EVIDENCE OF TREATMENT SPILOVERS WITHIN MARKETS

Marc Ferracci, Grégory Jolivet, and Gerard J. van den Berg*

Abstract—This paper provides a method to infer the presence of treatment spillovers within markets where a fraction of agents is treated. We model individual outcomes as functions of the assigned treatment status and the distribution of assigned treatments in a market. We develop a two-step identification and estimation method, focusing first on the treatment distribution among individuals within markets and then on the treatment distribution across markets. We apply our approach to training programs for unemployed individuals in France using rich administrative data. Our results provide evidence of interactions within local labor markets as potential individual outcomes vary with the proportion of treated individuals.

1. Introduction

TREATMENT evaluation methods borrowed from the statistical literature under the banners “matching” and “propensity score matching” have now become an integral part of the applied econometrician’s toolkit.1 These techniques, revolving around the Rubin model (1974), require two assumptions. The first is a conditional independence assumption (CIA) to control for confounding factors that drive both the assignment to treatment and the potential outcomes. The second assumption, referred to as the stable unit treatment value assumption (SUTVA), rules out any influence of an individual’s treatment status on another individual’s potential outcome (see Neyman, 1990, or Rubin, 1986). Recently, the latter has been receiving increasing attention from economists, as many settings in social sciences in general and economics in particular are likely to generate interactions between individuals. This includes peer effects, neighborhood effects, network effects and the dispersion of information, and various kinds of equilibrium effects of policy interventions.

In this paper, we consider a population of individuals divided into groups that we will call markets. In each market, a fraction of individuals is exposed to some treatment. The size of this fraction may vary across markets and, crucially, may causally influence individual potential outcomes. We develop and apply a method to infer the presence of such treatment spillovers, and thus the violation of SUTVA.

More generally, in a population of interest, we allow a policy to operate at two levels: individuals and markets. We build on the standard potential outcome model (Rubin, 1974) and allow an individual’s outcome to be a function not only of his own treatment status but also of the distribution of the treatment across individuals in his market. In cases where the individual treatment consists of the exposure to a particular policy intervention, like participation in a training program, it is of obvious interest to policymakers to know whether there is a causal effect of the fraction of treated individuals in a market (or any other feature of the treatment distribution) on individual outcomes.

Our method to detect the presence of treatment spillovers and social interactions consists of two steps, focusing first on the treatment distribution among individuals within markets and then on the treatment distribution across markets. Identification is delivered by a CIA at each step. The first step consists of estimating mean potential outcomes within each market separately. We define the market treatment as the relevant moment of the distribution of the treatment within the market (e.g., the proportion of treated). Using a second CIA, controlling for confounders driving the assignment to market treatments and potential outcomes, we can recover the causal effect of the treatment distribution within a market on the average treatment effect in the market. Interactions between agents across markets are ruled out. In this paper, evaluation is based on matching estimators at each step, but one could consider other techniques.

In the second part of the paper, we apply our approach to detect the presence of interactions within local labor markets where unemployed workers can participate in training programs. In this setting, markets are defined by occupational, spatial, and temporal indicators. The fraction of treated individuals in a market may influence individual probabilities to leave unemployment. If many individuals are treated, there may be crowding out among workers applying for the same vacancies. Moreover, there may be a response from the other side of the market. For example, firms may post more vacancies if the number of trained job seekers increases. The net effect of these two (and perhaps other) sources of interactions on individual outcomes and the average treatment effect is difficult to predict. In the empirical analysis, we set out to determine the shape of this effect as a function of the fraction of treated individuals, and we examine whether this shape is consistent with the absence of interaction effects.

We use a unique matched set of French administrative registers and survey data. The registers follow individuals over time and record unemployment and training spells. In surveys from the French national unemployment agency, firms report their job-opening expectations for the coming year at precise geographical and occupational levels. Caseworkers use this labor demand information to assign unemployed workers to training programs. At the market level, we need to ensure that no unobserved confounder drives both the proportion of individuals going through a training program.
EVIDENCE OF TREATMENT SPILLOVERS WITHIN MARKETS

and the potential unemployment duration outcomes. For this, it is useful to point out that the French training system is partly run by regions. We observe each region for a number of years, under several treatment regimes (i.e., with different proportions of treated). This allows us to control for a fixed unobserved region effect in the distribution of treatment across markets. Since the treatment at the market level is continuous, we base our second-step estimation on Hirano and Imbens (2004) and match markets on the generalized propensity score. The empirical results show that the potential individual outcomes are not invariant to the proportion of treated. The average probability of moving into work within a year if treated decreases steadily with the proportion of treated, while the average probability if not treated is convex, first decreasing and then increasing. We thus find empirical evidence of SUTVA violation and treatment spillovers.

II. The Model

The economy is segmented into local labor markets. We assume that an individual \( i \) belongs to only one market, so in obvious notation, \( M_i = m \) if \( i \) resides in market \( m \). When convenient, we also denote \( i \)'s market by \( m(i) \). We consider a binary treatment indicator \( Z_i \), which is equal to 1 in case of treatment and 0 otherwise. In our empirical application, treatment is defined as participation in a training program during the first \( d_z \) months of unemployment (\( d_z = 6 \) for the benchmark analysis).

The individual outcome is denoted by \( Y_i \). In the empirical application, \( Y \) is a binary variable equal to 1 if the worker leaves unemployment within \( d_Y \) months (\( d_Y = 12 \) for the benchmark analysis). We allow \( Y_i \) to depend on individual \( i \)'s treatment status \( Z_i \), but also on a function of the vector \( Z_{-i} \) of treatments received by all other individuals in market \( M_i \). The most general approach would be to write \( Y_i = Y (Z_i, Z_{-i}) \), but the treatment variable then becomes a high-dimensional object that is difficult to study. Throughout this paper, we allow only for distributional effects of the treatment. This means that an individual's outcomes are not affected by market size or permutation of the treatments among other individuals in the same market. Since \( Z \) is a binary variable, we thus consider the proportion \( Q_m \) of individuals treated in market \( m \),

\[
Q_m = E(Z | M = m), \tag{1}
\]

and assume that individual outcomes can be written as

\[
Y_i = Y_i (Z_i, Q_{m(i)}). \tag{2}
\]

In the context of our empirical application, we can think of several channels through which \( Q \) may affect \( Y \). If training programs have a positive effect on productivity, then an increase in \( Q \) implies that competition for jobs will be tougher, so treatment may have a crowding-out effect. It could also be that this increase in the average productivity among job seekers will stimulate labor demand as the average return to a job vacancy will increase.2

Note that we allow for treatment externalities across individuals within a market (through \( Q \)), but we assume that there are no spillovers between markets, that is, the distribution of outcomes \( Y \) in a market does not depend on values of \( Q \) or \( Z \) in other markets. In effect, this dictates the appropriate operationalization of the "market" concept.3

In line with the above notation, we define \( Y_i (z, q) \) as the potential outcome for individual \( i \) that applies if we assign values \( z, q \) to \( Z_i \) and \( Q_{m(i)} \), respectively. The evaluation of the causal effect of \( q \) on the average treatment effect in the market is based on the averages of these individual potential outcomes:

\[
Y_{z,q} = E_M \{ E [Y_i (z, q) | M] \}, \quad \forall (z, q). \tag{3}
\]

The expectation in equation (3) is taken over individuals in a given market and then over all markets in the economy, where we weigh the average across markets by market size (defined by the number of eligible individuals in the market). For notational convenience, we omit the latter weights (this applies to all expressions below). Accordingly, \( Y_{z,q} \) is the average potential outcome corresponding to the individual treatment status \( z \) as a function of the market treatment dimension \( q \) (the fraction of treated in the market), if the market invariably has the same composition as the population of eligible individuals.

The functions \( Y_{z,q} \) directly lead to average treatment effects,

\[
\delta_{z,q} = Y_{z,q} - Y_{z',q'}. \tag{4}
\]

The set \( \{ \delta_{z,q} \} \) fully describes the effects of the treatment.4

The SUTVA implies that an individual's potential outcomes should not depend on other workers' treatment status, so the proportion of treated in a given market should not affect individual outcomes. If we find that \( Y_{1,q} \) or \( Y_{0,q} \) is not constant when \( q \) varies, then the SUTVA does not hold. To infer whether treatment spillovers are present, we therefore aim to examine whether estimates of \( Y_{1,q} \) or \( Y_{0,q} \) vary with \( q \). Examining actual values of the parameters \( Y_{z,q} \) may be less relevant for policy evaluation. Because the effect of the treatment may vary across individuals, these parameters are informative only about the magnitude of effects of policies based on assigning the treatment at random, which is often not a policy option of interest.

2 Lechner (2002) studies whether the treatment participation probability is a source of heterogeneity in the treatment effect but does not discuss violations of the SUTVA.

3 This is in line with Manski's (2013) assumption of constant treatment response: an individual’s outcomes depend only on treatments in a subpopulation, called the individual’s reference group. In this paper, the reference group is the worker’s labor market, and it is assumed to be nonmanipulable (workers cannot move across markets).

4 Note that \( \delta_{z,q} = -\delta_{z',q'} \).
Standard matching applications usually compare pairs of treated and untreated individuals facing the same local market conditions. Ideally, the treatment and control groups are in the same market. Results of such studies are valid only for a given value of \( q \) and for the composition of the subpopulation of individuals in the market under consideration. It is not possible to extrapolate the results to other values of \( q \) if the SUTVA is violated. For instance, if a given study finds a positive average treatment effect (ATE) and then advocates an increase in the proportion of individuals being treated, it could be that the resulting ATE under the new value of \( q \) is smaller or larger than the one initially estimated.

### III. Identification and Inference

We aim to identify and estimate the average potential outcomes \( Y_{i,q} \) on a nonexperimental data set. For a given individual \( i \), we observe only \( Y(Z_{i}, Q_{m(i)}) \). The identification problem thus has two dimensions: we do not observe the outcome of an individual with \( Z_{i} = z \) had he received \( 1 - z \), and we do not know what would have happened in a market \( m \) with \( Q_{m} = q \) had this market been assigned a treatment policy regime \( q' \neq q \). We follow a two-step method for identification and estimation that explicitly accounts for these two levels.\(^5\)

#### A. Identification

**Identification of mean potential individual outcomes in each market.** Consider a given market \( m \). This market has been assigned a treatment status \( Q_{m} \), which is now fixed, say, to value \( q \). We are interested in estimating the two average potential individual outcomes \( E[Y_{i}(1,q) \mid M_{i} = m] \) and \( E[Y_{i}(0,q) \mid M_{i} = m] \) for this market. For a proportion \( q \) of workers, we observe \( Y(1,q) \) but not \( Y(0,q) \), and for the \( 1 - q \) remaining workers, we observe \( Y(0,q) \) but not \( Y(1,q) \).

In this paper, we follow the matching literature (Rubin, 1974) and assume that we can observe a set of individual characteristics, denoted as \( X \), that allows the following conditional independence assumption:

\[
Y_{i}(z,q) \perp Z_{i} \mid X_{i}, M_{i} = m, \forall z, q, m, i. \tag{5}
\]

\(^5\)The advantages of a two-dimensional evaluation approach for the study of social spillovers were recognized in some previous studies. Philipson (2000) devises two-stage randomization schemes for experimental studies. Hudgens and Halloran (2008) define a range of causal effects using potential outcomes if there are two groups, and they develop estimators for the case of fully randomized assignments. Manski (2013) also defines such causal effects with social interactions and considers various restrictions on the type of interactions and their implications for the identification of treatment effects. A recent paper by Graham, Imbens, and Ridder (2011) considers group treatments that consist of the assignment of a within-group composition and their effect on a group-level outcome. As in our setting, Hudgens and Halloran (2008) and Manski (2013) allow for only distributional effects of the treatment on individual potential outcomes. The equality \( \kappa_{q,q'} = \kappa_{q} + \kappa_{q'} - \kappa_{q,q'} \) is given in Hudgens and Halloran’s (2008) decomposition of the treatment effect into a direct effect (that of receiving the treatment, keeping interactions fixed, \( \kappa_{q} \)) and an indirect effect (due to interactions only, \( \kappa_{q,q'} \)).

In words, assumption (5) states that within each market \( m \), for any value \( q \) that has been assigned to this market, the assignment of treatment across individuals is independent of potential outcomes given individual characteristics \( X \). The CIA, equation (5), is similar to the identifying assumption in a large number of empirical evaluations of labor market programs, in that all variables driving both potential outcomes and treatment status are assumed to be included in the vector \( X \).

In addition to the CIA, identification requires a support overlap condition (often called common support assumption). This states that there is no value of \( X \) for which the probability of being treated is 0 or 1, or that the analysis concerns only values of \( X \) for which this probability is strictly between 0 and 1. In practice, this is implemented by imposing constraints on \( X \) such that this probability exceeds a threshold value strictly above 0 and falls short of another threshold value strictly below 1. The choice of thresholds depends on the application and sample size at hand. Section IVC explains how we impose the support overlap condition in our application. With CIA and support overlap, the two mean potential individual outcomes in a market are identified.

**Identification at the market level.** The identification challenge at the market level is due to the fact that \( E[Y_{i}(z,q) \mid M_{i} = m], z = 0, 1 \) is not observed for \( q \neq Q_{m} \). In line with the matching literature, we assume that conditionally on a set of market characteristics, the market potential outcomes are independent of the treatment status. Formally, if we assume that there is a vector \( W_{m} \) such that

\[
E[Y_{i}(z,q) \mid M_{i} = m] \perp Q_{m}, \forall m, z, q. \tag{6}
\]

then we can identify \( Y_{z,q} \). This also requires a support overlap condition. Since \( Q \) is a continuous variable, the implementation differs somewhat from the individual level (see Flores et al., 2012, and Kluve et al., 2012). In section IVC, we explain how we impose this condition in our application.

Note that \( E[Y_{i}(z,q) \mid M_{i} = m] \) is our market potential outcome of interest. Alternatively, one could be interested in the average of \( Y \) for a specific subgroup of individuals within each market (with a given value of \( X \)) or, if \( Y \) is not a binary variable, in other moments of the distribution of \( Y_{i}(z,q) \) in each market (e.g., the variance of individual outcomes). In that case, we can still use our approach provided we substitute the moment of interest for \( E[Y_{i}(z,q) \mid M_{i} = m] \) in assumption (6).

As was the case for matching at the individual level, the validity of assumption (6) depends on the empirical application. In the case of training and labor markets, one may think of a two-stage assignment process. First, a central planner (e.g., the government) decides on the assignment of training resources across local labor markets in order to optimize a given criterion (e.g., maximization of the employment rate). To this end, the central planner has information on some market characteristics \( W \) that influence potential outcomes. The probability that a market receives a given \( Q \) then depends
on $W$ and possibly on a set of variables independent of the outcomes. Once a market is assigned $Q$, the local planner assigns individual workers to training programs, based on $X$.

Empirically, the challenge is thus to include in $W$ the variables that are relevant for the assignment of training across markets and that influence the performance of the market participants. In general, it makes sense to include aspects of the within-market distribution of $X$, that is, of the confounders used in the first step. Without this, the CIA, equation (6), entails that the allocation of treatments to markets is not driven by the market composition in terms of characteristics that affect potential outcomes and that possibly cause self-selection into treatments. This scenario is often not realistic. For example, if a market contains many workers not sufficiently motivated to search for a job without any supplementary training, then the policymakers may choose to allocate a relatively high budget for training for that market.

Here, motivation may be captured by observable individual past labor market outcomes, which can be included in $X$. If markets are observed for a number of years, then it may also make sense to include binary market indicators in $W$ to account for unobserved heterogeneity across markets. Finally, it may be useful to include information on market agents who are not potentially subject to a treatment. For example, in our application in the next section, we also include indicators of local labor demand in $W$.

A summary of the two-step identification approach. With assumptions (5) and (6), we can write:

\[
Y_{z,q} = EM \left\{ E \left[ Yi(z,q) | M \right] \right\} \\
= EW \left[ EMW \left\{ E \left[ Yi(z,q) | M \right] | W \right\} \right] \\
= EW \left[ EMW \left\{ E \left[ Yi(z,q) | M \right] | W, Q = q \right\} \right] \\
= EW \left[ EMW \left\{ EX|M \left\{ E \left[ Yi(z,q) | X, M \right] \right\} | W, Q = q \right\} \right],
\]

(7)

where equation (6) allows going from the second to the third line, while equation (5) allows going from the fourth to the last line. We have dropped the subscript $m$ for notational convenience. The term on the first line is our target. The term on the last line is an observable quantity. Notice again that the expectations over markets are weighted by the market size (as defined by the measure of eligible individuals in the market) such that the terms in the above equalities capture averages over the full population of eligible individuals. Following the set of equalities (7) from top to bottom illustrates the assignment process that we have in mind. First, a social planner decides to allocate treatment intensity $Q$ across markets, where market assignment of $Q$ is independent of potential market outcomes given a set of market characteristics $W$. Then, in a given market $m$, individuals are assigned to treatment status $Z$. Within each market, selection into treatment is independent of potential individual outcomes given individual characteristics $X$.

Our identification strategy follows the equalities in equation (7). In a given market, conditional on $X = x$, it is a pure coincidence whether the individual receives the treatment $z$. The average potential outcome after treatment for all individuals in that market with the same $X$, and with the observed market value of $Q$, is then equal to the average outcome of those with $X$ who were treated in that market. This can be aggregated over all $X$ in the market. Next, conditional on $W$, it is a pure coincidence whether the market receives a high value of $Q$. In any other market with $W = w$, the average potential outcomes (for $z = 0$ and $z = 1$) if it had been exposed to $q$ are then equal to the average potential outcomes in the markets that actually received the value $q$. This can be aggregated over $W$ to obtain population-level averages.

B. Statistical Inference

In line with the identification strategy outlined above, the estimation procedure consists of two steps, one at the individual level and the other at the market level. In this section, we present the main features of this procedure. Details specific to our application are discussed in section IV.C.

First step: Matching individuals within markets. We want to estimate the average individual potential outcomes within each market. To this end, we use a weighted regression estimator that combines two methods: the approach based on regressing the outcome on the treatment variable and the covariates (Rubin, 1977) and the weight estimator developed by Hirano, Imbens, and Ridder (2003).

The latter involves computing an estimate of the average outcome for the treated as the market average of $Z \times Y / \hat{p}(X, m)$ where $\hat{p}$ is an estimate of the propensity score. The propensity score $p(x, m)$ is the probability of being treated in market $m$ conditionally on having individual characteristics $x$. Likewise, the outcome for the nontreated is estimated by the average of $(1 - Z) \times Y / (1 - \hat{p}(X, m))$. The weighted estimator can thus be seen as a weighted regression of $Y$ on $Z$ using the weights $[Z / \hat{p}(X, m) + (1 - Z) / (\hat{p}(X, m))]^{1/2}$. The weighted regression estimate we use is just an extension of this weight estimator where we augment the weighted regression by adding the covariates $X$. This type of estimator has been considered by Robins and Ritov (1997) and was implemented in Hirano and Imbens (2001). For each market $m$, we run the following regression,

\[
Y_i = \mu_m + X'_i \times \beta_m + \tau_m \times Z_i + u_i, \quad i \in m,
\]

(8)

\footnote{Rosenbaum and Rubin (1983) show that assumption (5) implies

\[
Y_i(z, Q_m) \perp Z_i \mid p[X, m], \quad \forall z, i, m, M_i = m.
\]

The propensity score should also be a function of the market treatment intensity $Q$. However, since we are at the individual step of the estimation procedure, $Q_m$ is fixed for each market $m$. We thus write the propensity score simply as a function of $X$ and $m$.}
using the weights \([Z_i / \hat{p}(X_i, m) + (1 - Z_i) / \hat{p}(X_i, m))^{1/2}\). We denote the resulting estimates as \((\hat{\mu}_m, \hat{\rho}_m, \hat{\tau}_m)\) and compute the market outcomes as:

\[
\hat{E} \left[ Y_i (z, q) | M_i = m, Q_m = q \right] = \frac{1}{N_m} \sum_{i \in m} \hat{\mu}_m + X'_i \hat{\rho}_m + z \times \hat{\tau}_m, \quad z = 0, 1. \tag{9}
\]

Second step: Matching markets. We now need to average the estimates obtained from equation (9) across markets, assuming that the market treatment \(Q\) is randomly assigned conditionally on \(W\). The units of observation are no longer individuals \(i\) but markets \(m\). We adopt the approach suggested by Hirano and Imbens (2004). They extend the propensity score result of Rosenbaum and Rubin (1983) to the case of a continuous treatment and show that the unconfoundedness assumption (6) leads to

\[
E \left[ Y_i (z, q) | M_i = m \right] \perp 1\{Q_m = q\} \mid g(Q = q | W_m), \quad \forall m, z, q.
\]

The function \(g(\cdot | w)\) is the conditional density of the treatment \(Q\). It is the counterpart of the propensity score in the case of a continuous treatment and is thus called the generalized propensity score (GPS). The estimation procedure devised by Hirano and Imbens (2004) then works as follows:

1. We estimate the GPS. We regress \(Q\) on the market variables \(W\), which yields the index \(\hat{\gamma} W\) and the distribution of the residuals \(\psi\). We then compute the estimate of the GPS as: \(\hat{\gamma}_m (q) = \hat{\psi} (q - \hat{\gamma} W_m)\).

2. We estimate the function that links the market outcome, \(E \left[ Y(z, Q) | M \right]\), to the market treatment \(Q\) and the GPS. We thus regress \(E \left[ Y(z, Q) | M \right]\) on a flexible (in general quadratic) function of \(Q\) and \(\hat{\gamma}(Q)\) and denote as \(\hat{H}\) the resulting function: \(\hat{E} \left[ Y(z, Q) | M \right] = \hat{H}_z (Q, \hat{\gamma}_M (Q))\).

3. For each value of \(z\) and \(q\), we compute our estimate of \(Y_{z,q}\) as the (weighted) average of \(\hat{H}_z (q, \hat{\gamma}_M (q))\) over markets: \(\hat{Y}_{z,q} = E_M \left[ \hat{H}_z (q, \hat{\gamma}_M (q)) \right]\).

Standard errors. While in theory, second-step standard errors can be calculated, Hirano and Imbens (2004) suggest bootstrapping to compute confidence intervals in practice when effecting the estimation of a continuous treatment. Moreover, because our method is sequential, the second-step standard errors would have to be corrected for the standard errors from the first stage. Clearly, with bootstrapping it is easy to take care of this correction.\(^9\)

\(^7\) We thus need to estimate a propensity score for each market \(m\), some of which may be small. In section C, we show how to choose specifications to overcome this issue.

\(^8\) For the case with a multivariate but not necessarily continuous treatment, see Imbens (1999).

\(^9\) We do not use a nearest-neighbor matching estimator at either stage so the bootstrap may be used to estimate the standard error of our estimates (see Abadie & Imbens, 2008, for a case where bootstrapping fails to work for nearest-neighbor matching estimators).

IV. Application to Training Policies in France

We first give an overview of the French unemployment institutions and training system. Then we present our data at the individual and at the market levels and discuss the econometric model along two lines: the two conditional independence assumptions and the model specification.

A. Training Programs for Unemployed Workers in France

We present the general organization of the French training system for job seekers, as well as the assignment process of unemployed workers to training programs. In relation to our model, the former point helps to understand between-market variation in \(Q\), while the latter helps to interpret within-market variation in \(Z\).

General organization. The French training system is run by three bodies: the national state, the social partners (trade unions and employer organizations), and the administrative regions.\(^10\) For simplicity of exposition, we simply refer to the public employment service (PE) and do not distinguish among its components. The PE role consists of counseling and monitoring unemployed job seekers as well as paying unemployment benefits. Any job seeker who wishes to enter a public training program must consult her local PE agency. The local PE agencies are also in charge of prescribing and buying specific training courses for eligible job seekers. Finally, the administrative regions express their needs for skills at the local level to PE agencies, based on the vacancies that are to be opened every year. The PE agencies then have to assign job seekers to training programs suited to the vacancies. In all cases, training capacities should be calibrated to fit job vacancies. Consequently, the probability that an unemployed person is trained depends on local labor market conditions. In the framework of our evaluation model, this is a source of variation in \(Q\) between regions and occupations.

Training programs. A meeting with a PE caseworker (typically 30 minutes long) is compulsory for all newly registered unemployed workers and recurs at least every six months. Depending on the individual’s profile, the caseworker can schedule follow-up interviews between two compulsory meetings, and interviews can be requested at any time by the unemployed workers themselves. Apart from a wide range of counseling measures, training programs may be proposed to job seekers during or between interviews. In theory, the job seekers are allowed to accept or refuse any program they are proposed, but a refusal can lead to a cut in unemployment benefits. In practice, however, sanctions for refusing a training program are rarely imposed.

Each year since 2001, the PE local agencies conduct a survey, called BMO, on the predicted job vacancies at the local level (see section IVB for more details on this survey). In

\(^10\) There used to be two main institutions: one run by the state the other by the social partners, which were merged in December 2008.
particular, the BMO survey is designed to help caseworkers assign unemployed workers to training programs in accordance with local labor demand. Self-selection may play a part in training participation. Field studies (Fleuret, 2006) show that low-skilled workers are less likely to accept training, although they are more likely to be proposed such programs by caseworkers. In section IVC we use a rich set of individual characteristics (in particular, detailed information on an individual’s unemployment and training history) to control for this feature.

Job seekers may also find training programs by themselves. In that case, they can benefit from some public funding to cover the costs of the program. A PE caseworker then has to assess whether the program is relevant given the job seeker’s professional project and the local labor demand.

A data set provided by PE describes the content of training programs. Unfortunately, due to the lack of common identifiers, we cannot merge this data set with the one we use in our estimation. Training programs can be of four different types. While the distribution across types is not uniform, the mass is not concentrated on a single type. For instance, out of the 593,126 programs that took place between 2005 and 2007, 17.9% were of the general type, 37.5% of the personal type, 29.9% were service oriented, and 14.7% were production oriented.

B. The Data

Individual data. We use quarterly extracts from an administrative data set, FNA, that contains information on all workers entering unemployment who receive unemployment or welfare benefits. We use eight randomly selected extracts of the FNA. Each extract represents 2.5% of unemployed workers between 1990 and 2007. For each individual, the extracted file has information on all the unemployment spells that could have occurred since 1990. They contain the dates when workers are registered as unemployed by the public employment service, as well as the dates when workers enter and exit training. We start the analysis in 2002 because institutional changes to the unemployment benefit scheme and to the monitoring of job seekers were implemented in 2001. Also, the BMO data that we use at the market level (see below) have been collected since 2001 only.

We define the treatment variable as follows: \( Z_i = 1 \) if \( i \) enters a training program within \( d_Z \) months after the beginning of the unemployment spell, 0 otherwise. As for the outcome variable, \( Y_i = 1 \) if \( i \) leaves unemployment within \( d_Y \) months after the beginning of the unemployment spell, 0 otherwise.

In our benchmark estimation, we set \( d_Z = 6 \) months and \( d_Y = 12 \) months. We choose these values because long-term unemployment (longer than a year) is a key statistic in the public debate on unemployment in France. Our observations consist of unemployment spells starting in either 2002 or in 2004. We use the following covariates for the estimation of the propensity score at the individual level: age, gender, duration of affiliation to the unemployment insurance system, unemployment benefits, reference wage (i.e., wage of the previous job, if any), and the time of year (month) when the unemployment spell started. To control for individual labor market histories, we also build two sets of covariates, related respectively to the periods \([t_0 - 2 \text{ years}, t_0]\) and \([t_0 - 7 \text{ years}, t_0 - 2 \text{ years} ]\), where \( t_0 \) is the starting date of the unemployment spell. For each period, we control for the following characteristics: number of unemployment spells, time spent unemployed and time spent in training.

Market data. We define a market \( m \) by three characteristics: a region \( r \), an occupation \( o \), and a time indicator \( t \). The time indicator is a dummy equal to 1 for unemployment spells starting in 2002 and 0 for those starting in 2004.

The regions we consider are defined by PE. We had to exclude three regions from our analysis: Aquitaine, Corsica (due to matching issues with the BMO), and Limousin (since the number of observations per market is too small). We also exclude all overseas departments, again because we cannot match them with the BMO survey. In the end, we consider 22 regions.

We have precise information on the occupation of the job that individuals are looking for. The corresponding variable is coded by a letter (seven categories, broadly defining the general occupation) and two digits. Since we need many observations per market and want these markets to be as isolated as possible from one another, we keep only the broad definition of occupations (seven categories): office worker, executive, industrial worker, construction worker, social worker, technical worker, and employee (e.g., sales).

An important market variable for our analysis is the labor market tightness indicator, constructed from the yearly BMO survey conducted by PE. From 2001 onward, the BMO collects firms’ job opening predictions at a very detailed level. More precisely, the BMO is conducted every year in November on all the firms affiliated to PE. It is thus possible to define the local labor market tightness \( \theta \) as the number of predicted vacancies over the number of unemployed workers. This variable is important because the BMO survey is conducted to help caseworkers assign unemployed workers

11 We have precise information on the occupation of the job that individuals are looking for. The corresponding variable is coded by a letter (seven categories, broadly defining the general occupation) and two digits. Since we need many observations per market and want these markets to be as isolated as possible from one another, we keep only the broad definition of occupations (seven categories): office worker, executive, industrial worker, construction worker, social worker, technical worker, and employee (e.g., sales).

12 We will use a time dimension to define markets. Hence, since we assume that markets are segmented, we do not consider the inflow in 2003 and 2005 in order to minimize the risk of overlap.

13 For instance, in November 2007, 1,524,557 firms were asked how many vacancies they were planning to post during the year 2008. The nonresponse rate in the BMO survey is high (more than 75% on average every year). The actual figures of vacancies are then recovered on the basis of the size, activity, and location of the respondent firms. Note that although this would be an issue if we wanted to evaluate labor demand precisely, the BMO predictions of vacancies are the actual numbers that PE observes when deciding on the allocation of training. Hence the nonresponse rate is not an issue for the assignment to training by PE caseworkers.
to the training programs that match the skills needed on local labor markets.

The labor market tightness indicator $\theta$ is computed as follows. The number of predicted vacancies is constant within a calendar year (due to the design of the BMO survey), but the unemployment stock is not. Therefore, we can compute the ratio of vacancies to unemployed for each pair $(r, o)$ and each month. Then for a market $m = (r, o, t)$ with $t = 2002$ (resp. $t = 2004$), we compute the labor market tightness $\theta_m$ as the average of all the monthly ratios in 2002 and 2003 (resp. 2004 and 2005).

Descriptive statistics. If we define markets as year specific in addition to region and occupation specific, then the data cover 308 of such markets ($22 \times 7 \times 2$ occupation categories $\times 2$ periods). The smallest (largest) of these has 324 (24,262) individuals. The dispersion in size stems from heterogeneity across regions, from rural regions like Auvergne to the Paris region, but also from heterogeneity across occupations. We will see below that the number of observations per market is insufficient for a fully nonparametric analysis with respect to region and occupation indicators. The fractions of treated individuals are rather low, ranging from 2% to 14%. Only twelve markets have a fraction above 10%. The population-wide fraction of treated does not change from 2% to 14%. Only twelve markets have a fraction above 5 percentage points. Thus, this variation can be relatively large given the support of $Q$.

C. Econometric Implementation

Let us first sum up the notations. Individuals are denoted by $i$. An individual $i$ is in market $m = (r, o, t)$ if $R_i = r$, $O_i = o$, $T_i = t$, in which case $m(i)$ (or $M_i$) equals $m$. The individual treatment is $Z_i$ and the outcome is $Y_i$. The confounders needed to write the independence assumption (5) form the vector $X_i$. The market treatment variable is $Q_m$, the average of $Z_i$ over $M_i = m$. The independence assumption (6) at the market level involves the market confounders $W_m$, which, as we explain below, consist of a local labor market tightness $\theta_m$, the market averages of the individual characteristics $\bar{X}_m$, and region dummies. The market outcome is the average of $Y_i$ over $M_i = m$. Individual outcomes depend on both $Z$ and $Q$: $Y_i = Y_i(Z_i, Q_m(i))$.

Assignment to treatment at the individual level. Assumption (5) is the typical CIA made in evaluation studies using matching estimators on nonexperimental data. In addition to standard variables such as age, gender, and reference wage (see section IVB), two features are crucial in making this assumption credible. First is the need to control for local labor market conditions (see Heckman, Ichimura, & Todd, 1997). Second, it is key to assess a worker “employability” as this feature may drive the caseworker’s effort to assign him to a training program (Sianesi, 2004). We address the first issue by controlling for local labor markets, that is, by using region, occupation, and year as confounders.

The employability issue is tackled using information on individuals’ unemployment and training histories. We start our observation window for this in 2002. We also observe unemployment and training spells that the individuals may have experienced between 1990 and 2001, and one might want to use all this information to incorporate an unobserved individual effect in the CIA, equation (5), running the first estimation step with fixed- or random-effect techniques. However, we cannot follow this route because of a labor market policy reform in 2001 that changed the assignment process, so we do not expect the parameters to be constant during the 1990–2007 period. Still, we can use individual histories between 1990 and 2001 to compute the confounders described in section IVB and thus capture part of the individual heterogeneity driving unemployment duration and assignment to training.

Allocation of treatment across markets. We now consider the CIA at the market level, equation (6). We use three types of confounders to justify this assumption. First, we control for local labor demand using the information of the BMO survey on the number of job vacancies in any local labor market for any year. As we mentioned in section IVA, the purpose of the BMO survey is to help PE caseworkers allocate training in accordance with firms’ demand for each type of job.

Second, as motivated in section IIIA, we control for indicators of the distribution of $X$ in each market $m$. Third, we account for time-invariant unobserved market characteristics. Recall that a market unit is a triplet (region, occupation, year). Local unemployment agencies are region specific. We may thus use region fixed effects to capture heterogeneity across unemployment agencies. One may wonder whether this can be extended by using region $\times$ occupation fixed effects. However, we observe each region/occupation pair only twice (in 2002 and in 2004), and this does not leave a substantial amount of residual variation in the fraction of treated to accommodate a meaningful support overlap condition. We thus use region dummies as confounders in our benchmark specification, and we rely on the local labor demand indicator to account for the allocation of $Q$ within regions across occupations. In a robustness check, we show that our results are not affected by interacting region fixed effects with aggregated occupation effects.

One may argue that the policymaker assigns a budget $B_m$ to each market and that the mapping from this budget $B_m$ to the fraction of treated individuals $Q_m$ also depends on the decisions of eligible individuals in the market to participate in the program. Such decisions depend on $X$. However, as we have seen, the decision to enroll is critically affected by the caseworker. We therefore feel that it is reasonable to capture $B_m$ by $Q_m$. 14

14 More generally, this argument suggests including a number of statistics of the within-market distribution of $X$ in the vector $W_m$. For example, one
Support overlap conditions. We need to ensure that at each step, there is sufficient overlap in the distribution of confounders across treatment statuses. At the individual level, within each market we consider the intersection of the supports of the estimated propensity score \( \hat{p} \) for treated and nontreated workers. Around 7.5% of our observations at the individual level, are not in this intersection. Our estimation results will discard these observations. This trimming of the data follows the strategy of Dehejia and Wahba (1999).

At the market level, the treatment status is continuous. Two recent papers (Flores et al., 2012, and Kluve et al., 2012) operationalize support overlap conditions for a continuous treatment (duration of exposure to training). They use quantiles of the observed treatment values \( Q \) to cut the range of \( Q \) into intervals, and they consider the GPS at the medians of these intervals. Their sample sizes are larger than the number of markets in our case (308), whereas our data contain many confounders. We use quantiles to cut the range of \( Q \) into three equal-probability intervals \( I_1, I_2, \) and \( I_3 \), and we keep a market \( m \) if the probability that it receives a treatment in either of these quantiles is above a given threshold \( \alpha \). Formally, we keep market \( m \) if the minimum of \( \int_{j} g_m(q) dq \) for \( j \in \{1, 2, 3\} \) is above \( \alpha \). In the benchmark estimation, we use \( \alpha = 0.1\% \). In a series of checks conducted in section V, we show that our results are robust to imposing a stricter support condition that is, by considering four intervals instead of three or by increasing \( \alpha \) to 2.5\%. With our benchmark support condition, we lose 14.3\% of our markets.

Benchmark specification details. In the within-market step, we regress the binary treatment variable \( Z_i \) on the covariates \( X_i \) for each market \( m \). We might have too few observations per market to run a fully nonparametric estimation with respect to the market indicators. Therefore, we include two of the three market characteristics among the vector of regressors in the estimation of the propensity score. More precisely, for each region \( r \), we regress \( Z_i \) on \( X_i, O_i, T_i \) by logit. Given the French unemployment institutions, we think that there is a stronger case for flexibility with respect to the region indicator. Once we have estimated \( \hat{p}(X_i, m(i)) \) for all \( i \), we impose the support overlap condition. We then estimate the market outcomes \( E[Y_i(z, Q_m)|M_i = m] \) using the weighted regression method presented in section IIIIB.

The between-market step starts with estimation of the GPS. We regress a logistic transformation of \( Q_m \) on \( W_m = (X_m, \theta_m, \text{region dummy}) \). The vector \( X_m \) contains the market average of individual confounders \( X \) except age squared, month of entry into unemployment and the labor market history variables.\(^{15}\) We then predict the GPS \( g(Q = q|W) \) assuming that the residuals of the previous regression are normally distributed. Next, we impose the corresponding support overlap condition and compute \( \hat{Y}_{i,q} \) as explained in section IIIB.\(^{16}\) We bootstrap this two-step procedure 500 times to produce confidence intervals. We resample within the whole population. For each value of \( Q \), we produce 95\% confidence intervals centered around the estimator with a normal approximation.\(^{17}\)

V. Estimation results

A. Empirical Evidence of Spillovers within Local Labor Markets

Average potential outcomes. If the SUTVA holds, then \( Y_{1,q} \) and \( Y_{0,q} \) should not vary with \( q \). Figures 1a and 1b show the estimates of these functions of \( q \). The dashed lines delimit the 95\% confidence intervals. Each dot in the graphs corresponds to a value of \( q \) actually observed in the data, provided it satisfies the support overlap condition. This gives values in the range 0.02 to 0.12. Note that the scale of the \( y \)-axis is different between figures 1a and 1b. Also, note that \( Y_{0,q} \) is more precisely estimated than \( Y_{1,q} \). All the results discussed in this section follow the benchmark analysis described in the previous section.

Clearly, neither \( \hat{Y}_{1,q} \) nor \( \hat{Y}_{0,q} \) remains constant when \( q \) varies. This is the main empirical finding of this paper: the SUTVA is violated since an individual’s potential outcomes depend on whether many or few people in his market are treated.

Taking a closer look at figures 1a and 1b, we note differences between the patterns of \( \hat{Y}_{1,q} \) and \( \hat{Y}_{0,q} \). The average potential outcome if treated, \( \hat{Y}_{1,q} \), steadily decreases with \( q \). If a worker is assigned to a training program within six months from the start of his unemployment spell, the probability that he leaves unemployment within a year is around 52\% when the proportion of treated \( q \) is 3\%, 48\% if \( q = 6\% \), and 44\% if \( q = 8\% \), and it falls below 40\% when \( q \) is closer to and above 10\%.

The average potential outcome if nontreated, \( \hat{Y}_{0,q} \), shows a convex pattern, decreasing while \( q \) is below 5.5\% and increasing after that. The magnitude of the effect of \( q \) on the average potential outcome for the nontreated, in figure 1b, may not look impressive at first, but note that the range of values for \( q \) is rather narrow. The probability of leaving unemployment within a year if one is not assigned to the training program falls from 63\% to 58\% as the proportion of treated \( q \) goes from 2.5\% to 5.5\%. Increases in \( q \) beyond 5.5\% then have a positive effect on the probability of leaving unemployment, as \( Y_{0,q} \) goes back to 63\% when \( q = 8.3\% \) and goes above 65\% when \( q > 9\% \).

\(^{16}\) In step 2, we use a logistic transformation of the market outcome to keep it in \([0, 1]\).

\(^{17}\) The depicted confidence intervals are pointwise. To assess whether nonparametrically estimated functions are monotone, uniform confidence bands would be superior. However, in our setting, this would involve a dramatic increase in the computational burden, which is already substantial.
There are multiple explanations for why an increase in treatment intensity has a positive or a negative effect on the individual potential outcomes under either treatment status. Assume, for instance, that training increases a worker’s productivity. An increase in $q$ can have both a positive and a negative effect on labor demand—positive if firms respond to the increase in average worker productivity by posting more job vacancies, and negative if, for example, (some of) the extra surplus generated by the new matches goes to the worker (through, e.g., bargaining over wages or on-the-job search). A short-term negative locking-in effect of training programs on job search, resulting from workers being unable to devote much time and effort to finding a job while in a training program, may decrease the search effort in the short run and thus the probability of filling a vacancy. The negative patterns observed in figures 1a and 1b may also reflect crowding out, which can affect both nontreated and treated workers as the average productivity among competitors for a job increases with $q$. To assess how these positive and negative interaction effects add up and lead to the patterns observed in figures 1a and 1b is beyond the scope of the paper.

**Average treatment effects.** The above empirical results can be recast in terms of $\hat{\delta}_{q,q} = \hat{Y}_{1,q} - \hat{Y}_{0,q}$ as a function of $q$. Figure 2 plots this function.

For values of $q$ below 5.5%, the average treatment effect does not vary much and stays around −10 percentage points (i.e., the probability of leaving unemployment within a year is 10 percentage points higher without treatment). The small but significantly negative average treatment effect $\hat{\delta}_{q,q}$ at $q$ close to 0 may be a combination of a negligible average causal effect on the exit rate to work after training (see Card, Kluve, & Weber, 2010, for a survey) and a locking-in effect that prevents exits to work during the training spell. At the same time, the small value of the average treatment effect and its insensitivity to $q$ at $q \leq 5.5\%$ should not be interpreted as evidence that the training policy has no effect on workers. After all, at these values of $q$, the average potential outcomes if treated and if nontreated both decrease with the proportion of treated. The insensitivity of the average treatment effect at $q \leq 5.5\%$ is due to the fact that at small values of $q$, these two functions decrease at a similar rate. If $q$ exceeds 5.5%, this pattern changes significantly, and as a result $\hat{\delta}_{q,q}$ displays a decreasing shape.

Figure 2 illustrates the discussion on average treatment effects in section II. Consider an econometrician who, in order to estimate the effect of the treatment, conducts a controlled experiment in a group of workers and thus allocates $Z$ randomly across individuals. This approach will yield an unbiased estimate of the average treatment effect conditionally on $q = E(Z)$, but since the SUTVA is violated, it is not clear whether this effect remains the same should the proportion of treated increase or decrease. From figure 2, an increase in the proportion of treated leads to a more negative average treatment effect as soon as the fraction of treated exceeds 5.5%. Indeed, in that region, we observe substantial changes in the treatment effect even if $q$ varies within a narrow range.

**B. Robustness Checks**

In this section we confront our main empirical findings to a series of robustness checks. The first step of our estimation does not depart significantly from the standard literature on evaluation of labor market policies. Hence, most of the robustness checks are motivated by the second step, at the market level. The results are summarized in figure 3, where
Figure 3.—Robustness Checks, \( \hat{Y}_{1,q} \) (Dotted) and \( \hat{Y}_{0,q} \) (Solid) versus \( q \)

Adding confounders to the estimation of the GPS

3a: benchmark

3b: adding indiv. histories

3c: adding reg./occ. effects

3d: \( \alpha = 1\% \)

3e: \( \alpha = 2.5\% \)

3f: 4 quantiles of \( Q \)

3g: propensity score non param. w.r.t. occup.

3h: \( (d_Z, d_Y) = (9, 15) \)

Figure 3a shows the benchmark results (i.e., the two curves from figures 1a and 1b now drawn in the same graph) for comparison.

The first series of alternative specifications pertain to the confounders used to ensure that the CIA at the market level, condition (6), is satisfied. The benchmark specification presented in section IVC does not include the market average of the unemployment and training history variables in the vector \( W \) of market confounders. Figure 3b shows the average potential outcomes when we account for these variables in the estimation of the GPS. \( Y_{1,q} \) now seems to display some concavity through a slightly increasing pattern for very low values of \( q \), but overall there are no significant differences with the benchmark results. However, including these new confounders reduces the support overlap so that in figure 3b, 23.4% of the market observations are lost to the support condition.

We also check whether the results are affected by allowing more time-invariant market characteristics. As discussed in section IVC, we control only for region fixed effects. Since our markets are defined by a triplet (region, occupation, year), one may ask whether occupation is a relevant fixed characteristic for the assignment of \( Q \) across markets, even though we control for the local labor market tightness \( \theta \). With only two observations (two years) for each region/occupation,
including a region/occupation fixed effect, would take out too much variation in the GPS. To overcome this issue, we aggregate the seven occupations into three groups and estimate the GPS in the second step with a fixed region/aggregate occupation effect. The support condition then takes out 17.9% of our markets. The results, shown in figure 3c, are again in agreement with the benchmark estimates.

We now take a closer look at the support overlap condition in the second step of the estimation. To show that this does not drive our results, we consider alternative specifications where we increase the threshold $\alpha$ to 1% or 2.5%. These stricter conditions take out many observations (33.8% when $\alpha = 1\%$ and 49% when $\alpha = 2.5\%$). Despite this, the results, as shown in figures 3d and 3e, are in agreement with the benchmark estimates. Another way to check the robustness to stricter support conditions is to split the range of $Q$ into more than three intervals. Figure 3f shows the estimated average potential outcomes when using four instead of three equal-probability intervals. Once again, the patterns do not change. A potential concern is that the patterns would change if we did not drop some observations due to a failure of the support overlap condition. In a previous version of this paper, we conducted the estimation with no support overlap condition at the market level. The results, not reported here but available from the authors on request, were qualitatively consistent with those shown in figures 1a and 1b.

We also check whether the specification used in the first stage of the estimation, at the individual level, influences the results. Since we do not have enough observations to separately estimate propensity scores for each region/occupation/year cell, we included occupation and year dummies in the set of regressors (together with the confounders $X$) and ran one estimation per region. We now consider a specification where, in the first step, propensity scores are estimated nonparametrically with respect to the occupation variable, whereas region and year dummies are included as regressors. Results are shown in figure 3g. They do not differ from the benchmark estimates.

Finally, we reconsider the choice of elapsed unemployment durations at which the treatment and outcomes variables are observed. In our benchmark results, treatment status is measured $d_Z = 6$ months after unemployment starts, whereas the outcome $Y$ equals 1 if the individual has left unemployment within $d_Y = 12$ months. As an alternative, we now set $(d_Z, d_Y) = (9, 15)$ months. Figure 3h shows that the qualitative results still hold: the SUTVA is violated, $Y_{1,q}$ is decreasing and $Y_{0,q}$ is decreasing for low values of $q$. One small difference we find with figure 3a is that $Y_{1,q}$ now has a stronger decreasing pattern when $q$ is low.

Finally, recall that one may use a different approach for inference in the first step as an alternative to the matching on observed characteristics. Although the main focus of this study is the SUTVA, not the CIA, and although in our empirical application we have a rich set of observed individual confounders, it may be interesting to examine an approach that deals with possible selection on unobservables in the first step. To this end, we consider the treatment and outcome variables as continuous duration variables possibly dependent on correlated unobserved heterogeneity terms, and we use the timing-of-events approach (Abbring & Van den Berg, 2003) for identification of a model with parametric treatment effects and mixed proportional hazard specifications. Preliminary investigations (in the working paper version of this paper) suggest that the resulting estimates of $Y_{z,q}$ show similar patterns to those discussed in this section.

VI. Conclusion

Evaluation with nonexperimental data usually requires two critical assumptions. First, the treated individuals and the comparison individuals should on average be as similar as possible except for their treatment status. Second, an individual’s outcome should not depend on other individuals’ treatment status. With matching methods, these are the CIA and the SUTVA. There is a tension between these assumptions. The more similar that two individuals are, the more likely it is that they act in each other’s proximity, which makes it more likely that their outcomes interfere. In this paper, we estimate the degree of interference by taking to nonexperimental data an augmented Rubin model (1974) that does not impose the SUTVA at the individual level. In our empirical analysis, we focus on training program participation by unemployed individuals. The results show that the estimated mean potential outcomes of treated and nontreated job seekers depend significantly on the proportion of individuals treated in the relevant local labor market. This draws attention to possible crowding-out effects of the large-scale introduction of training programs. In addition, we view the empirical results as supportive of the increasing research focus on interactions in general and spillover and equilibrium effects of active labor market policies in particular.

REFERENCES


Graham, Bryan, Guido Imbens, and Geert Ridder, “Measuring the Average Outcome and Inequality Effects of Segregation in the Presence

18 The first group consists of office/social workers and employees, the second of industrial and construction workers, and the last of executives and technical workers (who have supervision/management duties).
EVIDENCE OF TREATMENT SPILLOVERS WITHIN MARKETS 823

of Social Spillovers,” mimeograph, Department of Economics, University of California Berkeley.


——— “The Propensity Score with Continuous Treatments” (pp. 73–84), in A. Gelman and X.-L. Meng, eds., Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives (New York: Wiley, 2004).


