151.

_ , Ruth Miquel, and Conny Wunsch, "Long-run Effects of Public Sector Sponsored Training in West Germany," *Journal of the European Economic Association*, 2011, *9* (4), 742–784.

Liu, Songqi, Jing L. Huang, and Mo Wang, "Effectiveness of Job Search Interventions: A Meta-Analytic Review," *Psychological Bulletin*, 2014, *140* (4), 1009–1041.

Manski, Charles F., "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 2004, *72* (4), 1221–1246.

Mueller, Andreas I. and Johannes Spinnewijn, "The Nature of Long-Term Unemployment: Predictability, Heterogeneity, and Selection," *Journal of Political Economy*, 2025, *133* (12), 3846–3902.

_ , _ , and Giorgio Topa, "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias," *American Economic Review*, January 2021, *111* (1), 324–363.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science*, 2019, *366* (6464), 447–453.

Osikominu, Aderonke, "Quick Job Entry or Long-Term Human Capital Development? The Dynamic Effects of Alternative Training Schemes," *The Review of Economic Studies*, 01 2013, *80* (1), 313–342.

Schiprowski, Amelie, Julia Schmidtke, Johannes Schmieder, and Simon Trenkle, "The Effects of Unemployment Insurance Caseworkers on Job Search Effort," *AEA Papers and Proceedings*, May 2024, *114*, 567–71.

Spinnewijn, Johannes, "The Trade-Off between Insurance and Incentives in Differentiated Unemployment Policies," *Fiscal Studies*, 2020, *41* (1), 101–127.

van den Berg, Gerard J. and Johan Vikström, "Monitoring Job Offer Decisions, Punishments, Exit to Work, and Job Quality," *The Scandinavian Journal of Economics*, 2014, *116* (2), 284–334.

_ **and** _ , "Long-Run Effects of Dynamically Assigned Treatments: A New Methodology and an Evaluation of Training Effects on Earnings," *Econometrica*, 2022, *90* (3), 1337–1354.

VDAB, "Jaarverslag 2024," 2024.

Whitehouse, Justin, Qizhao Chen, Morgane Austern, and Vasilis Syrgkanis, "Inference on Optimal Policy Values and Other Irregular Functionals via Softmax Smoothing," 2026. arXiv:2507.11780.

Yadlowsky, Steve, Scott Fleming, Neil Shah, Emma Brunskill, and Stefan Wager, "Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects," *Journal of the American Statistical Association*, 2025, *120* (549), 38–51.

Appendix

A Proofs of Propositions

Proof of Proposition 1.

Proof. Step 1: The agency's problem. The agency operates under a fixed budget $\bar{B}$. Given a per-period targeting cost c, it can treat $\bar{B}/c$ individuals for one period and for one period only. Abstracting from distortionary taxation and the optimal overall size of the program, the agency allocates treatment across unemployed individuals to minimize the expected welfare losses associated with unemployment:

$$\begin{aligned} \min_{T} \quad & \int\int L^T(\pi,\tau)g(\pi,\tau)d\pi d\tau \\ & s.t.\ \bar{B} \geq \int\int B^T(\pi,\tau)g(\pi,\tau)d\pi d\tau, \end{aligned} \tag{A1}$$

where $L^T(\pi,\tau)$ and $B^T(\pi,\tau)$ are the type-level expected utility loss and budgetary cost defined in Section 2. Utility-loss levels u may differ across types and are observable, so that policies may condition on u; we make this dependence explicit in continuation losses, returns, the optimal targeting duration, and the treatment rule below, while leaving the u-dimension implicit in the type distribution $g(\pi,\tau)$ and in the full-spell objects (the budgetary cost B^T does not depend on u).

Step 2: Reformulation in terms of treatment indicators. A policy that treats each type at most once, for one period, is described by indicators $T_d(\pi,\tau) \in \{0,1\}$ with $\sum_{d=0}^{\bar{D}} T_d(\pi,\tau) \leq 1$, where $\bar{D}$ is the maximum duration of unemployment after which no further treatment is allowed. Treatment at duration d leaves survival up to d unaffected (no treatment occurs before d) and changes outcomes only for the $S_d(\pi,\tau)$ survivors: losses incurred before d are policy-invariant, and the loss from d onward changes from $L_d(\pi,u)$ to $L_d^T(\pi,\tau,u)$, where $L_d^T(\pi,\tau,u)$ and $L_d(\pi,u)$ denote the continuation losses from period d onward with and without treatment. Hence

$$L^T(\pi,\tau) \;=\; L(\pi) + \sum_{d=0}^{\bar{D}} T_d(\pi,\tau)\, S_d(\pi,\tau)\big(L_d^T(\pi,\tau,u) - L_d(\pi,u)\big), \qquad B^T(\pi,\tau) \;=\; \sum_{d=0}^{\bar{D}} c\, T_d(\pi,\tau)\, S_d(\pi,\tau),$$

since the treatment cost c is incurred at duration d only for those still unemployed at d. Substituting into (A1) and dropping the policy-invariant baseline loss $\int\int L(\pi)g(\pi,\tau)\, d\pi\, d\tau$ from the objective, the problem becomes

$$\min_{\{T_d\}} \quad \sum_{d=0}^{\bar{D}} \int\int T_d(\pi,\tau)\, S_d(\pi,\tau)(L_d^T(\pi,\tau,u) - L_d(\pi,u))g(\pi,\tau)d\pi d\tau \tag{A2}$$

$$s.t.\ \bar{B} \geq \sum_{d=0}^{\bar{D}} \int\int c\, T_d(\pi,\tau)\, S_d(\pi,\tau) g(\pi,\tau)d\pi d\tau,$$

$$T_d(\pi,\tau) \in \{0,1\}, \quad \sum_{d=0}^{\bar{D}} T_d(\pi,\tau) \leq 1.$$

In this notation, the return to targeting defined in equation (4) satisfies $-c\, R_d^T(\pi,\tau,u) = L_d^T(\pi,\tau,u) - L_d(\pi,u)$.

Step 3: Lagrangian and separation by type. Attach a multiplier $\lambda \geq 0$ to the budget constraint. The contribution of treating type (π,τ) at duration d to the Lagrangian is

$$S_d(\pi,\tau)\, g(\pi,\tau)\Big[\big(L_d^T(\pi,\tau,u) - L_d(\pi,u)\big) + \lambda c\Big] = -\, c\, S_d(\pi,\tau)\, g(\pi,\tau)\big[R_d^T(\pi,\tau,u) - \lambda\big]. \tag{A3}$$

The agency minimizes the Lagrangian, so each treated cell should make (A3) as negative as possible. Because the Lagrangian is an integral of terms that are separable across types, and the exclusivity constraint allows at most one nonzero indicator per type, the problem separates: for each type (π,τ) the agency solves

$$\max_d \quad \underbrace{S_d(\pi,\tau)\big[R_d^T(\pi,\tau,u) - \lambda\big]}_{\text{net welfare gain per cost, weighted by survival}}, \tag{A4}$$

treating at the maximizing duration if the maximized value is positive, and leaving the type untreated otherwise. The per-survivor return R_d^T governs whether a cell is worth treating relative to the shadow cost λ, but the survival rate $S_d(\pi,\tau)$ governs how many individuals that treatment reaches and hence how much aggregate welfare the cell delivers: it is not enough for the return to be high if almost no one survives to that duration to receive treatment. The optimal targeting duration therefore solves

$$d^\star(\pi,\tau,u;\lambda) \in \arg\max_d\ S_d(\pi,\tau)\big[R_d^T(\pi,\tau,u) - \lambda\big], \tag{A5}$$

trading off the survival rate, declining in d, against the net return $R_d^T - \lambda$, which may rise in d as the surviving unemployed become more disadvantaged. Since $S_{d^\star}(\pi,\tau) > 0$, the maximized value in (A4) is positive if and only if the return at the optimal duration exceeds the multiplier:

$$S_{d^\star}(\pi,\tau)\big[R_{d^\star}^T(\pi,\tau,u) - \lambda\big] > 0 \quad \Longleftrightarrow \quad R_{d^\star}^T(\pi,\tau,u) > \lambda. \tag{A6}$$

Step 4: Existence of the threshold and budget clearing. Let $E(\lambda)$ denote the expenditure induced by the policy just described,

$$E(\lambda)\ \equiv\ \int\int c\, S_{d^\star(\pi,\tau,u;\lambda)}(\pi,\tau)\, \mathbf{1}\big\{R_{d^\star}^T(\pi,\tau,u) > \lambda\big\}\, g(\pi,\tau)\, d\pi\, d\tau. \tag{A7}$$

$E(\lambda)$ is nonincreasing in λ for two reasons. First, the treated set shrinks: the maximized value in (A4) is decreasing in λ for every type, so any type left untreated at λ remains untreated at $\lambda' > \lambda$. Second, each treated type weakly delays treatment: for $\lambda' > \lambda$, optimality of $d^\star(\lambda)$ at λ and of $d^\star(\lambda')$ at λ' implies

$$S_{d^\star(\lambda)}\big[R^T_{d^\star(\lambda)} - \lambda\big] \geq S_{d^\star(\lambda')}\big[R^T_{d^\star(\lambda')} - \lambda\big] \quad \text{and} \quad S_{d^\star(\lambda')}\big[R^T_{d^\star(\lambda')} - \lambda'\big] \geq S_{d^\star(\lambda)}\big[R^T_{d^\star(\lambda)} - \lambda'\big];$$

adding the two inequalities yields $(\lambda' - \lambda)\big[S_{d^\star(\lambda)} - S_{d^\star(\lambda')}\big] \geq 0$, so $S_{d^\star(\pi,\tau,u;\lambda)}$ is nonincreasing in λ and each treated type's cost $c\,S_{d^\star}$ weakly falls. Moreover, $E(0)$ is the expenditure of the $\lambda = 0$ policy, which treats every type whose return is positive at some duration. Note that no sign restriction on returns is required: types whose return is nonpositive at every duration fail the criterion in (A6) for any $\lambda \geq 0$ and simply remain untreated. Finally, $E(\lambda) = 0$ for λ above the highest return. If $\bar{B} \geq E(0)$, the budget constraint is slack and the policy with $\lambda^\star = 0$ is optimal. Otherwise, continuity of $g(\pi, \tau)$ ensures that the set of types indifferent between treatment durations or exactly at the threshold has measure zero, so $E(\lambda)$ is continuous, and there exists $\lambda^\star > 0$ such that the budget binds, $E(\lambda^\star) = \bar{B}$. The same measure-zero property implies that the $\geq$ and $>$ cases in the treatment rule coincide, so the threshold $\bar{R} = \lambda^\star$ is well-defined.

Step 5: Global optimality. Let $T^\star$ denote the policy constructed above, which minimizes the Lagrangian pointwise at $\lambda^\star$ and exhausts the budget (complementary slackness: $\lambda^\star\big[E(\lambda^\star) - \bar{B}\big] = 0$). For any feasible policy T, writing $V(T)$ for the objective and $C(T)$ for the expenditure in (A2),

$$V(T^\star) \;=\; V(T^\star) + \lambda^\star\big[C(T^\star) - \bar{B}\big] \;\leq\; V(T) + \lambda^\star\big[C(T) - \bar{B}\big] \;\leq\; V(T),$$

where the first inequality holds because $T^\star$ minimizes the Lagrangian and the second because T is feasible ($C(T) \leq \bar{B}$) and $\lambda^\star \geq 0$. Hence $T^\star$ is optimal, establishing parts (i) and (ii) with $\bar{R} = \lambda^\star$: for each type, the agency selects the single targeting duration that maximizes the survival-weighted net return (A5) — the survival rate enters because the budget is spent on, and welfare accrues to, only those who remain unemployed at the treated duration — and it treats those types whose return at their optimal duration exceeds $\bar{R}$, the return forgone on the marginal treated type.

Step 6: Simplification. Suppose $\sigma_{\varepsilon_\pi} = \sigma_{\varepsilon_\tau} = \sigma_{\varepsilon_u} = 0$, and write $\pi_d \equiv \pi\,\theta_{\pi,d}$ and $\tau_d \equiv \tau\,\theta_{\tau,d}$ for the baseline job-finding probability and the treatment effect at duration d. Continuation losses then satisfy the recursions

$$L_d(\pi, u) = u_d + (1 - \pi_d)\,L_{d+1}(\pi, u), \qquad L^T_d(\pi, \tau, u) = u_d + (1 - \pi_d - \tau_d)\,L_{d+1}(\pi, u),$$

since a one-period treatment at d raises the exit hazard at d by τ_d and leaves hazards from $d+1$ onward unchanged. Subtracting, $L_d(\pi, u) - L^T_d(\pi, \tau, u) = \tau_d\,L_{d+1}(\pi, u)$ exactly, so $R^T_d = (\tau_d/c)\,L_{d+1}(\pi, u)$. With $\theta_{\pi,d} = \theta_{u,d} = 1$ for all d, the continuation loss is stationary and solves $L = u + (1 - \pi)L$, i.e., $L_{d+1}(\pi, u) = u/\pi$ for all d; with $\theta_{\tau,d} = 1$ the treatment effect is constant, $\tau_d = \tau$. Hence

$$R^T_d(\pi, \tau, u) \;=\; \frac{u}{c} \cdot \frac{\tau}{\pi} \qquad \text{for all } d,$$

and the threshold rule in part (i) reduces to a threshold rule in $u\,\tau/\pi$, and in τ/π when utility losses are identical across types. Since $R^T_d - \bar{R}$ is then constant in d while $S_d(\pi) = (1-\pi)^d$ is strictly decreasing, the survival-weighted excess return of every treated type is uniquely maximized at $d^\star = 0$: with returns constant over the spell, delaying treatment only shrinks the group that can be reached. $\square$

Proof of proposition 2. The proof of Proposition 2 proceeds in two steps. We first establish the result in the absence of heterogeneity, where the argument is straightforward. We then turn to the case with unobserved heterogeneity, which is considerably more involved.

Proof of proposition 2 without unobserved heterogeneity.

Proof. Suppose $\sigma_{\varepsilon_\pi} = \sigma_{\varepsilon_\tau} = \sigma_{\varepsilon_u} = 0$, so that $\pi_d = \pi\,\theta_\pi^d$, $\tau_d = \tau\,\theta_\tau^d$, and $u_d = u\,\theta_u^d$, and survivor averages coincide with individual characteristics. As in Step 6 of the proof of Proposition 1, a one-period treatment at duration d raises the exit hazard at d by τ_d and leaves hazards from $d+1$ onward unchanged, so the return is exactly

$$R^T_d \;=\; \frac{\tau_d}{c}\,L_{d+1}(\pi,u), \qquad L_{d+1}(\pi,u) \;=\; \sum_{\delta=d+1}^{\infty} u\,\theta_u^\delta \prod_{j=d+1}^{\delta-1}\bigl(1-\pi\,\theta_\pi^j\bigr),$$

where L_{d+1} is the continuation loss from period $d+1$ onward, linear in u.

Part (i). At $\theta_\pi = \theta_u = 1$ the continuation loss equals the per-period loss times the expected remaining duration of a constant-hazard spell, $L_{d+1}(\pi,u) = u\sum_{k=0}^{\infty}(1-\pi)^k = u/\pi$, so that

$$R^T_d \;=\; \frac{\tau\,\theta_\tau^d}{c}\cdot\frac{u}{\pi} \qquad\text{and}\qquad \frac{R^T_d(\pi',\tau',u')}{R^T_d(\pi,\tau,u)} \;=\; \frac{\tau'\,u'/\pi'}{\tau\,u/\pi},$$

since the factor θ_τ^d/c is common to both types. Log-linearizing yields part (i) with individual characteristics: because the factors θ_π^d, θ_τ^d, and θ_u^d are common to both types at the same duration, relative differences in (π_d,τ_d,u_d) coincide with relative differences in (π,τ,u). Away from $\theta_\pi = \theta_u = 1$, the ratio of continuation losses across types varies smoothly with (θ_π,θ_u), adding type-specific corrections of order $(\theta_\pi - 1)$ and $(\theta_u - 1)$, consistent with the neighborhood qualifier in the proposition.

Part (ii). Shifting the summation index by one period yields an exact self-similarity of the continuation loss:

$$L_{d+2}(\pi,u) = \sum_{\delta=d+1}^{\infty} u\,\theta_u^{\delta+1} \prod_{j=d+1}^{\delta-1}\bigl(1-(\theta_\pi\pi)\,\theta_\pi^j\bigr) = \theta_u\,L_{d+1}(\theta_\pi\pi,u):$$

one period further into the spell, the remaining problem is identical except that utility losses are scaled by θ_u and the entire hazard path is scaled by θ_π. Hence, exactly,

$$\frac{R^T_{d+1}}{R^T_d} = \theta_\tau\cdot\frac{L_{d+2}(\pi,u)}{L_{d+1}(\pi,u)} = \theta_\tau\,\theta_u\,\frac{L_{d+1}(\theta_\pi\pi,u)}{L_{d+1}(\pi,u)}.$$

To first order in $(\theta_\pi - 1)$, the last factor is governed by the elasticity of the continuation loss with respect to the hazard, which equals -1 at the benchmark, where $L_{d+1} = u/\pi$:

$$\frac{L_{d+1}(\theta_\pi \pi, u)}{L_{d+1}(\pi, u)} = 1 + (\theta_\pi - 1)\, \frac{\pi\, \partial_\pi L_{d+1}(\pi, u)}{L_{d+1}(\pi, u)} + O\big((\theta_\pi - 1)^2\big) \;\approx\; 2 - \theta_\pi.$$

Therefore

$$\frac{R^T_{d+1}}{R^T_d} \;\approx\; \theta_\tau\, \theta_u\, (2 - \theta_\pi) \;\approx\; \theta_\tau \big(1 + \theta_u - \theta_\pi\big),$$

where the last step drops $(\theta_u - 1)(1 - \theta_\pi)$, second order in $(\theta_\pi - 1, \theta_u - 1)$. Subtracting one and dropping the cross term $(\theta_\tau - 1)(\theta_u - \theta_\pi)$ yields

$$\frac{R^T_{d+1} - R^T_d}{R^T_d} \;\approx\; (\theta_\tau - 1) - (\theta_\pi - 1) + (\theta_u - 1) \;=\; \frac{\tau_{d+1} - \tau_d}{\tau_d} - \frac{\pi_{d+1} - \pi_d}{\pi_d} + \frac{u_{d+1} - u_d}{u_d},$$

which is part (ii) with individual characteristics in place of survivor averages. □

Proof of proposition 2 with unobserved heterogeneity.

Proof. Write $p \equiv \pi + \varepsilon_\pi$, $t \equiv \tau + \varepsilon_\tau$, and $w \equiv u + \varepsilon_u$, so that $P_d = p\, \theta_\pi^d$, $\tau_d = t\, \theta_\tau^d$, and $u_d = w\, \theta_u^d$, and let

$$L_{d+1}(p, w) \;=\; w\, \ell_{d+1}(p), \qquad \ell_{d+1}(p) \;\equiv\; \sum_{\delta=d+1}^{\infty} \theta_u^\delta \prod_{j=d+1}^{\delta-1} \big(1 - p\, \theta_\pi^j\big),$$

denote the individual continuation loss from period $d+1$ onward, which is linear in the individual loss level w, with $\ell_{d+1}(p)$ the continuation loss per unit of w. Survivor expectations are $\mathbb{E}_d[X] \;\equiv\; \mathbb{E}[X\, S_d(\pi, \varepsilon_\pi)]/\mathbb{E}[S_d(\pi, \varepsilon_\pi)]$, where the expectation is over $(\varepsilon_\pi, \varepsilon_\tau, \varepsilon_u)$ and the survival weights depend on ε_π only, with Var_d and Cov_d the corresponding variance and covariance. Throughout, "dispersion" refers to the standard deviations of $(\varepsilon_\pi, \varepsilon_\tau, \varepsilon_u)$ among survivors, and orders of approximation are joint in $(\theta_\pi - 1)$, $(\theta_u - 1)$, and the dispersion. Orders are understood pointwise in d: the constants in the error terms may grow with the duration (e.g., through $c_d(p)$ defined below), so the approximations hold for given durations rather than uniformly over the spell.

Step 1: The type-level return. Since targeting conditions only on observables, treating a type at duration d treats all its surviving ε-realizations. A one-period treatment at d raises the individual exit hazard at d by $\tau_d = t\, \theta_\tau^d$ and leaves hazards from $d+1$ onward unchanged, so it lowers the individual continuation loss by $t\, \theta_\tau^d\, w\, \ell_{d+1}(p)$. The type-level return defined in equation (4) — with the continuation losses L^T_d and L_d averaged over $(\varepsilon_\pi, \varepsilon_\tau, \varepsilon_u)$ among survivors at duration d — therefore equals

$$R^T_d \;=\; \frac{\theta_\tau^d\, \mathbb{E}\big[t\, w\, S_d\, \ell_{d+1}(p)\big]}{c\, \mathbb{E}[S_d]} \;=\; \frac{\theta_\tau^d}{c}\, \mathbb{E}_d\big[t\, w\, \ell_{d+1}(p)\big]. \tag{A8}$$

Step 2: Dynamic selection. Untreated survival satisfies $S_{d+1} = S_d\, (1 - P_d)$, so survivor expectations tilt as $\mathbb{E}_{d+1}[X] = \mathbb{E}_d[(1-P_d)\,X]/\mathbb{E}_d[1-P_d]$. Using $P_{d+1} = \theta_\pi\, P_d$,

$$\bar{P}_{d+1} = \theta_\pi \frac{\mathbb{E}_d[P_d(1-P_d)]}{1-\bar{P}_d} = \theta_\pi \frac{\bar{P}_d - \bar{P}_d^2 - \mathrm{Var}_d(P_d)}{1-\bar{P}_d} = \theta_\pi\, \bar{P}_d \left[1 - \frac{\mathrm{Var}_d(P_d)}{\bar{P}_d(1-\bar{P}_d)}\right],$$

an exact dynamic-selection identity: the observed change $\bar{\theta}_{\pi,d}$ compounds true duration dependence (θ_π) and dynamic selection (the variance term). Analogously, since $\bar{\tau}_d \equiv \mathbb{E}_d[\tau_d] = \theta_\tau^d\, \mathbb{E}_d[t]$ and $\bar{u}_d \equiv \mathbb{E}_d[u_d] = \theta_u^d\, \mathbb{E}_d[w]$, the observed changes in treatment effects and utility losses among survivors satisfy $\bar{\theta}_{\tau,d} \equiv \bar{\tau}_{d+1}/\bar{\tau}_d = \theta_\tau\, \mathbb{E}_{d+1}[t]/\mathbb{E}_d[t]$ and $\bar{\theta}_{u,d} \equiv \bar{u}_{d+1}/\bar{u}_d = \theta_u\, \mathbb{E}_{d+1}[w]/\mathbb{E}_d[w]$.

Step 3: Expansion of the individual continuation loss. Setting $k = \delta - d - 1$,

$$\ell_{d+1}(p) = \sum_{k=0}^{\infty} \theta_u^{d+k+1} \prod_{j=1}^{k} \big(1 - \theta_\pi^{d+j}\, p\big),$$

which equals $\sum_k (1-p)^k = 1/p$ at $\theta_\pi = \theta_u = 1$. For the θ_π-derivative, the $k=0$ term is independent of θ_π; for $k \geq 1$, differentiating the j-th factor gives $-(d+j)\theta_\pi^{d+j-1}p$, while the remaining $k-1$ factors each equal $(1-p)$ at $\theta_\pi = 1$ and the prefactor θ_u^{d+k+1} equals one at $\theta_u = 1$. Hence

$$\left.\frac{\partial \ell_{d+1}}{\partial \theta_\pi}\right|_{\theta_\pi=\theta_u=1} = -p\sum_{k=1}^{\infty}(1-p)^{k-1}\sum_{j=1}^{k}(d+j) = -\sum_{j=1}^{\infty}(d+j)(1-p)^{j-1} = -\frac{dp+1}{p^2},$$

where the second equality swaps the order of summation and evaluates the inner geometric sum, $\sum_{k=j}^{\infty}(1-p)^{k-1} = (1-p)^{j-1}/p$, and the third re-indexes $m = j-1$ and uses $\sum_{m=0}^{\infty}(1-p)^m = 1/p$ and $\sum_{m=0}^{\infty} m(1-p)^m = (1-p)/p^2$. For the θ_u-derivative, differentiating the prefactor θ_u^{d+k+1} gives $(d+k+1)$ at $\theta_u = 1$, so

$$\left.\frac{\partial \ell_{d+1}}{\partial \theta_u}\right|_{\theta_\pi=\theta_u=1} = \sum_{k=0}^{\infty}(d+k+1)(1-p)^k = \frac{dp+1}{p^2} = -\left.\frac{\partial \ell_{d+1}}{\partial \theta_\pi}\right|_{\theta_\pi=\theta_u=1}.$$

Hence, to first order in $(\theta_\pi - 1)$ and $(\theta_u - 1)$,

$$\ell_{d+1}(p) \;\approx\; \frac{1}{p} + c_d(p)\left[(\theta_u - 1) - (\theta_\pi - 1)\right], \qquad c_d(p) \equiv \frac{dp+1}{p^2}, \tag{A9}$$

and the key property, holding individual by individual, is that the coefficient grows by

$$c_{d+1}(p) - c_d(p) \;=\; \frac{1}{p}$$

across consecutive durations.

Step 4: Part (i), returns across types. Fix a duration d and evaluate to zeroth order in $(\theta_\pi - 1)$ and $(\theta_u - 1)$, so that $\ell_{d+1}(p) = 1/p$. Writing $P_d = \bar{P}_d(1+x)$ with $\mathbb{E}_d[x] = 0$ and expanding, $\mathbb{E}_d[1/P_d] = \big(1+\mathrm{CV}_d^2\big)/\bar{P}_d + O\big(\mathbb{E}_d[|x|^3]\big)$ with $\mathrm{CV}_d^2 \equiv \mathrm{Var}_d(P_d)/\bar{P}_d^2$, so that by (A8), using $1/p = \theta_\pi^d/P_d$ and

$\mathbb{E}_d[w] = \bar{u}_d/\theta_u^d$,

$$R_d^T = \frac{\theta_\tau^d}{c}\,\mathbb{E}_d[t]\,\mathbb{E}_d[w]\,\mathbb{E}_d[1/p]\,\big(1+\rho_d\big) = \frac{\bar{\tau}_d\,\bar{u}_d\,\theta_\pi^d}{c\,\bar{P}_d\,\theta_u^d}\,\big(1+\mathrm{CV}_d^2\big)\big(1+\rho_d\big) + O\big(\mathbb{E}_d[|x|^3]\big),$$

where $1+\rho_d \equiv \mathbb{E}_d[t\,w/p]/\big(\mathbb{E}_d[t]\,\mathbb{E}_d[w]\,\mathbb{E}_d[1/p]\big)$, so that ρ_d collects the pairwise survivor covariances between t, w, and $1/p$ (second order in the dispersion) and their third cross-moment (third order). Taking the ratio across two types at the same duration — the factor $\theta_\pi^d/(c\,\theta_u^d)$ is common to both —

$$\frac{R_d^T(\pi',\tau',u')}{R_d^T(\pi,\tau,u)} = \frac{\bar{\tau}_d'\,\bar{u}_d'/\bar{P}_d'}{\bar{\tau}_d\,\bar{u}_d/\bar{P}_d}\cdot\frac{\big(1+\mathrm{CV}_d'^2\big)\big(1+\rho_d'\big)}{\big(1+\mathrm{CV}_d^2\big)\big(1+\rho_d\big)}.$$

The survivor dispersion corrections CV_d^2 and ρ_d are second order in the dispersion and, unlike in part (ii) below, do not difference out across types, so the second factor is $1+O(\text{dispersion}^2)$. Log-linearizing the first factor yields part (i); reinstating the first-order terms in (A9) adds type-specific corrections of order $(\theta_\pi-1)$ and (θ_u-1), consistent with the neighborhood qualifier in the proposition.

Step 5: Part (ii), composition channel. We first track how the survivor average of $1/p$ drifts across durations. By the tilting identity and $P_d/p = \theta_\pi^d$,

$$\mathbb{E}_{d+1}[1/p] = \frac{\mathbb{E}_d[(1-P_d)/p]}{1-\bar{P}_d} = \frac{\mathbb{E}_d[1/p]-\theta_\pi^d}{1-\bar{P}_d},$$

so that, dividing by $\mathbb{E}_d[1/p]$ and using $\mathbb{E}_d[1/p] = \theta_\pi^d\,\mathbb{E}_d[1/P_d]$,

$$\frac{\mathbb{E}_{d+1}[1/p]}{\mathbb{E}_d[1/p]} = \frac{1-H_d}{1-\bar{P}_d}, \qquad H_d \equiv \frac{1}{\mathbb{E}_d[1/P_d]},$$

exactly, where H_d is the harmonic mean of the job-finding probability among survivors. Expanding as in Step 4, $H_d = \bar{P}_d\big(1-\mathrm{CV}_d^2\big) + O\big(\mathbb{E}_d[|x|^3]\big)$, so

$$\frac{\mathbb{E}_{d+1}[1/p]}{\mathbb{E}_d[1/p]} = 1 + \frac{\mathrm{Var}_d(P_d)}{\bar{P}_d(1-\bar{P}_d)} + O\big(\mathbb{E}_d[|x|^3]\big) = 1 + \Big(1-\frac{\bar{\theta}_{\pi,d}}{\theta_\pi}\Big) + O\big(\mathbb{E}_d[|x|^3]\big),$$

where the last equality uses the dynamic-selection identity of Step 2: the second-order composition effect is absorbed exactly by the observed change $\bar{\theta}_{\pi,d}$, and the error is third order (bounded by the third absolute moment of the survivor distribution of p).

Step 6: Part (ii), duration-dependence channel. Taking survivor expectations of (A9) at durations d and $d+1$,

$$\mathbb{E}_d[\ell_{d+1}] \approx \mathbb{E}_d[1/p] + (\theta_u-\theta_\pi)\,\mathbb{E}_d[c_d(p)], \qquad \mathbb{E}_{d+1}[\ell_{d+2}] \approx \mathbb{E}_{d+1}[1/p] + (\theta_u-\theta_\pi)\,\mathbb{E}_{d+1}[c_{d+1}(p)],$$

so that, dividing,

$$\frac{\mathbb{E}_{d+1}[\ell_{d+2}]}{\mathbb{E}_d[\ell_{d+1}]} \;\approx\; \frac{\mathbb{E}_{d+1}[1/p]}{\mathbb{E}_d[1/p]}\left[1+(\theta_u-\theta_\pi)\left(\frac{\mathbb{E}_{d+1}[c_{d+1}(p)]}{\mathbb{E}_{d+1}[1/p]}-\frac{\mathbb{E}_d[c_d(p)]}{\mathbb{E}_d[1/p]}\right)\right].$$

Since $c_d(p) = d/p + 1/p^2$, the coefficient ratio equals $d + \mathbb{E}_d[1/p^2]/\mathbb{E}_d[1/p]$, so its change across consecutive durations is

$$1 + \left(\frac{\mathbb{E}_{d+1}[1/p^2]}{\mathbb{E}_{d+1}[1/p]} - \frac{\mathbb{E}_d[1/p^2]}{\mathbb{E}_d[1/p]} \right) \;=\; 1 + O(\text{dispersion}^2) :$$

the d-dependent terms cancel exactly — the survivor-average counterpart of $c_{d+1}(p) - c_d(p) = 1/p$ — and the survivor drift of $\mathbb{E}[1/p^2]/\mathbb{E}[1/p]$ is second order in the dispersion, since tilting shifts survivor moments by covariance terms. Combining with Step 5 and dropping products of a second-order dispersion term with $(\theta_u - \theta_\pi)$ — third order jointly —

$$\frac{\mathbb{E}_{d+1}[\ell_{d+2}]}{\mathbb{E}_d[\ell_{d+1}]} \;\approx\; 1 + \left(1 - \frac{\bar{\theta}_{\pi,d}}{\theta_\pi}\right) + (\theta_u - \theta_\pi) \;\approx\; 1 + \theta_u - \bar{\theta}_{\pi,d},$$

where the last step replaces $1 - \bar{\theta}_{\pi,d}/\theta_\pi$ by $\theta_\pi - \bar{\theta}_{\pi,d}$, adding $(\theta_\pi - \bar{\theta}_{\pi,d})(\theta_\pi - 1)/\theta_\pi$ — second order in the dispersion times first order in $(\theta_\pi - 1)$, hence third order jointly. Finally, the expansion (A9) itself carries a Taylor remainder that is second order in $(\theta_\pi - 1, \theta_u - 1)$, and unlike the first-order coefficients, whose duration-differences cancel to $1/p$, its coefficients do not cancel across consecutive durations. The resulting deterministic error is second order in $(\theta_\pi - 1, \theta_u - 1)$ and proportional to $(\theta_\pi - 1)$: it vanishes identically at $\theta_\pi = 1$, where $\ell_{d+1}(p) = \theta_u^{d+1}/(1 - \theta_u(1-p))$ and the ratio equals θ_u exactly (see the Remark below); a direct second-order expansion shows it is of order $(\theta_\pi - 1)(\theta_\pi - \theta_u)$. Hence all error terms involving the unobserved heterogeneity are third order jointly, while the deterministic error is second order in $(\theta_\pi - 1, \theta_u - 1)$.

Step 7: Part (ii), heterogeneity in treatment effects and utility losses. Write

$$\begin{aligned}\mathbb{E}_d[t\,w\,\ell_{d+1}] = \mathbb{E}_d[t]\,\mathbb{E}_d[w]\,\mathbb{E}_d[\ell_{d+1}] + \mathbb{E}_d[\ell_{d+1}]\,\mathrm{Cov}_d(t,w) + \mathbb{E}_d[w]\,\mathrm{Cov}_d(t,\ell_{d+1}) \\ + \mathbb{E}_d[t]\,\mathrm{Cov}_d(w,\ell_{d+1}) + \mu_{3,d},\end{aligned}$$

where $\mu_{3,d}$ denotes the third central cross-moment of (t, w, ℓ_{d+1}). Relative to the product of the means, the covariance terms are second order in the dispersion and $\mu_{3,d}$ is third order. The changes of the covariance terms across consecutive durations are third order jointly through two channels: tilting the survivor measure changes a covariance by third central moments of the survivor distribution, while the shift of the function from ℓ_{d+1} to ℓ_{d+2} changes it by a covariance with $\ell_{d+2} - \ell_{d+1}$, which is second order in the dispersion times first order in the θ-deviations, since $\ell_{d+2}(p) - \ell_{d+1}(p) \approx (\theta_u - \theta_\pi)/p$ by (A9). Hence the covariance corrections cancel from the ratio to the stated order. Using (A8), $\bar{\theta}_{\tau,d} = \theta_\tau\,\mathbb{E}_{d+1}[t]/\mathbb{E}_d[t]$ and $\bar{\theta}_{u,d} = \theta_u\,\mathbb{E}_{d+1}[w]/\mathbb{E}_d[w]$ from Step 2, and Step 6,

$$\begin{aligned}\frac{R^T_{d+1}}{R^T_d} = \theta_\tau \cdot \frac{\mathbb{E}_{d+1}[t\,w\,\ell_{d+2}]}{\mathbb{E}_d[t\,w\,\ell_{d+1}]} \;&\approx\; \theta_\tau \cdot \frac{\mathbb{E}_{d+1}[t]}{\mathbb{E}_d[t]} \cdot \frac{\mathbb{E}_{d+1}[w]}{\mathbb{E}_d[w]} \cdot \frac{\mathbb{E}_{d+1}[\ell_{d+2}]}{\mathbb{E}_d[\ell_{d+1}]} \\ &\approx\; \bar{\theta}_{\tau,d} \cdot \frac{\bar{\theta}_{u,d}}{\theta_u}\left(1 + \theta_u - \bar{\theta}_{\pi,d}\right) \;\approx\; \bar{\theta}_{\tau,d}\left(1 + \bar{\theta}_{u,d} - \bar{\theta}_{\pi,d}\right),\end{aligned}$$

which is exact in θ_τ. The last approximation replaces $(\bar{\theta}_{u,d}/\theta_u)\big(1 + \theta_u - \bar{\theta}_{\pi,d}\big)$ by $1 + \bar{\theta}_{u,d} - \bar{\theta}_{\pi,d}$, an error of $\big(\bar{\theta}_{u,d}/\theta_u - 1\big)\big(1 - \bar{\theta}_{\pi,d}\big)$: by the tilting identity, $\bar{\theta}_{u,d}/\theta_u - 1 = -\mathrm{Cov}_d(w, P_d)/\big(\mathbb{E}_d[w](1 - \bar{P}_d)\big)$

is second order in the dispersion, while $1 - \bar{\theta}_{\pi,d}$ is first order in $(\theta_\pi - 1)$ plus second order in the dispersion, so the error is third order jointly. As with treatment effects, the survivor drift of utility losses is thus absorbed by the observed change $\bar{\theta}_{u,d}$ rather than adding to the approximation error. Subtracting one and dropping the cross term $(\bar{\theta}_{\tau,d} - 1)(\bar{\theta}_{u,d} - \bar{\theta}_{\pi,d})$ — second order only for θ_τ close to one, so that the difference form additionally requires θ_τ near one — yields

$$\frac{R^T_{d+1} - R^T_d}{R^T_d} \;\approx\; \big(\bar{\theta}_{\tau,d} - 1\big) - \big(\bar{\theta}_{\pi,d} - 1\big) + \big(\bar{\theta}_{u,d} - 1\big),$$

which is part (ii), since each term is the proportional change of the corresponding survivor average across consecutive durations, e.g., $\bar{\theta}_{\tau,d} - 1 = \big(\mathbb{E}_{d+1}[\tau_{d+1}] - \mathbb{E}_d[\tau_d]\big)/\mathbb{E}_d[\tau_d]$.

Step 8: Nesting. Without unobserved heterogeneity the survivor distribution is degenerate, so $\bar{P}_d = \pi_d$, $\bar{\tau}_d = \tau_d$, $\bar{u}_d = u_d$, $\bar{\theta}_{\pi,d} = \theta_\pi$, $\bar{\theta}_{\tau,d} = \theta_\tau$, $\bar{\theta}_{u,d} = \theta_u$, all dispersion terms vanish, and parts (i) and (ii) reduce to their deterministic counterparts. If ε_τ and ε_u are mean-independent of ε_π, then $\mathbb{E}_d[t] = \mathbb{E}[t]$ and $\mathbb{E}_d[w] = \mathbb{E}[w] = u$ do not vary with d, since the survival weights depend on ε_π only, so $\bar{\theta}_{\tau,d} = \theta_\tau$ and $\bar{\theta}_{u,d} = \theta_u$; in this case $\mathbb{E}_d[u_d] = u\,\theta_u^d$, recovering the formulation in which utility losses vary only observably. $\square$

Remark. When $\theta_\pi = 1$, the continuation loss has the closed form $L_{d+1}(p,w) = w\,\theta_u^{d+1}\big/\big(1-\theta_u(1-p)\big)$, so that the type-level return is, exactly,

$$R^T_d \;=\; \frac{\theta_u^{d+1}\,\theta_\tau^d}{c}\,\mathbb{E}_d\!\left[\frac{t\,w}{1-\theta_u(1-p)}\right]:$$

at the individual level, the return to treatment is proportional to $t\,w\big/\big(1-\theta_u(1-p)\big)$, the treatment effect times the utility-loss level, scaled by the inverse of the effective exit rate from the loss-weighted spell. If in addition the product $t\,w$ is mean-independent of ε_π — as when $(\varepsilon_\tau, \varepsilon_u)$ is independent of ε_π — an exact version of part (ii) follows: defining the generalized harmonic mean $\tilde{H}_d \equiv 1\big/\mathbb{E}_d\big[\big(1-\theta_u(1-p)\big)^{-1}\big]$, we have $R^T_{d+1}/R^T_d = \theta_\tau\,\theta_u\,\tilde{H}_d/\tilde{H}_{d+1}$ exactly, for any distribution of ε_π; at $\theta_u = 1$, $\tilde{H}_d$ reduces to the harmonic-mean job-finding probability among survivors, $1/\mathbb{E}_d[1/P_d]$. The arithmetic-mean statement in part (ii) approximates this exact statistic and is the empirically more convenient version.

B Data

B.1 Caseworker Assignment

Table B1: Caseworker Assignment

	N	Share
Full sample	149,629	1.00
Inbound call in weeks 1–3	25,808	0.17
Assigned to caseworker	11,894	0.08
Caseworker contact	10,811	0.07
Inbound call in weeks 4–5	30,764	0.21
Self-reliant	19,284	0.13
Assigned to caseworker	11,480	0.08
Caseworker contact	9,861	0.07
Considered for outbound call at 5 weeks	44,709	0.30
Prediction $>$ 65%	13,406	0.09
Prediction $<$ 65%	31,303	0.21
Calls in weeks 6–17	27,681	0.18
Self-reliant	8,628	0.06
Assigned to caseworker	15,544	0.10
Caseworker contact	13,004	0.09
Considered for call at 5 weeks & not assigned to caseworker by week 18	11,783	0.08
Calls in weeks 18–31	7,130	0.05
Self-reliant	2,160	0.01
Assigned to caseworker	4,265	0.03
Caseworker contact	3,405	0.02

Note: This table reports the number and share of unemployment spells for new unemployment spells occurring between 3 October 2020 and 9 September 2022. Individuals immediately reassigned to a caseworker from a previous spell, aged 60 and older or engaging in subsidized or temporary work by week 5 are excluded. The sample for "inbound call in weeks 4–5" is further restricted to spells lasting at least 3 weeks. The sample for "Considered for outbound call at 5 weeks" excludes job seekers who exited unemployment within 5 weeks, those who were assigned to a caseworker in weeks 1–3, those who made an inbound call in weeks 4–5 and those who were assigned to a caseworker through another mechanism. "Calls in weeks 6–17" includes both inbound and outbound calls. "Caseworker contact" refers to any contact with a caseworker at a regional office by week 17 or, for the last row, in weeks 18–31. "Considered for call at 5 weeks & not assigned to caseworker by week 18" also excludes job seekers who have exited unemployment by week 18. Note, differences arising between this table and Table 1 of Ernst et al. [2024] are due to job seekers younger than 26 and job seekers without a recorded risk score from the first week of unemployment being included in this table as well as due to the use of more comprehensive data on calls with the service line.

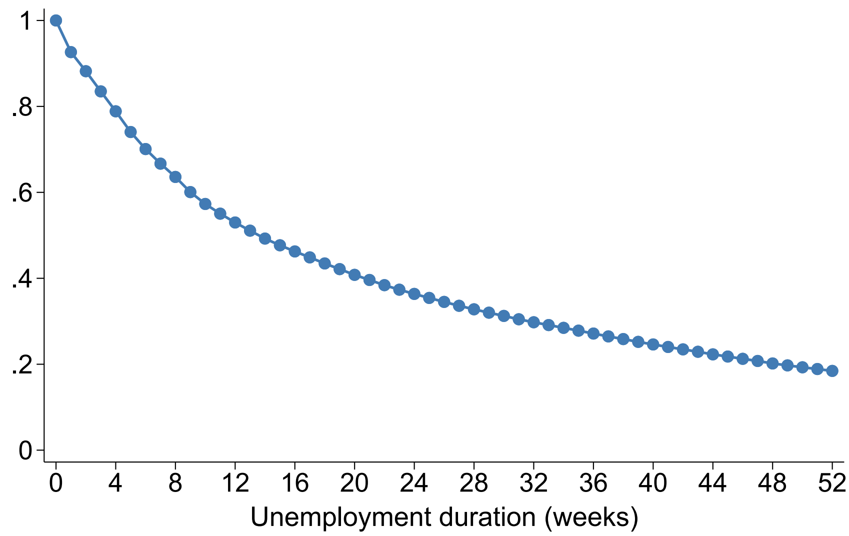
B.2 Descriptive Statistics

Table B2: Descriptive Statistics

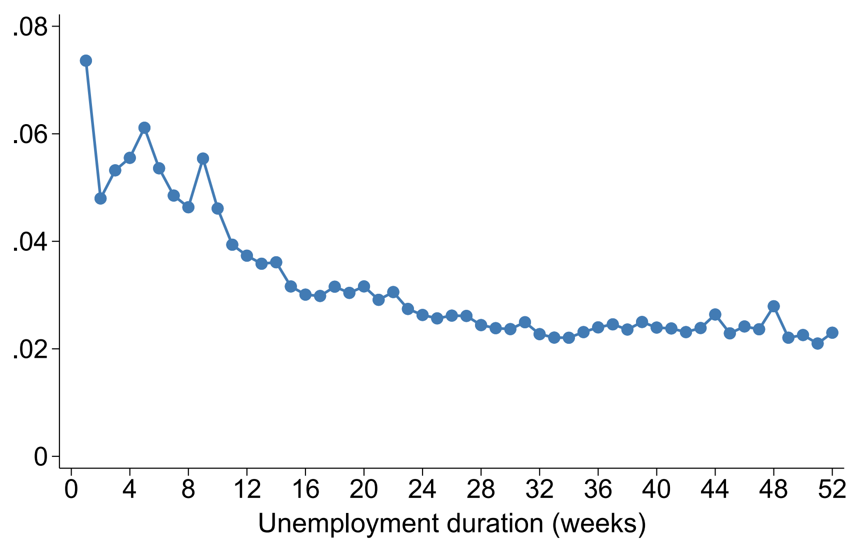
	Full Sample	Baseline RD	Het. RD	Dyn. RD
Share female	0.46	0.45	0.40	0.44
Share of Belgian origin	0.67	0.73	0.58	0.69
Share tertiary educated	0.25	0.26	0.14	0.23
Mean age	34.7	30.6	35.1	31.2
Share with very good Dutch ability	0.70	0.80	0.64	0.78
Mean previous unemployment duration	255	195	370	213
Mean days unemployed in last 5 years	342	218	711	250
Mean pred. job-finding rate (week 1)	0.615	0.637	0.619	0.625
N	781,944	78,675	130,981	38,209

Note: This table reports descriptive statistics for our full sample of unemployment spells starting between 1 September 2018 and 30 November 2022, as well as the estimation samples used for the three empirical strategies. "Mean previous unemployment duration" refers to the mean duration of a job seeker's previous unemployment spells, counting at most the last five spells.

B.3 Unemployment Dynamics



(a) Survival

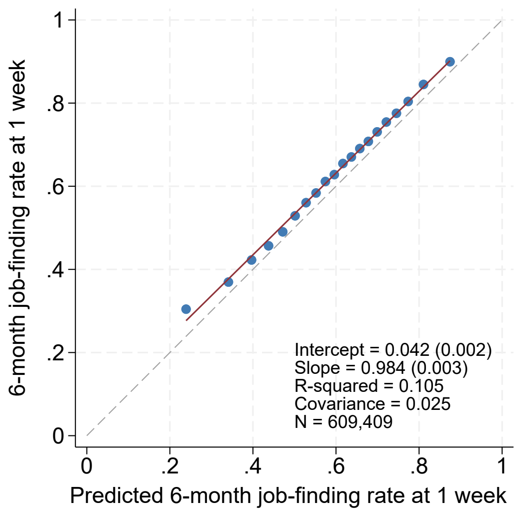


(b) Exit

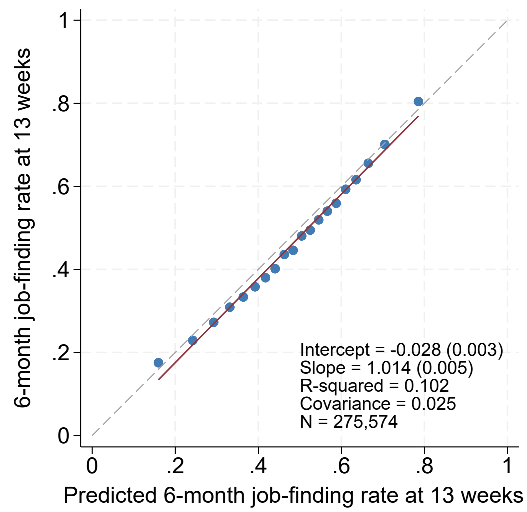
Figure B1: Unemployment Dynamics

Note: Panel (a) plots the proportion of job seekers remaining in unemployment at each duration from 0–52 weeks. Panel (b) plots the proportion of job seekers exiting unemployment each week, conditional on having been unemployed at the start of that week. The sample includes new unemployment spells beginning between 1 September 2018 and 30 November 2022 and is restricted to job seekers below the age of 60.

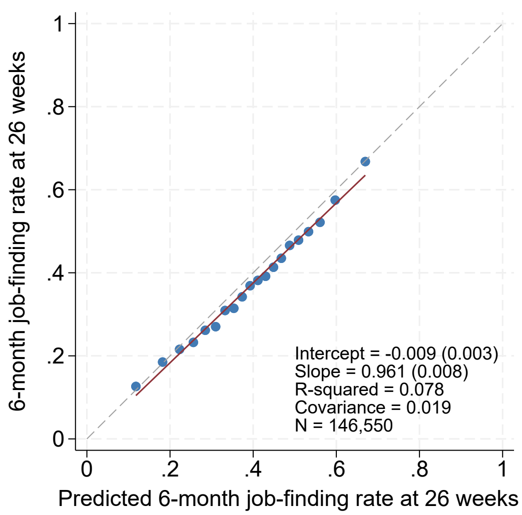
B.4 Risk Score Accuracy



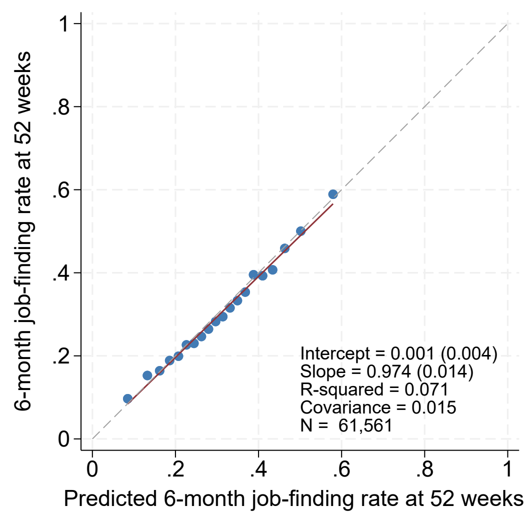
(a) At 1 Week



(b) At 3 Months



(c) At 6 Months



(d) At 12 Months

Figure B2: Employment Predictions vs. Outcomes

Note: This figure shows binned scatter plots comparing predicted 6-month job-finding rates and empirical 6-month job-finding rates.

B.5 Program Statistics

Table B3: Program Statistics

Category	Subcategory	Share	Duration unemployed at start (median days)	Program length (median days)	Mean predicted job-finding rate
Full sample		1.00			0.59
Job search information	Information	0.14	111	0	0.57
	Screening	0.08	125	0	0.56
	Advice	0.06	295	0	0.51
Job search orientation	Orientation	0.05	160	53	0.54
	Training	0.06	241	21	0.52
	Coaching	0.01	263	176	0.51
Job search reorientation	Coaching	0.06	246	138	0.51
	Outplacement	0.01	47	329	0.53
	Career guidance	0.00	25	20	0.63
Examination (e.g., disability check)		0.14	216	0	0.53
Training	General	0.09	190	66	0.46
	Vocational	0.20	216	64	0.57
	Language	0.04	208	46	0.47
	Online	0.18	151	1	0.57
	Formal education	0.00	76	281	0.40
	District works	0.01	352	193	0.46
	Other	0.09	222	170	0.51
Job application requirement		0.27	168	0	0.55
Job application assistance	App. submitted	0.04	229	64	0.55
	App. assistance	0.11	186	302	0.53
Non-VDAB employment service		0.17	150	17	0.53
Sanctions investigation		0.08	200	0	0.53
Other	Project line	0.01	93	0	0.53
	Certification	0.00	249	98	0.55
None		0.33			0.64

Note: This table shows descriptive statistics for participation in programs for all spells starting between 1 September 2018 and 30 November 2022 for individuals below the age of 60. We report the median duration of unemployment at which each category of program is started, the median duration of each program and the mean predicted 6-month predicted job-finding rate of participants in each program. Note that we only observe participation in programs until 30 November 2022. For "program length," we therefore further restrict the sample to only spells starting between 1 September 2018 and 30 November 2021.

C Baseline RD

C.1 Baseline RD: Distribution of Risk Scores

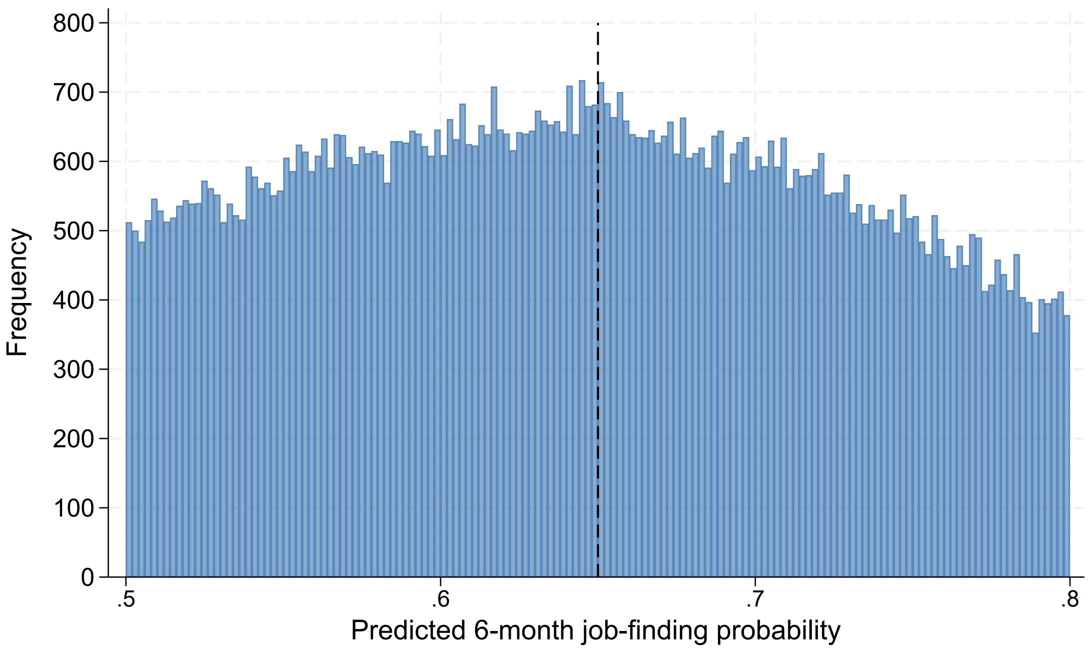


Figure C1: Distribution of Risk Scores

Note: This figure shows the frequency distribution of VDAB's predicted 6-month job-finding rates around the cutoff for the sample of spells used to estimate the effect of caseworkers on days worked per week. This includes spells beginning between 1 September 2018 and 9 September 2022.

C.2 Baseline RD: Balance Tests

This section reports how predetermined covariates vary around the baseline RD threshold. Figure C2 shows precisely estimated nulls for all available variables except tertiary education. In Figure C3, we show that the difference in the proportion of individuals with tertiary education is spurious, caused instead by clear nonlinearities in the relationship between the running variable and the proportion of job seekers with tertiary education. Applying a smaller bandwidth greatly reduces the point estimate and leaves the RD estimate statistically insignificant.

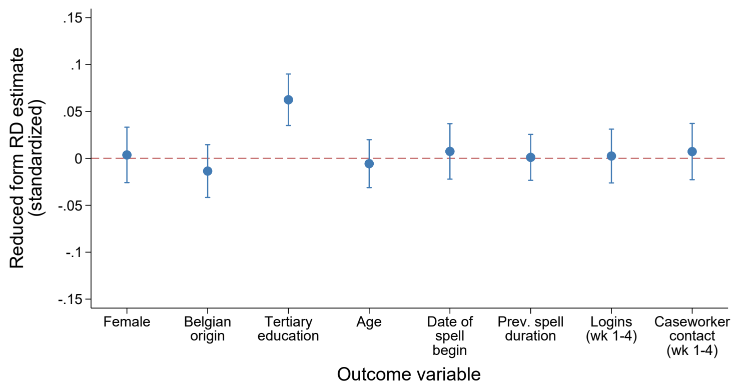
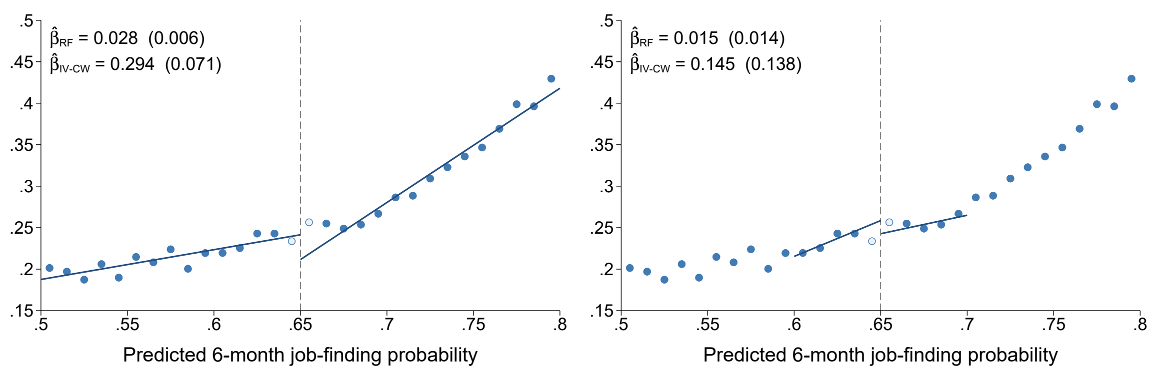


Figure C2: Balance Test

Note: This figure plots reduced form estimates of the effect of being below the cutoff on predetermined covariates. The same sample, controls and specification are used as in the baseline estimation. The covariates are all standardized within the estimation sample. We display 95% confidence intervals.



(a) 0.15 Bandwidth (Baseline)

(b) 0.05 Bandwidth

Figure C3: Tertiary Education

Note: This figure plots the proportion of job seekers with tertiary education. Panel (a) and (b) apply bandwidths of 0.15 and 0.05, respectively. The outcome is residualized on a set of controls and added to the mean. Excluding education, the same sample, controls and specification are used as in the baseline estimation. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

C.3 Baseline RD: Kernel, Bandwidth and Donut

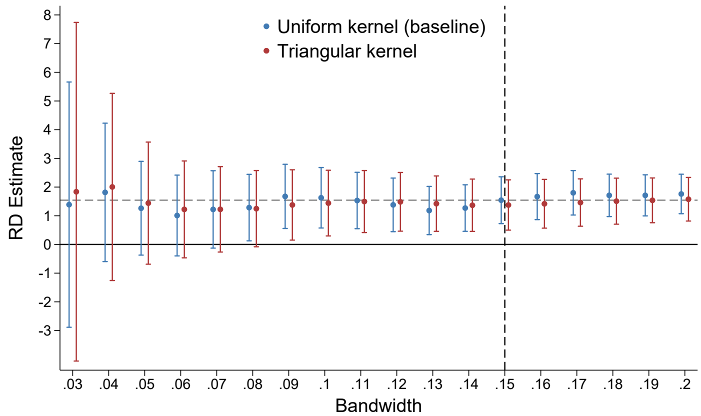


Figure C4: Estimates by Kernel and Bandwidth

Note: This figure plots the baseline RD estimate, estimated using different kernels and bandwidths. The vertical line represents the bandwidth used for the baseline specification. The horizontal gray line marks the baseline estimate. Apart from bandwidth restrictions, the same sample, controls and specification are used as in the baseline estimation. We display 95% confidence intervals for the estimated effects.

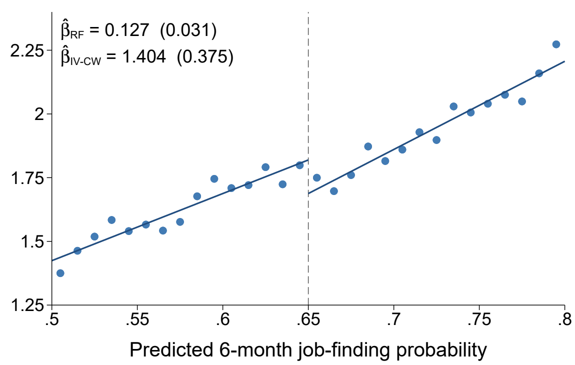


Figure C5: Baseline RD without Donut

Note: This figure plots days worked per week in weeks 6–17 after residualizing on a set of controls. The same sample restrictions, regression specifications and set of controls are used as in the baseline RD. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

C.4 Baseline RD: Sample Window

The baseline RD is estimated on a sample including unemployment spells starting between 1 September 2018 and 9 September 2022. In Figure C6, we check robustness of the baseline RD result to restricting this window. We first show robustness to only using risk scores from before the INOS system was introduced in October 2020 as well as afterwards. For the pre-INOS sample, we can only use one spell per individual as we only observed one of the risk scores used by the service line's outreach campaign, even if the individual had multiple spells. For the INOS sample, we observe the risk scores used in all spells. We then also show robustness to excluding the period over which the COVID-19 pandemic occurred. We find similar effects during the pandemic and in non-pandemic periods.

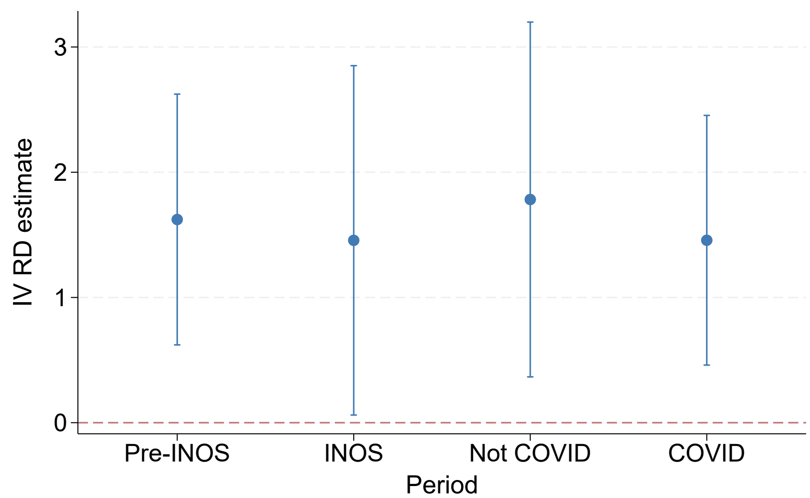
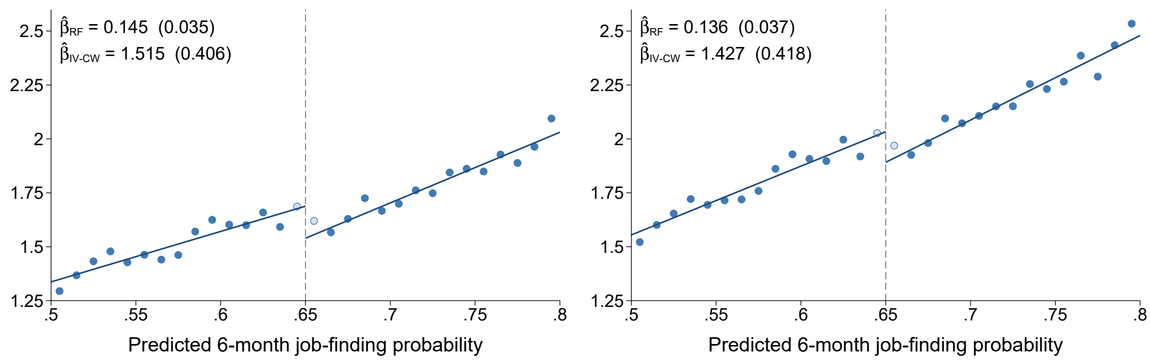


Figure C6: Baseline RD by Sample Window

Note: In this figure, we plot the baseline RD estimate of the effect of meeting with a caseworker on days worked per week in weeks 6–17 subject to different sample restrictions. The INOS sample includes spells, for which the service line recorded risk scores after the introduction of INOS, an information system for handling job seekers prior to their assignment to a caseworker. The pre-INOS sample includes spells, for which the service line recorded risk scores prior to the introduction of INOS. For some spells, risk scores were assigned both pre-INOS and using INOS. These spells appear in both samples. We define the COVID-19 pandemic as occurring between March 2020, when the number of reported cases first reached double digits in Belgium, and May 2022, after which the number of reported deaths per day remained below 15. Apart from these sample restrictions, the same sample, regression specifications and set of controls are used as in the baseline RD. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

C.5 Baseline RD: Definition of Employment

Throughout the paper, we define spells in temporary work and subsidized work-and-training programs as employment if they are directly followed by employment without a return into unemployment. One could argue, however, that these forms of employment could either constitute employment or unemployment. In Figure C7, we show that our results are mostly unaffected by the choice of definition.



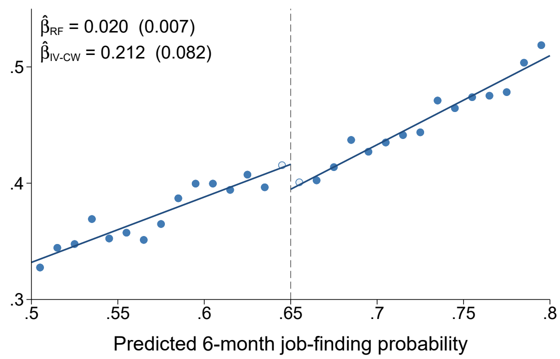
(a) Never counted as employment

(b) Always counted as employment

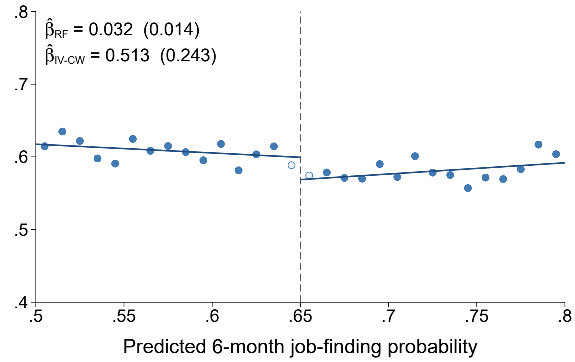
Figure C7: Results by whether temporary and subsidized work are counted as employment

Note: Panels (a) and (b) of this figure plot days worked per week in weeks 6–17 after residualizing on a set of controls. In panel (a), temporary work and subsidized work are never counted as employment. In panel (b), they are always counted as employment. The same sample restrictions, regression specifications and set of controls are used as in the baseline RD. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

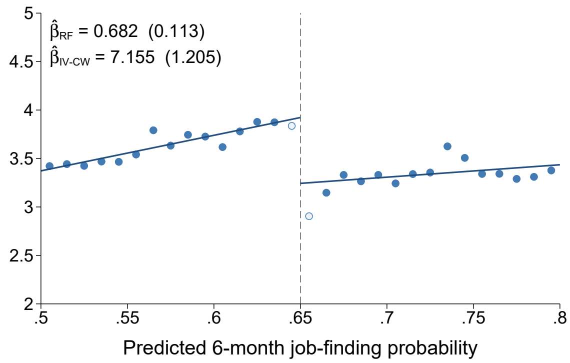
C.6 Baseline RD: Additional Outcomes



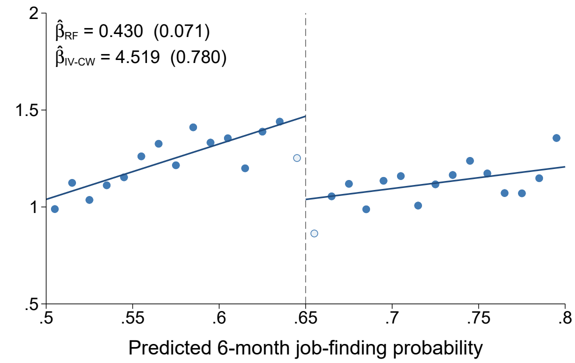
(a) Employment



(b) Employment Tenure



(c) Logins



(d) Self-Reported Job Applications

Figure C8: Other Employment Outcomes

Note: Panel (a) plots the proportion of employed job seekers after 17 weeks conditional on being unemployed after 5 weeks. Panel (b) plots the proportion of job seekers who are employed in an employment spell lasting at least two years, conditional on entering employment in weeks 6–17. Panel (c) plots the average number of logins to the VDAB website in weeks 6–17. Panel (d) plots the average number of job applications reported by job seekers in weeks 6–17. We residualize all outcomes on a set of controls and add the mean. The same sample restrictions, regression specifications and set of controls are used as in the baseline RD, with one exception: panel (b) imposes an additional sample restriction, excluding spells that began less than 2 years and 17 weeks before the end of our labor market history observation window, 31 July 2023. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

C.7 Baseline RD: Program Participation

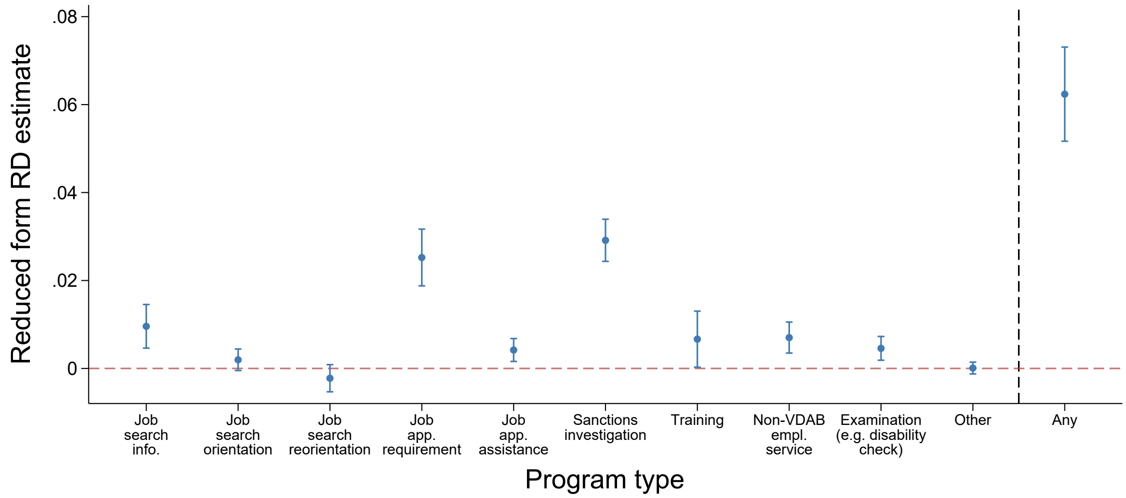


Figure C9: Effect on Program Participation (Weeks 6–17)

Note: This figure shows the reduced form effect of being below the cutoff on commencing participation in various programs in weeks 6–17 of unemployment. The same sample, specifications and controls are used as in the baseline regression. We report 95% confidence intervals. See Appendix B.5 for detailed statistics on various categories of programs.

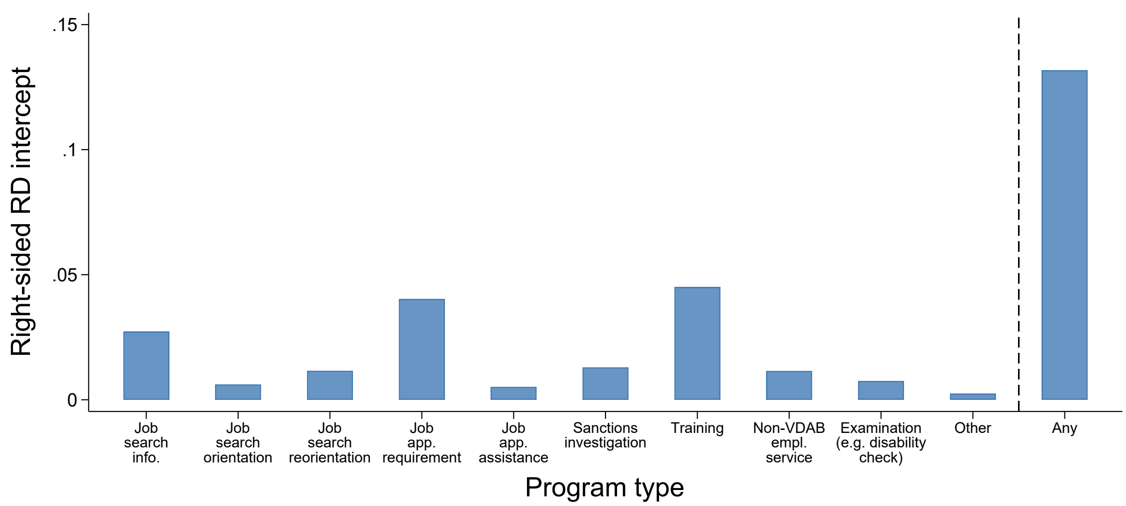


Figure C10: Program Participation Above the Threshold (Weeks 6–17)

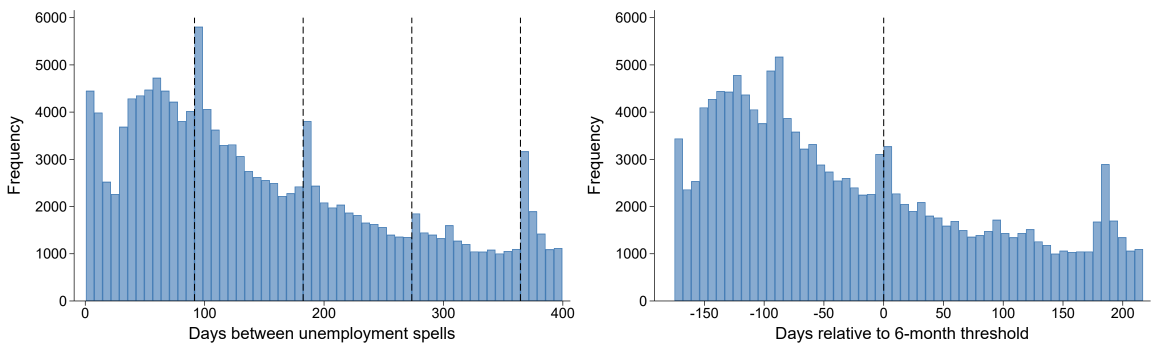
Note: This figure shows the right-sided intercept with the threshold, estimated using the baseline RD, where the outcome is participation in various programs in weeks 6–17 of unemployment. The same sample, specifications and controls are used as in the baseline regression.

D Heterogeneity RD

D.1 Heterogeneity RD: Bunching

In Panel (a) of Figure D1, we show that there is considerable bunching in the duration between unemployment spells. Specifically, many individuals re-enter unemployment exactly 3, 6, 9, or 12 months after the end of their previous unemployment spell. It is likely that these durations serve as reference points for employers when structuring employment contracts.

Re-assignment to one's previous caseworker is not, however, based on the exact number of days since previous unemployment. Instead it is determined by whether it has been more or less than exactly six months since one's previous unemployment spell. Panel (b) of Figure D1 shows how the bunching observed in the duration between spells translates into bunching in the running variable.



(a) Days Between Spells

(b) Distance to Threshold

Figure D1: Bunching in Duration Between Unemployment Spells

Note: Panel (a) plots the distribution of the days between unemployment spells. The vertical lines mark 3, 6, 9 and 12 months. Panel (b) plots the distribution of unemployment spell start dates relative to the day that is 6 months and 1 day after the end of the previous unemployment spell. Both plots use a bin width of 7 days and apply the same sample restrictions as used in the heterogeneity RD, except for those based on the running variable.

As bunching job seekers may differ from other job seekers, those bunching close to the threshold are excluded from the estimation sample in a 'donut' regression discontinuity design. We apply a novel data-driven approach to select the width of the donut. The approach involves predicting a counterfactual density by estimating the following model by Poisson regression,

$$\ln \mu_j = \beta_0 + \sum_{p=1}^{5} \beta_p \, z_j^p + \sum_{d=1}^{6} \eta_d \, I\{w_j = d\}, \tag{D1}$$

where μ_j is the expected number of job seekers in bin j, z_j is the distance between bin j and the threshold, and $I\{\cdot\}$ is an indicator function. The first two terms give a 5^{th} order polynomial in the running variable. The third term captures a strong weekly pattern in the data through fixed effects for the position of each bin within the week, $w_j \in \{0, 1, \ldots, 6\}$, defined as the remainder when the running variable is divided by seven. We estimate the model using a bin width of 1 day and excluding three weeks either side of the threshold in order to exclude the region distorted by bunching. The counterfactual density is obtained through the predicted bin counts $\hat{\mu}_j$.

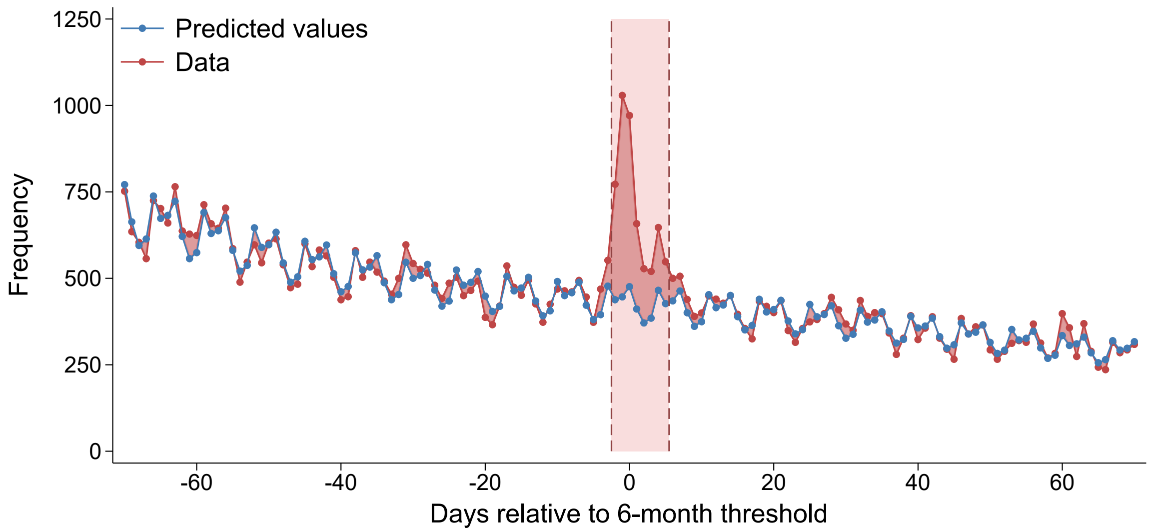


Figure D2: Data-driven Donut

Note: This figure plots the actual distribution of unemployment spells as well as the counterfactual distribution estimated using equation (D1). The highlighted region ($z_j \in [-2, 5]$) is the data-driven donut $\mathcal{D}$ where the actual density deviates from the counterfactual density by more than 20%.

Figure D2 plots the counterfactual density $\hat{\mu}_j$ alongside the actual density around the threshold. For each bin we define the relative deviation of the actual density from the counterfactual density,

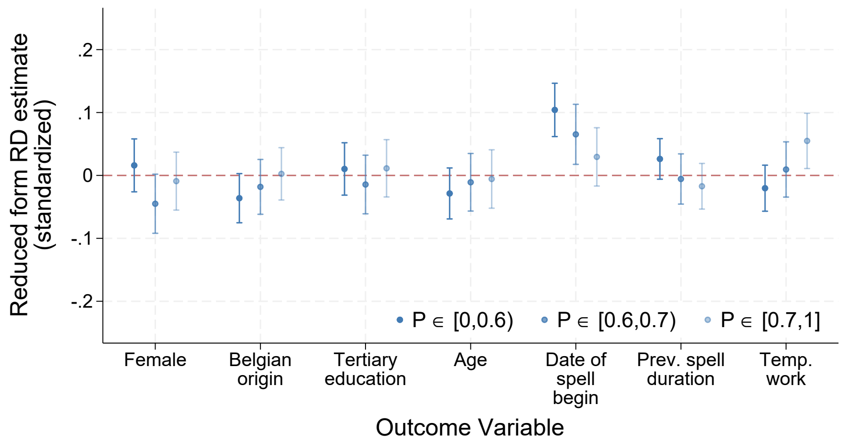
$$\Delta_j = \frac{c_j - \hat{\mu}_j}{\hat{\mu}_j}, \tag{D2}$$

where c_j is the observed number of job seekers in bin j. The data-driven donut $\mathcal{D}$ is then defined as the set of all bins within five weeks of the threshold whose actual density deviates from the counterfactual by more than 20%,

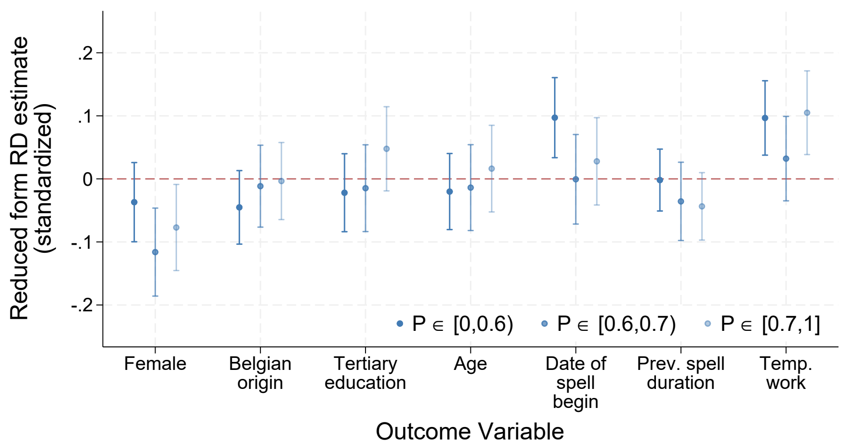
$$\mathcal{D} = \{\, j \ : \ -35 \leq z_j \leq 34 \ \text{ and } \ |\Delta_j| > 0.2 \,\}. \tag{D3}$$

As shown in Figure D2, the data-driven process selects a donut of $z_j \in [-2, 5]$. This donut is excluded from the estimation sample in the heterogeneity RD. In Appendix D.4, we show that the results are robust to using a wide range of donuts as well as no donut.

D.2 Heterogeneity RD: Balance Tests



(a) 20-Week Bandwidth (Baseline)

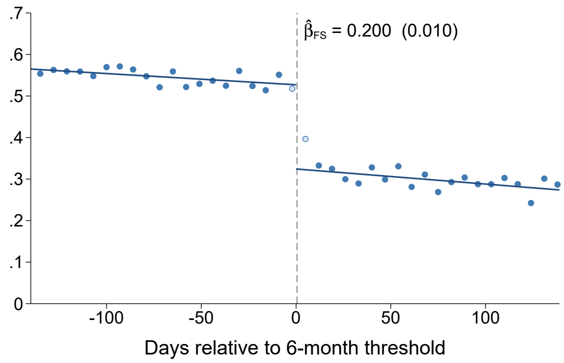


(b) 10-Week Bandwidth

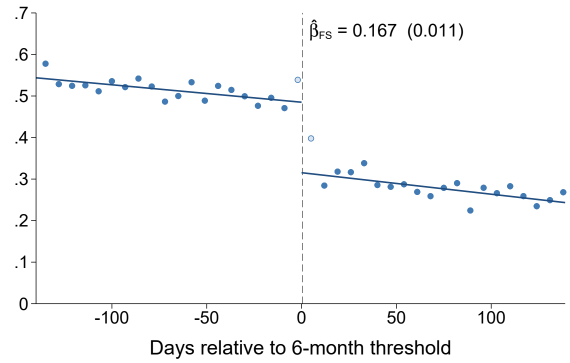
Figure D3: Balance Tests

Note: This figure plots reduced form estimates of the effect of being below the cutoff on predetermined covariates. The covariates are all standardized within the estimation sample. "Temp. work" indicates whether an individual had been engaged in temporary work directly before entering unemployment. Results are reported using 10-week and 20-week bandwidths. We display 95% confidence intervals.

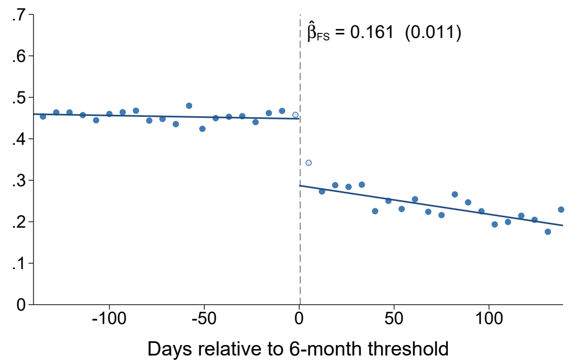
D.3 Heterogeneity RD: First Stage



(a) $P \in [0, 0.6)$



(b) $P \in [0.6, 0.7)$

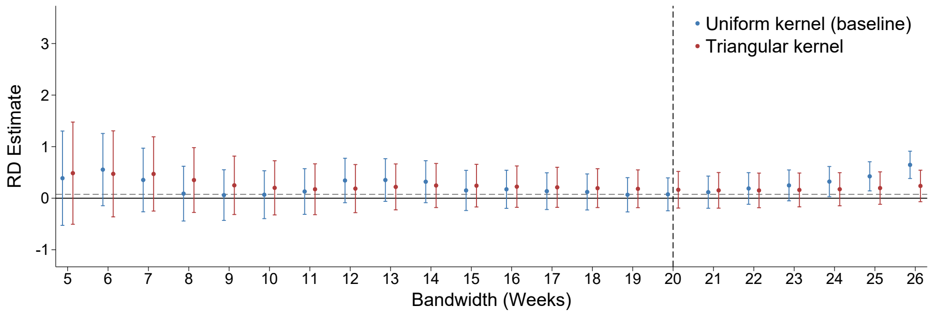


(c) $P \in [0.7, 1]$

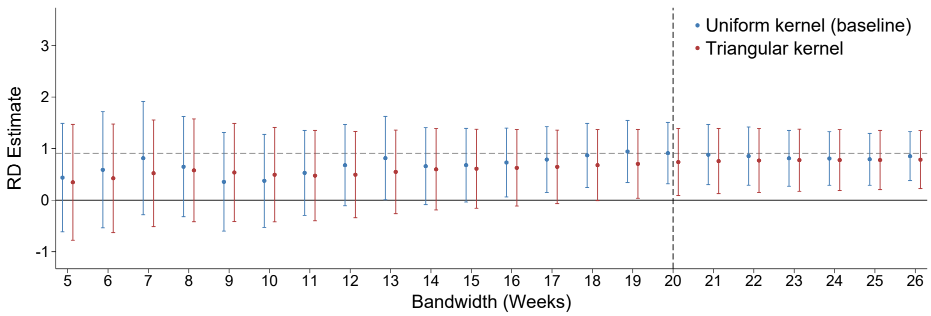
Figure D4: Caseworker Contact (Weeks 1–4)

Note: Panels (a) to (c) plot the proportion of job seekers who have contact with a caseworker in weeks 1–4 of unemployment for each risk group. The same sample restrictions, regression specifications and set of controls are used as in the baseline regression. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

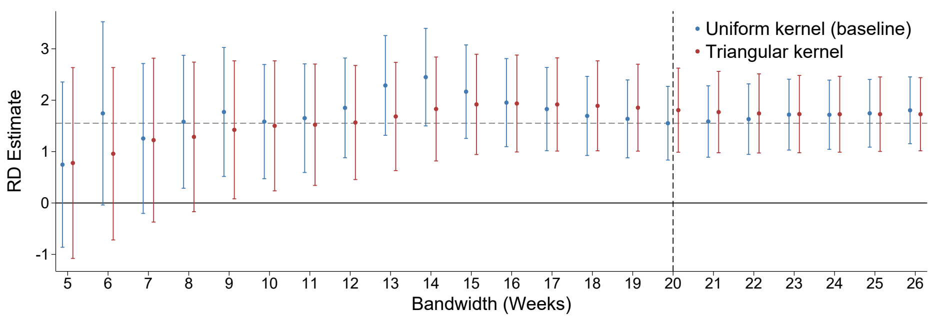
D.4 Heterogeneity RD: Kernel, Bandwidth and Donut



(a) $P \in [0, 0.6)$



(b) $P \in [0.6, 0.7)$



(c) $P \in [0.7, 1]$

Figure D5: Estimates by Kernel and Bandwidth

Note: These panels plot IV estimates of the effect of a caseworker on days worked per week in weeks 1–4 since unemployment, estimated using the heterogeneity RD, for different kernels and bandwidths. The vertical lines represent the bandwidth used for the baseline specification. The horizontal gray line marks the baseline estimate. Otherwise, the same sample restrictions, regression specifications and set of controls are used as in the baseline regression. We display 95% confidence intervals.

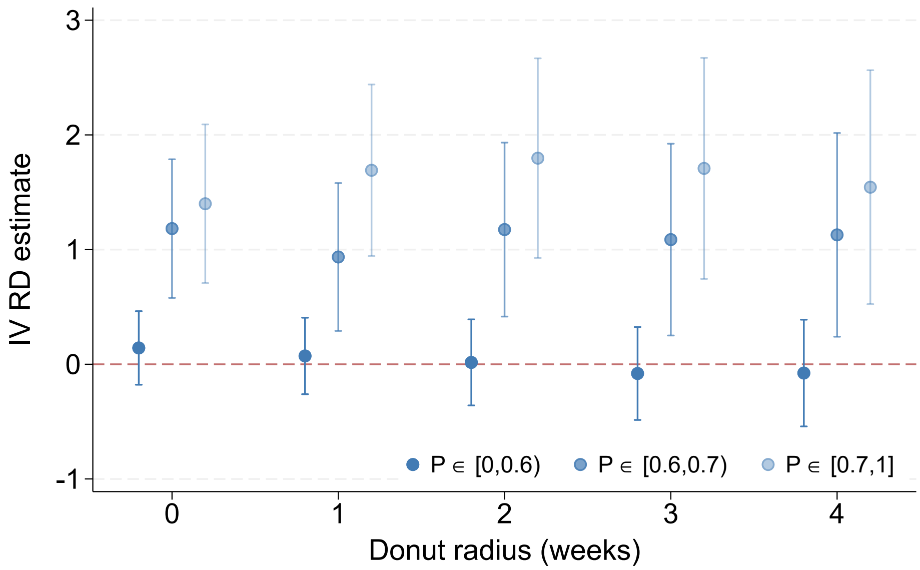


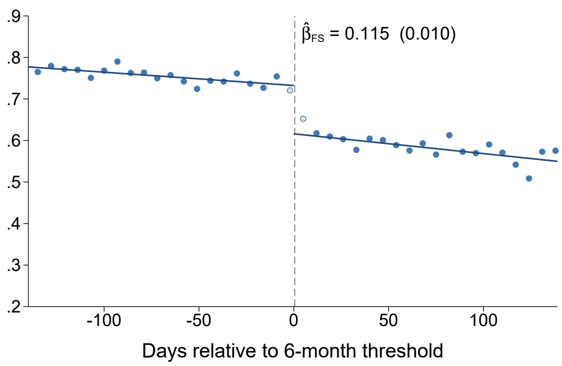
Figure D6: Estimate Robustness to Different Donuts

Note: These panels plot IV estimates of the effect of caseworkers on days worked per week estimated using the heterogeneity RD over weeks 1-4 since the start of unemployment. Robustness is shown to 0–4 week radii. We display 95% confidence intervals. Across all donut specifications, we reject the null hypotheses that the effect is equal in the lowest and highest risk groups with a p-value of, at most, 0.0009.

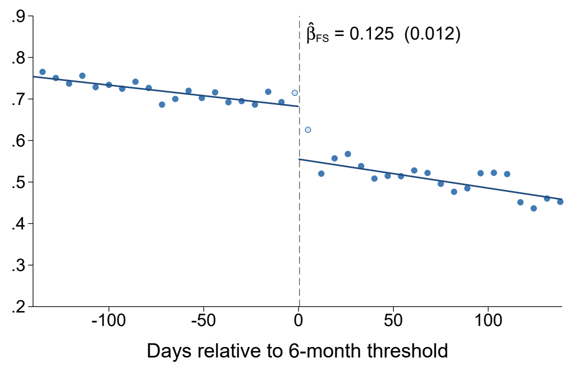
D.5 Heterogeneity RD: Robustness to Horizon

For our baseline estimates, we restrict the observation horizon to only four weeks. We choose this horizon, as after four weeks of unemployment, the service line begins interviewing job seekers and assigning them to caseworkers. By restricting the observation window to only four weeks, we therefore avoid dealing with multiple treatment waves occurring at different durations as well as declining compliance around the 6-month threshold.

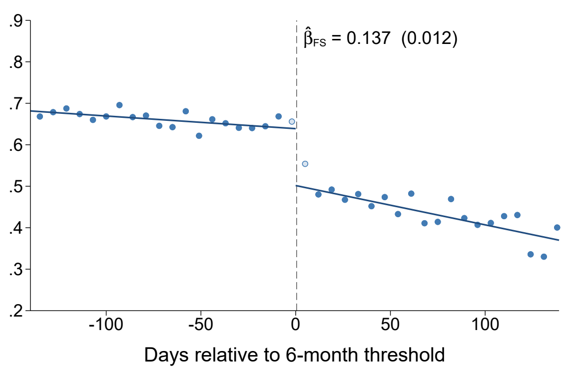
In Figure D7, however, we show that the first stage around the cutoff persists when using a much longer horizon, 16 weeks since unemployment. Even after the first round of service line contact, those below the 6-month threshold remain more likely to have been assigned to a caseworker. In Figure D8, we show that this continues to have an effect on employment over the 16-week horizon. Importantly, the effect on the high-risk group remains a rather precise zero while the effects on the low-risk group remain the largest.



(a) $P \in [0, 0.6)$



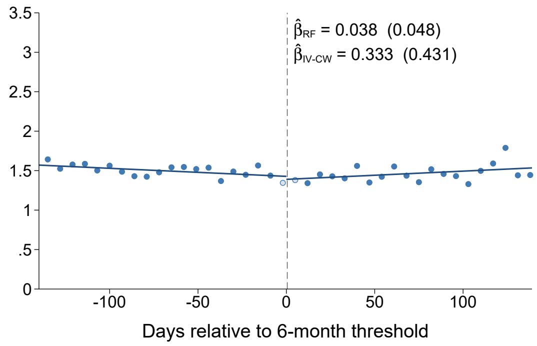
(b) $P \in [0.6, 0.7)$



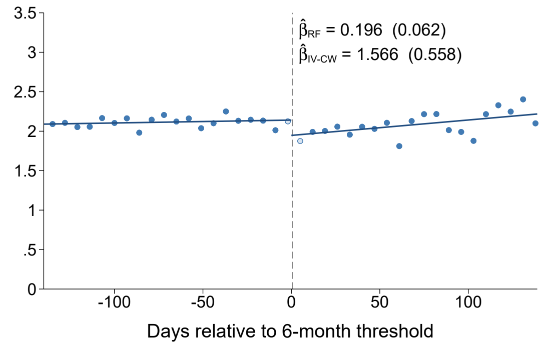
(c) $P \in [0.7, 1]$

Figure D7: Caseworker Contact (Weeks 1–16)

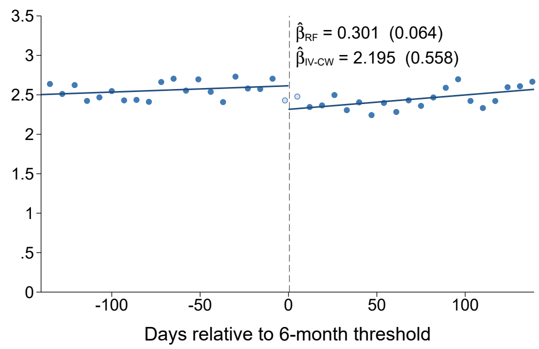
Note: Panels (a) to (c) plot the proportion of job seekers who have contact with a caseworker in weeks 1–16 of unemployment for each risk group. We exclude spells beginning less than 16 weeks before 5 January 2023. Otherwise, the same sample restrictions, regression specifications and set of controls are used as in the baseline regression. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.



(a) $P \in [0, 0.6)$



(b) $P \in [0.6, 0.7)$



(c) $P \in [0.7, 1]$

Figure D8: Days Worked per Week (Weeks 1–16)

Note: Panels (a) to (c) plot days worked per week by days relative to the threshold for each risk group. The outcome is residualized on a set of controls and added to the mean. The window of observation is weeks 1–16 since the start of unemployment. We exclude spells beginning less than 16 weeks before 5 January 2023. Otherwise, the same sample restrictions, regression specifications and set of controls are used as in the baseline regression. We report 95% confidence intervals. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

D.6 Heterogeneity RD: By Risk Ventiles

For our baseline analysis, we divide the sample into three groups by risk score. The groups are selected such that the medium-risk group has on average the same risk profile as the sample used in the Baseline RD, a risk score of around 0.65.

To show, however, that the result is not dependent on how the groups are selected, in this section, we split the sample into 20 risk score ventiles and estimate the RD for each ventile. The relationship between treatment effects and risk scores that we document remains clearly visible.

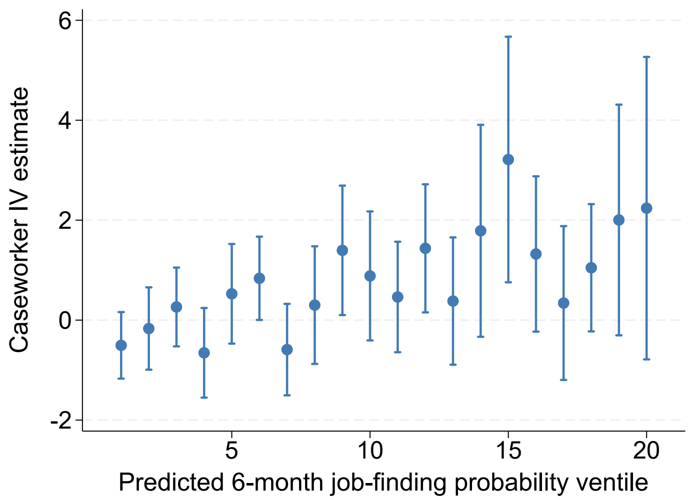


Figure D9: Caseworker IV Estimate by Risk Ventile

Note: Job seekers are split into 20 ventiles based on their predicted probability of finding a job within 6 months, P, that was predicted by the VDAB after 1 week of unemployment. The plot shows IV estimates of the effect of contact with a caseworker on days worked per week in weeks 1–4. The same sample restrictions, regression specifications and set of controls are used as in the baseline regression. We report 95% confidence intervals.

D.7 Heterogeneity RD: Other Outcomes

This section reports results for alternative employment outcomes. Figure D10, shows that the heterogeneity in treatment effects persists when using a binary indicator of employment as an outcome. The results reported in Figure D11 provide no strong evidence for treated job seekers being pushed into less stable employment. As the sample size is small when conditioning on employment within the first four weeks, we show results using both a 4-week horizon and a 16-week horizon.

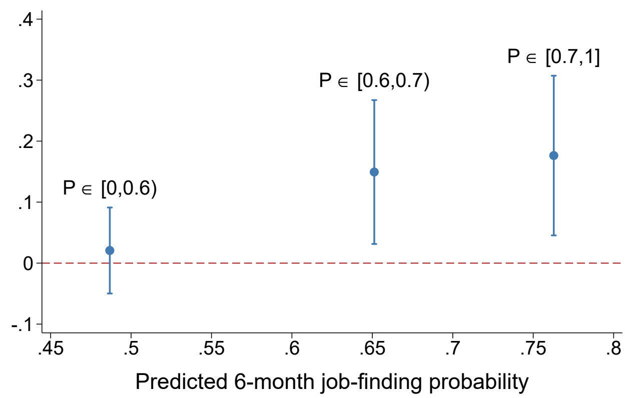


Figure D10: Employment After 4 Weeks

Note: This figure plots IV estimates of the effect of caseworkers on a binary outcome equal to 1 if a job seeker is employed four weeks after the start of their unemployment spell. These estimates are estimated using the heterogeneity RD, applying the same sample restrictions, regression specifications and set of controls as used in the baseline regression. We report 95% confidence intervals.

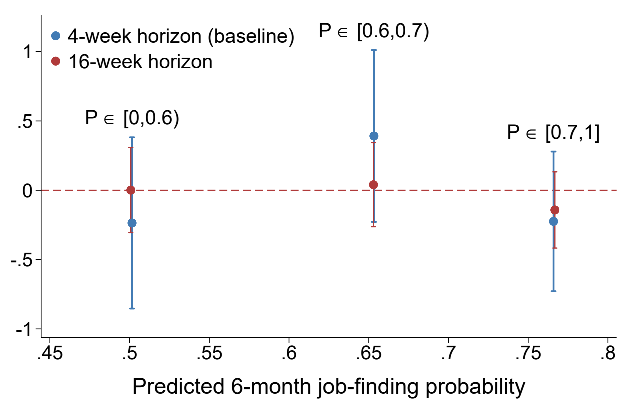
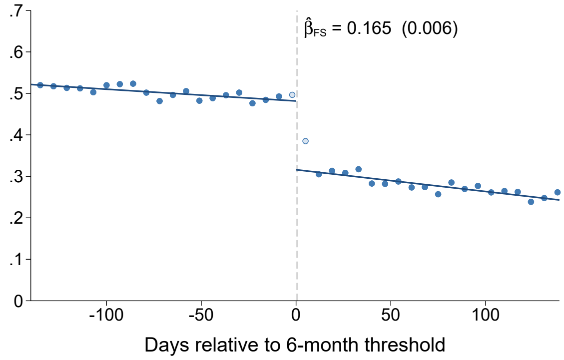


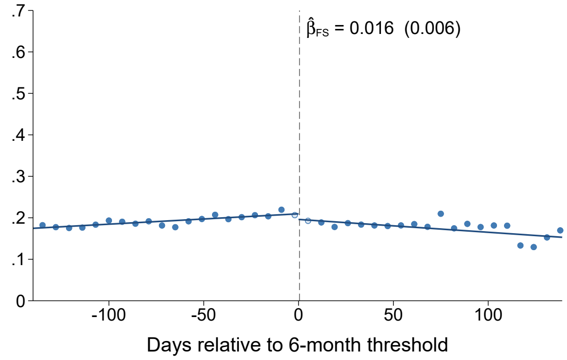
Figure D11: Employment Tenure

Note: This figure plots IV estimates of the effect of caseworkers on a binary outcome equal to 1 if, four weeks after the start of unemployment, a job seeker is employed in an employment spell lasting at least two years, conditional on being employed at four weeks. In red, we also report estimates for a 16-week horizon. These estimates are estimated using the heterogeneity RD, applying the same sample restrictions, regression specifications and set of controls as used in the baseline regression. We report 95% confidence intervals.

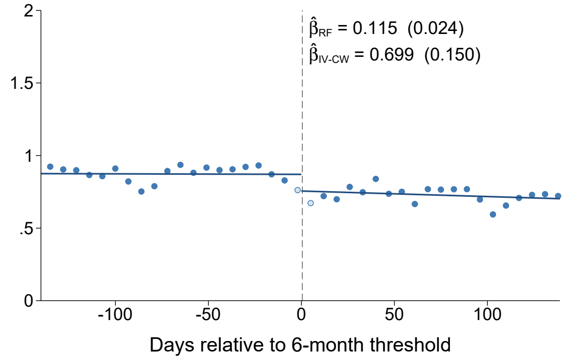
D.8 Heterogeneity RD: Placebo



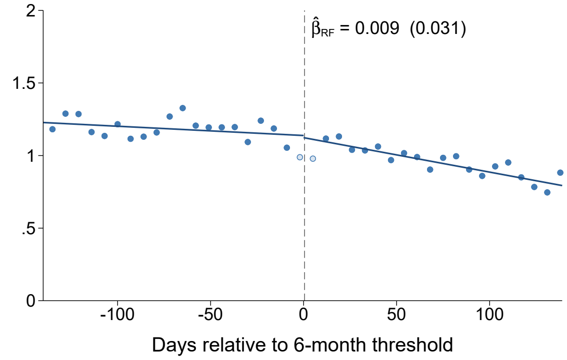
(a) Caseworker Contact (Baseline)



(b) Caseworker Contact (Placebo)



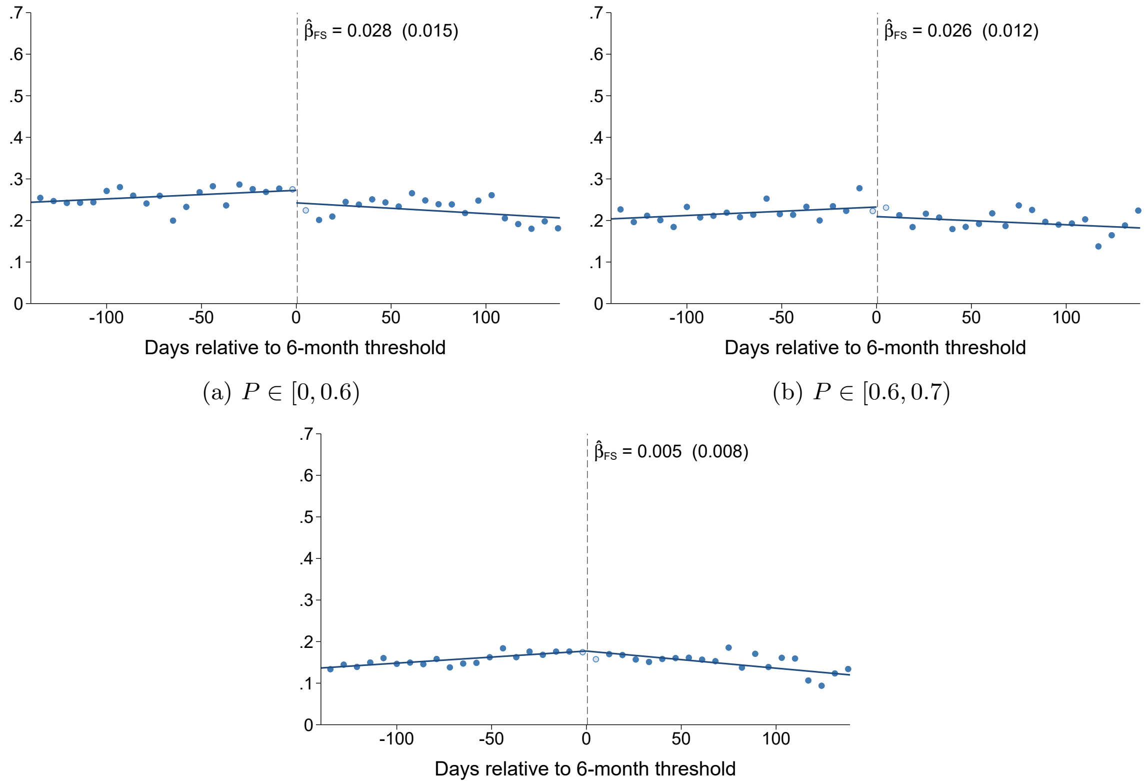
(c) Days Worked (Baseline)



(d) Days Worked (Placebo)

Figure D12: Placebo without Heterogeneity

Note: The baseline sample, used in panels (a) and (c), is restricted to job seekers who had contact with a caseworker in their previous spell. The placebo sample, used in panel (b) and (d), is restricted to job seekers who have not had contact with a caseworker in any spell ending less than a year before the start of the current spell. The outcome variable in panels (a) and (b) is contact with a caseworker in weeks 1–4 of unemployment, while the outcome variable in panels (c) and (d) is days worked per week in weeks 1–4 of unemployment. For panels (c) and (d), we residualize the outcomes on a set of controls and add the mean. The same sample restrictions, regression specifications and set of controls are used as in the baseline regression. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.



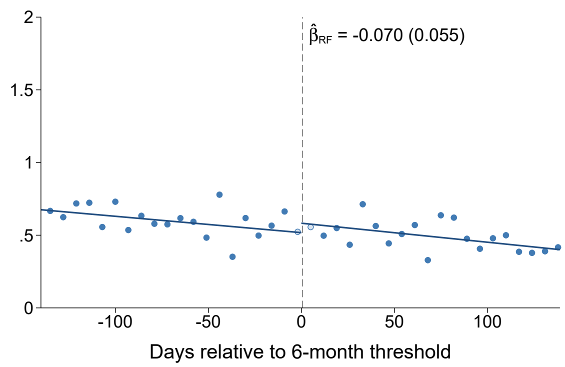
(a) $P \in [0, 0.6)$

(b) $P \in [0.6, 0.7)$

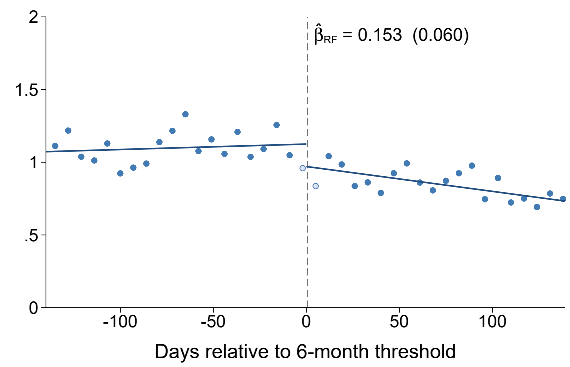
(c) $P \in [0.7, 1]$

Figure D13: Placebo Caseworker Contact (Weeks 1–4)

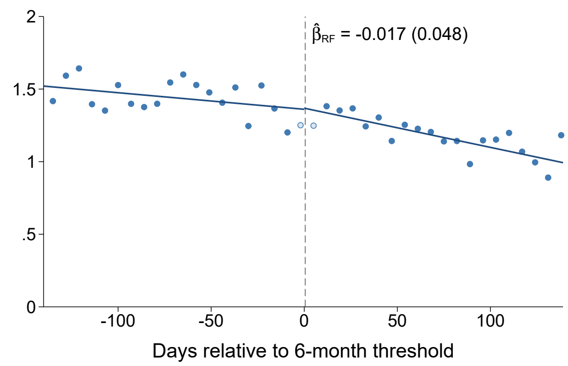
Note: The placebo sample is restricted to job seekers who have not had contact with a caseworker in any spell ending less than a year before the start of the current spell. Job seekers are then split into three groups by the probability of finding a job within 6 months P that was predicted for them after 1 week of unemployment. Panels (a) to (c) plot the proportion of job seekers who have contact with a caseworker in weeks 1–4 of unemployment for each risk group. The same sample restrictions, regression specifications and set of controls are used as in the baseline regression. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.



(a) $P \in [0, 0.6)$



(b) $P \in [0.6, 0.7)$



(c) $P \in [0.7, 1]$

Figure D14: Placebo Days Worked per Week (Weeks 1–4)

Note: The sample is restricted to job seekers who have not had contact with a caseworker in any spell ending less than a year before the start of the current spell. Job seekers are then split into three groups by the probability of finding a job within 6 months P that was predicted for them after 1 week of unemployment. The outcome variable is days worked per week in weeks 1–4 of unemployment. We residualize the outcome on a set of controls and add the mean. The same sample restrictions, regression specifications and set of controls are used as in the baseline regression. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

D.9 Heterogeneity RD: Program Participation

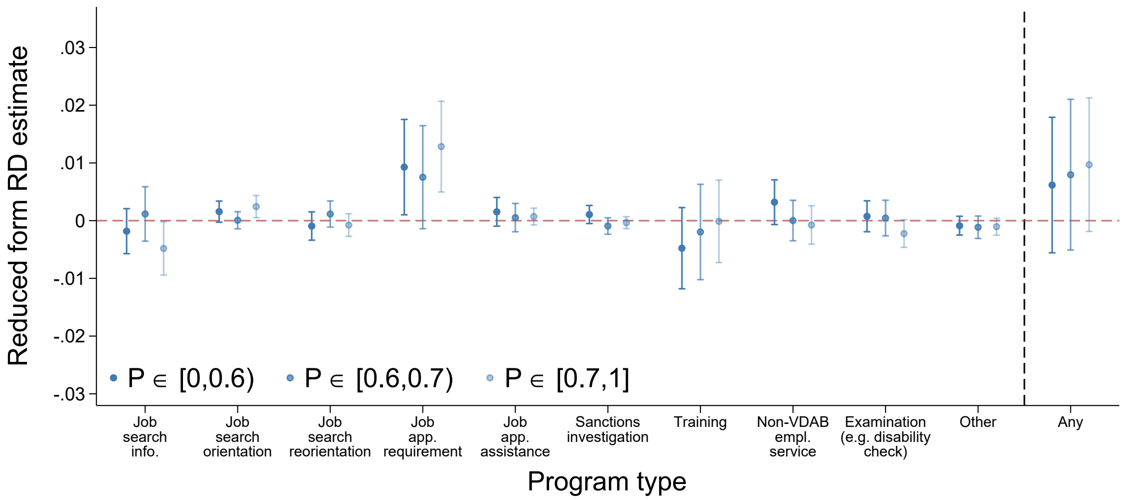


Figure D15: Effect of Being Below Threshold on Program Participation

Note: This figure plots reduced form estimates of the effect of being below the cutoff on participation in various programs in weeks 1–4 by risk group. The same sample restrictions, regression specifications and set of controls are used as in the heterogeneity RD. We display 95% confidence intervals for the estimated effects.

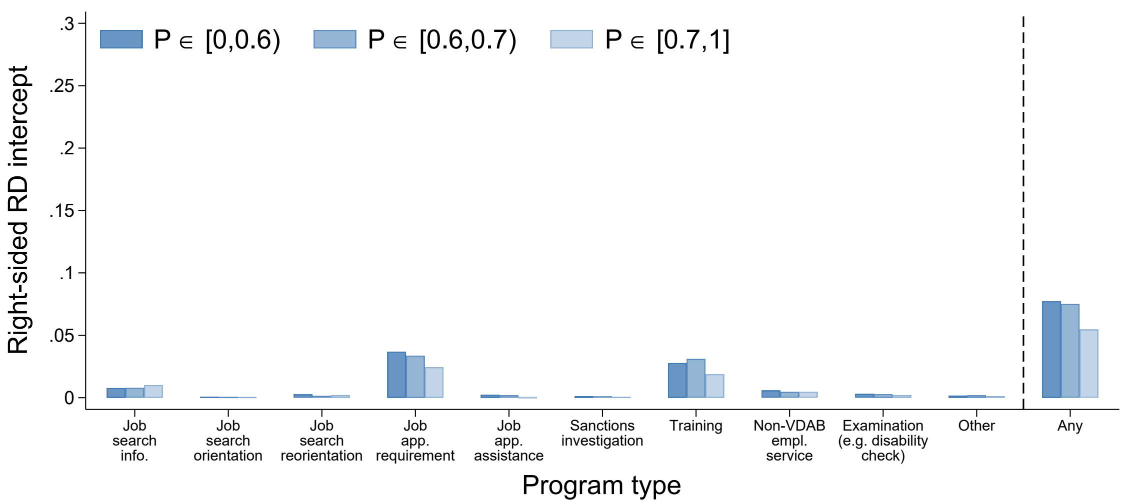


Figure D16: Program Participation Above the Threshold

Note: This figure plots estimates of the right-sided intercept with the threshold, estimated with the heterogeneity RD and using participation in various programs as the outcome (see Figure D15). The same sample restrictions, regression specifications and set of controls are used as in the heterogeneity RD. We display 95% confidence intervals for the estimated effects.

D.10 Comparing CATEs to RD Estimates

In this section, we explore why the heterogeneity in the estimated CATEs differs from the heterogeneity RD estimates. In particular, while the estimated treatment effect is close to zero for the high-risk group, the average estimated CATE for that group is 0.51.

In Figure D17, we show that this is primarily due to the simplifications made in order to estimate the CATEs in an RD context. Instead of allowing for a trend in the running variable, let alone differential trends for different risk groups, we simply apply a binary indicator for being below the cutoff as an instrument for treatment when estimating the CATEs. While we do reduce the bandwidth to 10 weeks either side of the threshold, the absence of a running variable nevertheless appears to affect the estimates.

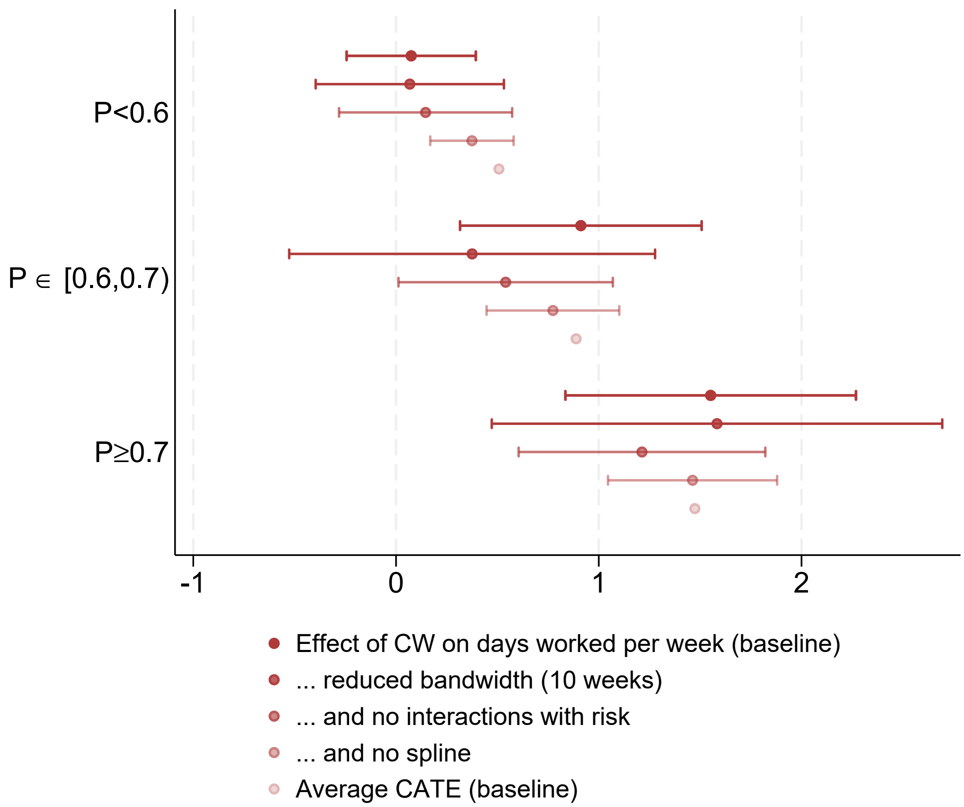


Figure D17: CATEs vs. RD Estimates

Note: For each risk group, this figure compares the estimated CATEs to the treatment effects estimated in the heterogeneity RD of the effect of a caseworker on days worked per week within four weeks of unemployment. For the second set of estimates, the bandwidth is reduced from 20 to 10 weeks. For the third set of estimates, the running variable and controls are no longer interacted with risk group. For the final set of RD estimates, the running variable is dropped from the regression model, such that we solely use a binary indicator of being below the threshold as an instrument for treatment. The final estimate is the mean estimated CATE in each risk group.

D.11 Variable Importance

Table D1: Variable Importance in Instrumental Forest

Variable	Baseline	No risk score
Risk score on day 8	0.375	
Age	0.190	0.113
Days unemployed in last 5 years	0.064	0.123
Education	0.053	0.074
Days unemployed in last 3 years	0.052	0.105
Average previous unemployment duration	0.050	0.132
Occupational employability index	0.048	0.109
Dutch ability	0.047	0.054
Quarter & year	0.047	0.075
Previous spell logins (weeks 1-4)	0.041	0.085
Arrondissement: Turnhout	0.015	0.037
Origin: not EU, UK, Turkey or Morocco	0.014	0.037

Note: This table reports variable importance of variables used in the instrumental forest. Variable importance is a depth-weighted measure of how often the forest splits on a variable. The first column reports variable importance from the baseline instrumental forest. In the second column, risk score is excluded from the set inputs used to estimate the forest. We only report variables used in the final model. An initial instrumental forest is estimated using a larger set of factors. The final instrumental forest is then estimated using only those factors with above mean variable importance in the initial model. The low-employment occupation index is calculated in two steps. First, for each occupation, the average 3-month job-finding rate among individuals who list that occupation as a "preferred occupation" is calculated. Then, for each individual, the average is taken of the first step averages of that individual's preferred occupations. Average previous spell duration is the mean duration of previous unemployment spells, counting at most five.

D.12 Descriptive Statistics by Risk Score and CATE

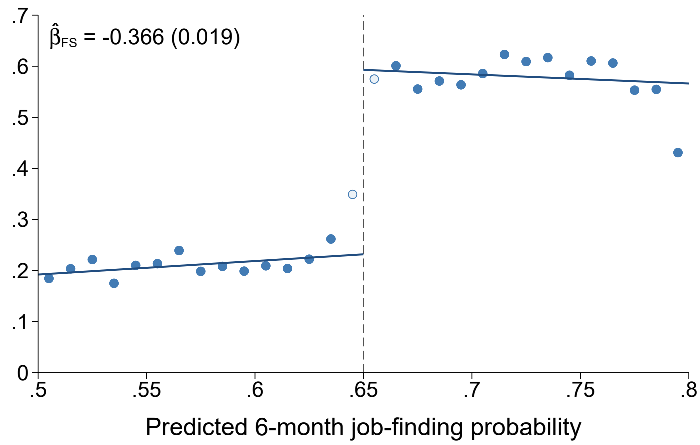
Table D2: Descriptive Statistics by Quadrant

	$P < P_{\text{med}}$		$P \geq P_{\text{med}}$	
	$\hat{\tau} \geq \hat{\tau}_{\text{med}}$	$\hat{\tau} < \hat{\tau}_{\text{med}}$	$\hat{\tau} \geq \hat{\tau}_{\text{med}}$	$\hat{\tau} < \hat{\tau}_{\text{med}}$
Female	34%	45%	34%	45%
Age ≤ 25	16%	24%	20%	44%
Age 26–44	47%	57%	52%	51%
Age 45+	37%	19%	28%	5%
Belgian-origin	49%	51%	56%	75%
Tertiary education	11%	16%	13%	17%
Advanced Dutch proficiency	55%	61%	60%	85%
Work disability	12%	11%	9%	8%
Low-employment occupation	22%	36%	22%	26%
Mean duration of prev. spells	402	363	239	158
N	8,342	16,320	16,321	8,342

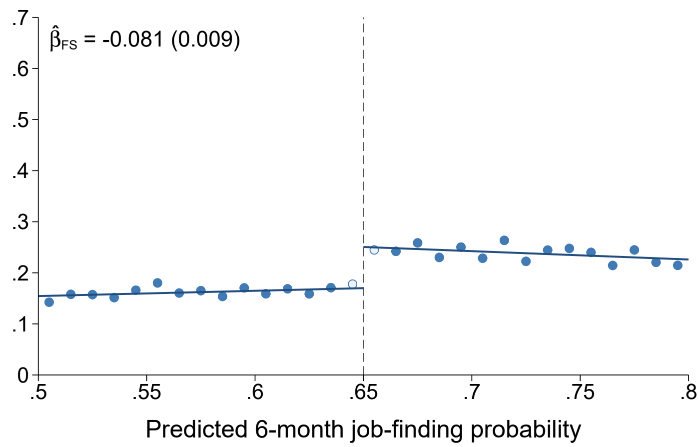
Note: The sample used to estimate the causal forest is divided into quadrants based on the median risk score and the median CATE. This table reports statistics for each quadrant. "Low-employment occupation" refers to the proportion with an occupational employability index in the lowest quartile. The occupational employability index is calculated in two steps. First, for each occupation, the average 3-month job-finding rate among individuals who list that occupation as a "preferred occupation" is calculated. Then, for each individual, the average is taken of the first step averages of that individual's preferred occupations. "Mean duration of prev. spells" is the mean duration of previous unemployment spells, counting at most five.

E Dynamics RD

E.1 Dynamics RD: RD Plot



(a) Contact with the Service Line



(b) First Contact with a Caseworker

Figure E1: First Stage (Weeks 20–31)

Note: These plots show the proportion of job seekers with contact with the service line or a caseworker in weeks 20–31. The sample is restricted to job seekers for whom the service line receives a risk score after 5 weeks of unemployment, who go on to complete at least 19 weeks of unemployment and who begin unemployment at least 31 weeks before 5 January 2023. Otherwise, the same sample restrictions, regression specifications and set of controls are used as in the baseline RD. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

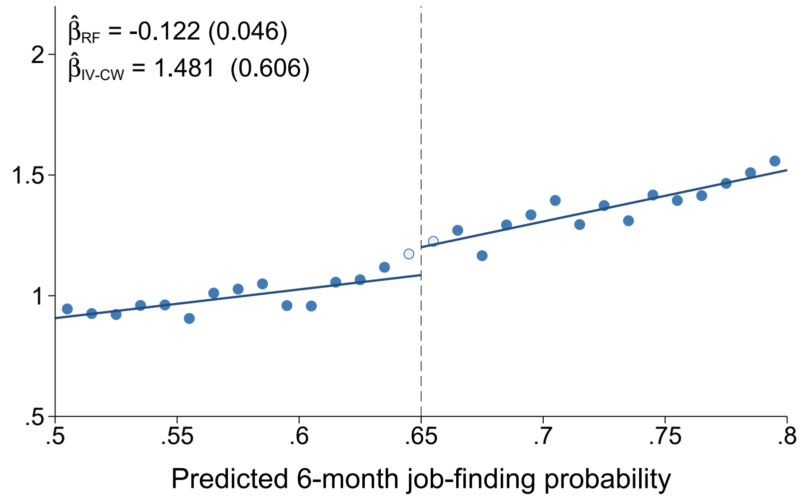


Figure E2: Days Worked per Week (Weeks 20–31)

Note: This figure plots days worked per week in weeks 20–31 of unemployment. The outcome is residualized on a set of controls and added to the mean. The sample is restricted to job seekers for whom the service line receives a risk score after 5 weeks of unemployment, who go on to complete at least 19 weeks of unemployment and who begin unemployment at least 31 weeks before 5 January 2023. Otherwise, the same sample restrictions, regression specifications and set of controls are used as in the baseline regression. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

E.2 Dynamics RD: Kernel and Bandwidth

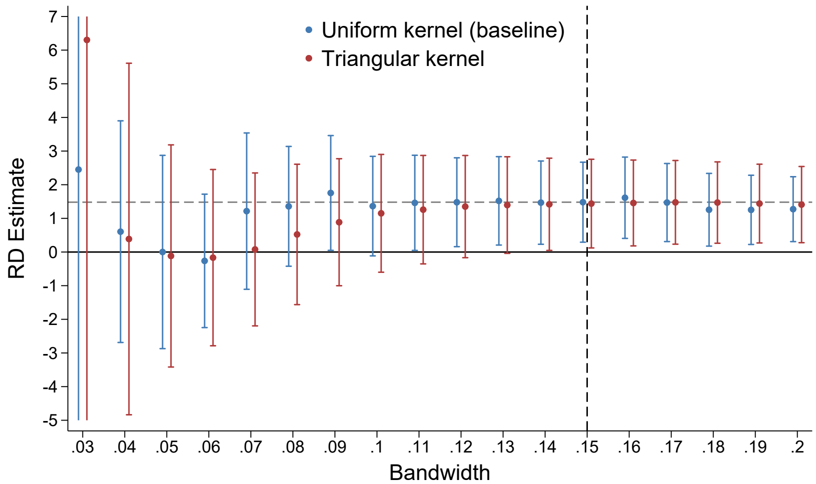


Figure E3: Estimates by Kernel and Bandwidth

Note: This figure plots IV estimates of the effect of caseworkers on days worked per week over weeks 20–31 of unemployment estimated using the dynamics RD with different kernels and bandwidths. The vertical line represents the bandwidth used for our baseline specification. The horizontal gray line marks our baseline estimate. Aside from kernel and bandwidth, the same sample restrictions, regression specifications and set of controls are used as in the baseline regression. We display 95% confidence intervals for the estimated effects.

E.3 Baseline and Dynamics RDs: Alternative Horizons

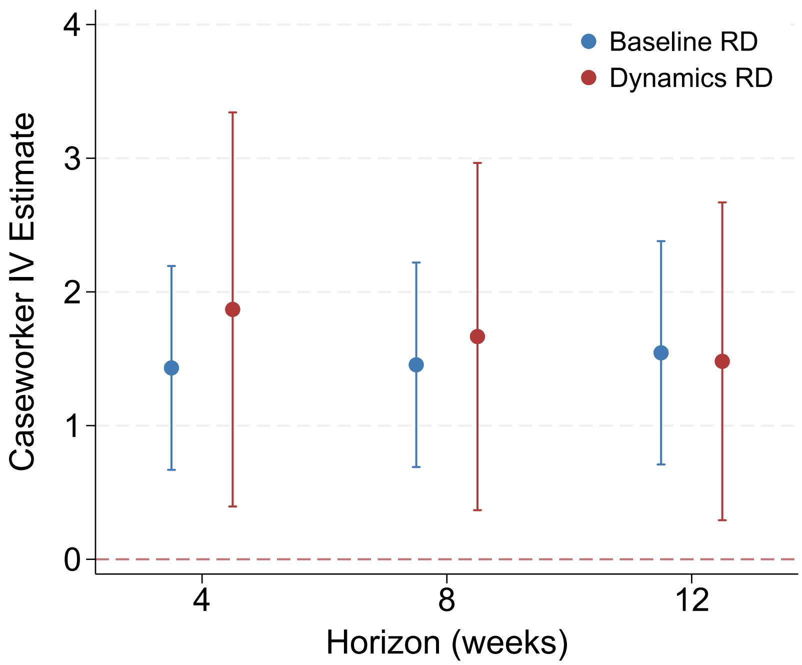


Figure E4: Estimates by Horizon

Note: This figure plots IV estimates of the effect of a caseworker on days worked per week estimated using the baseline RD and the dynamics RD. The effects are estimated over three horizons: 4 weeks, 8 weeks and 12 weeks (the baseline). We restrict the estimation sample uniformly for all estimates to spells starting between 1 September 2018 and 2 June 2022. Otherwise, the same sample, specifications and control variables are used as for the baseline estimates. We display 95% confidence intervals for the estimated effects.

E.4 Dynamics RD: Balance Tests

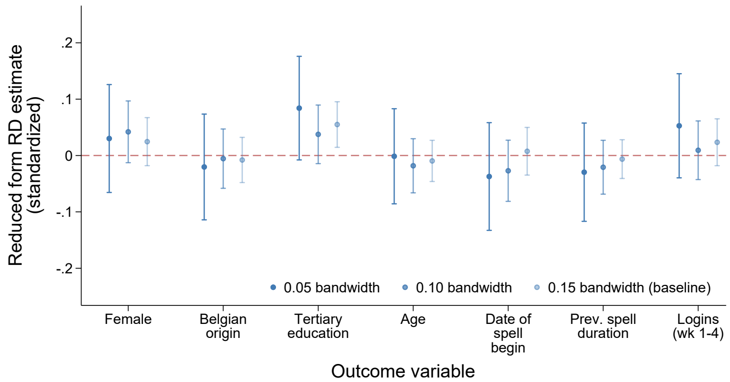
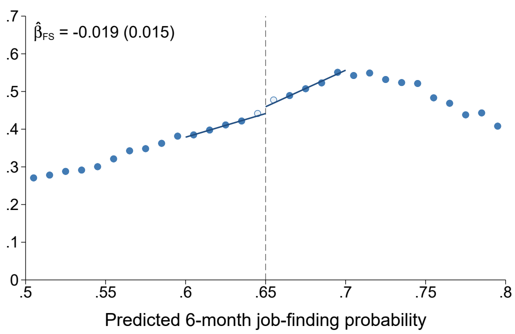
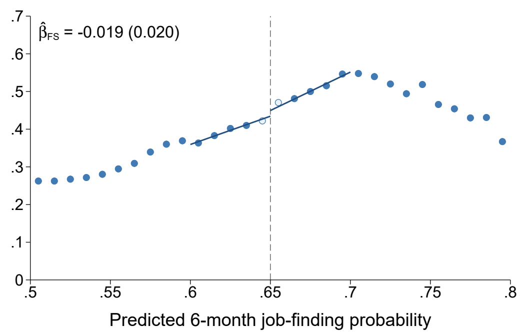


Figure E5: Balance Test

Note: This figure shows RD estimates of the effect of being below the cutoff on days worked per week alongside a set of pre-determined covariates. All outcomes are standardized. The same sample restrictions, regression specifications and set of controls are used as in the baseline dynamics RD regression. We show estimates using three bandwidth sizes and plot 95% confidence intervals.



(a) Full Sample



(b) Untreated at Week 20

Figure E6: Balance Test (CATEs)

Note: This figure plots predicted CATE by risk score and reports RD estimates for the effect of being below the cutoff. The same sample restrictions, regression specifications and set of controls are used as in the baseline dynamics RD regression, with two exceptions. First, a bandwidth of 0.05 is used rather than the baseline of 0.15, due to clear nonlinearities in the relationship between risk scores and CATEs. Second, in panel (b), the sample is restricted to individuals who did not meet with a caseworker in the first 19 weeks. Standard errors, reported in parentheses, are clustered at the individual level and are calculated using the nearest-neighbor variance estimator.

E.5 Dynamics RD: Program Participation

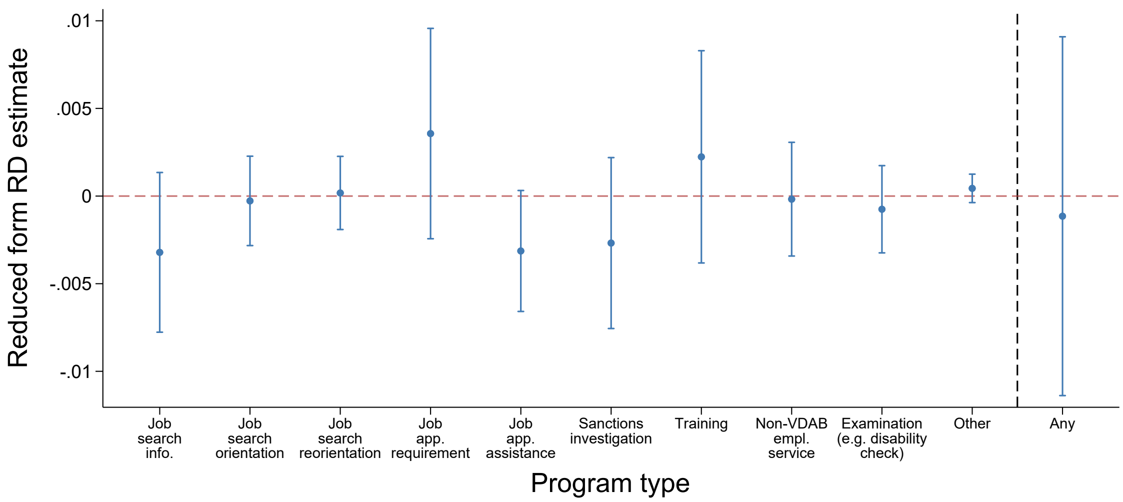


Figure E7: Effect on Program Participation (Weeks 20–31)

Note: This figure shows the reduced form effect of being *above* the cutoff on commencing participation in various programs in weeks 20–31 of unemployment. The same sample, specifications and controls are used as in the dynamics RD. We report 95% confidence intervals.

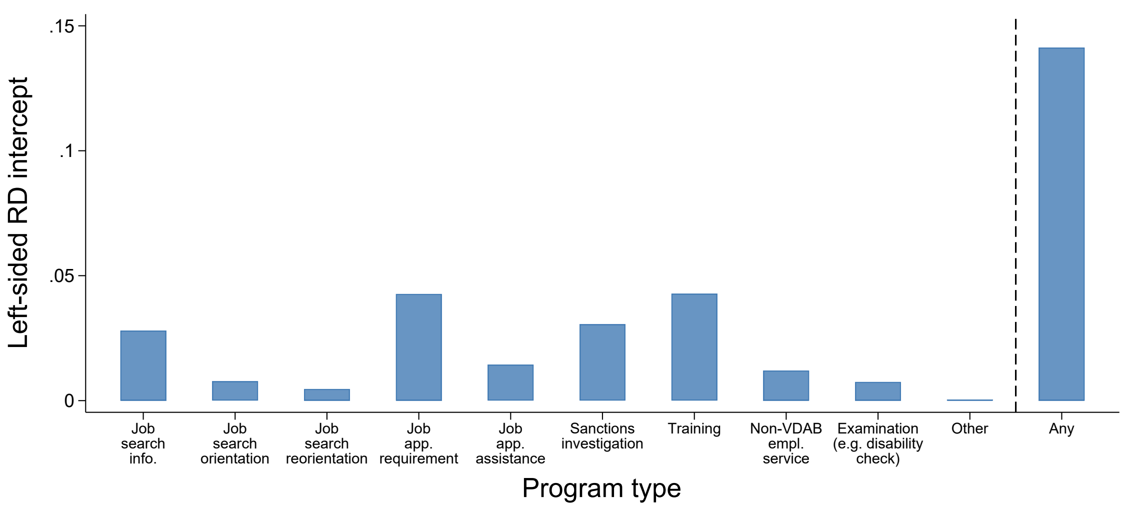
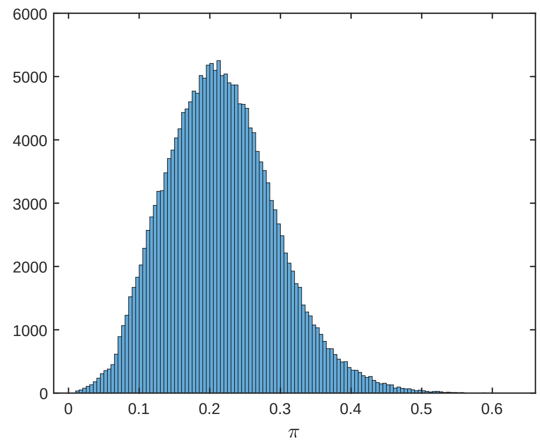


Figure E8: Program Participation Below the Threshold (Weeks 20–31)

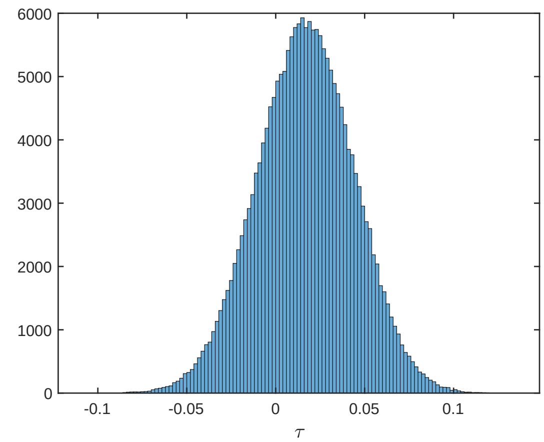
Note: This figure shows the left-sided intercept with the threshold, estimated using the dynamics RD, where the outcome is participation in various programs in weeks 20–31 of unemployment. The same sample, specifications and controls are used as in the dynamics RD. We report 95% confidence intervals.

F Calibration and Counterfactual Analysis

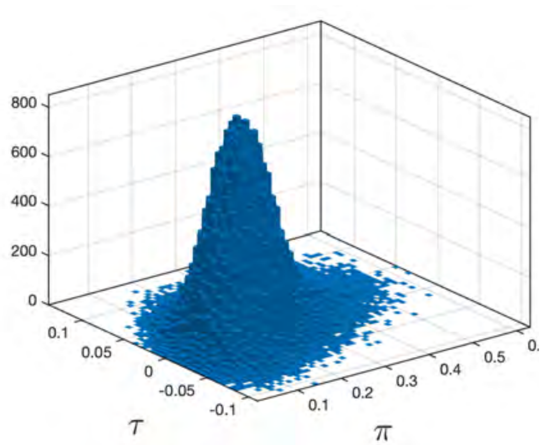
F.1 Additional Figures and Tables



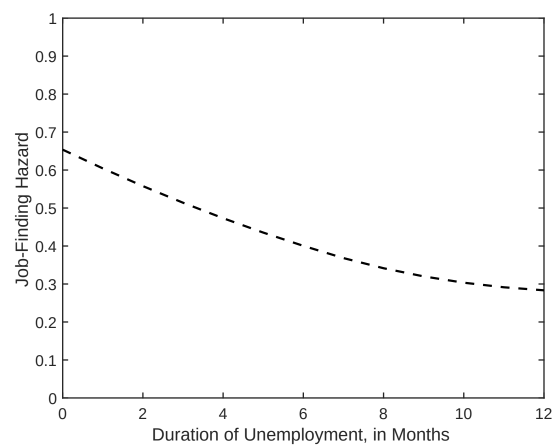
(a) Histogram of π



(b) Histogram of τ



(c) Histogram of π and τ



(d) Duration Dependence in Job Finding

Figure F1: Distribution of π and τ and Duration Dependence in Job Finding in Calibrated Model

Notes: The histograms show the distributions of the 200,000 individuals sampled for our model calibration.

Table F1: Shapley Decomposition of the Variance of the Targeting Index τ/π

$\text{Var}(\tau/\pi)$	0.0305
$V_\tau = \text{Var}(\tau \cdot \text{E}[1/\pi])$	0.0223
$V_\pi = \text{Var}(\text{E}[\tau] \cdot (1/\pi))$	0.0028
φ_τ	0.0250
φ_π	0.0055
Share φ_τ	0.821
Share φ_π	0.179

Notes: Based on model simulations with 200,000 individuals from the baseline calibration, using raw treatment effects (no floor or sample restriction). π is the one-month job-finding probability at the start of the unemployment spell absent treatment (observed part), and τ is the treatment effect of the observed type: the increase in the one-month job-finding probability in the month of treatment, with treatment at the start of the spell. V_τ fixes $1/\pi$ at its mean so that only τ varies, and V_π fixes τ at its mean so that only π varies. The Shapley contributions $\varphi_\tau = (V_\tau + \text{Var}(\tau/\pi) - V_\pi)/2$ and $\varphi_\pi = (V_\pi + \text{Var}(\tau/\pi) - V_\tau)/2$ sum exactly to $\text{Var}(\tau/\pi)$; the Share rows divide them by $\text{Var}(\tau/\pi)$.

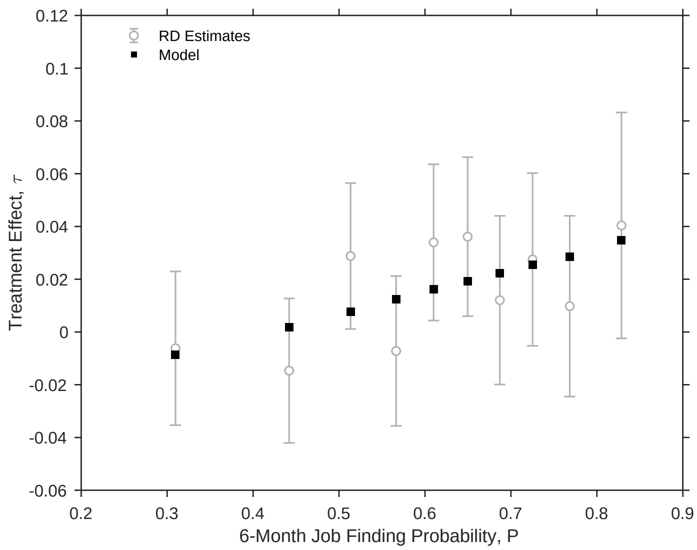
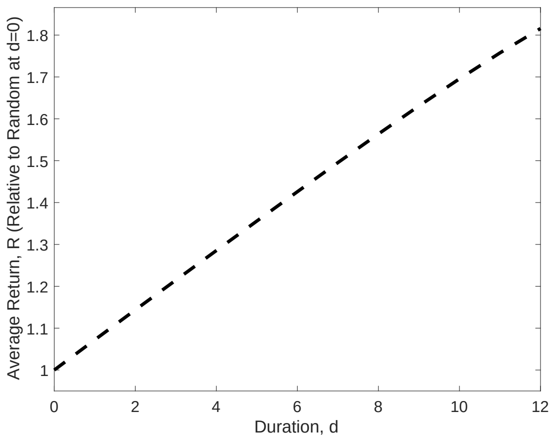
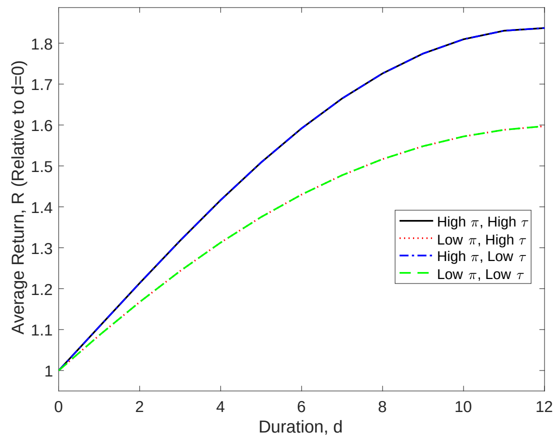


Figure F2: Treatment Effects by Decile of P Distribution, RD Estimates vs. Model

Note: The outcome variable for the RD estimates is the probability of working within the next four weeks.



(a) Random Targeting at d vs. Random at $d = 0$



(b) Targeting at d vs. $d = 0$, By Type

Figure F3: Welfare Gains of Targeting, By Duration

Note: Panel (a) plots the welfare gains from random targeting of surviving job seekers at duration d, relative to random targeting at $d = 0$. Panel (b) plots the welfare gains from targeting a given type at duration d, relative to targeting the same type at $d = 0$, for four types defined by combinations of high and low values of the baseline job-finding probability π and the treatment effect τ. High and low values correspond to the 75th and 25th percentiles of the respective calibrated distributions.

F.2 Algorithmic Bias

To assess algorithmic bias and its effect on the correlation between risk scores (P) and treatment effects (τ), we use the treatment effects of our IV estimates and treat 53% of the sample. For this exercise, we assume that the observed risk score (P) incorporates treatment effects. We then infer the underlying baseline job-finding probability π and construct a counterfactual risk score ($\widetilde{RS}$) as the 6-month job-finding probability in the absence of treatment.

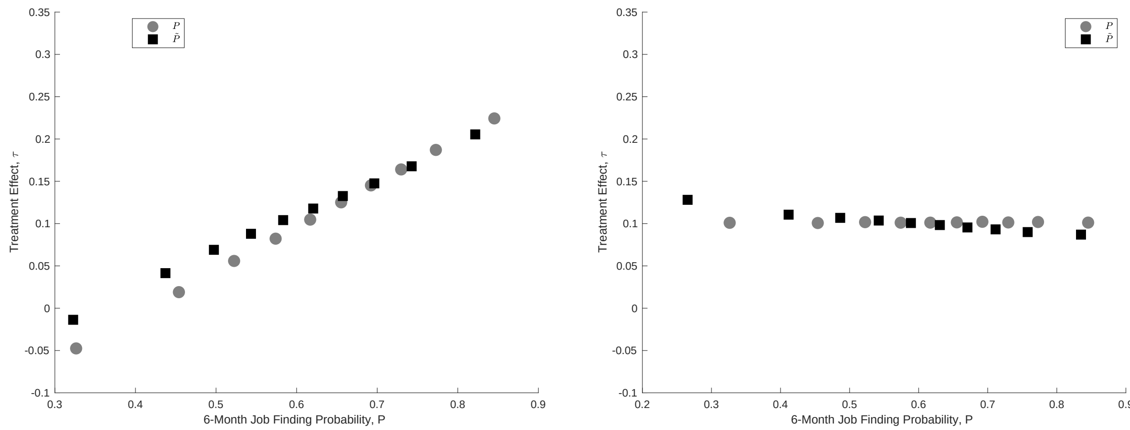
Table F2 and Figures F4 and F5 report the simulation results. In Panel A of the table and in Figure F4, treatment occurs in period 2 of the unemployment spell, and we report various moments for the baseline risk score P and the treatment-exclusive risk score $\widetilde{P}$. The last two columns of the table are the most relevant: they report the bivariate regression coefficient of τ on the risk score, which is the moment to which we calibrate the model. The results confirm the presence of a bias: the coefficient falls from 0.53 when using P to 0.43 when using $\widetilde{P}$. A similar bias arises in a version of the model with no correlation between the observed risk score and the treatment effects. The binned scatter plots in Figure F4 illustrate these patterns and show that, despite the bias, the positive relationship between the counterfactual risk score and the treatment effect remains strong in the baseline calibration.

Panel B of the table and Figure F5 confirm that these patterns are robust to treating individuals with $P < 0.65$ in month 2 and those with $P \geq 0.65$ in month 5, as in the Belgian policy. The results are very similar to those in Panel A and Figure F4.

Table F2: Algorithmic Bias Based on Model Simulations

	$\text{Cov}(P, \tau)$	$\text{Cov}(\widetilde{P}, \tau)$	$\text{Var}(P)$	$\text{Var}(\widetilde{P})$	$\text{Var}(\tau)$	$\frac{\text{Cov}(P,\tau)}{\text{Var}(P)}$	$\frac{\text{Cov}(\widetilde{P},\tau)}{\text{Var}(\widetilde{P})}$
Panel A: Treatment at 2 months							
Baseline Calibration	0.0118	0.0089	0.0226	0.0206	0.0130	0.5251	0.4304
Calibration with cov(P,τ)=0	0.0000	-0.0019	0.0226	0.0271	0.0068	0.0018	-0.0720
Panel B: Treatment at 2 months for $P < 0.65$ and at 5 months for $P \geq 0.65$							
Baseline Calibration	0.0118	0.0090	0.0226	0.0209	0.0130	0.5251	0.4328
Calibration with cov(P,τ)=0	0.0000	-0.0019	0.0226	0.0273	0.0068	0.0018	-0.0700

Notes: Based on model simulations with 200,000 individuals. P denotes the baseline risk score. $\widetilde{P}$ denotes the risk score without treatment at a given duration.

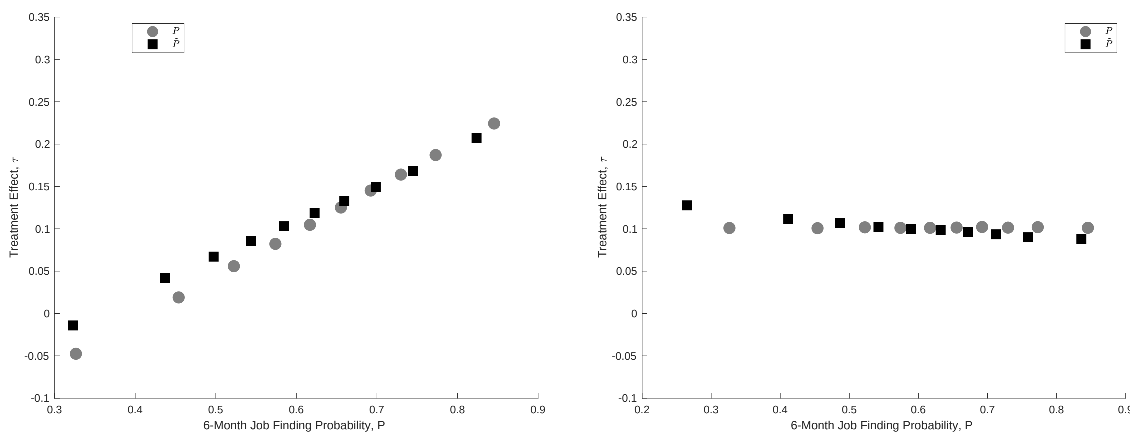


(a) Baseline Calibration

(b) Calibration with cov(P,τ)=0

Figure F4: Algorithmic Bias Simulations: Binned Scatter Plots (Treatment in Month 2)

Notes: Based on model simulations with 200,000 individuals. P denotes the baseline risk score. $\widetilde{P}$ denotes the risk score without treatment.



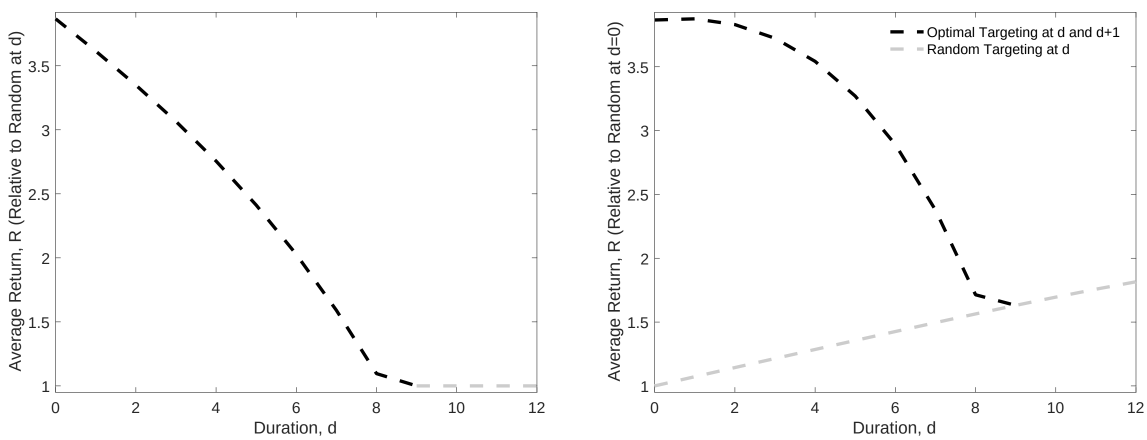
(a) Baseline Calibration

(b) Calibration with cov(P,τ)=0

Figure F5: Algorithmic Bias Simulations: Binned Scatter Plots (Treatment in 2 months for $P < 0.65$ and in 5 months for $P \geq 0.65$)

Notes: Based on model simulations with 200,000 individuals. P denotes the baseline risk score. $\widetilde{P}$ denotes the risk score without treatment.

F.3 Optimal Targeting With Targeting for Two Consecutive Periods



(a) Welfare Gain of Optimal Targeting at d and $d+1$ relative to Random Targeting at d

(b) Welfare Gain of Optimal Targeting at d and $d+1$ relative to Random Targeting at $d = 0$

Figure F6: Optimal Targeting when Targeting in Two Consecutive Periods

Figure F6 shows that allowing the policymaker to target individuals in two consecutive periods—rather than just one—increases the welfare gains from optimal targeting, holding the budget fixed (see Figure 9 for comparison for the optimal single period targeting). The reason is that the unemployment agency can now reach the highest-return individuals multiple times.

F.4 Robustness to Redistributive Welfare Weights

Our baseline welfare criterion is utilitarian and linear in unemployment losses: the planner values a reduction in expected losses equally across job seekers. To assess robustness to redistributive preferences, we replace it with a utilitarian criterion over concave transformations of individual outcomes, $V_i = \log(\bar{L} - L_i)$, where L_i denotes the expected unemployment loss of job seeker i and $\bar{L}$ is a welfare anchor calibrated below. The associated marginal welfare weight, $g_i = 1/(\bar{L} - L_i)$, increases with expected losses, so that employment gains accruing to job seekers with worse labor market prospects receive greater weight in the social objective.

Two implementation choices discipline this criterion. First, we anchor $\bar{L}$ to the marginal-utility gap between employment and unemployment: $\bar{L}$ is set such that the average unemployed job seeker — with expected loss $E[L]$ in the absence of targeting — carries twice the marginal welfare weight of an employed job seeker, $g(E[L])/g(0) = \bar{L}/(\bar{L} - E[L]) = 2$, which implies $\bar{L} = 2\,E[L]$. A twofold gap in marginal utilities is well within the range of direct evidence on the marginal value of resources in unemployment relative to employment: for Sweden, Landais and Spinnewijn [2021] estimate a marginal rate of substitution between unemployment and employment consumption of at least 1.6 using differences in marginal propensities to consume, and of 3.1 on average using revealed preferences from supplemental UI choices, while standard consumption-based calculations imply values of 1.1–1.5. Second, we cap the marginal welfare weight at ten times the employed weight — five times the weight of the average unemployed: welfare is linear in L above the loss level at which the cap binds ($\frac{9}{10}\bar{L}$, around the 89th percentile of expected losses), so that weights are bounded between one and ten and the criterion remains concave and well defined at all loss levels. The cap is essential because the distribution of expected losses is highly skewed: about ten percent of simulated losses exceed $2E[L]$ itself, where log welfare is not defined, and the marginal weight diverges as losses approach $\bar{L}$, concentrating virtually all social welfare on the extreme tail. The cap delivers a bounded, interpretable weight schedule and makes the criterion insensitive to how the extreme tail of losses is modeled. No cap is needed at the bottom, where weights are bounded below by the employed weight. The welfare criterion is thus

$$V(L_i) \;=\; \begin{cases} \log\big(\bar{L} - L_i\big) & \text{if } L_i \le L_{cap}, \\ \log\big(\bar{L} - L_{cap}\big) - \dfrac{L_i - L_{cap}}{\bar{L} - L_{cap}} & \text{if } L_i > L_{cap}, \end{cases} \qquad \bar{L} = 2\,E[L], \quad L_{cap} = \tfrac{9}{10}\,\bar{L},$$

which is continuous and concave, with marginal welfare weight $g(L_i) = \min\big\{1/(\bar{L} - L_i),\, 10/\bar{L}\big\}$, so that $g(L_i)/g(0) = \min\big\{\bar{L}/(\bar{L} - L_i),\, 10\big\}$ rises from one for the employed through two at the average loss to at most ten.

Welfare under the weighted criterion is computed exactly, without linearization, from the same simulations as in the baseline. For each individual — more precisely, for each draw of the unobserved heterogeneity within an observed type, each of which represents a distinct person — we simulate the expected present-value unemployment loss with and without treatment, L_i^1 and L_i^0: the survival-weighted sum of per-period utility losses over the spell, including the continuation loss at the model horizon. The welfare gain from treating person i is the exact difference $\log(\bar{L} - L_i^1) - \log(\bar{L} - L_i^0)$, rather than the first-order approximation $g_i\left(L_i^0 - L_i^1\right)$ with the weight evaluated at the untreated allocation; because the marginal weight declines as treatment reduces losses, the approximation would overstate the gains of high-loss individuals. Treatment costs remain in units of resources, so the return to a type is its expected welfare gain per unit of expected treatment cost, integrated over the draws of unobserved heterogeneity. Applying the concave transformation at the level of individual draws, rather than to the average loss of an observed type, matters for the same reason the criterion is redistributive in the first place: the planner values the distribution of welfare across persons, and averaging losses within observed types before applying the transformation would understate the dispersion that the welfare weights are meant to price.

We re-solve the optimal targeting policies under the weighted criterion: the characterization in Proposition 1 applies verbatim with returns defined as welfare gains per unit of treatment cost, and the simulation ranks individuals (and treatment durations) by these transformed returns. The definitions of the index policies (τ/π, τ only, π only) are unchanged; only their evaluation uses the weighted criterion, as does the random-targeting benchmark to which all welfare gains are normalized.

Table F3 repeats the sensitivity analysis of Table 2 under the weighted criterion. The conclusions are unchanged across all scenarios. Most importantly, redistributive preferences do not close the gap between risk-score targeting and the optimal policy. The welfare weights reward reaching job seekers with poor prospects, and risk-score targeting accordingly gains in absolute terms relative to the linear criterion; but the optimal policy benefits from the weights at least as much, since it reaches high-loss job seekers *and* selects the responsive ones among them. In the baseline, risk-score targeting realizes only 14 percent of the welfare gains of optimal targeting, down from 25 percent under the linear criterion, while the sufficient-statistic rule continues to capture nearly all of the optimal gains (97 percent). The optimally targeted population overlaps 94 percent with that under the linear criterion. Risk-score targeting narrows its gap to the optimal policy only in the scenarios that weaken the link between risk scores and treatment responsiveness, and only in the absence of any correlation (column 5) does it come close, attaining 80 percent of the optimal gains.

The last column of Table F3 replaces the log criterion with an exponential (CARA) transformation, $V_i = -\exp(\alpha L_i)/\alpha$, without any cap on the welfare weights. CARA welfare is defined at all loss levels and strictly concave throughout, and because marginal-weight ratios under CARA depend only on loss *differences*, the anchor pins down the single parameter directly: $\exp(\alpha E[L]) = 2$ implies $\alpha = \ln 2/E[L]$, with no reference level $\bar{L}$ and no cap needed on feasibility grounds. Weights then grow without bound in the tail, doubling with every additional $E[L]$ of losses, to roughly 900 times the employed weight at the largest simulated loss. The results confirm the robustness of the main conclusions while illustrating the role of the cap. The comparison to the optimal policy is essentially unaffected: the sufficient-statistic rule still captures 97 percent of the optimal gains, the optimally targeted population still overlaps 95 percent with the linear criterion, and risk-score targeting still forgoes the bulk of the attainable gains, realizing 17 percent of the welfare gains of optimal targeting. Only the comparison to random targeting changes, as the unbounded weights concentrate social welfare on the extreme tail of losses that risk-score targeting happens to reach; its distance to the optimal and the sufficient-statistic policies remains wide.

Table F3: Welfare Gains of Targeting Under Redistributive Welfare Weights, Across Alternative Calibration Scenarios

			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Linear (baseline)	Baseline	$\sigma_{\varepsilon_\pi} = 0$ $\theta_\pi = 1$	$\sigma_{\varepsilon_\tau} > 0$	$\theta_\tau < 1$	$\tau(IV)$	$corr(\pi, \tau) = 0$	Lower σ_{e_τ}	$\theta_u > 1$	At $d = 25$: no continuation loss	Target 30%	CARA (no cap)
Panel A. Targeting at $d = 0$												
Optimal Targeting vs. Random	2.85	6.55	12.11	4.21	6.55	2.85	2.22	6.55	7.14	3.97	10.37	9.90
Sufficient Statistics Targeting (τ/π) vs. Random	2.84	6.35	11.95	3.93	6.35	2.55	2.19	6.15	6.90	3.89	9.11	9.61
Targeting High Treatment Effects (τ) vs. Random	2.76	5.79	11.13	3.54	5.79	2.02	2.07	5.22	6.26	3.67	6.08	7.61
Targeting Low Risk Scores (π) vs. Random	0.70	0.89	-1.50	1.28	0.89	1.10	1.78	0.54	0.94	0.87	0.36	1.68
Panel B. Targeting at $d > 0$												
Optimal (Fixed Share) at $d = 3$ vs. $d = 0$	1.33	1.55	1.14	1.60	1.23	1.51	1.64	1.52	1.62	1.33	1.57	1.71
Optimal (Fixed Budget) at $d = 3$ vs. $d = 0$	0.43	0.25	0.66	0.44	0.17	0.55	0.75	0.22	0.25	0.32	0.93	0.40
Random at $d = 3$ vs. $d = 0$	1.21	1.64	-0.01	1.87	1.09	1.57	1.66	1.42	1.81	1.27	1.64	3.96
Panel C. Targeting Across Durations												
Global Optimal Targeting vs. Random at $d = 0$	2.88	6.58	12.11	4.57	6.55	2.89	2.22	6.61	7.18	3.98	10.54	9.91
Fraction Targeted at $d = 0$	0.85	0.85	1.00	0.80	0.96	0.70	0.87	0.79	0.81	0.87	0.61	0.85

Notes: This table repeats the sensitivity analysis of Table 2 with all policies evaluated — and the optimal policies re-solved — under the concave social welfare weights described in this appendix section: log welfare with the anchor $\bar{L}$ set such that the average unemployed job seeker carries twice the marginal welfare weight of an employed job seeker, and weights capped at ten times the employed weight. The first two columns report the baseline calibration under the linear criterion of Table 2 and under the welfare weights, respectively. Columns (1)–(9) apply the welfare weights to the calibration scenarios defined in columns (1)–(9) of Table 2. Column (10) reports the baseline calibration under the uncapped exponential (CARA) criterion with the same anchor, described in the text of this appendix section. The panels are defined as in Table 2, and welfare gains in each column are expressed relative to random targeting evaluated under the same welfare criterion. Negative entries indicate negative aggregate welfare gains of the respective policy.