The Scandinavian
Journal of Economics

Scand. J. of Economics 118(4), 941–974, 2016
DOI: 10.1111/sjoe.12166

Do Lower Caseloads Improve the Performance of Public Employment Services? New Evidence from German Employment Offices*

Jens Hainmueller

Stanford University, Stanford, CA 94305-6044, USA
jhain@stanford.edu

Barbara Hofmann

Institute for Employment Research, DE-90478 Nuremberg, Germany
barbara.hofmann@iab.de

Gerhard Krug†

Institute for Employment Research, DE-90478 Nuremberg, Germany
gerhard.krug@iab.de

Katja Wolf

Institute for Employment Research, DE-90478 Nuremberg, Germany
katja.wolf@iab.de

Abstract

The caseworker-to-clients ratio is an important, but understudied, policy parameter that affects both the quality and cost of public employment services that help job seekers find employment. We exploit a large-scale pilot by Germany's employment agency, which hired 490 additional caseworkers in 14 of its 779 offices. We find that lowering caseloads caused a decrease in the rate and duration of local unemployment as well as a higher re-employment rate. Disentangling the mechanisms that contributed to this improvement, we find that offices with lowered caseloads increased monitoring and imposed more sanctions but also intensified search efforts and registered additional vacancies.

Keywords: Caseworker; labor market policy; statistical matching; unemployment

JEL classification: *C*14; *H*43; *H*83; *J*68

I. Introduction

Active labor market policies (ALMP) are widely used in many countries to help unemployed workers find jobs, but there is mixed empirical evidence about the effectiveness of these policies. Some studies show that public

*We thank Michael Rosholm, Julia Schneider, Jeff Smith, and participants at an ASB/IAB workshop in Nuremberg, as well as JSM and EALE conferences, for helpful comments. We are grateful to Frank Sowa and Stefan Theuer for sharing information about the implementation of the pilot project. The usual disclaimer applies.

†G. Krug is also affiliated with University of Erlangen-Nuremberg (gerhard.krug@fau.de).

intervention in the labor market to match the unemployed with employers can help job seekers to find work more quickly (Yavas, 1994; Fougère *et al.*, 2009). In contrast, other studies suggest that ALMP can also result in inefficiencies, if, for example, the formal monitoring of job search efforts by caseworkers crowds out the informal private search by job seekers (van den Berg and van der Klaauw, 2006).

One of the most important unresolved questions about ALMP is how the behavior of caseworkers in public employment offices affects the re-employment chances of job seekers. Caseworkers play a crucial role in most labor market policies because they work directly with job seekers and try to help them find new employment. Differences in the quality, work conditions, and training of caseworkers can therefore affect the success of such policies (e.g., Behncke *et al.*, 2008; Rinne *et al.*, 2013). Consistent with this idea, Lagerström (2011) finds that caseworkers in Swedish employment offices vary dramatically in their effectiveness at bringing their clients back into regular employment.

Why is it that some caseworkers are more effective than others? Very few studies shed light on this important issue. For example, using data from Switzerland, Behncke *et al.* (2010) find that caseworkers who follow a less cooperative and less harmonious approach towards the unemployed increase the employment chances of their clients. In this study, we examine the effect of another potentially important factor: caseload (i.e., the ratio of caseworkers to unemployed clients). Caseload is an important policy parameter for at least two reasons. On the one hand, caseload influences the effectiveness of the assistance because it dictates how much time and effort a caseworker can devote to each client. On the other hand, the caseload is also a key driver of the administrative costs of the policy, as lower caseloads require that public employment offices hire more caseworkers. Despite the importance of this question for public policy, we have almost no empirical evidence about the effects of caseload on key indicators such as the unemployment rate, the duration of unemployment, and the outflow rate from unemployment to regular employment.

Our study contributes new evidence by drawing upon a large-scale pilot project of Germany's federal public employment agency (Bundesagentur für Arbeit, BA), which lowered the caseload in 14 of its 779 local employment offices in May 2007. The ratio of caseworkers to recipients of unemployment insurance (UI) benefits was set to 1:40 in the pilot offices, while it was about 1:100 in the non-participating offices (measured as full-time equivalents). Although not randomized, the BA chose the participating offices based on well-documented criteria that were mostly designed to achieve a representative sample. As a result, participating offices were fairly similar to non-participating offices prior to the pilot project. Our empirical strategy relies on a combination of matching and a

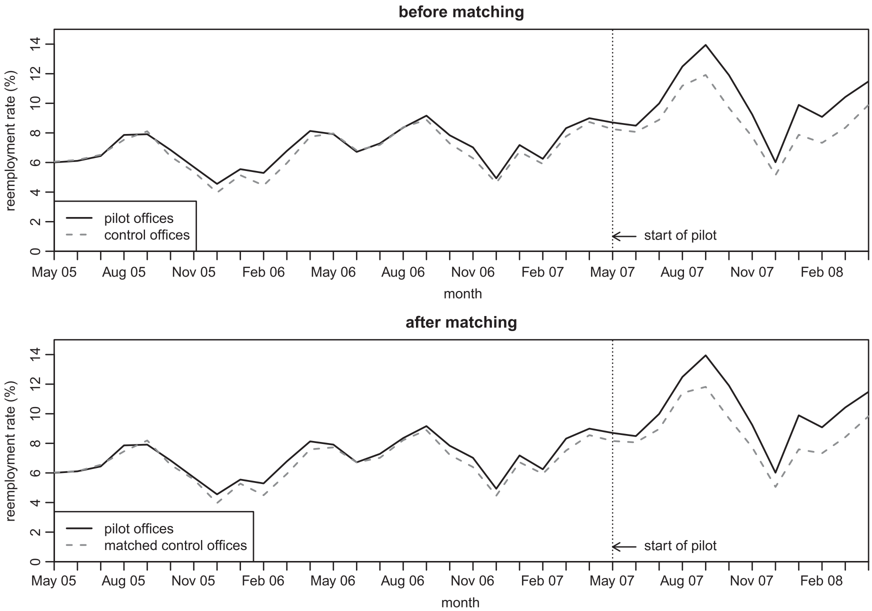

Fig. 1. Monthly re-employment rate before and after start of pilot project in pilot and control employment offices (May 2005–April 2008)

Notes: Monthly re-employment rate: number of re-employed individuals per month divided by stock of unemployed at the end of the preceding month.

Source: Calculations based on micro data of the German Federal Employment Agency (see text for details).

difference-in-differences (DiD) estimator to isolate the causal effects of the caseload decrease on several outcomes, including the unemployment rate, the cumulative unemployment duration, and the re-employment rate. We also conduct various robustness checks to corroborate the main findings, including tests for differential trends in the pre-program period. Moreover, we consider two intermediate outcomes (sanction rates and number of new vacancies registered) to shed light on the potential causal mechanisms through which caseload affects outcomes. We also consider potential negative side effects such as spillover into neighboring regions. Finally, we analyze the cost effectiveness of the pilot program to obtain a broader measure of the policy returns.

Overall, we find that the pilot project led to an improvement in the performance of participating local employment offices. Lowering of caseloads resulted in a decrease in the duration and rate of local unemployment and an increase in the re-employment rate. To preview the main result and identification strategy, the upper panel in Figure 1 shows the average monthly re-employment rate in the group of pilot offices and the comparison group of non-participating offices over a three-year period, including the two years prior to the start of the pilot in May 2007 and the year following the start of the pilot. The re-employment rate measures the share of BA clients who leave unemployment for unsubsidized regular employment; it is computed based on administrative micro data from the integrated employment biographies that capture re-employment based on unemployment records and subsequent employment records. We see that despite small differences in the level, there are no meaningful divergences in the re-employment trends in pilot and comparison offices for the two years before the pilot started. This supports our DiD identification assumption that pilot offices were not chosen strategically based on upward or downward pre-program trends, and therefore it is plausible that they would have followed parallel trends with the control offices in the absence of the pilot program. We also see that a sizable divergence in the monthly re-employment rates emerges shortly after the program started, as pilot offices show re-employment rates that are, on average, about 1 to 1.5 percentage points higher than in the control offices. This difference amounts to about an 11–15 percent increase over the average re-employment rate. The lower panel in Figure 1 shows that the results are very similar once we restrict the comparison group to a set of matched control offices that we matched on a comprehensive set of pre-treatment covariates to further improve the comparability. Overall, these findings show that lowering the caseload through the hiring of additional caseworkers led to considerably more re-employment of the unemployed into regular employment.

Disentangling the causal pathways that led to this improvement, we also find that lowering the caseload led to more proactive behavior in pilot offices. Compared to the control offices, pilot offices imposed more sanctions (e.g., on clients with low search effort) and registered more new vacancies. We also find that the pilot project did not result in negative spillovers into neighboring regions. If anything, the results suggest that neighboring employment offices benefited from the additional vacancies registered by caseworkers in the pilot offices, as all offices share a common vacancy database. Addressing cost effectiveness, we find that the costs imposed by the hiring of additional caseworkers were offset by the savings from increased effectiveness: net of additional salary costs, the pilot offices' UI benefit expenditures decreased by around 3.4 percent, and their clients' average earnings increased by around 6.7 percent.

II. The German Public Employment Service and the Pilot Project

Background

The pilot program we examine was conducted by Germany's federal public employment agency, which is in charge of three different groups of clients.

The first and largest group of clients is entitled to UI benefits, and clients in this group are registered either as unemployed or as participating in ALMP measures. A second group of clients consists of unemployed individuals who do not receive UI benefits (e.g., women who want to rejoin the labor force after long-term maternity leave). A third group consists of employed job seekers (i.e., individuals who are still employed but anticipate a transition into unemployment in the near future).[1] BA offices do not counsel welfare benefit recipients who receive means-tested unemployment benefits and who are typically long-term unemployed.

The BA is organized into 10 regional directorates, which contain 178 employment agencies. These employment agencies are further divided into 779 local employment offices, and this lowest level of aggregation is our unit of analysis. Each employment agency designates one of its local employment offices as its head office, but head offices only differ slightly in their structure from regular local employment offices. For example, the human resources departments are commonly located in the head office. Job seekers cannot choose the employment offices themselves but must register at the closest employment office. As of December 2006, the BA employed around 10,800 caseworkers, corresponding to an average caseload of around 140 recipients of UI benefits per caseworker.

The Pilot Project

In December 2006, the BA decided to run a large-scale pilot project, which consisted of a substantial increase in the number of caseworkers. The aim of the pilot project was to find out whether a lower caseload per caseworker leads to more re-employment and shorter unemployment durations. In particular, 490 additional caseworkers were placed in 14 out of the 779 local employment offices. We refer to these 14 offices as treated (or pilot) offices. From the group of non-treated offices, we drop some regions, as described in Section IV, and refer to the remaining offices as the control group.

On average, the number of additional caseworkers in pilot offices was about 36 per office, which amounts to about a 160 percent increase in the stock of caseworkers. The increases ranged from a minimum of 82 percent to a maximum of 340 percent (8–37 caseworkers) in one of the smaller treated employment offices. The hiring and training of the additional caseworkers started in late January and lasted until the end of April 2007. Unfortunately, we do not have systematic data on caseworker

[1] Note that individuals are required to register as job-seeking either three months before they become unemployed in the case of a temporary job, or as soon as possible if they are laid off.

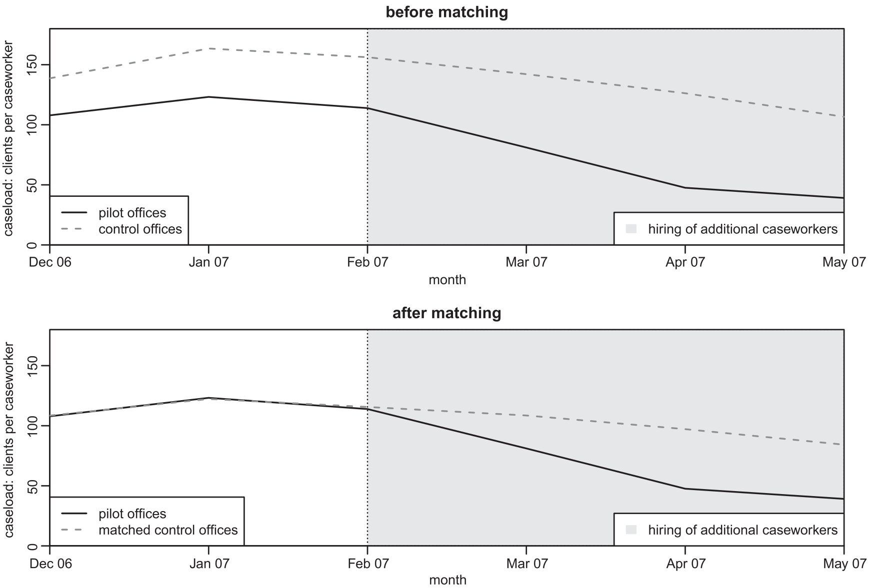

Fig. 2. Monthly caseload in pilot and control offices before the start of the pilot project (December 2006–May 2007)

Notes: Monthly caseload: number of caseworkers in full-time equivalents divided by the number of benefit recipients per local employment office. The upper panel refers to the raw data comparing the pilot offices with all other employment offices, and the lower panel refers to the matched data that compares the pilot offices with its nearest neighbors, based on the matching detailed in the text.

Source: German Federal Employment Agency.

characteristics. However, evidence from a companion qualitative research project suggests that the majority of new caseworkers were hired from the private sector or directly from university and did not have prior job experience in employment offices. All new caseworkers were trained for about four to 12 weeks. The type of training the new caseworkers received varied considerably, ranging from on-the-job training after a two-week computer course to an intensive training course over several weeks followed by mentoring from experienced caseworkers.

The goal of the pilot program was to achieve a caseload of one caseworker per 70 recipients of UI benefits by May 2007. However, because of a significant decrease in nationwide unemployment in the first months of 2007, the actual average caseload at the official start of the pilot project was 1:40 in the treated local employment offices and 1:100 in control offices.[2] The upper panel of Figure 2 depicts the average monthly caseload from December 2006 to May 2007 in the pilot offices compared to the

[2] Taking all BA clients (i.e., not only recipients of UI benefits) into account, the caseload was 1:80 and 1:170 in the pilot project and control offices, respectively.

control offices. The trends highlight three important aspects. First, even without any sample adjustments, the caseload in the pilot and control offices follows quite parallel trends in the pre-program period up to February 2007, when the hiring of additional caseworkers began. This lends initial support for our DiD strategy, which relies on the idea that the two groups would have continued on parallel trends in the absence of the treatment. Second, as a result of the hiring of additional caseworkers, the caseloads drop much more rapidly in the pilot offices compared to the control offices from March until May. This is the differential variation in caseload that we exploit for the identification. Third, even though the two groups exhibit parallel trends in the pre-program period, the caseloads in pilot offices were slightly lower, on average, compared to the control offices in December 2006, when the pilot was decided. However, after controlling for the covariates, any significant differences in the caseloads before the start of the pilot project disappear. This is shown in the lower panel of Figure 2, which depicts the average monthly caseloads in the pilot offices and a set of matched control offices that we identified based on our nearest-neighbor matching strategy that adjusts for a wide range of background covariates (below we describe the details of the matching). After the matching, the treated and control groups now follow virtually identical trends in the pre-program period.

The BA used a set of well-documented criteria to select offices for the pilot project, which gives us inferential leverage to control for the selection based on variables that proxy for the selection criteria. The first and most important criterion was regional dispersion: at least one local employment office had to be chosen from each of the 10 regional directorates. Figure A1 in the Appendix depicts the locations of the pilot offices and confirms this geographic restriction. This balanced geographic dispersion of pilot offices is beneficial for the analysis because it increases the regional overlap between treated and control offices. The second criterion involved the structure of the employment offices: both head and regular employment offices had to be represented among the pilot offices. The third criterion was a pragmatic one: pilot offices had to be able to provide facilities to accommodate the additional staff. Fourth, labor market conditions played a role in the selection process. In order to test the lowering of caseloads in a representative sample of employment offices, the BA chose to obtain a stratified sample by restricting the choice of pilot offices to six of 12 subclasses, following a classification of employment agencies developed by Blien *et al.* (2004). This classification groups employment offices into 12 different subclasses based on their labor market conditions. While the six subclasses that were chosen for the pilot program span a wide range from "good" to "challenging" labor market conditions, these subclasses on average reflect slightly better conditions compared to the population of all

12 subclasses. Therefore, pilot offices might have had slightly better labor market conditions, on average, compared to all other offices, but not compared to other offices in the chosen subclasses. Finally, anecdotal evidence suggests that private information about the capacity of the local employment offices also played some, albeit limited, role in the selection. The direction of this last selection criterion is unclear. Based on the anecdotal evidence that we gathered from BA officials, it seems more likely that officials picked particularly "needy" offices (which faced a high share of individuals at risk of long-term unemployment) to help them with additional caseworkers, which could lead to a possible downward bias in the estimates. However, we also need to consider the possibility that when faced with a choice between offices that were equally suitable based on the four criteria detailed above, officials could have used their knowledge about the capacity to pick the more promising offices for the pilot project. Indeed, if treatment were implemented in areas where unemployment was expected to fall, that would threaten our identification strategy and would lead to an upward bias. Below, we conduct a variety of checks to investigate this possibility. For example, if such selection had happened, we should find that unemployment in the neighboring offices fell as well (recall that treatment was implemented at the smallest administrative level). Also, we should find that pilot offices and control offices exhibit non-parallel trends in unemployment in the run-up to the start of the program. We find no support for these conjectures, which supports the idea that selection was not biased towards upward or downward trending offices.

As part of the pilot project, the selected pilot offices received the additional caseworkers. Beside the decrease in the caseload, there were two additional changes in the pilot offices. First, they had to sign new target agreements, which were monitored by the BA using a separate performance management tool. The main performance indicators in these target agreements were the unemployment duration and the re-employment rate (Hofmann *et al.*, 2012). Second, the pilot offices had to raise the share of the caseworkers who are responsible for the demand side of the labor market (i.e., caseworkers who concentrate on acquiring job vacancies from firms) to at least 30 percent. In contrast, in the non-pilot offices, about 20 percent of caseworkers, on average, focused on the demand side. We consider these changes as part of our treatment. To the best of our knowledge, no other major changes accompanied the participation in the pilot project. In particular, the ALMP budget – not including the caseworkers' wages – was not affected by the pilot project, so participating offices did not receive a higher or lower budget to implement ALMP measures.[3]

[3] This was confirmed in interviews with BA officials and using administrative data on the employment agency level where the budget is set. For example, in the employment agencies

Literature Review and Expected Effects of Lower Caseloads

Only a few empirical studies have looked at the effect of caseloads on the success of labor market policy. Koning (2009) examines the relationship between the caseload and various outcome measures, such as the outflow rate for the unemployed and the benefit denial rate of UI using data from local employment offices in the Netherlands. The main finding is that lower caseloads have a positive effect on the outflow rate of the short-term unemployed, whereas the effect on the long-term unemployed is insignificant. Note that the identification in Koning (2009) relies on variation between employment offices and over time. In contrast to our study, he cannot take advantage of exogenous variation of the caseloads, and his results might therefore be biased. Hill (2006) studies the effect of different types of case management within welfare-to-work programs in the US. She finds that the caseload is negatively (but not significantly) related to earnings and significantly positively related to social benefit receipts over a two-year period. In other words, her results suggest that lower caseloads would lower benefit receipts. Similarly to Koning (2009), Hill (2006) uses regional variation without an exogenous change of the caseload for identification of the caseload effect.

In contrast, Schiel *et al.* (2008) find some evidence for positive effects of a lower caseload on the employment chances of the long-term unemployed in an experiment administered in four BA offices. Also, Jerger *et al.* (2001) find an increased employment probability, but no effect on the stability of their employment relationships, using observational data from a local pilot project for social benefit recipients in Mannheim, Germany.

Existing studies also tell us very little about the mechanisms through which changes in the caseload bring about the observed effects on labor market outcomes. Labor market policies are chiefly concerned with three functions: income support to job seekers during the time of unemployment; improvement of skills or employability through training programs; and job brokerage and placement services (OECD, 2001). While lower caseloads might help employment services to better fulfill each of these functions, caseworkers in our pilot project were mostly concerned with the last function, and we therefore focus on this channel.

To fulfill the job brokerage function, placement services act as intermediaries in the labor market, matching supply to demand. Regarding the supply side, caseworkers are to a large degree concerned with counselling

with (without) treated offices, the budget decreased by 16.4 percent (12.6 percent) between 2006 and 2007, and the difference was not statistically significant. The decrease between 2006 and 2008 amounted to 14.7 percent (treated) and 11.1 percent (non-treated), and was also not statistically significant. Using the employment agencies with matched control offices, the picture is very similar.

as well as monitoring the unemployed. Counselling implies that unemployed clients are actively supported in their job search efforts (e.g., helped with their applications, given practice for job interviews, or helped to use the Internet for job search). In their meta analysis, Card *et al.* (2010) find a positive effect of placement service counselling efforts on re-employment chances. Analyzing four different counselling schemes for French job seekers, Crépon *et al.* (2005) find an increase in unemployment–employment transitions as well as a lower recurrence into unemployment. Van den Berg and van der Klaauw (2006) show that the effectiveness of the counselling depends on its intensity.

In contrast, monitoring implies that caseworkers observe the client's job search efforts (e.g., by requesting proofs of submitted applications or other job search activities). Evaluating a combined counselling and monitoring program, Gorter and Kalb (1996) find that intensified counselling and monitoring have a positive effect on the number of applications by the job seekers and reduce the time taken to find a job. Interestingly, they do not find any differences in the "success" probability of the applications between treated and controls, suggesting that the quality of applications did not change. Results from van den Berg and van der Klaauw (2006) indicate that monitoring is only effective in situations where individual employment prospects are low. In contrast, Kluve (2010) reports that monitoring seems to be more effective when employment prospects are good, and Ashenfelter *et al.* (2005) find little or no effects of monitoring job search activity on unemployment duration. Others have suggested that monitoring is more effective when it is combined with a credible threat of sanctions to incentivize job seekers (Abbring *et al.*, 2005; Lalive *et al.*, 2005; van den Berg and Vikström, 2014). In line with this, Black *et al.* (2003) provide experimental evidence on a job search assistance and monitoring program in the US, and their findings suggest that the threat of entering the program already raises exit rates from UI benefit receipts. Regarding the allocation of unemployed to training programs, the findings of Lechner and Smith (2007) suggest that allocation based on caseworkers' assessment is inferior to allocation based on statistical treatment rules.

When caseload is lower, one might expect that caseworkers are better able to track and inform themselves about the job seekers' search efforts. They might also be able to better tailor a placement strategy suitable for the individual job seeker. For example, if caseworkers know that an individual's search efforts are sufficiently high but still ineffective, they can choose the counselling strategy. If, in contrast, the search effort is low, they can choose to put more emphasis on monitoring, and if the search effort stays low, eventually impose a sanction. Therefore, given their lower caseload, we expect intensified monitoring and counselling efforts in the pilot offices. Because we cannot test this expectation directly, we estimate the pilot project's effect on the frequency of sanctions imposed.

As Behncke *et al.* (2008) argue, the effect of counselling and other strategies employed by caseworkers also depends on the demand side. They find that caseworkers who have better contacts with local firms are more successful in providing placement services. This effect might be particularly pronounced for the low-skilled unemployed. Because the pilot offices in our application had to raise the share of caseworkers responsible for the demand side to at least 30 percent, some part of a potential effect might be driven by intensified contacts with local firms. Because firms in Germany are not obligated to register vacancies at the local employment office, we use the pilot project's effect on the number of newly registered vacancies as a measure of intensified contacts with local firms.

In contrast to many evaluations of active labor market policies, we analyze the effect of our pilot program at the local level (i.e., the level of the local employment office). This allows us to address the question of regional spillover effects: does lowering the caseload in the pilot offices have an effect on other offices in close regional proximity? We can distinguish between positive and negative spillover effects. Negative spillover effects can arise if caseworkers from the pilot offices place their clients in vacancies in neighboring regions. In other words, in a regional labor market with a limited number of vacancies, pilot offices that fill a vacancy might do so at the expense of job seekers from neighboring regions. However, there can also be positive spillover effects. For example, if the additional caseworkers in the pilot offices are effective in registering many new vacancies, this will also benefit the neighboring offices because all employment offices share the same database of vacancies. We examine spillover by analyzing the effect of the pilot project on the performance of neighboring employment offices. To the best of our knowledge, this question has not been addressed by previous research on caseworkers.

III. Empirical Strategy

Method

To identify the causal effect of the pilot project, we use the potential outcome framework (Rubin, 1974; Holland, 1986), where causal effects are defined in terms of counterfactuals. Let $d_i \in \{0, 1\}$ be a binary treatment indicator that takes the value of one if office i was chosen as a pilot office, and zero otherwise. Let y_i and y_i^d denote realized and potential outcomes, respectively. For our estimand, we focus on the average treatment effect on the treated (ATT), which is defined as the difference in the expected outcomes under the treatment and control conditions for the treated offices.

Table 1. *List of control variables*

Control variable	Description
Eastern Germany	Dummy for Eastern Germany
Employment growth	Employment growth between the years 2004 and 2006
Commuting streams	Indicator of net commuting streams, $(no.incomers - no.outgoers)/no.employees)$
Population density	Population density in 2005
Growth of vacancies (R)	Growth rate vacancy rate (04/2006–04/2007)
Growth of unemployed	Growth rate of the stock of all unemployed (12/2005–12/2006)
WBR/unemployed (R)	Share of welfare benefit recipients of all registered unemployed
Number UI clients	Number of UI benefit recipients (12/2006)
% UI benefit recipients	Share of UI benefit recipients among all clients (12/2006)
% of subgroups	Shares of different types of unemployed among all unemployed ("activating"; "advancing"; "caring"; "market"; below 25, above 50 years; without school degree; male; German citizen)
UR	Average unemployment rate in 2006
Seasonal variation	Standard deviation of monthly unemployment rate in 2006
Δ regional–local UR	Difference local and regional unemployment rate (mean 2006)
Average wage	Average wages of full-time employed (06/2006)
Growth average wage	Growth rate of average wages between the years 2000 and 2006
Vacancy rate (R)	Vacancy rate (12/2006)
Sanction rate	Mean sanction rate (05/2006–04/2007)
ALMP rate	Mean share of ALMP measure participants on all job seekers (05/2006–04/2007)
UI expenditures	Expenditures UI 05/2006–04/2007
New vacancies (R)	Number of new vacancies acquired (05/2006–04/2007)
Caseload (R)	Caseload (12/2006)

Notes: The variables are measured at the level of the local employment office or, if indicated by "(R)", at the level of the employment agency.
Source: Data Warehouse (DWH) of the BA, the BA human resource department, and the Federal Statistical Office.

The ATT is identified under two assumptions: common support and selection on observables. The common support assumption is relatively innocuous in our data: given the favorable treated-to-controls ratio and the strict selection criteria, which included a focus on geographical dispersion, the covariate characteristics of the 14 participating offices are well within the common support of the characteristics of the non-participating offices. To render the selection on observables assumption plausible, we control for a battery of pre-treatment covariates x that capture the various aspects of the assignment mechanism. These variables are listed in Table 1.

In particular, we control for several indicators of the local labor market conditions, such as the average unemployment rate in 2006, the standard deviation of the monthly local unemployment rate in 2006 (to capture seasonal variation), and the absolute number of recipients of UI benefits. Because an unemployed person's job search is often not restricted to a single local employment office, we adjust for conditions in neighboring regions by controlling for the labor market situation of the whole employment agency (recall that local offices are nested in regional agencies).

Additionally, to account for private information about the performance of the local employment office, we control for several indicators that capture the characteristics of the office's client base (including indicators for age, education, gender, citizenship, and profiling types[4]). To measure most of these covariates as well as the outcome variables that we describe below, we draw upon the DWH of the BA, an unusually rich, centralized database that collects data about all its clients based on their administrative records. In sum, the variables we control for should reflect, to a major extent, the selection process described in Section II. Thus, in our application, it is very plausible that the conditional independence assumption (CIA) holds.

Nevertheless, even balancing these observed covariates, there might still be unobserved differences between the treatment and control groups. To further remove the effects of unobserved confounders, we rely on a combination of matching and a DiD estimation (Heckman *et al.*, 1998; Smith and Todd, 2005), which exploits the panel structure of our data. In particular, as the outcome variables are observed both in the pre-treatment period right before the start of the program y_{bef} and in the post-treatment period y_{aft}, we focus on the difference-in-differences ATT (ATT^{DiD}) defined as

$$ATT^{\text{DiD}}(\mathbf{x}) \equiv E[(y^1_{\text{aft}} - y^0_{\text{bef}}) - (y^0_{\text{aft}} - y^0_{\text{bef}})|\mathbf{x}, d = 1].$$

Apart from adjusting for the observed confounding variables through the matching, this design also accommodates the presence of time-invariant unobserved confounders. The key identifying assumption of the DiD-matching estimator is the parallel trends assumption defined as

$$E[y^0_{\text{aft}} - y^0_{\text{bef}}|\mathbf{x}, d = 1] = E[y^0_{\text{aft}} - y^0_{\text{bef}}|\mathbf{x}, d = 0],$$

which implies that conditional on the observed covariates, the average outcome for the treated offices in the absence of the treatment would have followed the same trend as in the control offices. To further strengthen this common trends assumption, we augment the covariate set by adding pre-treatment trends for the following variables: employment growth between 2004 and 2006, wage growth between 2000 and 2006, the growth of vacancies, and the growth of the unemployment rate during the year prior to the start of the pilot project. Below, we also conduct additional tests, which suggest that the common trends assumption is plausible in our application.

Using the identification assumption above, the ATT^{DiD} can be estimated as

$$\widehat{ATT}^{\text{DiD}} = \frac{1}{n^1} \sum_{i \in I_1 \cap CS} \left[(y^1_{\text{aft},i} - y^0_{\text{bef},i}) - \sum_{j \in I_0 \cap CS} w(i, j)(y^0_{\text{aft},j} - y^0_{\text{bef},j}) \right].$$

[4] The BA used a profiling scheme that grouped unemployed clients into four types according to their re-employment chances: "activating", "advancing", "caring", and "market".

where CS refers to the common support, and I_0 and I_1 denote the control group and the group of the pilot offices, respectively. The number of pilot offices in the common support region is denoted by n^1, and $w(i, j)$ is the weight of office j if it is matched to the pilot office i.

We consider various matching techniques, including propensity score matching (Rosenbaum and Rubin, 1983), Mahalanobis distance matching (Rubin, 1980), and genetic matching (GM; Diamond and Sekhon, 2013). Because GM maximizes covariate balance directly, it results in higher levels of covariate balance compared to the other matching methods. Therefore, it is our preferred matching method. See the Appendix for more details.

To account for any bias that might result from discrepancies remaining after the matching for our preferred specification, we employ additional regression adjustment in the matched sample (Abadie and Imbens, 2006, 2011). Thus, we estimate the treatment effect using regression adjustment with the covariates $\mathbf{x}$ in the matched data set. For robustness, we also estimate the DiD regressions without any matching, and the results are similar to the results obtained from the combination of matching and regression. We cluster the standard errors at the level of the employment agencies.

Outcome Variables and Data Sources

We use three outcome variables to measure the effect of lowering the caseload: the local unemployment rate, the duration of local unemployment, and the local re-employment rate. We compute these outcomes at the employment office level using the following three indicators:

1. unemployment rate – the number of unemployed BA clients relative to the total labor force, measured in April 2008;
2. unemployment duration – the mean duration of completed unemployment spells of BA clients between May 2007 until April 2008 (measured in days);[5]
3. re-employment rate – the share of BA clients who leave unemployment for unsubsidized employment ($\geq$ seven days), aggregated over the months May 2007 until April 2008.[6]

[5] After the expiration of the entitlement to UI benefits (usually after 12 months), the BA offices are no longer responsible for counseling the unemployed. In other words, unemployment durations are censored after 12 months to focus on the part of unemployment spells where we expect a treatment effect.

[6] Our findings are qualitatively robust using the mean monthly re-employment rate defined as the number of BA clients who leave unemployment for unsubsidized employment in a given month divided by the stock of BA clients in the previous month.

We chose the three outcome variables because they cover different aspects of the offices' performance: while the unemployment rate and, to a lower extent, the re-employment rate partly reflect the inflow into unemployment, the cumulative unemployment duration focuses on the outflow. Further, lowering the unemployment rate is a central goal of the BA and of ALMP more generally. However, a decrease in the unemployment rate or the cumulated unemployment duration could be driven by an increase of exits to non-participation (i.e., by more unemployed withdrawing from the labor market). Thus, the third indicator is of major interest because it measures whether individuals leave unemployment for regular employment.

All three outcome variables focus on the two groups of BA clients who are unemployed: the unemployed who receive UI benefits (comprising the major part) and those who do not receive any benefits. We drew the first two indicators from the DWH of the BA. The third indicator we computed based on administrative micro data from the integrated employment biographies (IEB; Version V11.00.00–131009) of the Institute for Employment Research (IAB). These latter data, which are based on social security records, allow us to measure re-employment based on linked unemployment and subsequent employment records.[7]

We restrict our analysis to the post-treatment period defined as the first year following the start of the pilot project, so the unemployment rate is measured in April 2008, and the unemployment duration and re-employment rate are measured over the May 2007 to April 2008 time period. After the end of the one-year period, the onset of the European financial crisis might have induced differential changes in the economic conditions of different labor market regions. Such changes could jeopardize our identification strategy by threatening the common trends assumption. Furthermore, from December 2008 onwards, the German government and the BA hired about 1,000 additional caseworkers nationwide, which contaminates our control group for a long-term assessment because it also benefited from a sudden (although smaller) decrease of the caseload.

Table 2 shows the details of the construction of the outcome variables. Note that for the DiD analysis, we compute each outcome measure for the pre- and the post-treatment periods, respectively, where the post-treatment period for each outcome refers to the timing defined above (e.g., April 2008 for the unemployment rate), and the pre-treatment period is defined based on the same timing but lagged by one year such that we capture the outcomes prior to the start of the pilot program (e.g., April 2007 for the unemployment rate).

[7] One well-known limitation of social security records in Germany and many other countries is that they exclude civil servants and self-employed individuals.

Table 2. *Construction of outcome variables*

Indicator	DiD measure
Unemployment rate (%)[a]	$\text{UE rate}_{\text{April 08}} - \text{UE rate}_{\text{April 07}}$
Unemployment duration (days)[a]	$\dfrac{\text{UE duration} \sum_{\text{May 07}}^{\text{April 08}}}{\text{UE exits} \sum_{\text{May 07}}^{\text{April 08}}} - \dfrac{\text{UE duration} \sum_{\text{May 06}}^{\text{April 07}}}{\text{UE exits} \sum_{\text{May 06}}^{\text{April 07}}}$
Re-employment rate (%)[b]	$\dfrac{\text{Re-employed unemployed BA clients} \sum_{\text{May 07}}^{\text{April 08}}}{\text{Stock UE April, 30, 2007 + entrants UE} \sum_{\text{May 07}}^{\text{April 08}}} - \dfrac{\text{Re-employed unemployed BA clients} \sum_{\text{May 06}}^{\text{April 07}}}{\text{Stock UE April, 30, 2006 + entrants UE} \sum_{\text{May 06}}^{\text{April 07}}}$

Notes: UE denotes unemployment (on the basis of unemployed BA clients). Variables are measured at the level of the local employment office.
Source: The data sources are (*a*) the DWH of the BA and (*b*) the IEB of the IAB.

Table 3. *Average outcomes before and after the pilot project*

	Period: before		Period: after		DiD effects	
Offices	Pilot	Non-pilot	Pilot	Non-pilot	ATT	ATT_{RA}
Unemployment rate (%)	3	3.36	1.92	2.67	−0.38	−0.46
					(0.07)	(0.06)
Unemployment duration (days)	160	160.04	127.63	136.71	−9.04	−9.00
					(1.79)	(1.66)
Re-employment rate (%)	31.7	29.71	34.56	31.09	1.47	1.47
					(0.55)	(0.51)

Notes: For the unemployment rate, the before and after measures refer to April 2007 and April 2008. For the unemployment duration and re-employment rate, the before and after measures are cumulated from May 2006 to April 2007 and from May 2007 to April 2008, respectively. ATT denotes DiD estimate without covariates. ATT_{RA} denotes regression-adjusted DiD estimate that controls for covariates. $N = 14$ for pilot offices and $N = 684$ for the control offices. Standard errors (in parentheses) are clustered by employment agency.

IV. Empirical Analysis

Estimates before Matching

Table 3 lists the mean values of the outcome variables before and after the start of the pilot project in the treatment and control offices. The last two columns refer to the DiD estimates (without and with regression adjustment).

The first two columns confirm that even without any sample adjustments, pilot and control offices have very similar outcomes in the pre-treatment period. None of the differences is statistically significant at conventional levels. This close similarity in performance in the pre-treatment period is consistent with the fact that the pilot offices were chosen to be representative of the population of all employment offices. The next two columns show the post-period outcomes. As indicated above, the German economy experienced an economic boom in 2007, and consequently both groups of employment offices – irrespective of treatment status – performed better in the year after the start of the pilot project compared to the year before the start of the pilot project. Compared to the year before, more unemployed individuals were re-employed, the average duration of unemployment was shortened, and accordingly the number of unemployed receiving UI benefits decreased.

However, as becomes clear in the last two columns from the DiD estimates, the pilot offices improved their performance more significantly compared to the control offices. The uncontrolled estimates suggest that the hiring of additional caseworkers lowered the unemployment rate by 0.38 percentage points, lowered the average unemployment duration by about 9.04 days, and increased the re-employment rate by about 1.47 percentage points. These uncontrolled DiD effect estimates correspond to a 13, 6, and 5 percent change over the baseline level for the pilot offices, respectively. All estimates are highly statistically significant ($p < 0.001$). As demonstrated in the last column, these effect estimates are very similar once we include the full set of covariates in our DiD regression. Taken together, these results suggest that the hiring of additional caseworkers improved the performance of the employment services.

Matching

We now consider whether the estimates are robust when we use matching to reduce model dependency. We imposed three matching restrictions. First, because head local employment offices differ from the regular local employment offices, we matched exactly on the organizational structure; that is, a head (regular) local employment office could only be matched to a head (regular) local employment office. Second, potential regional spillover effects might bias the estimation of the effect of interest. To account for this, we restricted the donor pool for each pilot office such that it excluded all the local employment offices that are located in the same employment office district as the pilot office ("neighboring offices"). Below, we use these neighboring offices to explicitly analyze potential spillover effects. Third, we excluded a very small set of local employment offices from the pool of potential controls because of their geographic peculiarity (i.e., the islands Borkum, Rügen, Norderney, Westerland, and Juist) or because of border changes of administrative districts (Lauterecken). This leaves us with a pool of potential controls that consists of 684 local employment offices.

Table A1 in the Appendix shows the covariate balance in the unmatched data. For each covariate, we display the means in the treatment and unmatched control groups as well as the p-values of covariate-by-covariate paired t-tests of mean differences and bootstrapped Kolmogorov–Smirnov tests for the equality of distributions. We see that the means for only two of the 30 covariates significantly differ at the 0.05 level. Pilot offices have more clients and therefore also higher sums for UI expenditures. We also see differences in the distributions of a few covariates: the population density, the number of clients, the sanction rate, the ALMP rate, and the UI expenditures. Overall, the samples are fairly balanced, as there are not many more significant differences than we would expect by chance.

To remove these differences in the covariate distributions, we apply the GM algorithm using a 1:3 matching with replacement. Table A2 in the Appendix shows the covariate balance in the matched data. We find that GM effectively balances the covariate distributions; none of the balance tests for the means and the distributions remains significant at the 95 percent level, and the means are close in terms of economic significance. As a robustness check, we also reran the matching with propensity score matching and Mahalanobis distance matching, both using three nearest neighbors. The balance results are presented in Tables A3 and A4 in the Appendix. Both methods perform slightly worse than GM, in the sense that they leave a few covariates imbalanced. As an additional balancing check, we also tested whether a variable that we did not match on was balanced as well. For this test, we examined the inflow into unemployment over the past 12 months at the start of the pilot project. It turned out that the differences between treated and matched control offices on this variable was not statistically significant (inflow of 6,000 versus 5,701, respectively).

To further remove the potential effects of the small remaining differences in the matched samples, we also conduct an additional regression adjustment on the matched sample using the covariates that were included in the matching procedure.

Treatment Effects

Table 4 presents the effect estimates for the three outcomes that measure the performance of the local employment offices. The DiD estimates are shown in the last two columns (raw ATT^{DiD} and the regression-adjusted $ATT_{\mathrm{RA}}^{\mathrm{DiD}}$). We refer to the $ATT_{\mathrm{RA}}^{\mathrm{DiD}}$ results primarily, as this is our preferred specification. Generally, it is worth stressing that the estimated effects for the matched sample are similar to those without matching (Table 3). For all three outcomes, relative performance improved much more significantly

Table 4. *Average treatment effect of the pilot project*

	$y_{\text{aft}} - y_{\text{bef}}$		DiD effects	
	Pilot offices	Control offices	ATT	ATT_{RA}
Unemployment rate (%)	−1.07	−0.68	−0.39 (0.07)	−0.39 (0.06)
Unemployment duration (days)	−32.37	−25.39	−6.98 (2.23)	−6.09 (2.41)
Re-employment rate (%)	2.86	1.49	1.36 (0.62)	1.87 (0.75)

Notes: ATT estimates are based on the matched data without regression adjustment, and ATT_{RA} estimates are based on the matched data with regression adjustment using the covariates that were used in the matching. Baseline mean of outcome variables: 3.00 (unemployment rate); 160.00 (unemployment duration); 31.70 (re-employment rate). $N = 56$ in the matched data. Standard errors (in parentheses) are clustered by employment agency.

in the pilot offices compared to the matched control offices. We find that the decrease in the caseload led to a drop in the unemployment rate and the cumulative unemployment duration, while the re-employment rate increased. The effect estimates are sizable, and while the re-employment rate effect is significant at the 95 percent level, the other two effects are significant at the 99 percent level. We interpret our results as short- to medium-run effects. For example, the regression-adjusted DID effect estimates indicate that the pilot project resulted in a 0.39 decrease in the unemployment rate. This constitutes about a 13 percent decrease compared to the baseline rate of 3.00 in pilot offices. Similarly, the pilot project lowers the cumulative unemployment duration by about six days; this constitutes about a 4 percent decrease compared to the baseline level in pilot offices. Moreover, the increased number of caseworkers not only led to a reduction in unemployment but also to an increase in re-employment in regular employment, as the cumulative re-employment rate increases by 1.87, about a 6 percent increase compared to the baseline level.[8]

Sensitivity

Our findings so far indicate that the reduction of the caseload led to a sizable increase in performance. How robust are these results? As a first sensitivity test, Table 5 replicates the estimates using an alternative identification where we condition on lagged outcomes (Angrist and Pischke, 2008, p. 243). So, instead of using the DiD identification, we regress the post-treatment outcomes on the treatment indicator and the pre-treatment

[8] Using the mean monthly re-employment rate instead, we find a treatment effect of 1.08 percentage points without regression adjustment (standard error, 0.3) and 1.23 (0.18) with regression adjustment. Given a baseline of 7.5 percent monthly re-employment rate, these effects amount to increases of 14 and 16 percent, respectively.

Table 5. *Sensitivity: conditioning on lagged outcomes*

	ATT	ATT^{RA}
Unemployment rate (%)	−0.42 (0.07)	−0.44 (0.05)
Unemployment duration (days)	−7.12 (2.29)	−4.44 (2.34)
Re-employment rate (%)	1.53 (0.71)	2.31 (0.69)

Notes: Estimates in the first column regress y_{aft} on y_{bef}. Estimates in the second column also add the matching covariates to the specification. $N = 56$ in the matched data. Standard errors (in parentheses) are clustered by employment agency.

outcomes (with and without the additional matching covariates), assuming that the treatment indicator is orthogonal to the potential outcomes once we condition on the pre-treatment outcomes and covariates. The results are very similar to the DiD estimates above.

As a second check, Table A5 in the Appendix tests whether our results are sensitive with respect to different matching algorithms. The first column presents the results when we use genetic matching with 1:1 and 1:2 instead of 1:3 matching, as above. The point estimates are fairly similar across these models; the only differences are that the re-employment rate effect size decreases and is no longer significant using 1:2 genetic matching, and that the point estimates for the unemployment duration are no longer significant at conventional levels given the reduced sample size (the point estimates are still positive and sizable). The next two columns replicate the models using Mahalanobis distance and propensity score matching. The results are again fairly similar, with the exception that for Mahalanobis distance matching, the effect on the mean unemployment duration is not significant at conventional levels (the point estimate is slightly bigger). Taken together, these robustness checks indicate that, with the possible exception of the mean unemployment duration, the effects are robust in sign, significance, and size using alternative matching algorithms or alternative estimators.

Common Trends

As a next check, we now address the common trends assumption. Recall that the previous models already accounted for pre-existing trends by including covariates that measure pre-treatment changes in key covariates, such as the employment growth rate or the growth of the unemployment rate. Now, we replicate the same DiD models as above to explicitly test whether the trends in the outcome variables in the year prior to when the pilot project was decided differ between treated and matched controls. We compare the period from January 2005 to December 2005 (pre-pseudo treatment) to the period from January 2006 to December 2006 (post-pseudo

Table 6. *Common trends test: effect on outcomes in pre-project period*

	$y_{\text{aft}} - y_{\text{bef}}$		DiD effects	
	Pilot offices	Control offices	ATT	ATT_{RA}
Unemployment rate (%)	−0.94	−1.06	0.12 (0.11)	0.19 (0.12)
Unemployment duration (days)	−25.85	−24.34	−1.52 (2.64)	2.74 (2.21)
Re-employment rate (%)	5.00	4.16	0.83 (0.56)	−0.39 (0.44)

Notes: January 2005 to December 2005 (pre-pseudo treatment) to January 2006 to December 2006 (post-pseudo treatment). ATT estimates are based on the matched data without regression adjustment, and ATT_{RA} estimates are based on the matched data with regression adjustment using the covariates that were used in the matching. $N = 56$ in the matched data. Standard errors (in parentheses) are clustered by employment agency.

treatment). The results, which are presented in Table 6, suggest that in the year before the start of the pilot project, the trends of the outcome variables are very similar in the treated and control offices, the point estimates are small and with varying signs, and none of them is statistically significant at conventional levels. This strongly corroborates the common trends assumption and confirms that the pilot project did not target offices with particularly promising pre-treatment trends.

Cost-Effectiveness

A politically highly relevant question is whether the positive effects of the increase in the number of caseworkers is economically efficient. To address this question, we turn to an assessment of the cost-effectiveness of the pilot project. We focus on the same period that we surveyed above, the 12 months after the start of the pilot. During this period, the personnel costs of the pilot project amounted to a total of around 34.51 million euros. To approximately assess cost-effectiveness, we use two different outcome variables. First, we compare UI expenditures between pilot and control offices.[9] To calculate the absolute costs, in the pilot offices we also add the additional salary costs of the newly hired caseworkers to the UI expenditures. Second, we calculate the sum of earnings over the first 12 months after re-employment of all the employment spells that we counted in the nominator of the re-employment rate.

To estimate the net effect of the pilot project, we use the same specification as above with the difference between the cumulated costs and the cumulated earnings in the first 12 months of the project and the 12 months before the project start as our dependent variable. The estimates reported in Table 7 show that in the post-pilot period, costs decreased and earnings

[9] The advantage of using real UI expenditures instead of estimated elasticities, as in Koning (2009), is that we avoid additional assumptions.

Table 7. *Cost-effectiveness: effect on UI expenditures and on earnings in million euros*

	$y_{\text{aft}} - y_{\text{bef}}$		DiD effects	
	Pilot offices	Control offices	ATT	ATT_{RA}
Costs	−7.97	−6.11	−1.86 (1.33)	−1.22 (0.69)
Earnings	1.5	0.36	1.13 (0.92)	3.09 (0.97)

Notes: The pre-(post) project period is May 2006 (2007) to April 2007 (2008). ATT estimates are based on the matched data without regression adjustment, and ATT_{RA} estimates are based on the matched data with regression adjustment using the covariates that were used in the matching. Baseline outcomes: 35.91 million euros (UI expenditures) and 46.17 million euros (earnings). $N = 56$ in the matched data. Standard errors (in parentheses) are clustered by employment agency.

increased in both the treated and the control offices. However, the DiD effect estimates show that the cost reductions were much larger in the pilot offices. In particular, the regression-adjusted specification suggests that the costs decreased by 1.22 million euros more due to the pilot project, an average reduction of around 3.4 percent over the baseline of 35.91 million euros of average costs of UI benefit expenditures in the pilot offices in the pre-pilot year. Similarly, the DiD effect estimates show that the earnings increased more in the pilot offices. The regression-adjusted specification suggests that the earnings increased by 3.09 million euros due to the pilot project, which corresponds to an average earnings increase of around 6.7 percent over the baseline of 46.17 million euros of average earnings in the pilot offices in the pre-pilot year.[10] Based on a 3 percent contribution rate in 2007 and 2008, an earnings increase of an average of 3.09 million euros amounts to about 90,000 euros increased UI contributions per treated office per year.

Taken together, these results suggests that the extra costs of the new caseworkers are compensated for by the achieved reduction in benefit expenses (a result consistent with Koning, 2009). It is worth noting that, on the one hand, the actual costs might be higher because additional capital costs are not taken into account, and therefore the estimates might slightly overstate the effect on cost savings. On the other hand, the estimates on benefits to the public budget (through additional UI contributions) might be conservative because we do not take into account additional taxes that are associated with faster re-employment in regular jobs. Overall, the results

[10] We could observe an earnings increase for three reasons: (1) the number of employed individuals increased; (2) the wages increased; (3) the re-employment spells were more stable and longer. While our main results presented above show an increased re-employment rate, additional analyses (not shown) also revealed no significant effects on average daily earnings and on average longer employment spells within the first 12 months after re-employment (controlling for the covariates).

presented in this section point to the cost-effectiveness of the improved caseload.

Causal Mechanisms

In this section, we address the question of how lowered caseload might have influenced the performance of the pilot offices. As described above, one of the main tasks of caseworkers is to monitor unemployed clients. An increase in monitoring due to a lower caseload might increase the number of imposed sanctions. To examine this channel, we consider the effect of the pilot project on the mean sanction rate during the first 12 months after the start of the pilot project.[11] A second potential mechanism involves the effects of lower caseload on the demand side of the labor market. Given that one aspect of the pilot project was an increase in the share of demand side oriented caseworkers, we might expect that the pilot offices were able to increase the number of vacancies through more intensive search efforts and contacts with local firms. To examine this channel, we investigate how the pilot project affected the number of freshly acquired vacancies in the caseworkers' vacancies pool. Unfortunately, we do not have reliable data on the number of vacancies at the local level because many local employment offices share a joint firm service on the level of the employment agency. Remember that we have around 178 employment agencies and 779 local employment offices. The 14 treatment offices were located in 13 different employment agencies. Therefore, we use data on the number of vacancies at the level of employment agencies.

The findings for the mechanism tests are provided in Table 8 (based on the benchmark model used above with the regression-adjusted ATT specification). We find sizable and significant effects for both intermediate outcomes. While in both pilot and control offices the sanction rate increased during the observation period, the pilot offices exhibit a relatively stronger increase in the sanction rate of about 0.86 percentage points. This constitutes a 63 percent increase over the baseline level. Moreover, we find that, aggregated over the period of 12 months after the start of the pilot, 3,358 more vacancies on average were registered in the agencies with pilot offices compared to the control group, a 6.3 percent increase over the baseline.

[11] Unemployment sanctions can be imposed as a result of refusing a training measure or a vacancy referral due to missing an appointment with the caseworker (or the medical service of the employment agency), or of failing to document sufficient search effort. To build the mean sanction rate, we add all types of sanctions imposed per month and divide it by the stock of unemployed in that month and build the average over the respective 12 months period.

Table 8. *Intermediate effects of the pilot project on the sanction rate and vacancy acquisition*

	$y_{\text{aft}} - y_{\text{bef}}$		DiD effects	
	Pilot offices	Control offices	ATT	ATT_{RA}
Sanction rate (%)	1.72	0.9	0.82 (0.2)	0.86 (0.17)
New vacancies acquired (#)	3506.36	−328.5	3834.86 (2188.78)	3538.48 (1310.12)

Notes: ATT estimates are based on the matched data without regression adjustment, and ATT_{RA} estimates are based on the matched data with regression adjustment using the covariates that were used in the matching. $N = 56$ in the matched data. Standard errors (in parentheses) are clustered by employment agency.

In sum, these results indicate that both the monitoring and the demand side channels contributed to the overall effect of the pilot project. It led to an increase in the number of vacancies that firms try to fill using the public employment service (demand side) as well as more intense monitoring (supply side).

Spillover Effects

In this section, we consider potential adverse effects of the pilot project, which is an important aspect, but many analyses on labor market programs do not pay attention to it. In a recent study, Crépon *et al.* (2013) have presented experimental evidence on the displacement effects of a job placement assistance program in France. In our application, one potential concern is that the pilot project induced negative regional spillover effects for the offices that are close to the pilot offices and, in particular, for local employment offices that are located in the same employment agency district. Recall that such spillover effects are theoretically ambiguous. Because they operate in the same regional labor market, pilot project offices might have improved their performance at the expense of their neighboring offices if, for example, the pilot offices place their clients in jobs that would otherwise be filled by clients from control offices. However, by acquiring more vacancies that enter the shared database of all BA employment offices, pilot offices could also have induced positive spillover effects for their neighbors because control offices now benefit from the increased opportunities to place their own clients.

To estimate potential average spillover effects, we apply the same matching procedure as described above, but instead of using the pilot offices as treated, we consider the neighboring offices as receiving treatment and discard the pilot offices from the pool of potential controls. In other words, we redefine the treatment as being adjacent to a pilot office, and therefore spillover effects should be captured by the effects of this treatment on our

Table 9. *Spillover effects on neighboring offices*

	$y_{\text{aft}} - y_{\text{bef}}$		DiD effects	
	Neighboring offices	Control offices	ATT	ATT_{RA}
Unemployment rate (%)	−0.74	−0.72	−0.02 (0.08)	−0.04 (0.05)
Unemployment duration (days)	−26.59	−23.56	−3.03 (1.79)	−2.92 (1.72)
Re-employment rate (%)	1.54	1.62	−0.07 (0.43)	0.43 (0.41)
New vacancies acquired (#)	4131.19	−237.35	4368.55 (2595.77)	4022.22 (2178.22)
Sanction rate (%)	0.73	0.71	0.03 (0.12)	0.06 (0.08)

Notes: For the unemployment rate, the before and after measures refer to April 2007 and April 2008. For the unemployment duration and re-employment rate, the before and after measures are cumulated from May 2006 to April 2007 and May 2007 to April 2008, respectively. ATT_{RA} estimates are based on regression adjustment using the covariates that were used in the matching. $N = 188$ in the matched data. Standard errors (in parentheses) are clustered by employment agency.

outcomes. We use the same covariates and the same matching algorithm as before. After matching, there were no significant differences between the new treatment and control group.

Table 9 displays the results. We do not find significant negative spillover effects: the reduction of the caseload in the pilot offices did not cause a deterioration of performance in the neighboring offices. If anything, we find positive spillover effects on the mean unemployment duration (significant at the 10 percent level). Most likely, these positive spillover effects occurred because by increasing the number of caseworkers who concentrate on acquiring job vacancies from firms, the pilot offices increased the number of registered vacancies that all the other employment offices could draw upon, given the shared database.

Table 9 also reports the treatment effect on the sanction rate. We do not find any effects on the sanction rate of the neighboring offices. Given no spillover effects on sanctions, one might argue that the estimates of the spillover effects approximate the impact that is attributable to vacancies.[12] However, because some of the additional vacancies were likely to be outside the commuting region for job seekers of the neighboring offices (but within the commuting region for job seekers of the treated offices), in fact, the presented spillover effects will be a lower bound of the impact of the additional vacancies.

In sum, the results of our spillover analysis suggest that – given no increased sanctions in the neighboring offices – additional vacancies do not lead to more re-employment but those who are re-employed return to work faster. Do these findings mean that it is not a lack of vacancies that

[12] We thank an anonymous referee for this suggestion.

keeps the remaining individuals from returning to work? Admittedly, that interpretation is somewhat speculative.

V. Conclusion

The caseworker-to-clients ratio is an important policy parameter for the public employment service, but little evidence exists about its precise effects. We exploit a pilot project of the BA, which significantly reduced the caseload in 14 out of its 779 local employment offices, to study the effectiveness of adding additional caseworkers. At the pilot project's start, the average ratio of caseworker to the number of recipients of UI benefits was 1:40 in pilot offices compared to 1:100 in non-participating offices. Using a combination of matching and DiD estimators, we find that lowering the caseload resulted in a sizable decrease in the unemployment rate and in the cumulated unemployment duration, and an increase in the re-employment rate for the period of 12 months after the start of the pilot project. These results are robust across various specifications. Disentangling the causal pathways, we find that the lower caseloads led to more proactive behavior on the part of the pilot offices, as they increased the monitoring and registered more new vacancies. Additional analysis indicates that the pilot project had no negative regional spillovers on the outcomes for neighboring offices. Addressing the cost effectiveness, we find that the costs imposed by the hiring of additional caseworkers were offset by the savings on UI expenditures and additional UI contributions. More precisely, net of the additional salary costs, the pilot offices' UI benefit expenditures decreased by around 3.4 percent, and their clients' average earnings increased by around 6.7 percent. The latter corresponds to roughly 90,000 euros in additional UI contributions, on average, per treated office per year.

Taken together, our results suggest that assigning more caseworkers – potentially in combination with higher caseworker monitoring – can be effective in lowering the local unemployment rate. This result alone is an important finding for policymakers because lowering the unemployment rate is a major objective of active labor market policy. Moreover, we found a reduced average search duration and an increased re-employment rate as a result of the pilot program. This is a second important implication: the unemployed clients who were counselled in employment offices with lower caseloads were more successful in finding a job. Overall, the policy implication of our study is that lower caseloads can improve the effectiveness of the public employment service.

As is common for impact assessments of this type, one potential concern might be the so-called "Hawthorne" effects (Mayo, 1945). In particular, we might be worried that caseworkers in the pilot offices exerted extra effort, first, because they were aware of the project and, second, because they were subject to increased monitoring by the BA headquarters. Note, however, that we found treatment effects on the unemployment rate, on earnings, and on UI expenditures, which are outcomes that were not directly part of the new target agreements for the treated offices. This suggests that not all of the increase in the effectiveness was due to increased monitoring by the BA through the target agreements. Another possibility is that the control offices might have lowered their performance, as all offices could potentially gain from a "successful" pilot project. This interpretation does not seem plausible for at least two reasons. First, the control offices we used in our analysis were not aware of being in the control group. Second, all offices also had target agreements to reach and therefore had a strong incentive to perform.

How should we judge the external validity of our findings? On the one hand, we could argue that given the fact that most additional caseworkers were relatively unexperienced and had to be trained first, our results might understate the effect that we would observe if more experienced caseworkers were hired. In addition, we found no negative spillover effects on neighboring regions. Thus, the positive effect from the pilot project is not at the expense of other regions. On the other hand, there is no guarantee that the same effects would prevail if the caseload were decreased in all local employment offices. We also emphasize that our study period was restricted to a year of relatively good economic conditions. Future research needs to clarify to what extent the positive effects of lower caseloads depend on favorable economic conditions.

Appendix

In this Appendix, we present additional results referenced in the main text.

Regional Dispersion of Pilot Project Offices

Figure A1 shows the regional dispersion of the local employment offices that participated in the pilot project.

Matching Methods and Balance Results

We consider various matching techniques including propensity score (PS) matching (Rosenbaum and Rubin, 1983), Mahalanobis distance (MD) matching (Rubin, 1980), and genetic matching (GM; Diamond and Sekhon, 2013). GM is based on a generalization of MD matching where each treated

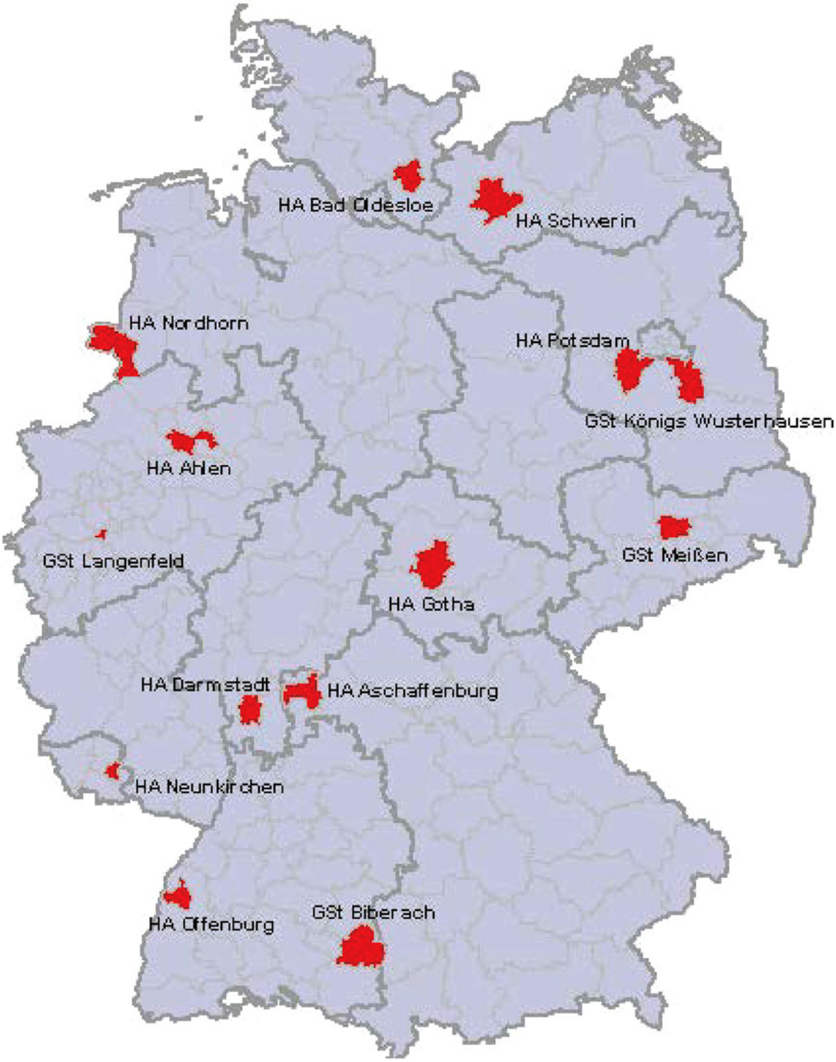

Fig. A1. Regional dispersion of pilot project offices
Notes: Pilot project offices in red/dark gray. HA denotes head local employment office, and GSt denotes regular local employment office.

unit is matched to m nearest neighbors according to the following generalized distance metric:

$$Dist(\mathbf{x}_i, \mathbf{x}_j) = \left[(\mathbf{x}_i - \mathbf{x}_j)'(S^{-1/2})' V S^{-1/2}(\mathbf{x}_i - \mathbf{x}_j)\right]^{1/2}.$$

Here, V is a $(k \times k)$ positive definite weight matrix with zero in all elements except the main diagonal, and $S^{1/2}$ is the Cholesky decomposition

Table A1. *Balance before matching*

	Mean Tr.	Mean Co.	T pval	KS pval
Eastern Germany	0.36	0.23	0.37	
Employment growth	−0.01	−0.02	0.12	0.29
Commuting streams	−0.10	−0.19	0.09	0.13
Population density	365.19	379.67	0.89	0.05
Growth of vacancies (region)	0.04	0.02	0.50	0.67
Growth of unemployed	−0.14	−0.16	0.65	0.76
WBR/unemployed (region)	0.61	0.61	0.98	0.46
Number UI clients	3882.76	2453.75	0.02	0.00
Share UI benefit recipients	0.66	0.66	0.83	0.47
Share type "activating"	0.14	0.15	0.14	0.28
Share type "advancing"	0.23	0.20	0.11	0.20
Share type "caring"	0.39	0.39	0.96	0.19
Share type "market"	0.10	0.09	0.77	0.68
Share type missing	0.10	0.10	0.66	0.69
Share above 50 years	0.41	0.40	0.34	0.55
Share below 25 years	0.11	0.12	0.86	0.98
Share without school degree	0.06	0.06	0.87	0.95
Share male	0.48	0.47	0.08	0.26
Share German	0.93	0.92	0.84	0.86
UR	0.11	0.12	0.26	0.63
Seasonal indicator	0.01	0.01	0.55	0.43
Δ regional–local UR	−0.00	−0.00	0.51	0.88
Average wage	84.41	82.58	0.60	0.89
Growth average wage	0.10	0.09	0.34	0.56
Vacancy rate (region)	0.10	0.09	0.89	0.58
Sanction rate	0.01	0.01	0.07	0.01
ALMP rate	0.24	0.23	0.21	0.03
UI expenditures	35,912,441.86	22,647,758.76	0.02	0.00
New vacancies	53,002.93	46,881.31	0.61	0.66
Caseload (region)	113.66	121.82	0.15	0.29

Notes: Mean Tr. denotes average for pilot offices. Mean Co. denotes average for control offices. T pval is the two-sided p-value from paired two sample t-tests. KS pval is the p-value from bootstrapped Kolmogorov–Smirnov tests. If not mentioned otherwise, the variables are measured at the level of the local employment office.
Source: DWH of the BA, the BA human resource department, and the Federal Statistical Office.

of S, the variance-covariance matrix of $\mathbf{x}$, the $(N \times k)$ matrix of covariate characteristics. Note that this metric generalizes the conventional MD metric that we obtain when setting each of the k parameters in the diagonal of V equal to 1.

In GM, the weights in the diagonal of V are chosen by an optimization algorithm such that covariate balance between the treatment and control groups is maximized based on an overall balance score. We define the balance score in the objective function as the lowest p-value across covariate-by-covariate paired t-tests for differences in means and bootstrapped Kolmogorov–Smirnov tests for the equality of distributions (the tests are computed for all covariates that are included in the matching).

Table A2. *Balance after 1:3 GM*

	Mean Tr.	Mean Co.	T pval	KS pval
Eastern Germany	0.36	0.24	0.19	
Employment growth	−0.01	−0.02	0.66	0.50
Commuting streams	−0.10	−0.08	0.77	0.24
Population density	365.19	358.14	0.91	0.15
Growth of vacancies (region)	0.04	0.03	0.73	0.17
Growth of unemployed	−0.14	−0.16	0.53	0.39
WBR/unemployed (region)	0.61	0.61	0.92	0.72
Number UI-clients	3882.76	3556.72	0.49	0.38
Share UI-benefit recipients	0.66	0.66	0.98	0.74
Share type “activating”	0.14	0.15	0.29	0.14
Share type “advancing”	0.23	0.22	0.34	0.23
Share type “caring”	0.39	0.39	0.99	0.11
Share type “market”	0.10	0.09	0.38	0.18
Share type missing	0.10	0.10	0.97	0.17
Share above 50 years	0.41	0.40	0.37	0.14
Share below 25 years	0.11	0.12	0.73	0.69
Share without school degree	0.06	0.06	0.69	0.56
Share male	0.48	0.48	0.79	0.21
Share German	0.93	0.93	0.99	0.54
UR	0.11	0.12	0.12	0.37
Seasonal indicator	0.01	0.01	0.96	0.24
Δ regional–local UR	−0.00	−0.00	0.43	0.11
Average wage	84.41	84.18	0.91	0.53
Growth average wage	0.10	0.10	0.87	0.24
Vacancy rate (region)	0.10	0.08	0.25	0.22
Sanction rate	0.01	0.01	0.38	0.09
ALMP rate	0.24	0.24	0.91	0.17
UI expenditures	35,912,441.86	32,679,677.93	0.45	0.08
New vacancies	53,002.93	35,433.21	0.12	0.08
Caseload (region)	113.66	120.07	0.14	0.39

Notes: Mean Tr. denotes average for pilot offices. Mean Co. denotes average for control offices. T pval is the two-sided p-value from paired two sample t-tests. KS pval is the p-value from bootstrapped Kolmogorov–Smirnov tests. If not mentioned otherwise, the variables are measured at the level of the local employment office.
Source: DWH of the BA, the BA human resource department, and the Federal Statistical Office.

Diamond and Sekhon (2013) present evidence from Monte Carlo simulations that show the good properties of this balance score. Because GM maximizes covariate balance directly, it results in higher levels of covariate balance compared to the other matching methods. It is therefore our preferred matching methods.

For robustness, we also employ alternative matching methods such as PS and MD matching. These lead to similar results.

Table A1 shows the covariate balance before matching. Tables A2, A3, and A4 show the covariate balance after GM, PS matching, and MD matching, respectively. Table A5 shows the DiD effect estimates from the various methods.

Table A3. *Balance after 1:3 PS matching*

	Mean Tr.	Mean Co.	T pval	KS pval
Eastern Germany	0.36	0.24	0.49	
Employment growth	−0.01	−0.01	0.58	0.16
Commuting streams	−0.10	0.04	0.10	0.00
Population density	365.19	605.26	0.15	0.00
Growth of vacancies (region)	0.04	0.04	0.98	0.41
Growth of unemployed	−0.14	−0.13	0.86	0.14
WBR/unemployed (region)	0.61	0.62	0.68	0.25
Number UI-clients	3882.76	5823.89	0.20	0.16
Share UI-benefit recipients	0.66	0.66	0.88	0.71
Share type "activating"	0.14	0.15	0.30	0.08
Share type "advancing"	0.23	0.22	0.51	0.28
Share type "caring"	0.39	0.37	0.45	0.20
Share type "market"	0.10	0.10	0.61	0.33
Share type missing	0.10	0.11	0.25	0.01
Share above 50 years	0.41	0.40	0.50	0.07
Share below 25 years	0.11	0.11	0.46	0.34
Share without school degree	0.06	0.06	0.63	0.51
Share male	0.48	0.49	0.52	0.38
Share German	0.93	0.90	0.17	0.11
UR	0.11	0.11	0.93	0.28
Seasonal indicator	0.01	0.01	0.65	0.25
Δ regional–local UR	−0.00	0.00	0.35	0.01
Average wage	84.41	90.68	0.24	0.03
Growth average wage	0.10	0.10	0.50	0.40
Vacancy rate (region)	0.10	0.11	0.55	0.26
Sanction rate	0.01	0.01	0.59	0.08
ALMP rate	0.24	0.25	0.51	0.28
UI expenditures	35,912,441.86	54,192,490.05	0.24	0.01
New vacancies	53,002.93	46,450.05	0.69	0.28
Caseload (region)	113.66	118.74	0.49	0.19

Notes: Mean Tr. denotes average for pilot offices. Mean Co. denotes average for control offices. T pval is the two-sided p-value from paired two sample t-tests. KS pval is the p-value from bootstrapped Kolmogorov–Smirnov tests. If not mentioned otherwise, the variables are measured at the level of the local employment office.
Source: DWH of the BA, the BA human resource department, and the Federal Statistical Office.

Table A4. *Balance after 1:3 MD matching*

	Mean Tr.	Mean Co.	T pval	KS pval
Eastern Germany	0.36	0.29	0.32	
Employment growth	−0.01	−0.02	0.48	0.24
Commuting streams	−0.10	−0.10	0.97	0.55
Population density	365.19	319.50	0.59	0.16
Growth of vacancies (region)	0.04	0.03	0.76	0.08
Growth of unemployed	−0.14	−0.16	0.64	0.51
WBR/unemployed (region)	0.61	0.61	0.89	0.22
Number UI-clients	3882.76	3409.22	0.37	0.24
Share UI-benefit recipients	0.66	0.65	0.51	0.72
Share type "activating"	0.14	0.14	0.67	0.23

(*Continued*)

Table A4. *Continued*

	Mean Tr.	Mean Co.	T pval	KS pval
Share type "advancing"	0.23	0.22	0.37	0.34
Share type "caring"	0.39	0.40	0.61	0.00
Share type "market"	0.10	0.09	0.46	0.01
Share type missing	0.10	0.09	0.87	0.08
Share above 50 years	0.41	0.41	0.32	0.14
Share below 25 years	0.11	0.11	0.64	0.23
Share without school degree	0.06	0.05	0.31	0.35
Share male	0.48	0.47	0.09	0.01
Share German	0.93	0.94	0.33	0.14
UR	0.11	0.13	0.06	0.03
Seasonal indicator	0.01	0.01	0.59	0.23
Δ regional–local UR	−0.00	0.00	0.26	0.02
Average wage	84.41	82.15	0.33	0.39
Growth average wage	0.10	0.10	0.84	0.54
Vacancy rate (region)	0.10	0.08	0.14	0.02
Sanction rate	0.01	0.01	0.09	0.01
ALMP rate	0.24	0.23	0.29	0.04
UI expenditures	35,912,441.86	30,861,699.95	0.32	0.00
New vacancies	53,002.93	41,303.98	0.32	0.20
Caseload (region)	113.66	119.69	0.22	0.13

Notes: Mean Tr. denotes average for pilot offices. Mean Co. denotes average for control offices. T pval is the two-sided p-value from paired two sample t-tests. KS pval is the p-value from bootstrapped Kolmogorov–Smirnov tests. If not mentioned otherwise, the variables are measured at the level of the local employment office.
Source: DWH of the BA, the BA human resource department, and the Federal Statistical Office.

Table A5. *Sensitivity: DiD effects using different matching methods and specifications*

Matching	1:1 GM	1:2 GM	1:3 MD	1:3 PS
Unemployment rate (%)	−0.35	−0.26	−0.61	−0.40
	(0.11)	(0.09)	(0.11)	(0.06)
Unemployment duration (days)	−3.75	−8.36	−9.47	−7.75
	(2.91)	(2.89)	(5.56)	(1.74)
Re-employment rate (%)	1.45	0.74	1.92	1.62
	(0.69)	(0.78)	(0.66)	(0.41)
N	28	42	56	56

Notes: Effect estimates are based on regression adjustment of unemployment rate, unemployment duration, and re-employment rate, except for the first column, where no regression adjustment is used because the number of covariates exceeds N. Standard errors (in parentheses) are clustered by employment agency.

References

Abadie, A. and Imbens, G. W. (2006), Large Sample Properties of Matching Estimators for Average Treatment Effects, *Econometrica 74*, 235–267.

Abadie, A. and Imbens, G. (2011), Bias Corrected Matching Estimators for Average Treatment Effects, *Journal of Business and Economic Statistics 29*, 1–11.

Abbring, J. H., van den Berg, G. J., and van Ours, J. C. (2005), The Effect of Unemployment Insurance Sanctions on the Transition Rate from Unemployment to Employment, *Economic Journal 115*(505), 602–630.

Angrist, J. D. and Pischke, J.-S. (2008), *Mostly Harmless Econometrics*, Princeton University Press, Princeton, NJ.

Ashenfelter, O., Ashmore, D., and Deschenes, O. (2005), Do Unemployment Insurance Recipients Actively Seek Work? Evidence from Randomized Trials in four U.S. States, *Journal of Econometrics 125*, 53–75.

Behncke, S., Froelich, M., and Lechner, M. (2008), Public Employment Services and Employers - How Important are Networks with Firms?, *Zeitschrift für Betriebswirtschaft 1*, 151–177.

Behncke, S., Froelich, M., and Lechner, M. (2010), Unemployed and Their Caseworkers: Should They be Friends or Foes?, *Journal of the Royal Statistical Society*, Series A *173*, 67–92.

Black, D. A., Smith, J. A., Berger, M. C., and Noel, B. J. (2003), Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System, *American Economic Review 93*(4), 1313–1327.

Blien, U. et al. (2004), Typisierung von Bezirken der Agenturen für Arbeit, *Zeitschrift für Arbeitsmarktforschung 37*, 146–175.

Card, D., Kluve, J., and Weber, A. (2010), Active Labour Market Policy Evaluations: A Meta-Analysis, *Economic Journal 120*(548), F452–F477.

Crépon, B., Dejemeppe, M., and Gurgand, M. (2005), Counseling the Unemployed: Does it Lower Unemployment Duration and Recurrance?, IZA Discussion Paper 1796.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013), Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment, *Quarterly Journal of Economics 128*, 531–580.

Diamond, A. and Sekhon, J. S. (2013), Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies, *Review of Economics and Statistics 95*, 932–945.

Fougère, D., Pradel, J., and Roger, M. (2009), Does Job Search Assistance Affect Search Effort and Outcomes?, *European Economic Review 53*, 846–869.

Gorter, C. and Kalb, G. (1996), Estimating the Effect of Counseling and Monitoring the Unemployed Using a Job Search Model, *Journal of Human Resources 31*, 590–610.

Heckman, J. J., Ichimura, H., and Todd, P. (1998), Matching as an Econometric Evaluation Estimator, *Review of Economic Studies 65*, 261–294.

Hill, C. J. (2006), Casework Job Design and Client Outcomes in Welfare-to-Work Offices, *Journal of Public Administration Research Theory 16*, 263–288.

Hofmann, B., Krug, G., Sowa, F., Theuer, S., and Wolf, K. (2012), Wirkung und Wirkmechanismen Zusätzlicher Vermittlungsfachkräfte auf die Arbeitslosigkeitsdauer: Analysen auf Basis eines Modellprojektes, *Zeitschrift für Evaluation 11*, 7–38.

Holland, P. W. (1986), Statistics and Causal Inference, *Journal of the American Statistical Association 81*(396), 945–960.

Jerger, J., Pohnke, C., and Spermann, A. (2001), Gut Betreut in den Arbeitsmarkt?: Eine Mikrooekonometrische Evaluation der Mannheimer Arbeitsvermittlungsagentur, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 34*, 567–576.

Kluve, J. (2010), The Effectiveness of European Active Labor Market Policy, *Labour Economics 17*, 904–918.

Koning, P. (2009), The Effectiveness of Public Employment Service Workers in the Netherlands, *Empirical Economics 37*, 393–409.

Lagerström, J. (2011), How Important Are Caseworkers – And Why? New Evidence from Swedish Employment Offices, IFAU Working Paper Series 2011:10.

Lalive, R., van Ours, J. C., and Zweimüller, J. (2005), The Effect of Benefit Sanctions on the Duration of Unemployment, *Journal of the European Economic Association 3*, 1386–1417.
Lechner, M. and Smith, J. (2007), What is the Value Added by Caseworkers?, *Labour Economics 14*, 135–151.
Mayo, E. (1945), *The Social Problems of an Industrial Civilization*, Graduate School of Business Administration, Harvard University, Boston, MA.
OECD (2001), *Labour Market Policies and the Public Employment Service, Proceedings of the Prague Conference, July 2000*, OECD Publishing, Paris.
Rinne, U., Uhlendorff, A., and Zhao, Z. (2013), Vouchers and Caseworkers in Training Programs for the Unemployed, *Empirical Economics 45*, 1089–1127.
Rosenbaum, P. R. and Rubin, D. B. (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika 70*, 41–55.
Rubin, D. B. (1974), Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies, *Journal of Educational Studies 66*, 688–701.
Rubin, D. B. (1980), Bias Reduction Using Mahalanobis-Metric Matching, *Biometrics 36*, 293–298.
Schiel, S., Schröder, H., and Gilberg, R. (2008), Mehr Vermittlungen durch mehr Vermittler? Ergebnisse des Modellversuchs "Förderung der Arbeitsaufnahme" (FAIR), in T. Kruppe (ed.), *IAB-Bibliothek*, Volume *312*, Bertelsmann, Bielefeld.
Smith, J. A. and Todd, P. E. (2005), Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?, *Journal of Econometrics 125*, 305–353.
van den Berg, G. J. and van der Klaauw, B. (2006), Counseling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment, *International Economic Review 47*, 895–936.
van den Berg, G. J. and Vikström, J. (2014), Monitoring Job Offer Decisions, Punishments, Exit to Work, and Job Quality, *Scandinavian Journal of Economics 116*, 284–334.
Yavas, A. (1994), Middlemen in Bilateral Search Markets, *Journal of Labor Economics 12*, 406–429.

First version submitted April 2014;
final version received September 2015.