Treatment Evaluation in the Case of Interactions within Markets

Marc Ferracci
Grégory Jolivet
Gerard J. van den Berg

January 2010

Treatment Evaluation in the Case of Interactions within Markets

Marc Ferracci
University of Marne-la-Vallée
and CREST-INSEE

Grégory Jolivet
University of Bristol

Gerard J. van den Berg
University of Mannheim, VU University Amsterdam,
IFAU and IZA

ABSTRACT

Treatment Evaluation in the Case of Interactions within Markets[*]

We extend the standard evaluation framework to allow for interactions between individuals within segmented markets. An individual's outcome depends not only on the assigned treatment status but also on (features of) the distribution of the assigned treatments in his market. To evaluate how the distribution of treatments within a market causally affects the average effect within the market, averaged over the full population, we develop an identification and estimation method in two steps. The first one focuses on the distribution of the treatment within markets and between individuals and the second step addresses the distribution of the treatment between markets. We apply our method to data on training programs for unemployed workers in France. We use a rich administrative register of unemployment and training spells as well as the information on local labor demand that is used by unemployment agencies to allocate training programs. The results show that the average treatment effect on the employment rate causally decreases with respect to the proportion treated in the market. Our analysis accounts for unobserved heterogeneity between markets (using the longitudinal dimension of the data) and, in a robustness check, between individuals.

JEL Classification: C13, C14, C21, C31, J64

Keywords: treatment evaluation, equilibrium effects, matching estimators

Corresponding author:

Gerard J. van den Berg
University of Mannheim
Department of Economics
L7, 3-5
68131 Mannheim
Germany
E-mail: gerard@uni-mannheim.de

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 one is a conditional independence assumption (CIA thereafter) in order 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 thereafter), rules out any influence of an individual's treatment status on another individual's potential outcome (see Neyman, 1923, or Rubin, 1986). Recently, the latter has been receiving increasing attention from economists, as many features pertaining to social sciences in general and economics in particular are likely to generate interactions between individuals with different treatment statuses. This includes peer effects, neighborhood effects, network effects and the dispersion of information, and various kinds of equilibrium effects of policy interventions in e.g. the labor market or the education system (below we discuss the literature in more detail).

In the first half of this paper, we develop a general framework to estimate average treatment effects that takes account of possible interactions between individuals, and that, indeed, estimates the magnitude of the interaction effects. The "matching" approach to evaluation constitutes the point of departure for our framework. Accordingly, it is designed for cases where the data are not based on randomized experiments. Consider a population divided into "markets". A market can be defined as a set of agents whose outcomes may depend on the treatment statuses of the other agents in the set, whereas there is no such interaction with agents outside of the set. Conceptually, the individual treatment operates at two levels: individuals and markets. If a CIA at the individual level holds within a given market, either using an experimental design, or by conditioning on a relevant set of confounders, or by exploiting some other feature of the data (for instance a panel), then average potential outcomes of treated and non-treated individuals can be estimated. However, if the SUTVA does not hold, these are only valid for the actual observed distribution of the treatment within this market. We deal with this by extending the standard potential outcome model (the Rubin model, 1974) to allow an individual's outcome to be a function not only of his own treatment status but also of the distribution of the treatment in his market. We specifically aim at the estimation of (*i*) the causal effect of the fraction of treated individuals in a market (or any other feature of the treatment distribution) on the average treatment effect in the market, and (*ii*) the causal effect of treatment on the individual outcome. We aim

---

[1]See e.g. Heckman, LaLonde and Smith (1999) for a review.

to estimate the former causal effect as a function of the fraction of treated in the market, if the market invariably has the same composition as the population of individuals. In cases where the individual treatment consists of the exposure to a particular policy intervention, like participation in a training program, this function is of obvious interest for policy makers. It captures information that is necessary for any decision on the extent to which a program should be rolled out in the eligible population. Accordingly, it is important to have an estimation method for this function that is valid in case of non-experimental data, and it is important to be precise about the underlying assumptions for such a method.

Our method involves two consecutive steps: one within each market, and one across all markets. To achieve identification, we need a CIA at each step. The first step consists of the estimation of mean potential outcomes within each market separately. The market outcomes are the mean potential outcomes estimated in the first step. We define the market treatment as some relevant moment(s) of the distribution of the treatment. For instance, with a binary treatment, we only need to consider the proportion of treated in each market. Using a second CIA to control for confounders driving the assignment to market treatments and potential outcomes, we can then recover the causal effect of the treatment distribution within a market on the average treatment effect in the market, if the market has the same composition as the population of individuals. As with "matching" techniques in general, our approach deals with heterogeneity in treatment effects. In our framework, this concerns heterogeneity across individuals as well as across markets.

In fact, the approach is sufficiently flexible to leave room for a wide set of alternative techniques within each step. If a stage involves non-experimental data, we can also resort to structural models or econometric techniques (for instance panel data techniques) to estimate treatment effects at this stage. This may be useful if the corresponding CIA is hard to justify. However, depending on the technique, it may restrict the extent to which effect heterogeneity is allowed.

In the second part of the paper, we apply our approach to study the effects of participation in a training program for unemployed workers in France on their probability to move to employment. In this setting, markets are defined by occupational, spatial, and temporal indicators. One may expect the fraction of treated individuals in a market to influence the magnitude of the individual effect of participation. First, if many individuals are treated then there may be crowding out among trained individuals applying for the same vacancies whereas the reverse may occur for untrained individuals. This may cause the average effect to decline as a function of the fraction of treated individuals. Secondly, if many individuals are treated then there may be a response from the other side of the market. For example, firms may substitute vacancies for non-trained workers by vacancies for skilled workers. Whether

this causes the average effect to increase or decrease as a function of the fraction of treated individuals depends on the firms' production function and on the wage determination in the market (see below for some literature). In the empirical analysis we set out to determine the precise shape of the average effect as a function of the fraction of treated individuals.

We use a register data set containing detailed information on unemployment and training spells in France at the individual level. These data are non-experimental. To produce our non-parametric benchmark results for within-market average effects, we follow the matching literature and make use of a rich set of covariates to ensure conditional independence between treatment and potential outcomes. We use the longitudinal dimension of our data to compute indicators of an individual's unemployment and training histories. While we evaluate the training programs over the period 2002-2006, our data go back as far as 1990, so we know how much time each worker spent in unemployment and/or in a training program in the years preceding his current unemployment spell. We view this information as important in order to capture individual confounders that may affect the actual treatment assignment. We follow Hirano and Imbens (2001) and use an augmented version of the weight estimator of Hirano, Imbens and Ridder (2003) for the estimation.

At the market level, we need to ensure that no unobserved confounder drives both the proportion of individuals going through a training program and the average unemployment duration when treated or non-treated. To this end, we use two distinctive features of our data. First, we have unique information on local labor demand. We merge our data with a survey from the French national unemployment agency where firms report their job opening predictions for the coming year at very precise geographical and occupational levels. This survey contains the information actually used by caseworkers to assign unemployed workers to training programs. Second, we follow local labor markets (defined as a pair occupation/region) through time, so we observe each market under several treatment regimes (i.e. with different proportions of treated). This allows us to control for a fixed unobserved market 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.

We also consider non-matching estimators at the individual level, to address a concern that we may not capture all selective unobserved heterogeneity at the individual level. Using the Timing-of-Events approach devised by Abbring and Van den Berg (2003), we are able to account for unobserved individual features driving assignment to training and unemployment durations. We can then compute the potential individual outcomes in each market and run our second step. This latter approach yields results that are qualitatively similar to the ones using the matching technique.

We should mention that we can only make some speculative remarks about the mechanisms through which the treatment and its distribution impact individuals and markets. To investigate this further may require the estimation of a structural economic model and/or additional data specific to the empirical application, and we feel that this would be beyond the scope of this paper. However, and this is an additional key contribution of this paper, by showing evidence of treatment spillovers on a very large scale we provide empirical support to the literature on equilibrium effects of treatments.

We end this section by discussing some relevant previous and concurrent literature on the evaluation of treatments in the presence of interactions or interference. During the last decade, interest in such evaluations has been growing in economics as well as in epidemiology and sociology. Somewhat loosely, one may distinguish between three bodies of work, where in fact these are highly related. The first of these concerns studies of the effect of an intervention program on group-level outcomes using data from field experiments in which the assignment of programs to groups (i.e., markets) is randomized or natural experiments. In this setting, ideally, the average effect can be directly estimated from sample averages. For example, Angelucci and De Giorgi (2009) analyze population effects of cash transfers to some households in the context of the Progresa program in Mexico. Duflo, Dupas and Kremer (2008) exploit randomization across schools and across classes in Kenya to study the effect of tracking. Frölich and Michaelowa (2005) evaluate the effects of textbooks on pupils and their classmates in five African countries. Miguel and Kremer (2004) use the random allocation of a medical treatment across villages and distance between these villages to reveal treatment spillovers. The approach has also been used to study neighborhood effects (see Oakes, 2008, for an overview). Evaluation studies of active labor market policies (ALMP hereafter) have not followed this approach for lack of experimental data. A notable exception is Blundell, Costa Dias, Meghir and Van Reenen (2004) which uses eligibility rules and the timing of implementation across regions to detect equilibrium effects of a job search program in the United Kingdom. Lechner (2002) studies whether the treatment participation probability is a source of heterogeneity in the treatment effect, but this link is not discussed in light of the SUTVA.

A second body of work considers group treatments that consist of the assignment of a within-group composition, and their effect on a group-level outcome (see Graham, Imbens and Ridder, 2009). For example, one may be interested in the causal effect of the classroom gender composition on the average school grades outcome. In this case the fraction of individuals with a certain characteristic is assigned to a group, but these characteristics are seen as inalienable, so that the individual effect of a counterfactual switch of the individual characteristic is not of primary interest. Graham, Imbens and Ridder (2009) provide a

careful discussion of the CIA needed in case of non-experimental data. Graham, Imbens and Ridder (2006) demonstrate how to use estimates to assess optimality of allocation schemes. These studies emphasize the importance of conditioning on group-level confounders.

Recently, some studies have recognized the benefits of a two-dimensional evaluation approach, in cases where every individual is potentially subject to an individual treatment as well as to interactions, and where potential outcomes depend on both, and where the interest is in the individual counterfactual effect as well as in the interaction effects. 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. (Hudgens and Halloran, 2008, also contains a short overview of older studies with two levels based on linear regression model specifications.) Manski (2009) also defines such causal effects with social interactions, and he considers various restrictions on the type of interactions and their implications.

Clearly, our paper builds on these three bodies of work as we develop a two-dimensional evaluation approach in non-experimental contexts. In our setting, the interest is in the effect of the fraction of treated, as well as the effect of treatment itself, on the individual outcome. Our estimation method is non-parametric, and the empirical application is highly relevant for labor market policy.

We should also mention a fourth body of work, which involves the estimation and calibration of structural equilibrium models that aim at quantifying specific pathways through which the treatment may impact the market or the economy. Heckman, Lochner and Taber (1998) study equilibrium effects of tuition policies. Lise, Seitz and Smith (2005) and Cahuc, Crépon, Guitard and Gurgand (2008) evaluate ALMP, and Albrecht, Van den Berg and Vroman (2009) evaluate equilibrium effects of adult education programs. Topa (2001) uses the structural econometric approach to address social network interactions and spillovers in the labor market.

The paper is organized as follows. Section 2 describes the statistical model and defines the outcomes of interest. Section 3 presents our two-step identification and estimation strategy. Section 4 introduces our empirical application: the institutional setting for training programs in France, the data, and the econometric specification. Section 5 presents our benchmark estimation results. Section 6 contains our main robustness check where we use the Timing-of-Events model to explicitly account for individual unobserved heterogeneity. Section 7 concludes. Additional robustness checks are in the Appendix.

2 A model to evaluate treatment effects at the individual and market level

2.1 Treatments, outcomes, and markets

The economy is segmented into M isolated markets. We assume that an individual belongs to only one market and we denote the population of market m by N_m. An observation consists of a pair (i, m) where i denotes the individual and $m = m(i)$ the market where i evolves (we will also use the notation $i \in m$ when $m(i) = m$).

We consider a binary treatment and define the dummy Z equal to one for individuals who have received the treatment and to zero for the others.[2] We also define an indicator of the treatment at the market level: let P_m be any function of the vector $[Z_i]_{i \in m}$ such that (i) P_m is the same for all individuals in a given market, and (ii) Z_i is not a deterministic function of $P_{m(i)}$. The former restriction states that the distribution of the treatment at the market level is the same for all individuals in a given market. The latter restriction rules out all cases, such as $P_m = [g(Z_i)]_{i \in m}$ with g strictly monotone, for which Z does not bring additional information with respect to P_m. Notice that P_m can be multi-dimensional. In our empirical application, we will take P_m to equal the proportion of actually treated in a given market:

$$P_m = E_{i \in m}(Z_i) = \frac{1}{N_m} \cdot \sum_{i \in m} Z_i. \tag{1}$$

We will often refer to this specification of P for illustration purposes. For notational convenience, we write $P_m = 0$ if no treatment is introduced in market m.

The individual outcome is denoted by Y_i and depends on the treatment through Z but also through the market variable P:

$$Y_i = Y_i\left(Z_i, P_{m(i)}\right). \tag{2}$$

If the SUTVA holds (see Rubin, 1986), then an individual's potential outcome is not affected by the treatment status of other individuals. In that case, Y_i does not depend on $P_{m(i)}$. More in general, we allow for treatment externalities across individuals within a market (through P), but we assume that there are no spillovers between markets, i.e. the distribution of outcomes Y in a market does not depend on the values of P or Z in other markets. In effect, this dictates the appropriate operationalization of the "market" concept.

---

[2]What follows could be extended to the case of a continuous treatment.

2.2 Potential outcomes and average treatment effects

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

$$Y_{z,p} = E_m E_{i \in m} [Y_i(z, p)], \quad \forall (z, p). \tag{3}$$

The expectation in (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 the market size as defined by the measure of eligible individuals in the market. For notational convenience we omit the latter weights (this applies to all expressions below). Accordingly, $Y_{z,p}$ as a function of p captures the average potential outcome corresponding to the individual treatment status z as a function of the market treatment dimension p (the fraction of treated in the market), if the market invariably has the same composition as the population of eligible individuals. These functions $p \to Y_{z,p}$ for $z = 0, 1$ are the objects in which we are primarily interested. Obviously, if the SUTVA holds, and if p is a scalar, we have $dY_{1,p}/dp = dY_{0,p}/dp = 0$.

The functions $Y_{z,p}$ directly lead to average treatment effects,

$$\delta_{p,p'}^{z,z'} = Y_{z,p} - Y_{z',p'}. \tag{4}$$

The set $\{\delta_{p,p'}^{1,0}, \delta_{p,p'}^{0,0}, \delta_{p,p'}^{1,1}\}$ fully describes the effects of the treatment.[3] For instance, $\delta_{p,p'}^{1,1}$ (resp. $\delta_{p,p'}^{0,0}$) gives the variation in the expected outcome of treated (resp. non treated) individuals when going from p to p'. The parameter $\delta_{p,p'}^{1,0}$ compares the average outcome of being treated when $P = p$ with that of not being treated when $P = p'$. At the limit $p' = 0$, no treatment is introduced, so $\delta_{p,0}^{1,0}$ is the average effect of being treated with $P = p$ as compared to the situation where no treatment is introduced at all. This latter parameter and $\delta_{p,0}^{0,0}$ are of importance for policy makers, as they indicate the "pure" effect on treated and non-treated individuals of the introduction of a treatment with a market dimension p.

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. This literature is therefore concerned with the average treatment effect $\delta_{p,p}^{1,0}$. In principle, the results are only valid for a given value of p,[4] and for the composition of the sub-population of individuals in the market under consideration. It is not possible to

---

[3] Note that $\delta_{p,p'}^{0,1} = -\delta_{p'p}^{1,0}$.

[4] Note that $\delta_{p,p}^{1,0} = \delta_{p,0}^{1,0} - \delta_{p,0}^{0,0}$. We have $\delta_{p,0}^{0,0} = 0$ only if the counterfactual $Y_{0,p}$ gives an accurate picture of what would have happened had the treatment not been introduced at all, i.e. if it equals $Y_{0,0}$. If the SUTVA does not hold, this will not be the case.

extrapolate the results to other values of p 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 p differs from the one initially estimated. Formally, defining P with (1), if one finds $\delta_{p,p}^{1,0} > 0$ and consequently P is increased from p to p', the new ATE $\delta_{p',p'}^{1,0}$ can be smaller or larger than $\delta_{p,p}^{1,0}$.

3   Identification and estimation

Our objects of interest are (functions of) the average potential outcomes $Y_{z,p}$. For a given individual i we only observe $Y\left(Z_i, P_{m(i)}\right)$ so we face a standard missing observation problem. In our setting, this problem has two dimensions: first, we do not observe the outcome of an individual with $Z_i = z$ had he received $1 - z$, and second, we do not know what would have happened in a market m with $P_m = p$ had this market been assigned to a treatment policy regime $p' \neq p$. We derive a two-step method for identification and estimation that explicitly accounts for these two levels. First we present two unconfoundedness assumptions that allow identification of $Y_{z,p}$ from the data. After that we present the two-step estimation method.

3.1   Identification

We make two unconfoundedness assumptions: one at the individual level and one at the market level. The first assumption states that, conditionally on a set of market characteristics denoted by W, the allocation of P across markets is independent of the average potential outcomes of treated and non-treated workers within each market:

$$E_{i \in m}[Y_i(z,p)] \quad \perp \quad P_m \quad | \quad W_m, \quad \forall m, z, p. \tag{5}$$

Basically, this assumption takes markets as units. P is a market treatment, and $E_{i \in m}[Y(z,p)]$ are the market outcomes of interest.

   The second unconfoundedness assumption states that within each market m, the allocation of treatment Z across individuals does not depend on the potential outcomes conditionally on a set of individual characteristics denoted by X:

$$Y_i\left(z, P_m\right) \quad \perp \quad Z_i \quad | \quad X_i, \quad \forall z, i \in m. \tag{6}$$

Assumption (6) is the usual CIA that economists have used in the evaluation of treatments using matching estimators.

In an experimental setting, it is possible to ensure that (5) and (6) hold. With observational data, it is an empirical question whether the set of conditioning variables for the two CIA's is sufficient. As we shall see, one may use longitudinal dimensions of the data (at both the market and the individual level) to control for time-invariant unobserved characteristics. We return to this issue in Sections 4 and 6.

Under assumptions (5) and (6), we can identify (functions of) the average potential outcomes $Y_{z,p}$ by way of

$$Y_{z,p} = E_W \left[E_m \left\{ E_{X|m} \left[E_{i \in m}(Y_i | X_i = x, Z_i = z) \right] | W_m = w, P_m = p \right\} \right]. \tag{7}$$

This equality follows from:

$$
\begin{aligned}
Y_{z,p} &= E_m \left\{ E_{i \in m}[Y_i(z,p)] \right\} \\
&= E_W \left[E_m \left\{ E_{i \in m}[Y_i(z,p)] | W_m = w \right\} \right] \\
&= E_W \left[E_m \left\{ E_{i \in m}[Y_i(z,P_m)] | W_m = w, P_m = p \right\} \right] \\
&= E_W \left[E_m \left(E_{X|m} \left\{ E_{i \in m}[Y_i(z,P_m) | X_i = x] \right\} | W_m = w, P_m = p \right) \right] \\
&= E_W \left[E_m \left(E_{X|m} \left\{ E_{i \in m}[Y_i(Z_i,P_m) | X_i = x, Z_i = z] \right\} | W_m = w, P_m = p \right) \right],
\end{aligned}
$$

where (5) allows to go from the second to the third line while (6) allows to go from the fourth to the last line. The last line equals the right-hand side of equation (7), which is an observable quantity.

Notice again that the expectations over markets and over W 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.

It may be useful to provide some intuition for the above result. Our line of reasoning follows the above set of equalities in reverse order, from the bottom to the top. In a given market, conditional on $X = x$, it is a pure coincidence whether the individual receives the treatment z or not. The average potential outcome after treatment for all individuals in that market with $X = x$ and with a given P is then equal to the average outcome of those with $X = x$ who were actually treated in that market. For a given P and X, any alternative subset of individuals receiving the treatment in this market would produce the same average outcome as the average outcome among the actual treated. This can be aggregated over all X in the market. Next, conditional on $W = w$, it is a pure coincidence whether the market receives a high value p of P or not. In any other market with $W = w$, the average potential outcomes for $z = 0$ and $z = 1$ if it had been exposed to p are then equal to the average potential outcomes in the market that actually received the value p. This can be aggregated over W to obtain population-level averages.

This line of reasoning carries a caveat for empirical applications, namely that it is important that W_m includes indicators of the distribution of X in the market m. Without this, the CIA (5) entails that the allocation of treatments to markets is not driven by the market composition in terms of characteristics that affect potential outcomes and, possibly, cause self-selection into treatments. The latter scenario is often not realistic. Consider the training program setting of our empirical analysis. In reality, if a market contains many discouraged workers who are not sufficiently motivated to search for a job without any supplementary training, then the policy makers may choose to allocate a relatively high budget for training for that market. If individual past labor market outcomes capture the degree of discouragement, then those past outcomes (X) may affect potential outcomes and, possibly, the treatment status, and the distribution of X may affect P.

If the evaluation framework is dynamic, with treatments and outcomes being outcomes of stochastic processes over time, then identification also requires a "no-anticipation" assumption. See Crépon, Ferracci, Jolivet and Van den Berg (2009) for details. This assumption is often not explicitly addressed in studies based on the "matching" approach. We follow this convention. We return to this in Section 6.

3.2   Estimation

Let the data consist of a sample of individuals i. For each individual, we observe his treatment status Z_i, his outcome Y_i and his observed characteristics X_i. Moreover, we observe the individual's market $m(i)$ as well as the characteristics of this market $W_{m(i)}$ and the market dimension of the treatment $P_{m(i)}$. To make the presentation clearer, we now assume that $P_{m(i)}$ is a scalar. Our main purpose is to compute the estimate defined by (7). To this end, we suggest a two-step estimation method based on assumptions (5) and (6). The first step consists in estimating the average individual outcome $E_{i \in m}\left[Y_i\left(z, P_m\right)\right]$ for all markets and $z = 0, 1$. The second step considers the quantities $\widehat{E}_{i \in m}\left[Y_i\left(z, P_m\right)\right]$ estimated in the first step and evaluates the effect of treatment P on these market outcomes.

First step: estimation at the individual level.   We want to estimate the average individual potential outcomes within each market. To this end, we first use the result of Rosenbaum and Rubin (1983) which states that assumption (6) implies:

$$Y_i\left(z, P_m\right) \quad \perp \quad Z_i \quad | \quad p(X_i, m), \quad \forall z, i, m = m(i) \tag{8}$$

where $p(x, m)$ is the propensity score: the probability of being treated conditionally on having individual characteristics x. We allow for one propensity score function for each market.

We use a weighted regression estimator which combines two methods: the approach based on regressing the outcome on the treatment variable and the covariates (see Rubin, 1977) and the weight estimator developed by Hirano, Imbens and Ridder (2003). The latter consists in computing an estimate of the average outcome for the treated as the market average of $Z \cdot Y / \widehat{p}(X, m)$ where $\widehat{p}$ is an estimate of the propensity score. Likewise, the outcome for the non treated is estimated by the average of $(1 - Z) \cdot Y / (1 - \widehat{p}(X, m))$. The weighted estimator can thus be seen as a weighted regression of Y on Z using the weights $[Z/\widehat{p}(X, m) + (1 - Z)/(\widehat{p}(X, m))]^{1/2}$. 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 estimators has been introduced by Robins and Ritov (1997) and was implemented in Hirano and Imbens (2001). For each market m, we run the following regression:

$$Y_i = \mu_m + X_i' \cdot \beta_m + \tau_m \cdot Z_i + u_i, \quad i \in m, \tag{9}$$

using the weights $[Z_i/\widehat{p}(X_i, m) + (1 - Z_i)/(\widehat{p}(X_i, m))]^{1/2}$. Denoting the resulting estimates as $(\widehat{\mu}_m, \widehat{\beta}_m, \widehat{\tau}_m)$, we compute the market outcomes as:

$$
\begin{aligned}
\widehat{E}_{i \in m}[Y_i(1, P_m)] &= \frac{1}{N_m} \cdot \sum_{i \in m} \widehat{\mu}_m + X_i' \cdot \widehat{\beta}_m + \widehat{\tau}_m \\
\widehat{E}_{i \in m}[Y_i(0, P_m)] &= \frac{1}{N_m} \cdot \sum_{i \in m} \widehat{\mu}_m + X_i' \cdot \widehat{\beta}_m.
\end{aligned}
\tag{10}
$$

The empirical implementation of these estimates raises dimensionality issues of two sorts. The first one is the degree of the factor series in the estimation of the propensity score. There is a conflict between the high order required by the efficiency of the estimator on the one hand (see Hirano, Imbens and Ridder, 2003), and the number of observations and the size of the X vector on the other. The second issue is prominent for our approach, since we need to estimate a propensity score for each market m. If some markets are small, it is not possible to run an estimation of the propensity score for each m. In Subsection 4.3 and Appendix B we show how to choose specifications to overcome this issue.

Second step: estimation at the market level. Next, we average the estimates $\widehat{E}_{i \in m}[Y(z, P_m)]$ across markets, assuming that the market treatment P 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 Robin (1983) to the case of a continuous treatment[5] and show that the unconfoundedness assumption (5) leads to:

$$E_{i\in m}[Y_i(z,p)] \quad \perp \quad \mathbf{1}\{P_m = p\} \quad | \quad f(P = p|W_m), \quad \forall m, z, p. \tag{11}$$

The function $f(\cdot|w)$ is the conditional density of the treatment P. It is the counterpart of the propensity score in the case of a continuous treatment and is thus called the generalized propensity score (GPS thereafter). We then proceed to estimation in three steps.

a) The first sub-step consists in estimating the GPS by regressing P on (a flexible function of) the market variables W. Denoting this estimate by $\widehat{f}$ we can define for each market m, the estimated value $\widehat{f}_m(p)$ of the GPS for any p:

$$\widehat{f}_m(p) = \widehat{f}\left(p|W_m\right).$$

b) We estimate the function that links the market outcome, $E_{i\in m}[Y(z,P)]$ to the market treatment P and the GPS. We thus regress $\widehat{E}_{i\in m}[Y(z,P)]$ on a flexible (in general quadratic) function of P and $\widehat{f}(P)$ and denote as $\widehat{Q}_z$ the resulting function:

$$\widehat{E}_{i\in m}[Y_i(z,P)] = \widehat{Q}_z\left(P_m, \widehat{f}_m\left(P_m\right)\right).$$

c) Lastly, for each value of z and p, we compute our estimate of $Y_{z,p}$ as the average of $\widehat{Q}_z\left(p, \widehat{f}_m(p)\right)$ over markets:

$$\widehat{Y}_{z,p} = E_m\left[\widehat{Q}_z\left(p, \widehat{f}_m\left(p\right)\right)\right].$$

An alternative single-step approach. The two-step method we have presented above involves the estimation of two propensity scores, one at the individual level (for the binary treatment Z) and one at the market level (for the continuous treatment P). It could thus be that the sensitivity issues raised by the estimation of the propensity score (specification, overlap) add up to yield imprecise and/or unstable estimation results. For some specific applications, for instance if the "markets" are classrooms or neighborhoods with few units, one might prefer a less demanding estimation procedure than our two-step method. In Appendix A we present an alternative single-step identification and estimation strategy with a single CIA.

---

[5] For the case with a multivariate but not necessarily continuous treatment, see Imbens (2001).

Discussion. We briefly return to the implications of using more conventional "matching" estimation methods if the real data-generating process is as described in Section 2 and Subsection 3.1, i.e. with multiple markets and two CIA. We have already seen at the end of Section 2 that within-market "matching" estimation with given P for each market separately does not allow for extrapolation to other values of P. For such an extrapolation we need to perform the second estimation step above.

Suppose one would only perform the second step, that is, perform matching at the market level, conditioning on the appropriate W_m, and using as market-specific outcomes simply the mean outcomes for $Z = 1$ and $Z = 0$ in the market, or their difference, or their mean weighted by P and $1 - P$. This basically captures the approach used in the literature if the within-market assignment is randomized or if the single treatment is a group composition. In our case, this method produces a reduced-form estimate of the over-all effect of P in the market on mean outcomes in the market. The within-market average outcomes do not equal the average potential outcomes because of the selectivity generated by X_i. Therefore we can not deduce the average individual effect of Z, nor do we know the population of individuals to which the estimate applies.

Now consider the estimation of the ATE without consideration of the existence of multiple markets. The perceived treatment is then one-dimensional and equal to Z, whereas P is unobserved. The objective would be to estimate average potential outcomes that could be expressed as $Y_{z=1}$ and $Y_{z=0}$, by assuming a CIA with confounders X_i. Obviously, this approach can not address the causal effect of P. Moreover, if $\text{var}(P_m) > 0$ then a conventional matching estimator will compare individuals with each other regardless of whether they operate in the same market. There is a selection problem if W_m is not included in X_i, because then W_m affects Z_i as well as the market-specific mean potential outcomes. In that case the average potential outcomes would be inconsistently estimated.

4 Application to training policies in France

In this section, we present our empirical application. We study the effect of training programs for unemployed workers in France on the probability of moving from unemployment to employment. 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 level. Lastly, we discuss the econometric model along two lines: the two conditional independence assumptions and the model specification.

14

4.1 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 theoretical model, the former point is key to understand between-market variation in P, while the latter helps to interpret within-market variation in Z. We also give some information on the content of training programs. Table 1 lists the acronyms we use below and summarizes features of the training system.

General organization. The French training system is run by three different bodies: the national state, the social partners (trade unions and employer organizations) and the administrative regions. The state funds training programs for the long-term (more than 12 months) unemployed who have exhausted their rights to unemployment benefits, as well as for welfare recipients. It also provides revenues to job seekers who are not eligible to unemployment benefits and who participate in state-appointed training programs. In addition, the state offers training to eligible and non-eligible unemployed through the public employment service called ANPE. The role of ANPE is to counsel the unemployed in their search activities and to monitor them. Any job seeker who wishes to enter a public training program must consult his local ANPE agency.

The social partners manage the institution in charge of the payment of unemployment benefits, called UNEDIC. UNEDIC provides all the funding for the unemployment benefits of eligible trainees. Besides, UNEDIC and its local agencies, called ASSEDIC, are now in charge of prescribing and buying specific training courses for eligible job seekers.

Finally, the administrative regions are also in charge of funding of training programs. Moreover, they express their needs for skills at the local level to ASSEDIC and ANPE agencies, based on the vacancies that are opened every year. The ANPE agencies are then asked by law to assign job seekers to training programs suited to the vacancies. For their part, ASSEDIC agencies are in charge of the assignment of eligible job seekers to the training programs they fund. In all cases, training capacities should be calibrated to fit open 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 P between regions and occupations.

Table 1: The French system of training for the unemployed

ANPE	-	*Agence Nationale Pour l'Emploi* Public employment service, counsels and monitors unemployed workers. Run by the state.
ASSEDIC	-	*Association pour l'Emploi dans l'Industrie et le Commerce* Local agencies of the UNEDIC.
BMO	-	*enquête Besoin de Main d'Œuvre* Survey conducted by the ASSEDIC every year since 2001. Collects firms' job opening predictions for the next year. Helps ANPE assign training programs.
FNA	-	*Fichier National des ASSEDIC* National register of unemployed workers.
UNEDIC	-	*Union Nationale interprofessionnelle pour l'Emploi Dans l'Industrie et le Commerce* Institution in charge of paying unemployment benefits. Run by the social partners (unions and employers). Pays unemployment and welfare benefits. (Since 2001) prescribes and buys some specific training courses. Conducts a yearly survey on local labor demand (BMO). These tasks are run at the local level by the ASSEDIC agencies.

Assignment to training programs. A meeting with an ANPE caseworker (typically 30 minutes long) is compulsory for all newly registered unemployed workers and recurs at least every 6 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 moment by the unemployed workers themselves. Apart from a wide range of counselling measures, training programs may be proposed to job seekers during interviews or in 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 not given.

Each year since 2001, the ASSEDIC conduct a survey, called BMO, on the predicted job vacancies at the local level (see Subsection 4.2 for more details on this survey). In particular, the BMO survey intends to give ANPE caseworkers some information to help them assign unemployed workers to training programs consistently with the open vacancies. Therefore, the need for specific skills should be correlated with the probability of being treated at the local level. In theory, the less employable persons have priority to enter

training programs. Yet, some recent field studies (see e.g. 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. This suggests that self-selection plays a significant part in training participation. In Subsection 4.3 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. In Section 6 we address the issue of unobserved heterogeneity in individual selection into training.

Job seekers may also find training programs by themselves. Some surveys indicate that those programs are generally oriented toward the acquisition of more general human capital. In that case the unemployed worker can benefit from some public funding[6] to cover the program's tuition costs. This requires that the program be validated by an ANPE caseworker who is in charge of checking that the program is somehow relevant given the job seeker's professional project and the local labor demand. A field study by Fleuret (2006) also shows that low-skilled workers are far less likely to ask for the validation of a training program than, for instance, executives. Finally, it turns out that ANPE caseworkers have much power in the assignment process, as they may either prescribe or validate the training programs.

Contents of training programs. Unfortunately, the data we use for the empirical analysis do not contain information on the content of training programs. However additional data provided by UNEDIC make it possible to describe this content with some precision.[7] These data give a set of details on training programs, including a grouping into four types of training: "*general*" (e.g. mathematics, economics, languages), "*personal*" (e.g. development of mental abilities, development of professional organization capacities), "*service oriented vocational skills*" (e.g. accounting, hotel business) and "*production oriented vocational skills*" (e.g carpentry, engineering). 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*". Additional information is given on the education level of the programs, showing that programs at the college level and above account for less than 25% of the total. Hence, about 75% of training programs are presumably oriented towards low-skilled workers. This is in line with the objectives of ANPE caseworkers when assigning people to training. [8]

---

[6]The funding may come from the administrative regions, the state or UNEDIC, depending on the eligibility of the job seeker, and on the content of the program.

[7]Due to the lack of common identifiers, we cannot merge this additional data set with the one we use in our estimation.

[8]We do not have information on what types of programs are most frequently proposed by ANPE caseworkers, and what types are mostly chosen by the unemployed.

4.2 The data

Individual data. We use quarterly extracts from the FNA, an administrative data set collected by UNEDIC. The FNA file contains information on all the workers entering unemployment who are unemployment benefits or welfare recipients. We use eight randomly selected extracts of the FNA. Each extract represents 2.5% of unemployed workers between 1990 and 2007.[9] For each individual, the extracted file mixes information collected by UNEDIC and by ANPE on all the unemployment spells that could have occurred since 1990. It contains the dates at which workers are registered and unregistered as unemployed by the public employment service, as well as the dates when people enter and exit training.

We start the analysis in 2002, for three reasons. First, between 1993 and 2001, the time profile of unemployment benefits was decreasing over the unemployment period. However, for unemployed workers who entered a training program, the unemployment benefits remained constant until the program stopped. Hence, the system was providing an eligibility incentive to enter a program regardless of a causal effect of this program on re-employment. A reform in 2001 re-introduced a constant benefit over the whole period of eligibility for unemployment insurance. Second, the reform also changed the assignment process, as monitoring by ANPE caseworkers has increased after 2001. Lastly, the BMO data we use at the market level (see the next paragraph) have been collected since 2001 only.

We define the treatment variable Z_i and outcome variable Y_i as follows:

- $Z_i = 1$ if i enters a training program within d_Z months as from the beginning of the unemployment spell, 0 otherwise.

- $Y_i = 1$ if i leaves unemployment within d_Y months as from 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 is a key statistic in the public debate on unemployment in France. Our observations consist of unemployment spells starting either in 2002 or in 2004.[10] 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 the 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

---

[9]There can be repetitions across extractions so we have slightly less than 20% of the whole unemployment population between 1990 and 2007.

[10]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.

unemployment spell. For each period we control for the following characteristics: number of unemployment spells, time spent unemployed and time spent in training. We will comment on these variables when discussing the CIA at the individual level in Subsection 4.3.

Market data. We define a market m by three characteristics : a region r, an occupation o and a time indicator t, $m = (r, o, t)$.

The regions we consider are defined by the ASSEDIC. They do not match precisely the administrative regions. We solve this issue by aggregating data to the department level. Departments are administrative sub-divisions of regions and their definition is the same in all data sets. 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. This results in 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 we want these markets to be as isolated as possible from one another, we only keep the broad definition of occupations (seven categories).

Finally, the time indicator is a dummy equal to one for unemployment spells starting in 2002 and to zero for those starting in 2004. Though limited to two periods, this panel dimension of our data potentially allows us to control for unobserved fixed region/occupation effects driving the allocation of P. We return to this in Subsection 4.3 when discussing the CIA at the market level.

The market characteristics W_m include the means of individual characteristics $\bar{X}_m$. As we have seen, the omission of these would be hard to defend. We also include a labor market tightness indicator. This is constructed from the yearly BMO survey conducted by UNEDIC. From 2001 onwards, 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 UNEDIC. For instance, in November 2007, 1 524 557 firms were asked how many vacancies they were planning to post during the year 2008.[11] It is thus possible to define the local labor market tightness θ as the number of predicted vacancies over the number of unemployed workers. This variable is a relevant confounder because the BMO

---

[11]The non-response rate 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 precisely labor demand, the BMO predictions of vacancies are the actual numbers that ANPE observes when deciding on the allocation of training. Hence the non-response rate is not an issue for the assignment to training by ANPE caseworkers.

survey is conducted to help caseworkers assign unemployed workers to the training programs that match the skills needed on local labor markets.

The labor market tightness indicator θ 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 θ_m as the average of all the monthly ratios in 2002 and 2003 (resp. 2004 and 2005).

Descriptive statistics. Table 2 gives the number of individuals in each market in 2002 (left column) as well as the proportion of treated individuals (right column) when we define treatment as entering a training program within $d_Z = 6$ months. There is heterogeneity in market size across regions, from the more rural regions like Auvergne to the Paris region (Ile de France), but also across occupations. Likewise, the proportion of treated varies across these two dimensions. One can also note that we have a large number of observations per market, although perhaps not large enough to run a fully non-parametric analysis with respect to region and occupation indicators. We discuss this issue in Subsection 4.3.

Importantly, we note that the proportions of treated are rather low in our data. In 2002, the maximum is .136 and only 6 markets have a proportion of treated higher than .1 (the numbers are similar in 2004). This will make it very difficult to predict the treatment effect when the proportion of treated is above .1. Our estimation results will thus focus on low values of P.

4.3 Econometric implementation

Let us first sum up the notations. Individuals (more precisely unemployment spells) 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 $m(i) = m$. The individual treatment is Z_i and the outcome is Y_i. The confounders needed to write the independence assumption (6) form the vector X_i. The market treatment variable is P_m, the average of Z_i over $i \in m$. The independence assumption (5) at the market level involves the market confounders W_m which consist of a local labor market tightness θ_m, the market averages of the individual characteristics $\overline{X}_m$ and, depending on the specification, some market indicators (region, occupation or both). The market outcome is the average of Y_i over $i \in m$. Individual outcomes depend on both Z and P: $Y_i = Y_i\left(Z_i, P_{m(i)}\right)$.

Table 2: number of workers (left) and proportion of treated (right) in each market in 2002

	Office worker		Executive		Industrial worker		Construction worker		Social worker		Technical worker		Employee (sales...)	
Alpes	3 086	.039	2 116	.054	3 187	.031	1 736	.033	1 245	.077	1 185	.058	8 254	.023
PACA	2 860	.048	2 366	.069	1 644	.032	1 806	.028	1 419	.070	1 118	.053	7 396	.033
Alsace	1 336	.061	1 066	.059	2 319	.025	953	.035	426	.099	533	.068	3 600	.036
Auvergne	678	.049	579	.100	951	.049	529	.051	396	.129	389	.093	2 817	.040
Basse Normandie	1 091	.051	638	.064	1 807	.043	764	.041	458	.085	608	.058	3 763	.042
Bourgogne	1 281	.046	709	.055	1 903	.043	772	.039	479	.098	680	.046	3 826	.034
Bretagne	2 406	.075	2 053	.084	2 547	.046	1 245	.063	1 459	.123	1 132	.096	7 259	.051
Centre	2 392	.048	1 365	.060	3 065	.032	1 101	.037	764	.092	974	.066	5 301	.041
Champagne	1 073	.075	493	.059	1 728	.036	781	.049	281	.096	455	.086	2 917	.039
Cote	2 217	.051	2 085	.046	980	.040	1 595	.023	899	.090	782	.046	6 882	.034
Franche Comte	744	.060	547	.075	1 567	.043	555	.047	333	.099	569	.088	2 241	.039
Haute Normandie	1 418	.063	871	.067	2 103	.034	1 143	.031	551	.073	722	.060	3 756	.034
Ile de France	13 148	.050	18 679	.048	6 246	.024	4 797	.021	3 107	.074	3 744	.055	22 946	.031
Languedoc	2 669	.058	2 093	.061	1 426	.046	1 808	.040	1 242	.072	951	.068	7 569	.036
Lorraine	1 841	.047	973	.073	2 978	.041	1 545	.046	564	.073	995	.070	5 329	.029
Midi	2 398	.055	2 037	.057	1 748	.039	1 394	.039	1 357	.082	957	.054	6 886	.028
Nord	2 099	.074	1 269	.072	2 640	.050	1 679	.043	761	.101	932	.079	5 395	.044
Pas de Calais	993	.067	450	.093	1 449	.057	995	.065	345	.136	601	.092	3 435	.058
Pays de Loire	2 994	.057	2 372	.075	3 731	.047	1 678	.051	1 629	.097	1 416	.078	8 028	.047
Picardie	1 508	.066	936	.076	2 153	.035	1 017	.038	481	.131	697	.059	3 833	.033
Poitou	1 502	.067	958	.067	1 361	.049	839	.037	636	.079	678	.065	4 505	.041
Vallées	3 157	.049	2 819	.055	3 014	.040	1 581	.044	1 278	.085	1 229	.061	7 050	.029

Note 1: $(d_Z, d_Y) = (6, 12)$.

Note 2: The numbers are computed after the first step of the estimation method. Therefore, the individuals whose predicted propensity score does not satisfy the support condition (see Subsection 4.3) are excluded.

Note 3: Region labels and definitions are those of the UNEDIC and differ slightly from the administrative ones.

21

Assignment to treatment at the individual level. We discuss our CIA at the individual level (6). This is the typical assumption made in evaluation studies using matching estimators on non-experimental data. Two features are crucial in making this assumption credible. First, one should control for local labor market conditions (see Heckman, Ichimura and 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 (see Sianesi, 2004). We address the first issue by controlling for local labor market indicators. Indeed, assumption (6) is made within each market m.

The "employability" issue is tackled using information on individuals' unemployment and training histories. We start our analysis in 2002 but we observe for any worker all the unemployment and training spells he may have experienced between 1990 and 2001. Ideally, one would want to use this information to incorporate an unobserved individual effect in the CIA (6) and run the first step of estimation using fixed- or random-effect techniques. However, we cannot follow this route because of the 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 Subsection 4.2 and thus capture part of the individual heterogeneity driving unemployment duration and assignment to training.

In Section 6 we use the longitudinal dimension of our data to show that our estimation results are robust to the presence of unobserved heterogeneity. To this end, we apply the "Timing-of-Events" approach (see Abbring and Van den Berg, 2003) and control for unobserved heterogeneity terms driving both duration until training and unemployment duration.[12]

Allocation of treatment across markets. We now focus on the CIA at the market level (5). Our identification strategy requires that P be allocated randomly across markets conditionally on a set of market characteristics. We use two very important features of our data to assess the validity of this assumption. First, we control for local labor demand using the information of the BMO survey. Indeed, the key feature one may think of when discussing treatment across labor markets is heterogeneity on the demand side of the market. This dimension is usually not accounted for as most matching studies make use of registered data on unemployed workers. The BMO solves this problem by giving the number of job vacancies in any local labor market for any year. Moreover, as we mentioned in Subsection 4.1, the

---

[12]Since the matching approach suggested in Section 3 is more general than the Timing-of-Events approach (it does not require duration data) and since the point of this paper is first to discuss the SUTVA rather than the CIA, our benchmark estimation results will be based, at the individual level, only on observed confounders.

purpose of the BMO survey is to help ANPE caseworkers allocate training in accordance with firm demand for each type of job.

The second key feature lies in the longitudinal dimension of our market data. Indeed the typical problem in evaluation studies is that one does not know what would have happened had a unit been given another treatment status. Here, a unit is a market i.e. a triplet (region, occupation, year). Two of these characteristics are fixed in time, and we have two observation for each region/occupation pair. Therefore, we observe each region/occupation pair under two treatment regimes: P in 2002 and P in 2004. If there is variation in P with time, we can introduce a market fixed effect among the confounders in (5) and thus control for unobserved market characteristics in the allocation of P.[13]

To assess the relevance of this strategy, we may examine the evolution of the proportion of treated through time. In Figure 1, we plot for each region/occupation pair the proportion of treated in 2004 against that in 2002, together with the 45° line. There is variation in P in most region/occupation cells and this variation can be relatively large (given the low values of P). Therefore, the data seem well-fitted for our approach. Yet, one may question the precision of our estimates as we only have two dates to compute a market fixed effect. We address this concern in Appendix B where we replace our region/occupation fixed effect with region and occupation dummies. The results barely change.
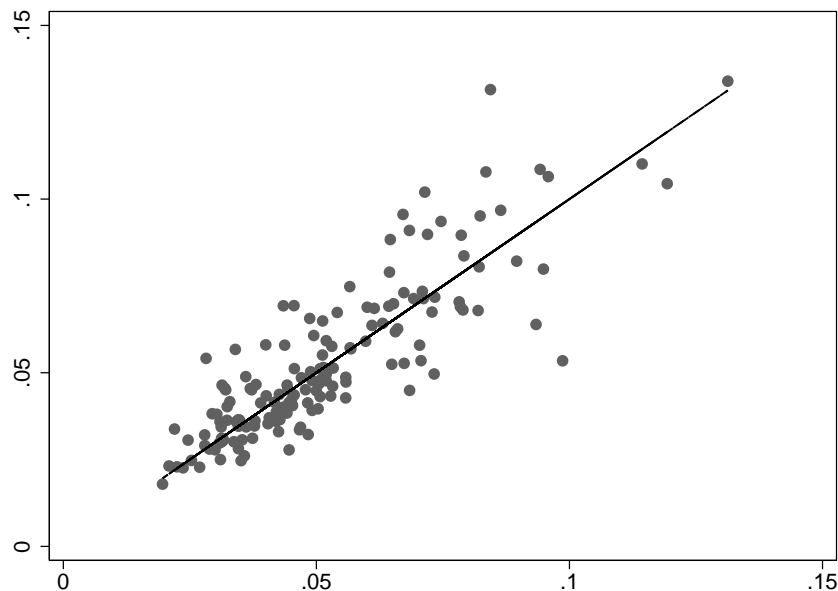
We should make a caveat about the operationalization of the cross-market treatment assignment as the fraction of treated eligible individuals in the market. One may argue that the policy maker assigns a budget B_m to each market, and that the mapping from this budget B_m to the fraction of treated individuals P_m also depends on the decisions of eligible individuals in the market to participate in the program. Such decisions depend on X_i. However, as we have seen, the decision to enroll is critically affected by the case worker. We therefore feel that it is reasonable to capture B_m by P_m. More in general, the above argument suggests to include many statistics of the within-market distribution of X in the vector W_m. For example, one may include a number of quantiles of the marginal distributions of the elements of X, and/or correlations of elements of X.

Specifications. The first step of the estimation method presented in Subsection 3.2 consists in regressing the binary treatment variable Z_i on the covariates X_i for each market m. As Table 2 in Subsection 4.2 shows, we might have too few observations per market to run a fully non-parametric estimation with respect to the market indicators. Therefore, we include two of the three market characteristics among the vector of regressors in the estima-

---

[13]We will make a slight abuse of language and refer to a region/occupation fixed effect as a market fixed effect. For example, executives in Ile-de-France (Paris region) in 2002 and those in 2004 are on different markets. Yet, our market fixed effect will be a dummy Ile-de-France/executives.

tion of the propensity score. More precisely, for each region r, we regress Z_i on X_i, O_i, T_i by logit. We believe that heterogeneity between regions calls for a more flexible specification. In Appendix B, we will show that our results are robust to changes in the specification.

Figure 1: The proportion of treated in 2002 (x-axis) and 2004 (y-axis)

Once we have estimated $\widehat{p}(X_i, m(i))$ for all i, we impose a common support condition. In each market m, we drop all treated (resp. non-treated) individuals whose predicted propensity score is outside the support of the scores of the non treated (resp. of the treated). We then estimate the market outcomes $E_{i \in m}[Y_i(z, P_m)]$ using the weighted regression method presented in Subsection 3.2.

The second step consists first in predicting the GPS. For the benchmark analysis, we regress P_m on $W_m = \left(\overline{X}_m, \theta_m, \theta_m^2\right)$ allowing for a market (actually region/occupation) fixed effect. In Appendix B, we will show that our results are robust when replacing this market fixed effect with region and occupation dummies. We estimate the distribution of the residuals with a kernel and predict the GPS $f(P = p|W)$. We then compute $\widehat{Y}_{z,p}$ as explained in Subsection 3.2. We bootstrap this two-step procedure a hundred times and consider the mean and standard errors of the resulting distributions.[14]

---

[14]We do not use a nearest-neighbor matching estimator at either stage so bootstrap may be used to estimate the variance of our estimates (see Abadie and Imbens, 2008, for a case where bootstrap fails to work for nearest-neighbor matching estimators).

5   Estimation results

5.1   The effect of the proportion of treated on individual outcomes

This section presents our main empirical result: the causal effect of the proportion of treated in a market on the average potential outcomes, for the underlying full population. As mentioned in Section 2, if the SUTVA holds then $Y_{1,p}$ and $Y_{0,p}$ should not vary with p. Figures 2a-2b display the estimates of these as functions of p. The dashed lines delimit the confidence interval. Each dot on the graphs corresponds to a value of p actually observed in the data. As mentioned in Section 4, we can only predict the two average potential outcomes of interest for values of p below 15%. Lastly, note that the scale of the y-axis is different between Figure 2a and Figure 2b ($Y_{0,p}$ is actually more precisely estimated than $Y_{1,p}$).

   The most striking features of these two graphs are that i) neither $\widehat{Y}_{1,p}$ nor $\widehat{Y}_{0,p}$ remains constant when p varies and ii) both decrease when p increases. These are the main empirical findings of this paper. The first point, i), shows that in our data, the SUTVA is violated, since an individual's potential outcome does depend on whether many or few people in his market are treated. The second result, ii), indicates that the proportion of treated in a market has a negative effect on the potential outcomes under treatment as well as under no treatment.

   Taking a closer look at Figures 2a-2b, we note a few differences between the patterns of $\widehat{Y}_{1,p}$ and $\widehat{Y}_{0,p}$. The former shows some concavity while the latter seems to be convex. The outcome of treated individuals is almost constant for the very low values of p and starts to decrease once p is above .03. This does not necessarily mean that the SUTVA is correct for low values of p. Indeed, the slope of the outcome of non treated is the steepest when p is small and becomes less negative as p increases. While $\widehat{Y}_{0,p}$ seems almost constant when p is around .1, we cannot assess whether it will then become increasing since we observe very few markets with values of P above .1 (and none above .15).

   The magnitude of the effect of p on the average potential outcomes in Figures 2a-2b may not seem impressive, but note that the range of values for p is small: from .021 to .136. Our results show that even small changes in p affect the potential outcomes.

   In Appendix B we show that our empirical results are robust to a series of tests. These concern changes in the specification of the econometric model at each step of the estimation (see also Section 6, where we address the crucial issue of unobserved individual heterogeneity and show that we get similar results when using the "Timing-of-Events" approach). We also show that choosing other dates for the observation of the treatment status and the outcome does not affect the results.
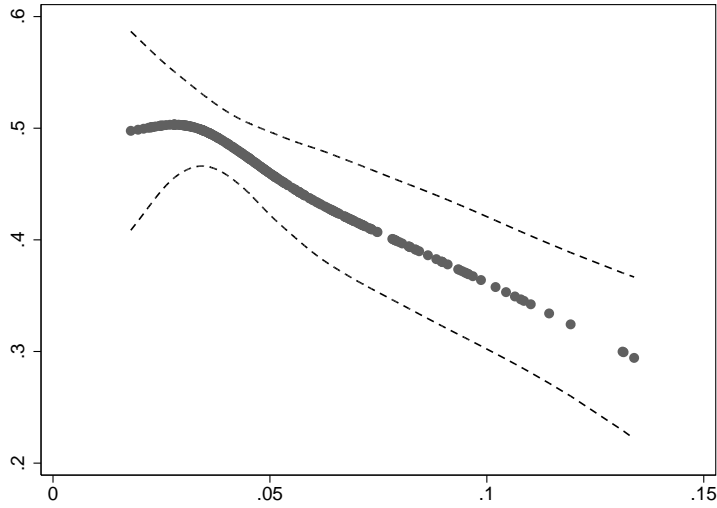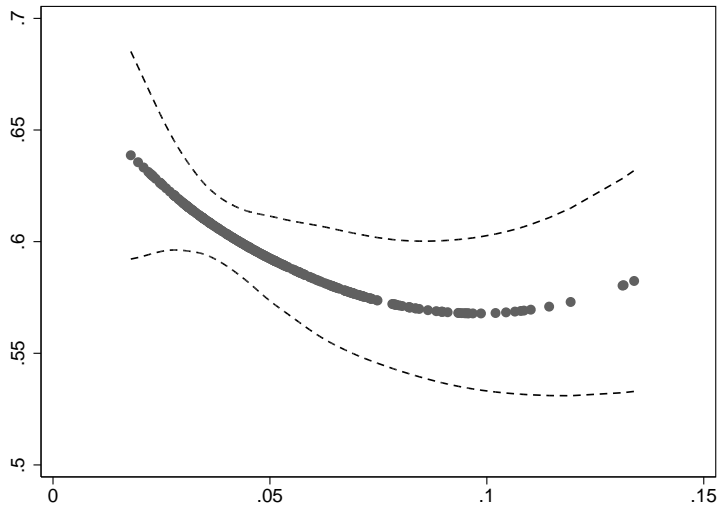
Figure 2a: $\widehat{Y}_{1,p}$ $vs.$ p

Figure 2b: $\widehat{Y}_{0,p}$ $vs.$ p

In sum, the potential outcomes for treatment and non-treatment decrease with p. For $\widehat{Y}_{0,p}$, the slope seems to vanish when p increases beyond .10, but there are very few markets with $P > .10$. For the range of common values (between .02 and .10), both average potential outcomes are significantly decreasing.

Our approach does not give an explanation for these results. In the remainder of this subsection we discuss some possible explanations based on the structural models mentioned in Section 1, but we leave a thorough investigation for further research. First, if training increases individual productivity, it could be that firms would prefer a worker who just exited a training program over a non-treated worker. This would make $Y_{0,p}$ decrease as p increases.

26

At the same time, if more people are treated, then a treated worker will be competing with more treated workers who, contrary to non treated, are as productive as him. Hence, $Y_{1,p}$ can be non-increasing with respect to p because of competition for jobs within the group of treated. This is the crowding-out argument mentioned in Section 1.

If the proportion of treated increases, and if training improves productivity, then firms may have an incentive to open more job vacancies, in particular more vacancies for skilled (trained) individuals. If search is undirected, i.e. if firms cannot target specifically treated or non-treated workers before a hiring interview, then an increase in labor demand may also benefit non-treated workers.
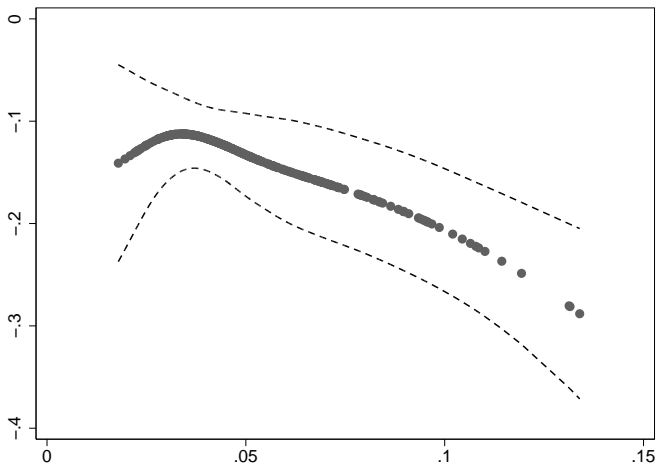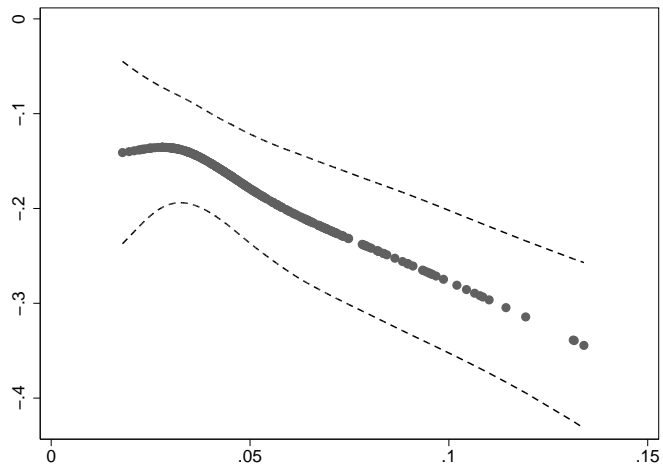
The size of a negative (resp. positive) effect of crowding-out (resp. a labor demand response) on outcomes depends on wage determination. If workers have some market power (through e.g. bargaining over wages or on-the-job search), the increase in productivity due to training might not convert fully into an increase in firms' profit, as workers can get a share of the additional surplus. This will affect both firms' response to an increase in the treatment probability p (potentially leading to less job creation), and their relative preferences for treated and non-treated workers (firms might not prefer to hire treated workers if wages are too high). In any case, our results seem to be consistent with an explanation that is primarily driven by crowding-out effects in a labor market with imperfect information and search frictions.

All these issues motivate further research. This is beyond the scope of the present paper. Recall that the main points of the paper are to show that such interactions can take place, to show that they affect the evaluation of the treatment, and to suggest an evaluation method when the SUTVA no longer holds.

5.2   Average treatment effects

Figures 2a-2b point at a significant effect of the market dimension of the treatment on the average individual potential outcomes. This feature implies that p is as important as z when assessing the effect of the treatment. To illustrate this, let us compute the average treatment effects. We show $\widehat{\delta}_{p,p}^{1,0} = \widehat{Y}_{1,p} - \widehat{Y}_{0,p}$ as a function of p in Figure 3a and $\widehat{\delta}_{p,\underline{p}}^{1,0} = \widehat{Y}_{1,p} - \widehat{Y}_{0,\underline{p}}$ in Figure 3b, where $\underline{p}$ if the lowest value of P observed in our data.[15] In order to draw comparisons between these two functions, we use the same scale for the y-axis in both graphs.

---

[15]Obviously, $\widehat{\delta}_{p,\underline{p}}^{1,0}$ is just $\widehat{Y}_{1,p}$ shifted downwards.

Figure 3a: $\widehat{\delta}_{p,p}^{1,0} = \widehat{Y}_{1,p} - \widehat{Y}_{0,p}$ $vs.$ p 　　　　Figure 3b: $\widehat{\delta}_{p,\underline{p}}^{1,0} = \widehat{Y}_{1,p} - \widehat{Y}_{0,\underline{p}}$ $vs.$ p

Figures 3a-3b illustrate the discussion on average treatment effects in Subsection 2.2. Consider an econometrician who, in order to estimate the effect of the treatment, proceeds to 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 $p = E(Z)$ but, since the SUTVA is violated, he cannot ensure whether this effect remains the same should the proportion of treated increase or decrease. Looking at Figure 3a, we see that, for values of p above .04, an increase in the proportion of treated leads to a more negative treatment effect. One can argue that the decrease in $\delta_{p,p}^{1,0}$ is small. Still, the violation of the SUTVA may be important for the evaluation of the treatment for two reasons. First, as this was the case for the outcomes in Figures 2a-2b, we observe changes in the treatment effect even though p varies within a narrow range. Secondly, even if the treatment effect was constant with p, the fact that the outcomes of treated and non-treated workers decrease with the proportion of treated indicates that the treatment is far from having no effect.

Figure 3b allows us to complete the evaluation of the treatment by comparing the outcomes of treated individuals in a market with $P = p$ with those of non-treated individuals in a market with the lowest value of P (ideally, we would like to have $\underline{p} = 0$). We see that treated individuals have a smaller probability of leaving unemployment within the year than workers who are not treated in a market where very few people are treated. This is as close as we can get to a "pure" effect of the treatment i.e. an effect where the counterfactual is the situation with no treatment at all. While the decrease in $\delta_{p,p}^{1,0}$ in Figure 3a was of a small magnitude, the fall in $\delta_{p,\underline{p}}^{1,0}$ shown in Figure 3b is substantial.

As a sensitivity analysis, we replicate the above estimations separately for small and large

markets as defined by the number of unemployed individuals. The results (not shown here) are very similar to those presented.

Overall, our empirical analysis paints a negative picture of training programs. Figure 3a shows that treated workers are always more likely than non-treated workers to experience long-term unemployment and Figures 2a-2b and 3b show that an individual's probability to leave unemployment within 12 months is negatively affected by others' treatment status. We shall now give a few precisions in defense of training programs. First, these programs can take time and while participating, an individual can devote less time and effort to job search. This is the well-known "locking-in" effect. Also, the purpose of training programs is not only to help workers re-access employment earlier but also to allow them to acquire some skills that will facilitate their access to more stable jobs. A recent study by Crépon, Ferracci and Fougère (2007) on French data showed, using a "Timing-of-Events" model, that training programs have almost no effect on unemployment duration but significantly increase the duration of the subsequent job. What had not been shown so far by the empirical literature on training programs is the global effects of these treatments on local labor markets i.e. the evolution of outcomes with p as shown in Figures 2a-2b and 3a-3b.

6 Allowing for unobserved individual heterogeneity

Obviously, testing the SUTVA requires validity of the CIA. However, without experimental data where assignment to treatment is random, we cannot ensure that no unobserved confounder drives both the outcomes and the treatment status. At the market level, since we observe markets (region/occupation) under different treatment regimes, we can control for an unobserved fixed effect in the allocation of P. In this section, we suggest a method to control also for unobserved heterogeneity at the individual level. To this end, we use the fact that both our outcome and treatment variables result from duration processes. Using the "Timing-of-Events" (ToE thereafter) approach, we can impose parametric assumptions on the distributions of these durations and control for fixed unobserved individual effects that drive them both (see Abbring and Van den Berg, 2003). Then, we will be able to compute our average outcomes at the market level, accounting for unobserved heterogeneity, and go to the second step of the estimation method we presented in Section 3. We do not use this method for our benchmark results because the matching approach we have presented in Section 3 is more general as it does not require duration data.

6.1 The "Timing-of-Events" approach

We consider two duration processes: T_Z and $T_Y(t_Z)$, both starting at date $t = 0$. The former is the duration until treatment while the latter is the potential unemployment duration if treated at date t_Z. The hazard rates of these two processes depend on individual observed heterogeneity, X, and on an unobserved heterogeneity term, denoted as v_Z for duration T_Z and as v_Y for T_Y. Importantly, v_Z and v_Y can be correlated and if this is the case, a CIA based solely on observed confounders will not permit identification of the treatment effect. The ToE literature, initiated by Abbring and Van den Berg (2003), suggests a way out of this issue, which consists in making the three following assumptions.

First, one has to make a "no-anticipation" assumption. This assumption is crucial for the interpretation of unemployment spells that end before treatment. Crépon, Ferracci, Jolivet and Van den Berg (2009) discuss this in detail for the "matching" approach. Denoting by $f_{A|B}$ the conditional density of A given B we assume:

$$f_{T_Y(t')|X,v_Y}(t) = f_{T_Y(t'')|X,v_Y}(t), \quad \forall t \leq \min(t', t''). \tag{12}$$

Assumption (12) means that there is a unique process ruling unemployment duration prior to treatment. It implies that an individual has the same probability of leaving unemployment today whether he is going to be treated tomorrow or a year from now.

We also need a CIA, but now, the outcomes and treatment of interest are potential durations and the conditioning involves unobserved confounders $v = (v_Z, v_Y)$:

$$T_Y(t_Z) \quad \perp \quad T_Z \quad | \quad X, v_Z, v_Y, \quad \forall t_Z. \tag{13}$$

Note that (13) implies (6) where the conditioning is on both X and (v_Z, v_Y). Lastly, if the CIA (13) involves unobserved individual characteristics, we need to put some structure on the hazard rates of T_Z and $T_Y(t_Z)$, denoted, respectively, by $h_{T_Z|X,v_Z}$ and $h_{T_Y(t_Z)|X,v_Y}$. The ToE approach requires mixed proportional hazard rates,

$$
\begin{aligned}
h_{T_Z|X,v_Z}(t) &= \lambda_Z(t) \cdot \phi_Z(X) \cdot exp(v_Z), \\
h_{T_Y(t_Z)|X,v_Y}(t) &= \lambda_Y(t) \cdot \phi_Y(X) \, exp(\delta_Z \mathbf{1}\{t > t_Z\} + v_Y),
\end{aligned}
\tag{14}
$$

where λ_Z and λ_Y are piecewise constant functions.[16] Abbring and Van den Berg (2003) show

---

[16] This means that we have an integer K_Z, intervals $(I_{Zk})_{0<k\leq K_Z} \subset \mathbb{R}^+$, $I_{Zk} \cap I_{Zk'} = \emptyset$ if $k \neq k'$, and positive scalars $(\lambda_{Zk})_{0<k\leq K_Z}$ such that: $\lambda_Z(t) = \sum_{k=1}^{K_Z} \lambda_{Zk} \cdot \mathbf{1}\{t \in I_{Zk}\}$. The same goes for λ_Y with different intervals and scalars.

that all parameters of this model are identified (up to a multiplicative constant in the λ's) and also that one can recover the distribution of (v_Z, v_Y).

So far in this section, we have not used market indicators. In theory, one could allow for all the parameters and functions of the above model to be market specific. However in practice the ToE approach demands a large number of observations. We thus consider one model for the whole economy and introduce region, occupation and year dummies in the regressors X and, importantly, in the treatment effect. Therefore, in equations (14), ϕ_Z and ϕ_Y are log-linear functions of (X, R, O, T) and the treatment parameter δ_Z is a linear function of (R, O, T). The treatment effect thus varies across markets. Note that this specification leads us to assume only one distribution of (v_Z, v_Y) in the economy (i.e. unobserved heterogeneity is not market-specific).

We assume that the pair (v_Z, v_Y) can take R values in $(\mathbb{R}^{*+})^2$. We denote by $q(v)$ the probability that (v_Z, v_Y) takes a given value v. The correlation between v_Z and v_Y will indicate whether conditional independence between treatment and outcomes requires unobserved confounders.

Once we know the parameters of the model and the distribution of heterogeneity, we can compute estimates of the two outcomes of interest: $\widehat{E}_{i \in m}[Y(1, P_m)]$ and $\widehat{E}_{i \in m}[Y(0, P_m)]$, using the estimated distributions of T_Z and T_Y (see the details in Appendix C).

6.2   Results

The first step now consists in maximizing the likelihood of observed durations (T_Y, T_Z) (possibly censored) over all spells in the economy. We first present the distribution of individual unobserved heterogeneity, allowing for $R = 4$ groups.
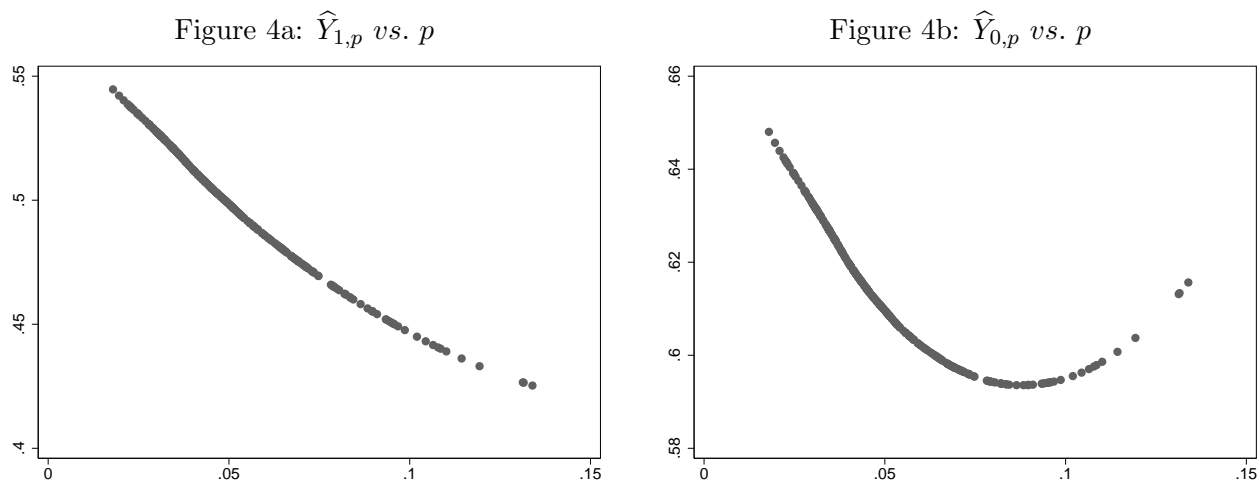
Table 3: Distribution of unobserved heterogeneity

group 1:	$exp(\widehat{v}_Z) = .000076$	$exp(\widehat{v}_Y) = .34$	$q(v) = .019$
group 2:	$exp(\widehat{v}_Z) = .086$	$exp(\widehat{v}_Y) = 15.1$	$q(v) = .413$
group 3:	$exp(\widehat{v}_Z) = .000039$	$exp(\widehat{v}_Y) = 11.9$	$q(v) = .357$
group 4:	$exp(\widehat{v}_Z) = .0080$	$exp(\widehat{v}_Y) = 3.7$	$q(v) = .211$

Table 3 indicates that unobserved heterogeneity plays a role in both hazard rates. Importantly, the two unobserved factors driving T_Z and T_Y are not independent as the covariance between $exp(\widehat{v}_Z)$ and $exp(\widehat{v}_Y)$ equals .122. We thus need to control for unobserved heterogeneity in the allocation of treatment across individuals. Note that v_Z and v_Y are positively

correlated so individuals with the highest propensity to leave unemployment tend to be assigned to treatment earlier.

The evaluation of the treatment shown in Section 5 might thus be biased because of heterogeneity. Does this affect our main results (violation of the SUTVA and negative effect of p on average outcomes)? We can answer this question by using the distributions of T_Z and $T_Y(t_Z)$ estimated in the first step to predict market average outcomes $\widehat{E}_{i \in m}[Y(1, P_m)]$ and $\widehat{E}_{i \in m}[Y(0, P_m)]$ and then running the second step of our estimation method (see Subsection 3.2). This two-step method where, in the first step, standard matching has been replaced by ToE, controls for unobserved fixed heterogeneity at both the individual (ToE) and market (fixed effects in the estimation of the GPS) levels. We show in Figures 4a-4b the estimation results on both outcomes.

Figure 4a: $\widehat{Y}_{1,p}$ vs. p Figure 4b: $\widehat{Y}_{0,p}$ vs. p

Results controlling for individual unobserved heterogeneity
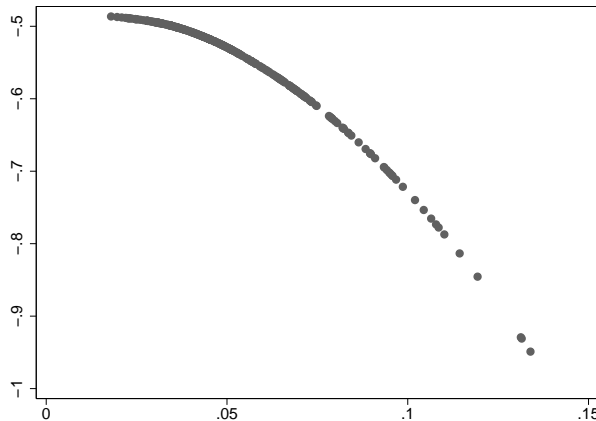
It is clear from Figures 4a-4b that controlling for unobserved individual heterogeneity does not affect our two main results: both $Y_{1,p}$ and $Y_{0,p}$ vary with p and these two functions show a decreasing pattern.[17] As it was the case in Figure 2b, $Y_{0,p}$ seems to be convex. However $Y_{1,p}$ no longer shows any concavity (see Figure 2a) and substantially decreases with p even for small values of p. These results are crucial for our analysis as they show that the SUTVA is violated even when one controls for unobserved heterogeneity at both levels of the analysis.

Lastly, we look at the effect of the treatment on the hazard rate of T_Y, δ_Z (see equation 14). To this end, we consider the δ_Z parameters we have estimated for each market in the

---

[17]We cannot directly compare the levels of $Y_{z,p}$ in Figures 2a-2b and 4a-4b because the ToE approach identifies hazard rates only up to a multiplicative constant.

first step using the ToE approach and use them as the market outcomes of interest in our second step. We thus have to assume that the CIA (5) holds when replacing $E_{i \in m}[Y(z, P_m)]$, $z = 0, 1$ with δ_{Zm}, where δ_{Zm} is the effect of treatment on the hazard rate of T_Y in market m.[18] We plot the estimated function $p \to \delta_Z(p)$ in Figure 5. Consistently with what we found using the matching approach (see Figures 3a-3b), $\delta_Z(p)$ shows a decreasing pattern, indicating that workers participating in training programs stay longer in unemployment as the proportion of treated in their local labor market increases.

Figure 5: $\widehat{\delta}_Z(p)$ vs. p

7 Conclusion

Evaluation with non-experimental 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, in order to mimic randomization. Secondly, an individual's outcome should not depend on other individuals' treatment statuses. With matching methods, these are the CIA and the SUTVA. There is a tension between these assumptions. The more similar two individuals are, the more likely it is that they act in each others' proximity, which makes it more likely that their outcomes interfere. In this paper we design and apply a method to overcome this and, indeed, to estimate the degree of interference. Our extension of the Rubin model (1974) does not impose the SUTVA at the individual level, thus allowing for interactions between individuals. In our empirical analysis, we focus on training program participation by unemployed individuals. The results show that the estimated mean potential outcomes of both treated and non-treated job seekers

---

[18]Note that the CIA at the market level would then involve the marginal effect of training on hazard rates. This seems less realistic than assuming that the outcome of interest for the authorities is the proportion of people still unemployed after a year.

depend significantly, and negatively, on the proportion of individuals treated in the relevant local labor market. Since the application builds on a general statistical model and a large-scale data set, we view these results as strong empirical support of the increasing focus on interactions, in economic research in general, and research on treatment evaluations in particular.

One of the main features of our approach is its flexibility with respect to the estimation techniques used at each stage. We exploit this feature in a sensitivity analysis, by using the Timing-of-Events method rather than propensity score matching, and we show that individual unobserved heterogeneity does not qualitatively affect our results. It would be interesting to take our two-step approach to other data sets, as the issue of interactions between units applies to many fields in social sciences and our model is general enough to be used in a large number of empirical applications. The first obvious application would be to use our approach with data from a natural or controlled experiment. Another application would consist in using a structural model to describe interactions within markets. For instance, we could build on the recent developments in the labor economic literature (see Lise, Seitz and Smith, 2005) to run a structural estimation of the treatment effect in the first stage and then resort to more reduced-form models at the second stage. One can also tackle the second stage with a theoretical model in order to explain the allocation of treatment across markets. For instance, it could be that the authorities decide to allocate treatment resources across regions in order to reach a social optimum. This approach would then merge the structural treatment literature with the recent contributions of Graham, Imbens and Ridder (2006) and Bhattacharya and Dupas (2009) which explicitly study the optimal allocation of types or treatments across groups in the presence of social interactions.

References

Abadie, A., and G. Imbens (2008): "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 76(6), 1537-57.

Abbring, J., and G.J. van den Berg (2003): "The non-parametric identification of treatment effects in duration models," *Econometrica*, 71, 1491-517.

Albrecht, J., Van den Berg, G.J., and S. Vroman (2009): "The Aggregate Labor Market Effects of the Swedish Knowledge Lift Program," *Review of Economic Dynamics*, 12(1), 129-146.

Angelucci, M., and G. De Giorgi (2009): "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?" *American Economic Review*, 99(1), 486-508.

Bhattacharya, D., and P. Dupas (2009): "Inferring Welfare Maximizing Treatment Assignment under Budget Constraints," NBER Working paper, 14447.

Blundell, R., Costa Dias, M., Meghir, C., and J. Van Reenen (2004): "Evaluating the Employment Impact of a Mandatory Job Search Program," *Journal of the European Economic Association*, 2(4), 569-606.

Cahuc, P., Crépon, B., Guitard, J., and M. Gurgand (2008): "Policy Evaluation in Equilibrium. The Case of Job Search Counseling in France," Crest Working Paper, 2008-13.

Crépon, B., Ferracci, M., and D. Fougère (2007): "Training the Unemployed in France: How Does It Affect Unemployment Duration and Recurrence?" IZA Working Paper, 3215.

Crépon, B., Ferracci, M., Jolivet, G., and G.J. van den Berg (2009): "Active Labor Market Policy Effects in A Dynamic Setting", *Journal of the European Economic Association*, 7, 595-605.

Duflo, E., Dupas, P. and M. Kremer (2008): "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," CEPR Discussion Paper, DP7043.

Fleuret, A. (2006): "L'Orientation des Chômeurs Vers la Formation de 2002 à 2004," *DARES Premières synthèses*, 29(2).

Frölich, M., and K. Michaelowa (2005): "Peer Effects and Textbooks in Primary Education: Evidence from Francophone Sub-Saharan Africa," IZA Working Paper, 1519.

Graham, B., Imbens, G., and G. Ridder (2006): "Complementarity and Aggregate Implications of Assortative Matching," NBER Working paper, 14860.

Graham, B., Imbens, G., and G. Ridder (2009): "Measuring the Average Outcome and Inequality Effects of Segregation in the Presence of Social Spillovers," mimeo.

Heckman, J., H. Ichimura, and P. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64(4), 605-54.

Heckman, J., LaLonde, R., and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in: O. Ashenfelter, and D. Card, eds., *Handbook of Labor Economics*, vol. 3A, ch. 31. North-Holland, New York, 1865-2097.

Heckman, J., L. Lochner, and C. Taber (1998): "General-Equilibrium Treatment Effects: A Study of Tuition Policy," *American Economic Review*, 88(2), 381-6.

Hirano, K., and G. Imbens (2001): "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology*, 2, 259-278.

Hirano, K., and G. Imbens (2004): "The Propensity Score with Continuous Treatments," *in* A. Gelman , and X.-L. Meng, eds., *Applied Bayesian Modeling and Causal Inference from Incomplete Data Perpectives*. New York: Wiley.

Hirano, K., Imbens, G., and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161-1189.

Hudgens, M.G., and M.E. Halloran (2008): "Toward Causal Inference With Interference," *Journal of the American Statistical Association*, 103, 832-842.

Imbens, G. (2001): "The Role of the Propensity Score in Estimating Dose-Response Functions," *Biometrika*, 87(3), 706-10.

Lechner, M. (2002): "Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies," *Review of Economics and Statistics*, 84(2), 205-220.

Lise, J., Seitz, S., and J. Smith (2005): "Equilibrium Policy Experiments and the Evaluation of Social Programs," NBER Working paper, 10283.

Manski, C.F. (2009): "Identification Of Treatment Response With Social Interactions," Draft, Northwestern University.

Miguel, E., and M. Kremer (2004): "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72(1), 159-217.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles," translated (with discussion) in *Statistical Science*, 5(4), 465-80.

Oakes, M. (2008): "Methodological Issues in Neighborhood Effects Research," Mimeo, University of Minnesota

Philipson, T. (2000): "External Treatment Effects and Program Implementation Bias," NBER Working Paper, 250.

Robins, J., and Y. Ritov (1997): "Towards a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-parametric Models," *Statistics in Medicine*, 16, 285-319.

Rosenbaum, P., and D. Rubin (1983): "The Central Role of the Propensity Score Matching in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

Rubin, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

Rubin, D. (1977): "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 62, 1-26.

Rubin, D. (1986): "Statistics and Causal Inference: Which Ifs Have Causal Answers," *Journal of the American Statistical Association*, 81, 961-2.

Sianesi, B. (2004): "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," *Review of Economics and Statistics*, 86, 133-55.

Topa, G. (2001): "Social Interactions, Local Spillovers and Unemployment," *Review of Economic Studies*, 68(2), 261-95.

APPENDIX

A A alternative single-step identification and estimation approach

The alternative approach consists in considering the pair $(Z_i, P_{m(i)})$ as an individual treatment. Then we could use $(X_i, W_{m(i)})$ to write an unconfoundedness assumption at the individual level:

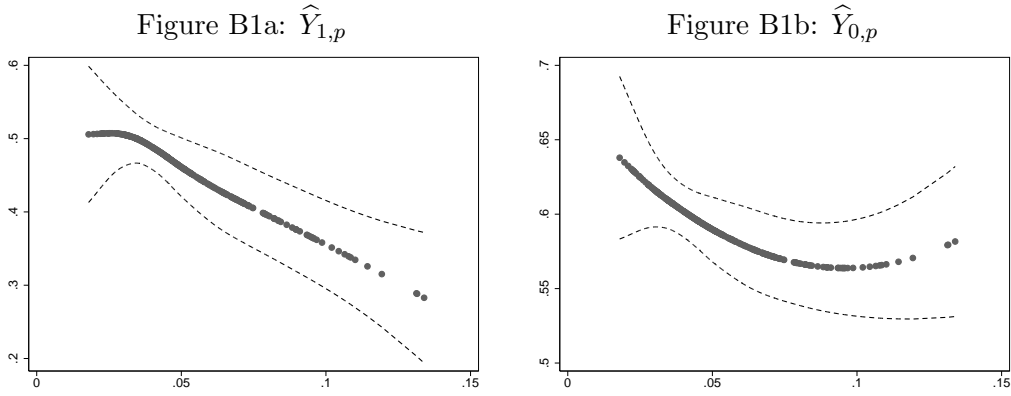$$Y_i(z,p) \quad \perp \quad (Z_i, P_{m(i)}) \quad | \quad X_i, W_{m(i)}, \quad \forall i, z, p, \tag{A1}$$

and thus identify $E[Y(z,p)]$ over the population of individuals.

Estimation is then straightforward as one can directly apply the method of Hirano and Imbens (2004) for a continuous treatment. Note that the estimation of the propensity score can be done in two steps by estimating first the density of P conditional on (X, W) and then the density of Z given $(X, W, P = p)$.

This approach differs from our two-step method not only because of the estimation procedure but also, and more crucially, because it is based on different independence assumptions. The main difference is that the allocation of P operates at the market level in our two-step approach and at the individual level in the alternative approach. We think of (5) and (6) as more intuitive assumptions than (A1) since they do not lead one to overlook the two dimensions of the treatment.
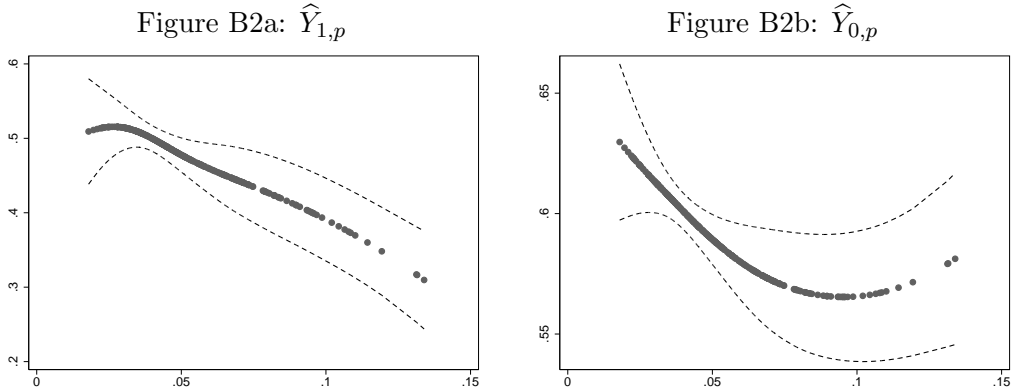
B Robustness checks: Econometric specification

Market specification in the first step. Our estimation method requires to estimate average outcomes $E_{i \in m}[Y_i(1, P_m)]$ and $E_{i \in m}[Y_i(0, P_m)]$ for each market m. This means that we should run the first step presented in Subsection 4.2 within any market. This demands far more observations per market than we have. Therefore, as mentioned in Subsection 4.3, rather than estimating one propensity score for each region/occupation/year cell, we include occupation and year dummies in the set of regressors (together with the confounders X) and run one estimation per region. This leads to the benchmark results presented in Section 5. We chose to be non parametric with respect to the region variables since administrative regions play an important role in the French training system (see Subsection 4.1). Yet, one could ask whether this choice has an influence on the results. Therefore, we proceed to an estimation in which, in the first step, propensity scores are now estimated non parametrically with respect to the occupation variable while region and year dummies are included as regressors. Results are shown in Figures B1a-B1b. They are almost similar to those of the benchmark estimation (shown in Figures 2a-2b).
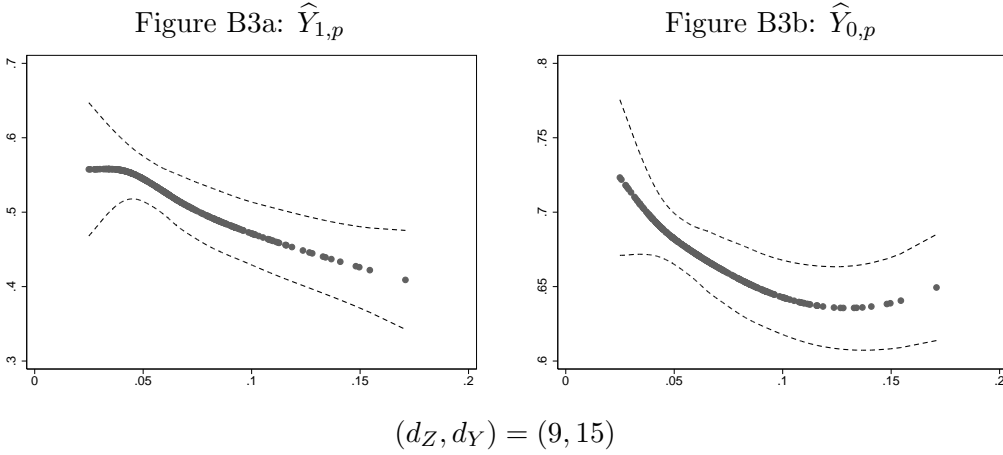
Figure B1a: $\widehat{Y}_{1,p}$ Figure B1b: $\widehat{Y}_{0,p}$

One estimation of the propensity score per occupation

Market specification in the second step. The CIA at the market level requires that P be allocated randomly across markets conditionally on characteristics W. An important feature of our estimation method is that we include in W a region/occupation fixed effect, using the fact that we observe each region/occupation twice (in 2002 and 2004). Since we only have two dates to estimate the market fixed effect, we could doubt the precision of our predictions. To address this issue, we re-run our method using this time both a region and an occupation dummy in the second step (in the estimation of the GPS). This specification controls for a region and an occupation fixed effect but does not allow for an interaction between these two effects. Results are shown in Figures B2a-B2b. The two main conclusions (the SUTVA is rejected and outcomes significantly decrease with p) are reinforced.

Figure B2a: $\widehat{Y}_{1,p}$ Figure B2b: $\widehat{Y}_{0,p}$

Region and occupation dummies in the estimation of the GPS

Changing the dates when treatment and outcome are measured. In our benchmark results, treatment status is measured $d_Z = 6$ months after unemployment starts whereas the outcome Y equals one if the individual has left unemployment within $d_Y = 12$ months. In Figure B3a-B3b, we show estimation results when setting $(d_Z, d_Y) = (9, 15)$ months. The two outcomes are still decreasing with p although $\widehat{Y}_{0,p}$ seems to be more convex than in Figure 2b.

Figure B3a: $\widehat{Y}_{1,p}$ Figure B3b: $\widehat{Y}_{0,p}$

$$(d_Z, d_Y) = (9, 15)$$

C The Timing-of-Events approach: derivation of the predicted outcomes

The two market outcomes of interest $E_{i\in m}\left[Y_i(1, P_m)\right]$ and $E_{i\in m}\left[Y_i(0, P_m)\right]$ can be derived from the distributions of T_Z and $T_Y(t_Z)$. Consider the probabilities of having left unemployment at d_Y when treated at t_Z, conditionally on $t_Z < d_Z$ and $t_Z \geq d_Z$ respectively. Using the CIA (13) we have:

$$Y_i(1, P_m) = \sum_{t_Z=1}^{d_Z-1}\sum_v q(v) \cdot \mathrm{Pr}_{im}\left(T_Y(t_Z) \leq d_Y\right) \cdot \frac{\mathrm{Pr}_{im}\left(T_Z = t_Z\right)}{\mathrm{Pr}_{im}\left(T_Z < d_Z\right)},$$

$$Y_i(0, P_m) = \sum_{t_Z=d_Z}^{+\infty}\sum_v q(v) \cdot \mathrm{Pr}_{im}\left(T_Y(t_Z) \leq d_Y\right) \cdot \frac{\mathrm{Pr}_{im}\left(T_Z = t_Z\right)}{\mathrm{Pr}_{im}\left(T_Z \geq d_Z\right)},$$

The probabilities Pr_{im} are market specific and conditional on X_i, v. Remember that d_Z and d_Y are the dates when the treatment and unemployment status are checked (respectively 6 and 12 months for our benchmark estimation results). We can rewrite these two equalities as follows:

$$Y_i(1, P_m) = \sum_{t_Z=1}^{d_Z-1}\sum_v q(v) \cdot \left[1 - S_{T_Y(t_Z)|X_i,v_Y}(d_Y+1)\right] \cdot \frac{f_{T_Z|X_i,v_Z}(t_Z)}{1 - S_{T_Z|X_i,v_Z}(d_Z)},$$

$$Y_i(0, P_m) = \sum_{t_Z=d_Z}^{+\infty}\sum_v q(v) \cdot \left[1 - S_{T_Y(t_Z)|X_i}(d_Y+1)\right] \cdot \frac{f_{T_Z|X_i}(t_Z)}{S_{T_Z|X_i}(d_Z)}.$$

The second equality involves a sum from d_Z to infinity. To simplify this expression, note that the no-anticipation assumption (12) and the MPH specification (14) allow us to write:

$$Y_i(0, P_m) = \sum_v q(v) \cdot \sum_{t_Z=d_Z}^{d_Y}\left[1 - S_{T_Y(t_Z)|X_i,v_Y}(d_Y+1)\right] \cdot \frac{f_{T_Z|X_i,v_Z}(t_Z)}{S_{T_Z|X_i,v_Z}(d_Z)}$$

$$+ \left[1 - S_{T_Y(d_Y+1)|X_i,v_Y}(d_Y+1)\right] \cdot \frac{S_{T_Z|X_i,v_Z}(d_Y+1)}{S_{T_Z|X_i,v_Z}(d_Z)}.$$

Lastly, we take the average within each market m to get $E_{i\in m}\left[Y_i(1, P_m)\right]$ and $E_{i\in m}\left[Y_i(0, P_m)\right]$.