Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://amstat.tandfonline.com/loi/ubes20

A Framework for Separating Individual-Level Treatment Effects From Spillover Effects

Martin Huber & Andreas Steinmayr

To cite this article: Martin Huber & Andreas Steinmayr (2019): A Framework for Separating Individual-Level Treatment Effects From Spillover Effects, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2019.1668795

To link to this article: https://doi.org/10.1080/07350015.2019.1668795

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2019, VOL. 00, NO. 0, 1–15
https://doi.org/10.1080/07350015.2019.1668795

Taylor & Francis
Taylor & Francis Group

A Framework for Separating Individual-Level Treatment Effects From Spillover Effects

Martin Huber[a] and Andreas Steinmayr[b,c,d,e]

[a]Department of Economics, University of Fribourg, Fribourg, Switzerland; [b]Department of Economics, University of Munich, Munich, Germany; [c]IfW Kiel, Kiel, Germany; [d]IZA, Bonn, Germany; [e]CESifo, Munich, Germany

ABSTRACT

This article suggests a causal framework for separating individual-level treatment effects and spillover effects such as general equilibrium, interference, or interaction effects related to treatment distribution. We relax the stable unit treatment value assumption assuming away treatment-dependent interaction between study participants and permit spillover effects within aggregates, for example, regions. Based on our framework, we systematically categorize the individual-level and spillover effects considered in the previous literature and clarify the assumptions required for identification under different designs, for instance, based on randomization or selection on observables. Furthermore, we propose a novel difference-in-differences approach and apply it to a policy intervention extending unemployment benefit durations in selected regions of Austria that arguably affected ineligibles in treated regions through general equilibrium effects in local labor markets.

ARTICLE HISTORY
Received November 2017
Accepted August 2019

KEYWORDS
Difference-in-differences; General equilibrium effects; Interaction effects; Interference effects; Spillover effects; Treatment effects

1. Introduction

Most studies on treatment evaluation implicitly or explicitly rule out general equilibrium, interference, or interaction effects related to individual treatment assignment. The stable unit treatment value assumption (SUTVA) formalizes the absence of any such spillovers between study participants (see, e.g., Rubin 1990). However, the satisfaction of SUTVA appears unrealistic in many scenarios including labor market, development, health, and educational interventions, see Heckman, Lochner, and Taber (1998) for a critical discussion. Considering for instance a training program, the share of individuals who receive some training in a region may have an impact on someone's employment probability even net of the individual training status, due to an increase in the regional supply of a particular skill. When assessing the effects of book provision to high school students, spillover effects may occur through sharing the books with peers in class who did not receive the books. The share of vaccinated individuals in a country might affect the health status of non-vaccinated subjects through the likelihood of transmitting some disease. In such cases, the overall treatment effect would be different from the (average) individual one, see, for instance, Sobel (2006) for a framework to characterize the bias in the presence of interactions. Spillover effects are also a likely reason why many interventions deemed successful in a small-scale randomized experiments, where the treatment group is small compared to the population, fail to produce similar effects when scaled-up to a larger group (Deaton and Cartwright 2016).

As first contribution, this article proposes a general non-parametric framework for separating individual-level treatment effects from spillover effects and systematically categorizes the effects considered in the previous literature on spillovers. Without imposing parametric restrictions, we discuss the identification of various effects under alternative assumptions like random assignment and selection on observables. As second contribution, we propose novel common trend assumptions that permit difference-in-differences(DiD)-based identification and illustrate the method reconsidering data from Lalive, Landais, and Zweimüller (2015). The latter study the spillover effects of an extension of unemployment benefits in selected regions of Austria and find that this policy decreased the job-search duration of ineligible individuals in treated regions. We use our framework to provide a sharper definition of the identified effects and apply our DiD methodology to assess the total effect, the total effects on eligibles, and the spillover effects on ineligibles in treated regions. Furthermore, we compare estimation based on (i) a common trend assumption within groups having the same eligibility status and (ii) a stronger common trend assumption across groups and discuss testable implications of the latter. The results suggest that the stronger assumption, as considered in some estimations of Lalive, Landais, and Zweimüller (2015), is most likely violated. Even though the estimates under common trends within groups are smaller in absolute terms, they qualitatively confirm results under the stronger assumption, namely a substantial positive total effect on eligibles and negative spillovers on ineligibles.

One crucial condition underlying all our approaches is the satisfaction of SUTVA on some aggregate level, see Hong and Raudenbush (2006), while it (in contrast to the standard literature) may be violated on the individual level. Throughout the article, we will refer to the aggregate entities as regions. Regional SUTVA allows for spillover effects between individuals within regions, but rules out such effects across regions. Given

CONTACT Andreas Steinmayr *andreas.steinmayr@econ.lmu.de* Department of Economics, University of Munich, Ludwigstraße 33, 80539, Munich, Germany.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/UBES*.

regional SUTVA, the total treatment effect may be split up into two causal mechanisms: (i) an individual effect and (ii) a within-region spillover effect that is driven by the treatment of other individuals in the region. The regional treatment may be defined as a binary variable, for example, whether a region is targeted by a treatment at all or not, or by a multivalued regional treatment intensity, in either case reflecting specific distributions of treated individuals in a region. The individual treatment is a binary indicator for whether an individual is treated as even in targeted regions, only a subgroup may actually be treated. As an important feature of our framework, individual and regional treatment effects may interact arbitrarily. This permits that spillover effects depend on the individual treatment status and that individual treatment effects depend on the regional treatment intensity. Albeit this makes the analysis more complex, it appears important in practice. For instance, a training may be more effective in a labor market where only few other individuals obtain a similar skill.

Our article is distinct from approaches of the literature on peer effects that typically rely on structural assumptions not imposed here. See, for instance, Graham (2008), who shows that conditional variance restrictions on outcomes point identify peer effects if outcomes are linear in average characteristics within some region or group, as discussed in Manski (1993). Our approach of defining and identifying effects is more closely related to nonparametric mediation analysis, which aims at disentangling the causal mechanisms through which a treatment affects an outcome (see, e.g., Robins and Greenland 1992; Pearl 2001; Robins 2003; Petersen, Sinisi, and van der Laan 2006; VanderWeele 2009; Hong 2010; Imai, Keele, and Yamamoto 2010; Huber 2014, among others). Our framework is at least in terms of notation also related to the dynamic treatment effects literature aiming to analyze sequences of treatments (see, e.g., Robins 1986, 1989; Robins, Hernan, and Brumback 2000; Lechner 2009; Lechner and Miquel 2010).

We use the principal stratification framework of Frangakis and Rubin (2002) to investigate effects for subpopulations or strata defined by the relation between the individual treatment state and regional treatment intensity, for example, eligibles and ineligibles. Principal stratification has been applied in the context of mediation analysis for instance by Rubin (2004) and VanderWeele (2008, 2012) and in the context of spillover effects by Forastiere, Mealli, and VanderWeele (2016). The latter study similarly to this article investigates stratum-specific causal mechanisms, however, the proposed identifying assumptions are different. While Forastiere, Mealli, and VanderWeele (2016) impose homogeneity assumptions on potential outcomes or effects across specific strata conditional on observables, our DiD approach relies on common time trends of potential outcomes within strata, but across regions (possibly conditional on observables).

Further strategies for evaluating spillovers include double randomization of both the regional and individual treatments, see, for example, Hudgens and Halloran (2008), Crépon et al. (2013), Baird et al. (2014), and Angelucci et al. (2018), or assuming selection on observables w.r.t. the regional and individual treatments, see Ferracci, Jolivet, and van den Berg (2014), which implies quasi-randomness given observed covariates. In contrast to identification within principal strata, neither of these approaches permits distinguishing effects between subpopulations receiving and not receiving the individual treatment under a specific regional treatment intensity. In so-called partial population experiments, see Moffitt (2001), regional treatment is randomized, while individual treatment is deterministically assigned based on an observed eligibility criterion, for example, a poverty index as in Angelucci and Giorgi (2009), implying that not everyone is eligible to treatment in treated regions. This permits identifying a subset of principal strata effects, namely the total effect on eligibles in treated regions as well as the spillover effect on the ineligibles in treated regions. See also Angelucci and Maro (2016), who discussed non- and quasi-experimental methods such as conditional independence, regression discontinuity, and instrumental variable assumptions aiming for the same principal strata effects. We complement these strategies by suggesting a DiD approach, one variant of which also permits identifying the spillover effect on eligibles in treated regions (rather than the total effect alone). Our method shares the feature of partial population experiments that individual treatment must be deterministic in observed covariates, but replaces the randomization of the regional treatment by a common trend assumption, which is weaker than the one imposed in Lalive, Landais, and Zweimüller (2015).

The remainder of this article is organized as follows. Section 2 proposes a general framework for defining spillover and individual treatment effects for various subpopulations and systematically reviews the effects considered in the previous literature. Section 3 provides identification results under various sets of assumptions related to randomization and selection on observables. It also suggests a novel DiD approach for the identification of spillover and individual treatment effects within principal strata. Section 4 presents an applications to a labor market intervention in Austria that extended unemployment benefits in selected regions. Section 5 concludes.

2. Definition of Effects

This section first introduces a general framework for defining individual and spillover effects based on a regional SUTVA. It then systematically categorizes the effects considered in several empirical studies according to this framework. Finally, it discusses further parameters not assessed in these studies that may nevertheless be of policy interest.

2.1. A General Framework for Individual and Spillover Effects

We denote by Z the regional treatment intensity, by D the individual treatment assignment, and by Y an individual level outcome. Within a region, Z might affect Y also other than through the individual treatment decision D, reflecting general equilibrium, interaction, or other spillover effects. The regional intensity of a training for job seekers, for instance, may affect the employment probability net of individual training participation through general equilibrium effects in the region: the larger the proportion of trained individuals, the lower may employment chances be for individuals in that labor market, conditional on their own treatment state. As a second example, the distribution of books to students in developing countries may have spillover effects on the academic performance of other students

through book sharing or more broadly through book-induced and performance-relevant interactions. We note that our definition of such effects may also include the impact inherent to the mere assignment of Z if it exists, for example, if book provision to students is organized through a school fair which has an effect by itself regardless of the actual distribution of books and the resulting student interactions.

The individual treatment is assumed to be binary, participation versus nonparticipation. The framework could also be extended to multivalued individual treatments, which is omitted for the sake of simplicity. Depending on the application, Z might be either binary or multivalued to reflect different distributions of treated individuals in a region. Even though we henceforth refer to Z as regional treatment intensity, which suggests considering different proportions of treated individuals across regions, we bear in mind that different values in Z may more generally reflect different choices of individuals to be treated in a region.

While the SUTVA is allowed to be violated on the individual level within some region, we assume throughout that SUTVA holds on the regional level. This rules out spillover effects across (rather than within) regions. For a more formal discussion, let $k \in \{1, \ldots, R\}$ index a specific region, with R being the number of regions. Using the potential outcome notation, see, for instance, Rubin (1974), let $D_{i,k}(z_1, \ldots, z_R)$ denote the potential individual treatment state of subject i in region k when setting intensities Z in regions 1 to R to the respective values $z_1, \ldots, z_R$. Furthermore, let $Y_{i,k}(z_1, \ldots, z_R, d)$ denote the potential outcome of individual i in region k when setting the regional treatment intensities (or distributions) to $z_1, \ldots, z_R$ and the individual treatment D to $d \in \{1, 0\}$. Regional SUTVA implies that the potential treatment state of some individual i in some region k is only affected by z_k, the intensity in the own region, but not by any other region. This rules out migration across regions, which is related to the intact clusters assumption of Hong and Raudenbush (2006). Likewise, regional SUTVA requires that no intensity other than that of region k influences the potential outcomes. This rules out treatment effects across regional borders. Assumption 1 formalizes these requirements.

Assumption 1 (SUTVA on the regional level). $D_{i,k}(z_1, \ldots, z_R) = D_{i,k}(z_k)$ and $Y_{i,k}(z_1, \ldots, z_R, d) = Y_{i,k}(z_k, d)$, for all $z_1, \ldots, z_R$ in the support of Z, $d \in \{1, 0\}$, and any subject i in any region k.

For notational convenience, we will henceforth keep the indices i and k implicit and write $Y_{i,k}(z_k, d)$ simply as $Y(z, d)$ and $D_{i,k}(z_k)$ as $D(z)$, which appears permissible after invoking regional SUTVA. Furthermore, we denote by T some target population of interest, which may, for instance, comprise all individuals receiving the individual treatment ($D = 1$). This allows defining average individual and spillover effects for some population $T = t$:

$$\begin{aligned} \delta_t(z) &= E[Y(z,1) - Y(z,0)|T = t] \text{ with } z \text{ in the} \\ &\quad \text{support of } Z, \\ \theta_t(z', z, d) &= E[Y(z', d) - Y(z, d)|T = t] \text{ with } z' \neq z \text{ and} \\ &\quad z', z \text{ in the support of } Z \text{ and } d \in \{0, 1\}. \end{aligned} \tag{1}$$

$\delta_t(z)$ is the impact of the individual treatment D given $Z = z$ and may thus, be a function of the regional treatment intensity. For instance, an individual training could be less effective if a larger share of labor market participants receive the same qualification. $\theta_t(z', z, d)$ is the spillover effect when comparing the regional treatment intensities z' versus z conditional on $D = d$ and may therefore, be a function of the individual treatment state. For instance, the spillover effect of providing some students with books on academic performance (Frölich and Michaelowa 2011) may be larger for students not receiving books ($D = 0$) than for students receiving books ($D = 1$). Only if there are no interaction effects between Z and D on Y are $\delta_t(z)$ and $\theta_t(z', z, d)$ not functions of z and d, respectively, and may be written as δ_t and $\theta_t(z', z)$. This is for instance satisfied under the constant unit-level treatment effect assumption of Robins (2003), requiring that $Y(1,1) - Y(0,1) = Y(1,0) - Y(0,0)$ for any individual. We will henceforth allow for interactions between Z and D. If the regional treatment intensity has only two levels, the spillover effect reduces to a binary comparison of $\theta_t(d) = E[Y(1, d) - Y(0, d)|T = t]$ with $z = 1$ and $z = 0$ denoting the higher and lower regional treatment intensity, respectively. We consider the binary case for most of the remainder of this article, but deviate whenever appropriate, as in parts of Sections 2.2, 2.3, and 3.

We introduce further notation for defining target populations determined by how the individual treatment state varies with the regional treatment intensity. Similar to the principal stratification approach of Frangakis and Rubin (2002) and the instrumental variable framework of Angrist, Imbens, and Rubin (1996), any individual i belongs to one of four compliance types $\mathcal{T}$ defined by the potential individual treatment states under $z = 1$ and $z = 0$: always takers ($\mathcal{T}_i = a : D_i(1) = D_i(0) = 1$) who are individually treated both under high and low regional treatment intensity, compliers ($\mathcal{T}_i = c : D_i(1) = 1, D_i(0) = 0$) who receive individual treatment under high, but not under low regional treatment intensity, defiers ($\mathcal{T}_i = d : D_i(1) = 0, D_i(0) = 1$) who behave opposite to the compliers, and never takers ($\mathcal{T}_i : D_i(1) = D_i(0) = 0$) who do not receive individual treatment under either regional intensity. $\mathcal{T}_i$ cannot be learned for any individual without further assumptions, because either $D_i(1)$ or $D_i(0)$ is observed, depending on Z.

2.2. Effects Considered in Empirical Examples

We consider empirical examples that are representative for identification approaches and effects typically investigated in the literature on spillover effects to provide a categorization according to the framework of Section 2.1. Our first example is PROGRESA, a partial population experiment in the sense of Moffitt (2001). It consists of a conditional cash transfer program for poor households in Mexico, in which treatment villages are randomly chosen, but only parts of the households in treated villages are actually offered the cash transfer as a function of a poverty index. For related examples, see Miguel and Kremer (2004), who study the spillover effects of a deworming treatment in Kenya, Baird et al. (2014), who assessed the spillover effects of a cash transfer program in Malawi, and Dahl, Løken, and Mogstad (2014), who estimated the peer effects of paid paternity leave in Norway using a regression discontinuity design. In PROGRESA, individual treatment D is the eligibility for cash transfers. The lower treatment intensity ($Z = 0$) corresponds to

zero such that no individual obtains a transfer. That is, $\Pr(D = 1|Z = 0) = 0$ and $D_i(0) = 0$ for all i, which rules out defiers and always takers. In treated villages ($Z = 1$), households below a particular poverty threshold were entitled to cash transfers ($D = 1$), while wealthier households were not ($D = 0$). Therefore, the types have a clear interpretation: never takers are noneligible, wealthier households, while compliers are poorer and eligible if the village is randomized in.

The design of PROGRESA allows identifying the spillover effect on the never takers in the absence of the individual treatment, $\theta_n(0) = E[Y(1,0) - Y(0,0)|\mathcal{T} = n]$, under regional SUTVA, because Z is random, type $\mathcal{T}_i$ of any individual i is deterministic in the observed poverty index, and D is deterministic in Z and the poverty index. This permits identifying never takers in both treated and nontreated villages, see the formal discussion in Section 3.1. Angelucci and Giorgi (2009), among others, used this strategy and found that PROGRESA cash transfers to eligible households indirectly increase the consumption of ineligible households. What they called the indirect treatment effect corresponds to $\theta_n(0)$ in this article. Angelucci and Giorgi (2009) also considered the (total) average effect of the policy intervention on the compliers (eligible households), denoted by $\Delta_c = E[Y(1,1) - Y(0,0)|\mathcal{T} = c]$. The latter parameter comprises both the individual and spillover effects on eligible households:

$$\begin{aligned}\Delta_c &= E[Y(1,1) - Y(1,0)|\mathcal{T} = c] + E[Y(1,0) \\ &\quad - Y(0,0)|\mathcal{T} = c] = \delta_c(1) + \theta_c(0) \\ &= E[Y(0,1) - Y(0,0)|\mathcal{T} = c] + E[Y(1,1) \\ &\quad - Y(0,1)|\mathcal{T} = c] = \delta_c(0) + \theta_c(1). \qquad (2)\end{aligned}$$

Equation (2) shows that the total effect on the compliers adds up to the individual treatment effect in villages receiving cash transfers ($\delta_c(1)$) and the spillover effect when not treated individually ($\theta_c(0)$). Alternatively, it adds up to the individual treatment effect in villages not receiving cash transfers ($\delta_c(0)$) and the spillover effect when treated individually ($\theta_c(1)$). That is, the two decompositions differ with respect to whether the interaction effects of Z and D are assigned to the individual or to the spillover effect. Arguably, $\delta_c(0)$ appears particularly interesting, because it corresponds to the individual effect if no one else received the treatment in the region. This corresponds to the effect we have in mind when imposing the individual level SUTVA, which rules out any spillover effects. However, whenever $Z = 0$ represents a regional treatment intensity of zero as in PROGRESA, $\delta_c(0)$ and $\theta_c(1)$ cannot be nonparametrically identified. The reason is that $\Pr(D = 1|Z = 0) = 0$ implies that individually treated do not exist for $Z = 0$, such that $E[Y(0,1)|T = t]$ cannot be inferred for any t. In a related study, Lalive and Cattaneo (2009) nevertheless decomposed individual and spillover effects among compliers, but require a tighter model that parametrically determines how spillover effects come about. Finally, Lalive, Landais, and Zweimüller (2015) evaluated the total effect on compliers and the spillover effect on never takers in treated regions, $\Delta_{c,Z=1}\,\theta_{n,Z=1}(0)$, however, based on DiD rather than assuming randomization of the regional treatment, see the discussion in Section 4. The contribution of the present article is to provide DiD-based identification of $\Delta_{c,Z=1}\,\theta_{n,Z=1}(0)$ under somewhat weaker assumptions and to provide conditions that permit identifying spillover effects on compliers, see Section 3.4.

As a further example, Crépon et al. (2013) assess a randomized job placement assistance program in France, where the treatment probability differs across regions, which corresponds to a multivalued Z. They find that the regional intensity of the program negatively affects the employment chances of individuals not taking the treatment. The analysis differs from PROGRESA in that not only Z, but also the individual treatment D is randomized. This implies that characteristics and effects do not vary across populations T defined in terms of Z, D, or $\mathcal{T}$ and correspond to those in the total population: $\delta_t(z) = \delta(z) = E[Y(z,1) - Y(z,0)]$ and $\theta_t(z',z,d) = \theta(z',z,d) = E[Y(z',d) - Y(z,d)]$. The authors consider four regional treatment intensities corresponding to the share of treated job seekers (with 0 corresponding to 0% and 1 to 100%). $z \in \{0, 0.25, 0.5, 0.75\}$ to assess heterogeneity in individual treatment effects $\delta(z)$ over positive values z. To see this, consider the following saturated potential outcome model: $E[Y(Z,D)] = \beta_0 + \beta_1 D + \beta_2 I\{Z = 0.25\} + \beta_3 I\{Z = 0.5\} + \beta_4 I\{Z = 0.75\} + \beta_5 DI\{Z = 0.25\} + \beta_6 DI\{Z = 0.5\} + \beta_7 DI\{Z = 0.75\}$. Under heterogeneous treatment effects, the β coefficients are means or mean effects. For instance, $\beta_0 = E[Y(0,0)]$. Therefore, differences in the interactions are identified over positive values z. For instance, $\beta_5 - \beta_6$ identifies $\delta(0.25) - \delta(0.5)$.

However, $\delta(0)$ is not nonparametrically identified in this setup because individually treated do not exist when the regional treatment intensity is exactly zero. See the formal discussion in Section 3.2 and in particular common support Assumption 6, which is not satisfied. The reason is that $E[Y(0,1)] = \beta_0 + \beta_1$ (and thus, $\delta(0)$) cannot be identified as D is collinear with its interaction with Z in a regression. For instance, when comparing $Z = 0.25$ versus $Z = 0$, β_1 cannot be separated from the interaction β_5. However, evaluations with treatment intensities close to zero in some region may come close to measuring $\delta(0)$. This might appear interesting because $\delta(0)$ gives the individual treatment effect that would occur under a satisfaction of individual-level SUTVA, that is, in the absence of any spillovers. Based on a selection on observables assumption implying quasi-randomization conditional of observed characteristics, Ferracci, Jolivet, and van den Berg (2014), for instance, considered the evaluation of $\delta(z)$ with the minimum regional intensity amounting to a treatment probability of just 2%. Analogously, the design of Crépon et al. (2013) permits identifying spillover effects $\theta(z',z,1)$ for $z' \neq z$ and $z', z \in \{0.25, 0.5, 0.75\}$, as well as any $\theta(z',z,0)$ for $z' \neq z$ and $z', z \in \{0, 0.25, 0.5, 0.75\}$. However, the spillover effect of some positive versus a zero regional treatment intensity under individual treatment assignment, $\theta(z,0,1) = E[Y(z,1) - Y(0,1)]$ for $z \in \{0.25, 0.5, 0.75\}$, remains unidentified. Again, close-to-zero regional treatment intensities permit approximating $\theta(z,0,1)$, see Baird et al. (2014). Table 1 summarizes the various effects which have been considered in previous studies based on the causal framework introduced in Section 2.1.

2.3. Further Effects

It is worth mentioning that the spillover and individual treatment effects $\theta(z',z,d)$ and $\delta(z)$ identified by double

Table 1. Summary of effects and empirical examples.

Parameter	Symbol	Description	Examples
$E[Y(1,1)-Y(0,0)\|\mathcal{T}=c]$	Δ_c	(Total) treatment effect on compliers	Angelucci and Giorgi (2009) and Lalive and Cattaneo (2009)
$E[Y(1,1)-Y(0,0)\|\mathcal{T}=c,Z=1]$	$\Delta_{c,Z=1}$	(Total) treatment effect on compliers in treated regions	Lalive, Landais, and Zweimüller (2015)
$E[Y(z,1)-Y(z,0)\|\mathcal{T}=c]$	$\delta_c(z)$	Individual treatment effect on compliers	Lalive and Cattaneo (2009)
$E[Y(1,d)-Y(0,d)\|\mathcal{T}=n]$	$\theta_n(d)$	Spillover effect on never takers	Angelucci and Giorgi (2009) and Lalive and Cattaneo (2009)
$E[Y(1,d)-Y(0,d)\|\mathcal{T}=n,Z=1]$	$\theta_{n,Z=1}(d)$	Spillover effect on never takers	Lalive, Landais, and Zweimüller (2015)
$E[Y(z,1)-Y(z,0)]$	$\delta(z)$	Individual treatment effect in the population	Crépon et al. (2013) and Ferracci, Jolivet, and van den Berg (2014)
$E[Y(z',d)-Y(z,d)]$	$\theta(z',z,d)$	Spillover effect in the population	Crépon et al. (2013) and Baird et al. (2014)

randomization as in Crépon et al. (2013) or selection on observables as in Ferracci, Jolivet, and van den Berg (2014) can be defined for potentially infeasible combinations of regional and individual treatments. For instance, $\theta(0.25,0,1) = E[Y(0.25,1) - Y(0,1)]$ is practically infeasible, as it yields the average spillover effect of 25% versus 0% treated when in fact setting the individual treatment to 1 for all individuals, that is, for 100% of the population. Such effects may nevertheless appear interesting if policy makers aim at maintaining random assignment of D beyond the experiment such that treated and nontreated are representative for the total population at any intensity z, because $\theta(z',z,d)$ and $\delta(z)$ can also be interpreted as effects on the average individual. However, most empirical problems are characterized by nonrandom selection into individual treatment assignment. Caseworkers in employment offices, for instance, typically assign active labor market programs to job seekers depending on education and previous labor market experience, among other factors. Under selection, the interpretation of such practically infeasible parameters as effects on the average individual appears less attractive, as the composition of treated changes with z and does not necessarily match the average individual.

We, therefore, discuss further effects that might be of policy interest, have not been considered in the review of Section 2.2, and are not defined in terms of infeasible combinations of regional and individual treatments. For a binary Z, for instance $\delta_{Z=1,D=1}(1) = [Y(1,1) - Y(1,0)|Z=1,D=1]$, the average effect of D on those individually treated in treated regions, appears relevant for judging whether D was effective among those who actually received it. We note that the latter parameter is a mixture of the impacts on compliers and always takers, as for either group $D(1)=1$, such that observing $Z=1$ implies $D=1$. If Z is such that $\Pr(D=1|Z=0)>0$ and treated exist under a low treatment intensity, then $\delta_{Z=0,D=1}(0) = E[Y(0,1)-Y(0,0)|Z=0,D=1]$, the effect of D on the individually treated in regions with low treatment intensity appears interesting, too, which is a mixture of impacts on always takers and defiers (if the latter exist). Furthermore, policy makers might also want to learn about $\theta_{Z=1,D=1}(1) = E[Y(1,1) - Y(0,1)|Z=1,D=1]$ and $\theta_{Z=0,D=1}(1) = E[Y(1,1) - Y(0,1)|Z=0,D=1]$, that is, the spillover effects on individually treated subjects in regions with high or low treatment intensity. Likewise, the spillover effects on nontreated individuals, $\theta_{Z=1,D=0}(0) = E[Y(1,0) - Y(0,0)|Z=1,D=0]$ and $\theta_{Z=0,D=0}(0) = E[Y(1,0) - Y(0,0)|Z=0,D=0]$ appear of policy interest.

Finally, we consider the spillover effect conditional on the potential individual treatment state under a particular regional treatment intensity, denoted by $D(z)$, rather than setting D to some value d for every individual. Analogous to the denomination of Pearl (2001) in the context of causal mediation analysis, we refer to this parameter as natural spillover effect. For illustration, we focus on the natural spillover effect in treated regions with $Z=1$, given $D(1)$, that is the individual treatments actually occurring in $Z=1$:

$$\theta_{Z=1}(D(1)) = E[Y(1,D(1)) - Y(0,D(1))|Z=1]. \tag{3}$$

In contrast to $\theta(z',z,d)$ and $\delta(z)$, this parameter is by definition not defined in terms of infeasible combinations of D and Z, but reflects the spillover effects actually occurring in treated regions on both individually treated ($D(1)=1$) and nontreated ($D(1)=0$) subpopulations. $\theta_{Z=1}(D(1))$ is thus a weighted average of the previously discussed spillover effects $\theta_{Z=1,D=1}(1)$ and $\theta_{Z=1,D=0}(0)$, where the weights correspond to the shares of the individually treated and nontreated in the treated regions, which follows from the law of total probability:

$$\begin{aligned}
\theta_{Z=1}(D(1)) &= E[Y(1,1) - Y(0,1)|Z=1,D(1)=1] \\
&\quad \cdot \Pr(D(1)=1|Z=1) \\
&\quad + E[Y(1,0) - Y(0,0)|Z=1,D(1)=0] \\
&\quad \cdot \Pr(D(1)=0|Z=1) \\
&= \underbrace{E[Y(1,1) - Y(0,1)|Z=1,D=1]}_{\theta_{Z=1,D=1}(1)} \\
&\quad \cdot \Pr(D(1)=1|Z=1) \\
&\quad + \underbrace{E[Y(1,0) - Y(0,0)|Z=1,D=0]}_{\theta_{Z=1,D=0}(0)} \\
&\quad \cdot \Pr(D(1)=0|Z=1).
\end{aligned}$$

Section 3.3 provides identification results based on selection on observables assumptions for the causal parameters considered in this section.

3. Identifying Assumptions

We formally discuss different sets of assumptions and their identifying power for various effects introduced in Section 2. We consider (i) randomization of Z and deterministic assignment of D as in PROGRESA, (ii) double randomization of Z and D as in Crépon et al. (2013), (iii) selection on observables w.r.t. Z and D

as in VanderWeele (2010), and (iv) DiD approaches as in Lalive, Landais, and Zweimüller (2015). Throughout, we assume that regional SUTVA holds, see Assumption 1 in Section 2.1.

3.1. Randomization of Z and Deterministic D

This section focusses on the identification of $\theta_n(0)$ and Δ_c in partial population experiments. We subsequently formalize the assumptions of random regional treatment assignment and deterministic individual treatment assignment, assuming that Z and D are binary to identify $\theta_n(0)$ and Δ_c.

Assumption 2 (*Random assignment of the regional treatment*). $\{Y(z',d), D(z)\} \perp Z$ for all $z', z, d \in \{0,1\}$ and Z not being degenerate.

Furthermore, D is assumed to be a deterministic function of the regional intensity and observed characteristics, denoted by X, which are measured at or prior to the assignment of Z.

Assumption 3 (*Deterministic individual treatment assignment*). $D = g(Z, X)$, with g being a known function.

In PROGRESA, for instance, the individual treatment is fully determined by Z and a poverty index. Specifically, $D = g(0,X) = 0$, while $D = g(1,X)$ might be either 1 or 0 depending on the score of the poverty index X. By Assumption 3, both $D(0)$ and $D(1)$ and thus, also the type $\mathcal{T}$ is identified for any subject. By Assumption 2, types have identical proportions in treated and nontreated regions. This implies that Z and X are independent, as no determinants of D must affect Z. Therefore, the spillover effect on never takers and the total effect on compliers is identified, given that these types exist in the population, as postulated in Assumption 4.

Assumption 4 (*Existence of never takers and compliers*). $\Pr(g(1,X) = g(0,X) = 0) > 0$ and $\Pr(g(1,X) - g(0,X) = 1) > 0$.

Proposition 1. Under Assumptions 2–4,

$$\begin{aligned}\theta_n(0) &= E[Y(1,0) - Y(0,0)|\mathcal{T} = n] = E[Y(1,0)|Z = 1, \\ &\quad \mathcal{T} = n] - E[Y(0,0)|Z = 0, \mathcal{T} = n] \\ &= E[Y|Z = 1, \mathcal{T} = n] - E[Y|Z = 0, \mathcal{T} = n], \qquad (4)\\ \Delta_c &= E[Y(1,1) - Y(0,0)|\mathcal{T} = c] = E[Y(1,1)|Z = 1, \\ &\quad \mathcal{T} = c] - E[Y(0,0)|Z = 0, \mathcal{T} = c] \\ &= E[Y|Z = 1, \mathcal{T} = c] - E[Y|Z = 0, \mathcal{T} = c]. \qquad (5)\end{aligned}$$

Disentangling the total effect on the compliers requires further assumptions, for instance, effect homogeneity in spillover effects across types. If one assumes for instance that $\theta_c(0) = \theta_n(0)$, then $\delta_c(1) = \Delta_c - \theta_n(0)$.

We refer to Forastiere, Mealli, and VanderWeele (2016) for a more thorough discussion of homogeneity assumptions on potential outcomes or effects across types for identifying spillover effects. Finally and as a deviation from the PROGRESA setting, consider the case that $g(0,X)$ may be 1 or 0 depending on X. That is, households with a very low poverty index always receive transfer payments (even if $Z = 0$), such that always takers exist, implying $\Pr(g(1,X) = g(0,X) = 1) > 0$. Together with Assumptions 2 and 3, spillover effects on always takers are in this case identified by $\theta_a(1) = E[Y(1,1) - Y(0,1)|\mathcal{T} = a] = E[Y|Z = 1, \mathcal{T} = a] - E[Y|Z = 0, \mathcal{T} = a]$. If defiers exist, their total effect is given by $\Delta_d = E[Y(1,1) - Y(0,0)|\mathcal{T} = d] = E[Y|Z = 0, \mathcal{T} = d] - E[Y|Z = 1, \mathcal{T} = d]$.

3.2. Randomization of the Regional and Individual Treatment

This section discusses the identification of $\delta(z)$, $\theta(d)$, and $\theta(D(z))$ based on double randomization. We subsequently maintain Assumption 2, but replace Assumption 3 by random individual treatment assignment conditional on Z:

Assumption 5 (*Random assignment of the individual treatment within regions*). $Y(z',d) \perp D|Z = z$, for all $z', z, d \in \{0,1\}$.

The individual-level treatment effect $\delta(z)$ is identified under double randomization if the following common support condition is satisfied:

Assumption 6 (*Existence of individuals with $D = 1$ and $D = 0$ conditional on $Z = z$*). $0 < \Pr(D = 1|Z = z) < 1$.

For the identification of $\delta(0)$, for instance, Assumption 6 is violated if $Z = 0$ implies that nobody is individually treated, that is $\Pr(D = 1|Z = 0) = 0$.

Proposition 2. Under Assumptions 2, 5, and 6, the individual effect among the individually treated is given by

$$\begin{aligned}\delta(z) &= E[Y|Z = z, D = 1] - E[Y|Z = z, D = 0] \qquad (6)\\ &= \frac{E[Y \cdot D|Z = z]}{\Pr(D = 1|Z = z)} - \frac{E[Y \cdot (1 - D)|Z = z]}{1 - \Pr(D = 1|Z = z)}.\end{aligned}$$

Identification of the spillover effect $\theta(d)$, on the other hand, requires the satisfaction of the following common support condition:

Assumption 7 (*Existence of regions with $Z = 1$ and $Z = 0$ conditional on $D = d$*). $0 < \Pr(Z = 1|D = d) < 1$.

Assumption 7 is closely linked to Assumption 6, albeit tailored to $\theta(d)$ rather than $\delta(z)$. By Bayes' theorem, $\Pr(D = 1|Z = 0) = 0$ for example implies that $\Pr(Z = 0|D = 1) = 0$ such that $\Pr(Z = 1|D = 1) = 1$ and Assumption 7 is violated for the identification of $\theta(1)$.

Proposition 3. Under Assumptions 2, 5, and 7, the spillover effect conditional on Z, D is given by

$$\begin{aligned}\theta(d) &= E[Y|Z = 1, D = d] - E[Y|Z = 0, D = d] \qquad (7)\\ &= \frac{E[Y \cdot Z|D = d]}{\Pr(Z = 1|D = d)} - \frac{E[Y \cdot (1 - Z)|D = d]}{1 - \Pr(Z = 1|D = d)}.\end{aligned}$$

Finally, for the identification of the natural spillover effect, $\theta(D(z))$, common support as postulated in Assumption 7 needs to hold for both $D = 1$ and $D = 0$:

Assumption 8 (*Existence of regions with $Z = 1$ and $Z = 0$ given $D = 1$ and $D = 0$*). $0 < \Pr(Z = 1|D = d) < 1$ for all $d \in \{1,0\}$.

It is easy to see from Bayes' theorem that Assumption 8 implies both Assumptions 6 and 7.

Proposition 4. Under Assumptions 2, 5, and 8, the natural spillover effect in treatment regions is given by

$$\theta(D(z)) = E\left[\left(\frac{Y \cdot Z}{\Pr(Z=1|D)} - \frac{Y \cdot (1-Z)}{1-\Pr(Z=1|D)}\right) \cdot \frac{\Pr(Z=z|D)}{\Pr(Z=z)}\right]. \quad (8)$$

Proof. See Appendix A.1. □

Albeit double randomization appears intuitively attractive, a conceptual shortcoming already raised in Section 2.3 w.r.t. $\delta(z)$ and $\theta(d)$ also applies to the natural spillover effect. $\theta(D(z))$ is only properly identified if the individual treatment remains randomly assigned beyond the experimental evaluation. If the assignment rule is, however, selective in real world applications, the distribution of $D(z)$ as well as $\theta(D(z))$ remain unknown. We refer to Imai, Tingley, and Yamamoto (2013) for a further discussion on this issue.

3.3. Selection on Observables

This section considers selection on observables assumptions for identifying $\delta_{Z=z,D=1}(z)$, $\theta_{Z=z,D=d}(d)$, and $\theta_{Z=1}(D(z))$, as well as a weighted version of the spillover effect. As a relaxation of the assumptions in Section 3.2, we assume that regional treatment assignment is quasi-random conditional on a set of observables, denoted by X, and that individual treatment assignment is quasi-random conditional on Z, X. Such or similar sequential exogeneity assumptions have among others been considered by Pearl (2001), Flores and Flores-Lagunes (2009), Hong (2010), VanderWeele (2010), Imai, Keele, and Yamamoto (2010), Tchetgen Tchetgen and Shpitser (2012), Vansteelandt, Bekaert, and Lange (2012), and Huber (2014) in the context of causal mediation analysis, while VanderWeele et al. (2013) extend the framework to distinguish between spillovers and other causal mechanisms. We replace Assumptions 2 and 5 by Assumptions 9 and 10.

Assumption 9 (Conditional independence of the regional treatment). $\{Y(z',d), D(z)\} \perp Z|X = x$ for all $z', z, d \in \{0,1\}$ and x in the support of X.

Assumption 9 states that the joint distribution of the potential outcomes and individual treatments are independent of the regional treatment intensity conditional on X. This rules out unobserved confounders affecting regional treatment assignment on the one hand and the potential outcomes and/or individual treatment under $z = 0$ on the other hand, when controlling for X.

This is known as conditional independence, selection on observables, or exogeneity in the treatment evaluation literature (see, e.g., Imbens 2004).

Assumption 10 (Conditional independence of the individual treatment). $Y(z',d) \perp D|Z = z, X = x$ for all $z', z, d \in \{0,1\}$ and x in the support of X.

Assumption 10 states that the individual treatment is conditionally independent of the potential outcomes conditional on the actual regional treatment intensity Z and covariates X. It rules out unobserved confounders jointly affecting the individual treatment and the potential outcomes under $z = 0$ after controlling for X and Z. As a conceptual improvement over the double randomization framework w.r.t. real world applications, individual treatment assignment may now be a function of X (in addition to Z). If, for instance, caseworkers assign a training program D to job seekers based on their individual characteristics and if all characteristics also affecting outcome Y are observed in X, identification is obtained despite nonrandom treatment assignment.

Assumption 11 (Common support restrictions).

(a) $0 < \Pr(D=1|Z=z, X=x) < 1$ for all x in the support of X,
(b) $0 < \Pr(Z=1|D=d, X=x) < 1$ for all x in the support of X,
(c) $0 < \Pr(Z=1|D=d, X=x) < 1$ for all $d \in \{1,0\}$ and x in the support of X.

Assumptions 11(a)–(c) are analogous to Assumptions 6–8, but are stronger in the sense that they are required to hold conditional on X. Similarly as before, Assumption 11(c) is stronger than (and implies) Assumptions 11(a) and (b). Note that Assumption 11(a) is necessarily violated if individual treatment assignment given Z depends deterministically on (an index of) X, as it is the case in PROGRESA.

Proposition 5. Under Assumptions 9, 10, and 11(a),

$$\delta_{Z=z,D=1}(z) = E\left[\left(\frac{Y \cdot D}{\Pr(D=1|Z=z,X)} - \frac{Y \cdot (1-D)}{1-\Pr(D=1|Z=z,X)}\right) \cdot \frac{\Pr(Z=z|X) \cdot \Pr(D=1|Z=z,X)}{\Pr(Z=z) \cdot \Pr(D=1|Z=z)}\right], \quad (9)$$

while under Assumptions 9, 10, and 11(b),

$$\theta_{Z=z,D=d}(d) = E\left[\left(\frac{Y \cdot Z}{\Pr(Z=1|D=d,X)} - \frac{Y \cdot (1-Z)}{1-\Pr(Z=1|D=d,X)}\right) \cdot \frac{\Pr(Z=z|X) \cdot \Pr(D=d|Z=z,X)}{\Pr(D=d) \cdot \Pr(Z=z|D=d)}\right], \quad (10)$$

and finally, under Assumptions 9, 10, and 11(c),

$$\theta_{Z=1}(D(z)) = E\left[\left(\frac{Y \cdot Z}{\Pr(Z=1|D,X)} - \frac{Y \cdot (1-Z)}{1-\Pr(Z=1|D,X)}\right) \cdot \frac{\Pr(Z=z|D,X)}{\Pr(Z=z|X)} \cdot \frac{\Pr(Z=1|X)}{\Pr(Z=1)}\right]. \quad (11)$$

The proofs for Equations (9)–(11) are provided in Appendix A.2 and are closely related to those in Huber (2014).

Under particular conditions, one can identify type-specific effects (Δ_c, $\theta_c(d)$, $\delta_c(z)$, $\theta_n(d)$, $\theta_a(d)$) despite the fact that in contrast to Assumption 3, Assumptions 9 and 10 do not permit learning the types from the data. However, the assumptions imply that heterogeneity in potential outcomes across types is exclusively driven by X. Formally, $Y(z',d) \perp D(z)|X = x$, see Imai, Keele, and Yamamoto (2010), that is, types and potential outcomes are conditionally independent given X. The additional assumption that D is monotonic in Z given X allows identifying the type proportions, see Abadie (2003):

Assumption 12 (Monotonicity). $\Pr(D(1) \geq D(0)|X = x) = 1$ for all x in the support of X.

In analogy to the concept of weighted treatment effects in Hirano, Imbens, and Ridder (2003), reweighing observations to appropriately average over X for replicating the covariate distribution in some target population yields the effects on compliers, always takers, and never takers.

Proposition 6. Under Assumptions 9, 10, 11(b), and 12 as well as $E[\omega(X)] > 0$, the weighted spillover effect is given by

$$\theta_\omega(d) = E\left[\left(\frac{Y \cdot Z}{\Pr(Z=1|D=d,X)} - \frac{Y \cdot (1-Z)}{1-\Pr(Z=1|D=d,X)}\right) \cdot \frac{\omega(X)}{E[\omega(X)]}\right], \quad (12)$$

where $\omega(X)$ is the weighting function, whose absolute value is assumed to be bounded from above. Setting $\omega(X) = 1 - \frac{D(1-Z)}{1-\Pr(Z=1|X)} - \frac{(1-D)Z}{\Pr(Z=1|X)}$ identifies $\theta_c(d)$, as this reweighs observations according to the distribution of X among compliers, see Abadie (2003). The weighting functions for always and never takers are $\frac{D(1-Z)}{1-\Pr(Z=1|X)}$ and $\frac{(1-D)Z}{\Pr(Z=1|X)}$, respectively.

3.4. Difference-in-Differences

Based on DiD methods, we subsequently discuss the identification of three effects considered in the empirical application of Section 4: the spillover effect among never takers in treated regions, $\theta_{Z=1,n}(0)$, the total effect among compliers in treated regions, $\Delta_{Z=1,c}$, and the (total) average treatment effect in treated regions (ATET), which is formally defined as $\Delta_{Z=1} = E[Y(1,D(1)) - Y(0,D(0))|Z = 1]$. In addition, two further parameters are considered, namely the spillover effects on always takers in treated regions, $\theta_{Z=1,a}(1)$, and on compliers under a multivalued Z, $\theta_{Z=2,c}(z' = 2, z = 1, d = 1)$.

A precondition for DiD-based identification is that the outcome is observed both in a baseline period prior to the assignment of Z and D and in a follow-up period after the treatments when the effects are evaluated, requiring either panel data or repeated cross sections. For this reason, we introduce a time index for the period in which the outcome is measured: Y_0 denotes the pretreatment outcome, while Y_1 is the outcome in the follow-up period. Note that in the previous discussion, $Y_1 = Y$. Likewise, $Y_1(z,d)$ denotes the potential outcome in the follow-up period for $Z = z$ and $D = d$, while $Y_0(z,d)$ denotes the potential outcome in the pretreatment period, that is, prior to the actual assignment of Z and D. Therefore, $Y_0(z,d)$ is defined in terms of (anticipating) treatments not yet realized, because any treatment state is assumed to be zero in the baseline period. Identification of type-specific effects is based on combining common trend assumptions on outcome changes over time with deterministic individual treatment assignment (Assumption 3). (We refer to Deuchert, Huber, and Schelker (2018) for alternative DiD approaches to the identification of type-specific effects when Z is in contrast to the present framework assumed to be random, while the compliance type is not directly observed, i.e., not a deterministic function of Z and X as imposed by Assumption 3.)

Identification of the total effect on compliers and the spillover effect on never takers rests on the same assumptions. Assumption 13 states that the mean potential outcomes in the absence of any regional and individual treatment within a type would change by the same magnitude from the baseline to the follow-up period across (actual) regional treatment intensities.

Assumption 13 (Common trends within never takers/compliers across regions). $E[Y_1(0,0) - Y_0(0,0)|Z = z, \mathcal{T} = \tau] = E[Y_1(0,0) - Y_0(0,0)|Z = z', \mathcal{T} = \tau]$ for all $z \neq z'$ and $\tau \in \{n, c\}$.

Assumption 14 rules out any average effects of Z or D on the outcome on never takers or compliers in the baseline period. Such effects could arise, for example, if some units changed their behavior already in the baseline period in anticipation of the individual or regional treatment to come.

Assumption 14 (No average anticipation effect). $E[Y_0(z,d) - Y_0(0,0)|Z = z', \mathcal{T} = \tau] = 0$ for all z, z', d in the support of D, Z and $\tau \in \{n, c\}$.

Note that Assumptions 13 and 14 do not restrict the proportions of types to remain constant across regions (as Z is not randomized) or over time. However, the types of interest must be observed for any time period and regional treatment state, which requires Assumption 4 to hold conditionally. Assuming that Z and D are binary, the spillover effect on never takers and the total effect on compliers in the treated regions are given by the following expressions.

Proposition 7. Under Assumptions 3, 4 given Z in either period, 13, and 14:

$$\begin{aligned}\theta_{Z=1,n}(0) &= E[Y_1|Z=1,\mathcal{T}=n] - E[Y_0|Z=1,\mathcal{T}=n] \\ &\quad - [E[Y_1|Z=0,\mathcal{T}=n] - E[Y_0|Z=0,\mathcal{T}=n]], \\ \Delta_{Z=1,c} &= E[Y_1|Z=1,\mathcal{T}=c] - E[Y_0|Z=1,\mathcal{T}=c] \\ &\quad - [E[Y_1|Z=0,\mathcal{T}=c] - E[Y_0|Z=0,\mathcal{T}=c]].\end{aligned} \quad (13)$$

Proof. See Appendix A.3. □

We also consider a further common trends condition as used, for instance, in Lalive, Landais, and Zweimüller (2015). Assumption 13′ states that the mean potential outcomes in the absence of any regional and individual treatment would change

by the same magnitude from the baseline to the follow-up period across (actual) regional treatment intensities and types.

Assumption 13′ (Common trends across types and regions). $E[Y_1(0,0) - Y_0(0,0)|Z = z, \mathcal{T} = \tau] = E[Y_1(0,0) - Y_0(0,0)|Z = z']$ for all $z \neq z'$ and $\tau \in \{n, c\}$.

Proposition 8. Under Assumptions 3, 4 given Z in either period, 13′, and 14,

$$\begin{aligned}\theta_{Z=1,n}(0) &= E[Y_1|Z=1,\mathcal{T}=n] - E[Y_0|Z=1,\mathcal{T}=n] \\ &\quad - [E[Y_1|Z=0] - E[Y_0|Z=0]]\,, \\ \Delta_{Z=1,c} &= E[Y_1|Z=1,\mathcal{T}=c] - E[Y_0|Z=1,\mathcal{T}=c] \\ &\quad - [E[Y_1|Z=0] - E[Y_0|Z=0]]\,. \end{aligned} \tag{14}$$

Proof. See Appendix A.3. □

Assumption 13′ is stronger than Assumption 13, implying a common trend in the mean potential outcomes of compliers and never takers in nontreated regions, that is across (rather than within) types. This yields the following testable implication (under Assumptions underlying Equation (13)): $E[Y_1 - Y_0|Z = 0, \mathcal{T} = c] = E[Y_1 - Y_0|Z = 0, \mathcal{T} = n]$. Identification using Assumption 13′ is feasible even if the variables X underlying the deterministic treatment assignment are not observed in nontreated regions, which is not the case for Assumption 13.

In many applications like PROGRESA and the one presented in Section 4, always takers and defiers can be ruled out based on the individual treatment rule of Assumption 3. The ATET is then identified (by the law of total probability) as weighted average of $\theta_{Z=1,n}(0)$ and $\Delta_{Z=1,c}$. The weights depend on the shares of compliers and never takers in treated regions in the follow-up period:

$$\begin{aligned}\Delta_{Z=1} &= \Delta_{Z=1,c} \cdot \Pr(\mathcal{T}=c|Z=1) + \theta_{Z=1,n}(0) \\ &\quad \cdot \Pr(\mathcal{T}=n|Z=1). \end{aligned} \tag{15}$$

If always takers exist, the spillover effect $\theta_{Z=1,a}(1)$ can be identified based on the following assumption.

Assumption 15 (Common trends within always takers across regions). $E[Y_1(0,1) - Y_0(0,0)|Z = z, \mathcal{T} = a] = E[Y_1(0,1) - Y_0(0,0)|Z = z', \mathcal{T} = a]$ for all $z \neq z'$.

Assumption 15 may appear nonintuitive, as pre- and post-treatment potential outcomes are defined on distinct d. However, adding and subtracting $Y_1(0,0)$ in the conditional expectations shows that the following two comprehensible conditions are sufficient for Assumption 15. First, a common trend assumption analogous to Assumption 13 must hold for always takers: $E[Y_1(0,0) - Y_0(0,0)|Z = z, \mathcal{T} = a] = E[Y_1(0,0) - Y_0(0,0)|Z = z', \mathcal{T} = a]$. Second, the average individual treatment effect on always takers under a low regional intensity must be constant across (actual) regional intensities: $E[Y_1(0,1) - Y_1(0,0)|Z = z, \mathcal{T} = a] = E[Y_1(0,1) - Y_1(0,0)|Z = z', \mathcal{T} = a]$. To see this, note that $E[Y_1(0,1) - Y_0(0,0)|Z = z, \mathcal{T} = a] = E[Y_1(0,1) - Y_0(0,0)|Z = z', \mathcal{T} = a]$ is the same as $E[Y_1(0,1) - Y_1(0,0) + Y_1(0,0) - Y_0(0,0)|Z = z, \mathcal{T} = a] = E[Y_1(0,1) - Y_1(0,0) + Y_1(0,0) - Y_0(0,0)|Z = z', \mathcal{T} = a]$, which holds if $E[Y_1(0,1) - Y_1(0,0)|Z = z, \mathcal{T} = a] = E[Y_1(0,1) - Y_1(0,0)|Z = z', \mathcal{T} = a]$, that is, mean effects are constant, and $E[Y_1(0,0) - Y_0(0,0)|Z = z, \mathcal{T} = a] = E[Y_1(0,0) - Y_0(0,0)|Z = z', \mathcal{T} = a]$, that is, common trends hold.

Assumption 15 seems tighter than Assumption 13, but the two are not strictly nested. Under the existence of always takers given Z in either period and the satisfaction of Assumptions 3, 14, and 15, $\theta_{Z=1,a}(1)$ is identified by (the proof is similar to that of (13) and omitted).

Proposition 9. Under Assumptions 3, 14, and 15, as well as the presence of always takers given Z in either period,

$$\begin{aligned}\theta_{Z=1,a}(1) &= E[Y_1|Z=1,\mathcal{T}=a] - E[Y_0|Z=1,\mathcal{T}=a] \\ &\quad - [E[Y_1|Z=0,\mathcal{T}=a] - E[Y_0|Z=0,\mathcal{T}=a]]\,. \end{aligned} \tag{16}$$

We subsequently propose a strategy for identifying spillover effects on compliers on top of the total effect, which requires multiple regional treatment intensities. Suppose that Z can take three values: 0 (no regional treatment), 1 (low intensity), and 2 (high intensity). Furthermore, we restate Assumption 3 more precisely by requiring $g(z', X) = g(z, X)$ for any $z', z > 0$: eligibility depends on the same criteria in all regions because compliers satisfy $D(0) = 0, D(1) = D(2) = 1$. Therefore, variation in regional treatment intensity does not originate from different eligibility criteria, but from differences in the distribution of X across regions. This also entails different complier shares, which is required for identifying spillover effects.

Assumption 16 (Effect homogeneity among compliers across regions). $E[Y_1(1,1) - Y_1(0,0)|Z = 2, \mathcal{T} = c] = E[Y_1(1,1) - Y_1(0,0)|Z = 1, \mathcal{T} = c]$.

By Assumption 16, the average total effect on compliers of a low versus no regional treatment is constant across (actually) high and low treatment regions.

This allows separating the total effect of $Z = 2$ versus $Z = 0$ on compliers in high treatment regions into the total effect of $Z = 1$ versus $Z = 0$, which by Assumption 16 equals the respective effect in low treatment regions, and the spillover effect of $Z = 2$ versus $Z = 1$.

Proposition 10. Under the existence of compliers given Z in either period and Assumptions 3, 13, 14, and 16, the spillover effect of $Z = 2$ versus $Z = 1$ is identified by

$$\begin{aligned}&\theta_{Z=2,c}(z'=2, z=1, d=1) \\ &\quad = E[Y_1|Z=2,\mathcal{T}=c] - E[Y_0|Z=2,\mathcal{T}=c] \\ &\qquad - [E[Y_1|Z=1,\mathcal{T}=c] - E[Y_0|Z=1,\mathcal{T}=c]]\,. \end{aligned} \tag{17}$$

Proof. See Appendix A.3. □

Finally, DiD assumptions might be relaxed to hold given observed covariates, denoted by W. This requires conducting DiD conditional on W and averaging over W to mimic the covariate distribution in period 1 of the respective type given $Z = 1$.

Proposition 11. Under the conditional validity (conditional on W in either period) of Assumptions 3, 4 given Z in either period,

13, and 14, the spillover effect on never takers and the total effect on compliers is identified:

$$\begin{aligned}\theta_{Z=1,n}(0) &= E_{W_1|Z=1,\mathcal{T}=n}[E[Y_1|W_1, Z=1, \mathcal{T}=n] \\ &\quad - E[Y_0|W_0, Z=1, \mathcal{T}=n] \\ &\quad - [E[Y_1|W_1, Z=0, \mathcal{T}=n] \\ &\quad - E[Y_0|W_0, Z=0, \mathcal{T}=n]]], \\ \Delta_{Z=1,c} &= E_{W_1|Z=1,\mathcal{T}=c}[E[Y_1|W_1, Z=1, \mathcal{T}=c] \\ &\quad - E[Y_0|W_0, Z=1, \mathcal{T}=c] \\ &\quad - [E[Y_1|W_1, Z=0, \mathcal{T}=c] \\ &\quad - E[Y_0|W_0, Z=0.\mathcal{T}=c]]], \end{aligned} \tag{18}$$

where W_1, W_0 denote the observed covariates in periods 0 and 1, respectively.

4. Application

To illustrate the proposed DiD approach in an empirical application, we reconsider the setting studied in Lalive, Landais, and Zweimüller (2015), henceforth LLZ. They study the effect of a large-scale extension of the duration of unemployment benefit eligibility on the job search behavior of eligible and ineligible workers in the same labor market. LLZ define as the *micro effect* changes in the search strategy of unemployed workers induced by changes in unemployment insurance generosity. They refer to *market externalities* as changes in equilibrium labor market conditions induced by changes in unemployment insurance policies. These effects correspond to our individual treatment and spillover effects.

The Regional Extension Benefit Program (REBP) in Austria extended unemployment benefit eligibility from one to four years for a large subset of workers in selected regions of Austria from June 1988 until August 1993. Individuals were eligible if they were above 50 years old at the start of their spell and had more than 15 years of continuous work history in the past 25 years. In line with our conceptual framework, we refer to individuals fulfilling the individual criteria as *compliers* ($\mathcal{T}=c$) and to those not fulfilling the criteria as *never takers* ($\mathcal{T}=n$). To identify spillover effects, LLZ restrict their sample to workers aged 46–54 as those are most likely to compete with eligible unemployed for the same vacancies. They further restrict the sample to men who never worked in the steel sector. We impose the same restrictions. The Austrian government enacted REBP in 28 labor market districts. To be eligible for the extended benefits, newly unemployed workers had to reside in of these districts for at least 6 months prior to the claim in addition to meeting the criteria described above. A reform in 1991 abolished REBP in 6 of the 28 regions and tightened eligibility criteria requiring that beneficiaries not only resided in but also had to be employed in REBP regions. We follow LLZ and exclude the regions where REBP was phased out early. Regional treatment (Z) is 1 for counties covered by REBP and 0 otherwise. D equals one if $Z=1$ and $\mathcal{T}=c$ and zero otherwise.

In the notation of the current article, LLZ estimate the total effect on compliers ($\Delta_{c,Z=1}$) and the spillover effect on never takers ($\theta_{n,Z=1}(0)$) in treated regions using a DiD identification strategy. We also aim for these parameters based on the identification result presented in Equation (13). In addition, we estimate the ATET, that is, the total effect in treated regions $\Delta_{Z=1}$. This effect captures the total increase in unemployment duration in the treated regions and is the relevant parameter for assessing the fiscal effects of the policy. We deviate from the analysis of LLZ in three dimensions. First, we investigate only the period 1988–1993, whereas LLZ in addition consider 1994–1997 when no new entries into the program occurred, but market externalities might have still persisted. Second, we only use the pretreatment period 1980–1987 as the untreated reference period while LLZ exploit both years before and after the program as untreated reference years. Third, we do not control for covariates in our baseline specification.

We make several advances relative to LLZ. First, our framework allows a refined interpretation of the effects in LLZ. Second, we estimate the total treatment effect in treated regions that is the relevant parameter to assess the fiscal consequences of the policy. Third, we discuss and show identification under different sets of common trend assumptions. Fourth, we discuss potential violations of regional SUTVA not considered by LLZ. In what follows, we assess the plausibility of the assumptions outlined in Section 3.4.

4.1. Assessment of Assumptions

4.1.1. Assumption 1: Regional SUTVA

Regional SUTVA rules out spillover effects across treated and nontreated labor markets. To ensure that control regions are not affected by spillovers from REBP, LLZ exclude non-REBP counties with more than 5% of new hires coming from REBP regions from the analysis. But LLZ keep REBP regions with a strong integration with non-REBP regions. In the median REBP county, 35% of new hires come from non-REBP regions. However, regional SUTVA implies that spillover effects are not directional in the sense that they can only occur in control regions integrated with treated regions. Integration of REBP regions with control regions also constitutes a violation of regional SUTVA. We would expect that spillover effects are biased toward zero in REBP regions that are integrated with non-REBP regions. Reduced labor supply of compliers is potentially offset by the inflow of workers from non-REBP regions. Thus, our estimated spillover effects might be biased toward zero in regions integrated with non-REBP regions.

A potential solution is to drop REBP regions that are highly integrated with non-REBP regions. However, due to a high integration, applying the rule to drop REBP counties with more than 5% of new hires coming from non-REBP would drop 96% of observations ins REBP counties from the sample. We thus, restrain from any further sample restrictions.

4.1.2. Assumption 3: Deterministic Individual Treatment Assignment

Individuals who fulfill certain observed criteria at the beginning of the unemployment spell are eligible for the extended benefits. This rule satisfies deterministic individual treatment assignment. However, the inflow into unemployment may not be exogenous to regional treatment status and meeting the individual eligibility criteria. Firms may lay off workers who are eligible for the extended benefits with higher probability.

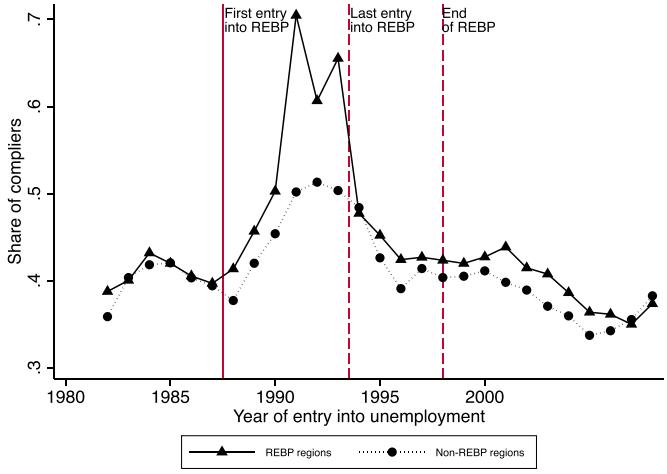

Figure 1. Share of compliers by REBP status of the county and year of entry into unemployment. NOTES: The figure depicts the share of compliers by REBP status of the county and year of entering unemployment. We refer to individuals fulfilling the individual criteria as *compliers* and to those not fulfilling the criteria as *never takers*. The sample includes all men between 46 and 54 years who became unemployed in a given year. Non-REBP counties with high labor market integration to REBP counties are excluded from the sample.

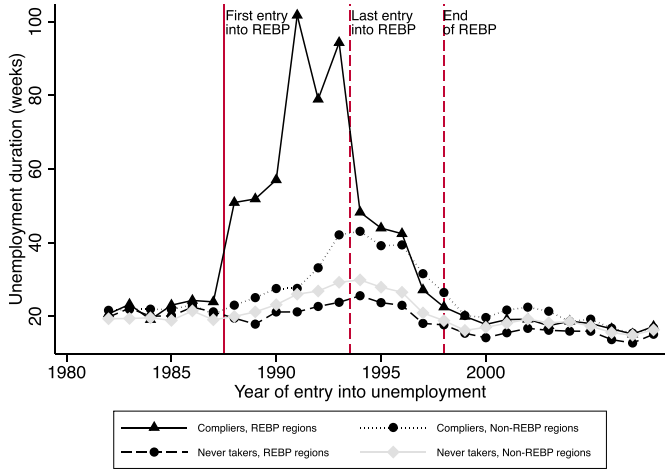

Figure 2. Unemployment duration by eligibility and REBP status of the county by year of entry into unemployment. NOTES: The figure depicts the average unemployment duration in weeks for four distinct groups by year of entering unemployment. The first group are individuals in REBP counties who fulfill the REBP eligibility criteria of being above 50 and having more than 15 years of work history in the past 25 years prior to becoming unemployed. The second group are individuals who would fulfill the eligibility criteria but do not live in REBP counties. The third group are individuals in REBP counties who do not fulfill the eligibility criteria (less than 50 and/or less than 15 years of continuous work history). The fourth group are individuals who do not live in REBP counties and do not fulfill the eligibility criteria. We refer to individuals fulfilling the individual criteria as *compliers* and to those not fulfilling the criteria as *never takers*. The sample includes all men between 46 and 54 years who became unemployed in a given year. Non-REBP counties with high labor market integration to REBP counties are excluded from the sample.

Characteristics of these marginal workers may differ from those of workers who would have been laid off also in the absence of REBP.

Figure 1 shows the share of compliers in REBP and non-REBP regions among 46 to 54-year-old entrants into unemployment. In the pre-REBP period, this share was about 40% in both types of regions. During the REBP period, the share of compliers increased above 50% in non-REBP regions but to more than 70% in REBP regions. In the post-REBP period, the share of compliers decreased to roughly 40% in both types of regions but remains somewhat higher in REBP regions. Overall, these patterns point to a substantial effect on unemployment inflow as a consequence of REBP (see Winter-Ebmer 2003).

These endogenous separations have several consequences. First, the level of the regional treatment intensity Z is higher as labor supply is further reduced. Second, the characteristics of the pool of compliers might change. LLZ show that log wage and tenure was slightly higher for compliers who entered during REBP period. There was no such difference for never takers. The size of the spillover effect is influenced by both, the level of the regional treatment as well as the characteristics, and related, the potential outcomes of the compliers. Thus, endogenous layoffs change the nature of the regional treatment Z. LLZ discuss potential consequences of endogenous layoffs in their Appendix A.4.

4.1.3. Assumption 13 and 13′: Common Trends

Figure 2 shows the average unemployment duration of compliers and never takers by residence in REBP and non-REBP counties by year of entry into unemployment. Unemployment duration was relatively stable and similar for all four groups between 1981 and 1986, suggesting that common trends across regions, see Assumption 13′ (henceforth A13′), might hold. However, during the REBP period, the unemployment duration of compliers in non-REBP regions rose sharply while the same was not true for never takers in non-REBP regions. The analysis in Table 2 confirms that the assumption of a common trend across never takers and compliers in non-REBP regions as implied by A13′ can be rejected. Thus, identification under the weaker Assumption 13 (henceforth A13), which invokes common trend within rather than across types, appears more credible. LLZ mix the usage of A13 and A13′ but do not discuss the implications of either assumption. The approach graphically showing yearly effects introduced in their Equation (2) is implicitly based on A13 while the approach showing average affects introduced in Equation (3) is based on A13′.

4.1.4. Assumption 14: No Anticipation Effects

Assumption 14 rules out average effects on the potential outcomes within compliance types in the baseline period in anticipation of the individual or regional treatment. Such anticipation effects may, for example, occur if individuals who became unemployed between the announcement and enactment of REBP changed their search efforts. The flat and equal pretreatment trends in Figure 2 do not point to anticipation effects.

4.2. Results

Table 2 presents the results for the total effect $\Delta_{Z=1}$, the total effect on compliers ($\Delta_{c,Z=1}$), and the spillover effect on never takers ($\theta_{n,Z=1}(0)$) in treated regions. All results are given separately under the two common trend assumptions A13 and A13′.

Panel 1 of Table 2 presents the estimated average effects for the period 1988–1993. The first two columns provide the total effect in treated regions. Columns 3 and 4 show the total effect on compliers in treated regions. Columns 5 and 6 provide the spillover effect on never takers in treated regions. The last

Table 2. Total effect on compliers and spillover effect on never takers.

	$\Delta_{Z=1}$		$\Delta_{c,Z=1}$		$\theta_{n,Z=1}(0)$		Test A13′
	A13′	A13	A13′	A13	A13′	A13	
Panel 1: Average effect 1988–1993							
Average effect	29.43***	28.95***	53.99***	51.50***	−5.80**	−3.39	4.90***
	(5.32)	(5.31)	(6.35)	(6.66)	(2.61)	(2.33)	(1.52)
Panel 2: Average effect 1988–1993—conditional on observables							
Average effect	27.88***	27.06***	51.74***	48.43***	−6.35***	−3.59*	5.79***
	(5.11)	(5.03)	(6.16)	(6.47)	(2.45)	(2.14)	(1.54)
Panel 3: By year of entering unemployment							
1988	8.84	8.75	25.67**	24.60**	−3.03	−2.42	1.69
	(7.25)	(7.27)	(12.00)	(12.10)	(4.97)	(4.96)	(1.04)
1989	8.05	7.99	25.07***	23.53***	−6.25	−5.08	2.71
	(6.21)	(6.28)	(8.12)	(8.65)	(5.60)	(5.26)	(1.71)
1990	11.41	11.34	27.92***	26.25***	−5.27	−3.73	3.21*
	(7.58)	(7.68)	(9.81)	(10.29)	(6.43)	(6.02)	(1.68)
1991	49.02***	48.67***	72.02***	70.94***	−5.88	−4.48	2.47
	(13.67)	(13.77)	(15.34)	(15.70)	(6.93)	(6.52)	(1.72)
1992	23.75***	23.45***	44.84***	42.52***	−8.75	−5.96	5.11**
	(8.45)	(8.57)	(10.43)	(10.92)	(6.41)	(6.00)	(2.09)
1993	43.93***	39.06***	63.20***	53.13***	−14.74**	−3.75	21.06**
	(10.88)	(11.24)	(11.27)	(13.53)	(6.93)	(6.73)	(9.11)
Observations	102,299	102,299	87,464	45,923	87,039	56,376	72,204

NOTES: Results are presented for the main REBP period 1988–1993 and the 1980–1987 pretreatment period. Panel 1 presents nonparametric estimates of the average effects over the period 1988–1993. Panel 2 presents semi-parametric estimates of the average effects over the period 1988–1993 conditional on education, family status, and tenure. Panel 3 presents nonparametric estimates separately by year of entering unemployment. Standard errors stem from a clustered bootstrap at the region × year level with 499 replications. The estimation samples includes male workers between 46 and 54 years that were not employed in the steel sector. All duration outcomes are expressed in weeks.

*Statistical significance at the 10% level.

**Statistical significance at the 5% level.

***Statistical significance at the 1% level.

column is on testing A13′ as outlined in Section 3.4. Under A13′, the unemployment duration of compliers and never takers in non-REBP counties has the same trend.

The average total effect on compliers in treated regions amounts to 53.99 weeks increased unemployment duration under A13′ and 51.5 under A13. The average spillover effect on never takers in treated regions corresponds to −5.8 under A13′ and −3.39 weeks under A13. The latter estimate is not statistically significant at conventional levels but of the same magnitude as the effect reported in LLZ under a somewhat different evaluation approach. Based on these estimates and the share of compliers in the treated regions in the treatment period (58.92%), the total effect of REBP in treated regions amounts to 29.43 under A13′ (column 1) and 28.95 under A13 (column 2). Column 7 shows that the unemployment duration of compliers increases significantly compared to the unemployment duration of never takers in non-REBP regions. Thus, A13′ is unlikely to hold and results based on A13 should be considered more credible.

Panel 2 presents estimates for the same effects as in Panel 1, however, based on conditional DiD as discussed at the end of Section 3.4 by controlling for the covariates education, family status, and tenure. We apply semiparametric inverse probability weighting, see Abadie (2005), using probit specifications for the conditional probabilities. Observations are reweighted to match the covariate distribution of the respective target population, that is, compliers in columns 3 and 4 and never takers in columns 5 and 6, in REBP counties in the treatment period. In column 7, the target population are unemployed meeting eligibility criteria in non-REBP counties in the treatment period. All results are similar to those in Panel 1. Panel 3 of Table 2 presents the estimates by year of entering unemployment. The total effect on compliers is largest in 1991 while the spillover effect on never takers is largest in 1992 and 1993.

5. Conclusion

Most contributions in the field of treatment evaluation rule out spillover effects related to individual treatment assignment. This can be formalized by the SUTVA, which assumes away any form of treatment-dependent spillover between study participants. However, such spillovers likely occur in many empirical problems as for instance the assessment of labor market, development, or educational interventions. This article suggests a general framework for separating individual level treatment effects and spillover effects under the assumption that SUTVA holds on an aggregate rather than individual level, for example, across regions rather than individuals. We use the framework to systematically categorize the individual-level and spillover effects considered in the previous literature. We discuss identifying assumptions under different designs, for instance, based on randomization or selection on observables, and also propose a novel DiD approach.

As an empirical illustration, we reconsider data from Lalive, Landais, and Zweimüller (2015) who studied the spillover effects of a large-scale extension of unemployment benefits in selected regions of Austria and find that this policy decreased the

job-search duration of ineligible individuals in treated regions. Our framework provides a sharper definition of the identified effects. Furthermore, we apply our DiD methodology to estimate the total effect, the total effects on eligibles, and the spillover effects on ineligibles in treated regions under somewhat weaker common trend assumptions than underlying some of the results in Lalive, Landais, and Zweimüller (2015). Even though the stronger common trend assumption is rejected by the data, both approaches qualitatively point into the same direction of a strong positive effect on the job-search duration of eligibles and a negative spillover effect on ineligibles.

A nontrivial question beyond the scope of this article is how the boundaries of aggregate units should be defined. Regional SUTVA is only plausible if the aggregate units coincide with relevant markets. Predefined regional or administrative entities formed for example by industry, education, or age as used in most empirical studies might only crudely approximate relevant markets. Nimczik (2018) provided a data-driven method to define labor markets and shows that traditional definitions perform quite poorly in separating distinct labor markets. Future research should therefore, make use of the increasing availability of microdata and advances in econometric modeling to implement data-driven approaches for defining relevant markets.

Appendix A: Proofs

A.1. Proof of Equation (8)

Under Assumptions 2, 5, and 8,

$$\begin{aligned}
&\theta_{Z=1}(D(z))\\
&\quad=\theta(D(z))=E_{D(z)}[E[Y(1,d)-Y(0,d)|D(z)=d]]\\
&\quad=E_{D|Z=z}[E[Y|Z=1,D]-E[Y|Z=0,D]]\\
&\quad=E_{D|Z=z}\left[\frac{E[Y\cdot Z|D]}{\Pr(Z=1|D)}-\frac{E[Y\cdot(1-Z)|D]}{1-\Pr(Z=1|D)}\right]\\
&\quad=E_{D|Z=z}\left[\frac{E[Y\cdot Z|D]}{\Pr(Z=1|D)}-\frac{E[Y\cdot(1-Z)|D]}{1-\Pr(Z=1|D)}\right]\\
&\quad=E\left[\left(\frac{Y\cdot Z}{\Pr(Z=1|D)}-\frac{Y\cdot(1-Z)}{1-\Pr(Z=1|D)}\right)\cdot\frac{\Pr(Z=z|D)}{\Pr(Z=z)}\right].
\end{aligned}$$

The second equality follows from Assumptions 2 and 5, the fourth from Bayes' theorem and the fifth from the law of iterated expectations. Also note that $E_{A|B}[C]$ denotes the expectation of C over A conditional on B.

A.2. Proof of Equations (9)–(11)

Under Assumptions 9, 10, and 11(b),

$$\begin{aligned}
&\delta_{Z=z,D=1}(z)\\
&\quad=E[Y(z,1)-Y(z,0)|Z=z,D=1]\\
&\quad=E_{X|Z=z,D=1}[E[Y(z,1)-Y(z,0)|Z=z,D=1,X]]\\
&\quad=E_{X|Z=z,D=1}[E[Y|Z=z,D=1,X]-E[Y|Z=z,D=0,X]],\\
&\quad=E_{X|Z=z,D=1}\left[\frac{E[Y\cdot D|Z=z,X]}{\Pr(D=1|Z=z,X)}\right.\\
&\qquad\left.-\frac{E[Y\cdot(1-D)|Z=z,X]}{1-\Pr(D=1|Z=z,X)}\right]\\
&\quad=E\left[\left(\frac{Y\cdot D}{\Pr(D=1|Z=z,X)}-\frac{Y\cdot(1-D)}{1-\Pr(D=1|Z=z,X)}\right)\right.\\
&\qquad\left.\cdot\frac{\Pr(Z=z|X)\cdot\Pr(D=1|Z=z,X)}{\Pr(Z=z)\cdot\Pr(D=1|Z=z)}\right],
\end{aligned}$$

where the second equality follows from the law of iterated expectations, the third from Assumptions 9 and 10, the fourth from probability theory, and the fifth from the law of iterated expectations and Bayes' theorem.

Under Assumptions 9, 10, and 11(c),

$$\begin{aligned}
&\theta_{Z=z,D=d}(d)\\
&\quad=E[Y(1,d)-Y(0,d)|Z=z,D=1]\\
&\quad=E_{X|Z=z,D=d}[E[Y(1,d)-Y(0,d)|Z=z,D=1,X]]\\
&\quad=E_{X|Z=z,D=d}[E[Y|Z=1,D=d,X]-E[Y|Z=0,D=d,X]]\\
&\quad=E_{X|Z=z,D=d}\left[\frac{E[Y\cdot Z|D=d,X]}{\Pr(Z=1|D=d,X)}-\frac{E[Y\cdot(1-Z)|D=d,X]}{1-\Pr(Z=1|D=d,X)}\right]\\
&\quad=E\left[\left(\frac{Y\cdot Z}{\Pr(Z=1|D=d,X)}-\frac{Y\cdot(1-Z)}{1-\Pr(Z=1|D=d,X)}\right)\right.\\
&\qquad\left.\cdot\frac{\Pr(Z=z|X)\cdot\Pr(D=d|Z=z,X)}{\Pr(D=d)\cdot\Pr(Z=z|D=d)}\right],
\end{aligned}$$

where the second equality follows from the law of iterated expectations, the third from Assumptions 9 and 10, the fourth from probability theory, and the fifth from the law of iterated expectations and Bayes' theorem.

Under Assumptions 9, 10, and 11(d),

$$\begin{aligned}
&\theta_{Z=1}(D(z))\\
&\quad=E_{X|Z=1}[E_{D(z)|Z=1,X}[E[Y(1,d)-Y(0,d)|D(z)=d,X]]]\\
&\quad=E_{X|Z=1}\left[E_{D|Z=z,X}\left[E[Y|Z=1,D,X]-E[Y|Z=0,D,X]\right]\right]\\
&\quad=E_X\left[E_{D|X}\left[\left(\frac{E[Y\cdot Z|D,X]}{\Pr(Z=1|D,X)}-\frac{E[Y\cdot(1-Z)|D,X]}{1-\Pr(Z=1|D,X)}\right)\right.\right.\\
&\qquad\left.\left.\cdot\frac{\Pr(Z=z|D,X)}{\Pr(Z=z|X)}\right]\cdot\frac{\Pr(Z=1|X)}{\Pr(Z=1)}\right]\\
&\quad=E\left[\left(\frac{Y\cdot Z}{\Pr(Z=1|D,X)}-\frac{Y\cdot(1-Z)}{1-\Pr(Z=1|D,X)}\right)\right.\\
&\qquad\left.\cdot\frac{\Pr(Z=z|D,X)}{\Pr(Z=z|X)}\cdot\frac{\Pr(Z=1|X)}{\Pr(Z=1)}\right],
\end{aligned}$$

where the first equality follows from the law of iterated expectations, the second from Assumptions 9 and 10, the third from probability theory, and the fourth from the law of iterated expectations and Bayes' theorem.

A.3. Proof of Equations (13) and (17)

Under Assumptions 3, 13, and 14,

$$\begin{aligned}
&E[Y_1|Z=1,\mathcal{T}=n]-E[Y_0|Z=1,\mathcal{T}=n]\\
&\quad-[E[Y_1|Z=0,\mathcal{T}=n]-E[Y_0|Z=0,\mathcal{T}=n]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_0(1,0)|Z=1,\mathcal{T}=n]\\
&\quad-[E[Y_1(0,0)|Z=0,\mathcal{T}=n]-E[Y_0(0,0)|Z=0,\mathcal{T}=n]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_0(0,0)|Z=1,\mathcal{T}=n]\\
&\quad-[E[Y_1(0,0)|Z=0,\mathcal{T}=n]-E[Y_0(0,0)|Z=0,\mathcal{T}=n]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_0(0,0)|Z=1,\mathcal{T}=n]\\
&\quad-[E[Y_1(0,0)|Z=1,\mathcal{T}=n]-E[Y_0(0,0)|Z=1,\mathcal{T}=n]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_1(0,0)|Z=1,\mathcal{T}=n]\\
&=\theta_{n,Z=1}(0),
\end{aligned}$$

where the first equality follows from the fact that never takers are identified by Assumption 3 and the observational rule ($Y|Z=z,\mathcal{T}=n$ corresponds to $Y(z,0)|\mathcal{T}=n$), the second from Assumption 14 (such that $E[Y_0(1,0)|Z=1,\mathcal{T}=n]=E[Y_0(0,0)|Z=1,\mathcal{T}=n]$),

and the third from Assumption 13. The proof for the total effect on the compliers in treated regions ($\Delta_{c,Z=1}$) is analogous and therefore, omitted.

Under Assumptions 3, 13′, and 14,

$$\begin{aligned}
&E[Y_1|Z=1,\mathcal{T}=n]-E[Y_0|Z=1,\mathcal{T}=n]\\
&-[E[Y_1|Z=0]-E[Y_0|Z=0]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_0(1,0)|Z=1,\mathcal{T}=n]\\
&-[E[Y_1(0,0)|Z=0]-E[Y_0(0,0)|Z=0]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_0(0,0)|Z=1,\mathcal{T}=n]\\
&-[E[Y_1(0,0)|Z=0]-E[Y_0(0,0)|Z=0]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_0(0,0)|Z=1,\mathcal{T}=n]\\
&-[E[Y_1(0,0)|Z=1,\mathcal{T}=n]-E[Y_0(0,0)|Z=1,\mathcal{T}=n]]\\
&=E[Y_1(1,0)|Z=1,\mathcal{T}=n]-E[Y_1(0,0)|Z=1,\mathcal{T}=n]\\
&=\theta_{n,Z=1}(0),
\end{aligned}$$

where the first equality follows from the fact that never takers are identified by Assumption 3 and the observational rule ($Y|Z=z,\mathcal{T}=n$ corresponds to $Y(z,0)|\mathcal{T}=n$), the second from Assumption 14 (such that $E[Y_0(1,0)|Z=1,\mathcal{T}=n]=E[Y_0(0,0)|Z=1,\mathcal{T}=n]$), and the third from Assumption 13′. The proof for the total effect on the compliers in treated regions ($\Delta_{c,Z=1}$) is analogous and therefore, omitted.

Under Assumptions 3, 13, 14, and 16,

$$\begin{aligned}
&E[Y_1|Z=2,\mathcal{T}=c]-E[Y_0|Z=2,\mathcal{T}=c]\\
&-[E[Y_1|Z=1,\mathcal{T}=c]-E[Y_0|Z=1,\mathcal{T}=c]]\\
&=E[Y_1(2,1)|Z=2,\mathcal{T}=c]-E[Y_0(2,1)|Z=2,\mathcal{T}=c]\\
&-[E[Y_1(1,1)|Z=1,\mathcal{T}=c]-E[Y_0(1,1)|Z=1,\mathcal{T}=c]]\\
&=E[Y_1(2,1)|Z=2,\mathcal{T}=c]-E[Y_0(0,0)|Z=2,\mathcal{T}=c]\\
&-[E[Y_1(1,1)|Z=1,\mathcal{T}=c]-E[Y_0(0,0)|Z=1,\mathcal{T}=c]]\\
&=E[Y_1(2,1)|Z=2,\mathcal{T}=c]-E[Y_0(0,0)|Z=2,\mathcal{T}=c]\\
&-[E[Y_1(1,1)|Z=1,\mathcal{T}=c]-E[Y_1(0,0)|Z=1,\mathcal{T}=c]\\
&+E[Y_1(0,0)|Z=1,\mathcal{T}=c]-E[Y_0(0,0)|Z=1,\mathcal{T}=c]]\\
&=E[Y_1(2,1)|Z=2,\mathcal{T}=c]-E[Y_0(0,0)|Z=2,\mathcal{T}=c]\\
&-[E[Y_1(1,1)|Z=2,\mathcal{T}=c]-E[Y_1(0,0)|Z=2,\mathcal{T}=c]\\
&+E[Y_1(0,0)|Z=2,\mathcal{T}=c]-E[Y_0(0,0)|Z=2,\mathcal{T}=c]]\\
&=E[Y_1(2,1)|Z=2,\mathcal{T}=c]-E[Y_1(1,1)|Z=1,\mathcal{T}=c]\\
&=\theta_{c,Z=2}(z'=2,z=1,d=1),
\end{aligned}$$

where the first equality follows from the fact that compliers are identified by Assumption 3 and the observational rule ($Y|Z=z,\mathcal{T}=c$ corresponds to $Y(z,1)|\mathcal{T}=c$ for $z>0$), the second from Assumption 14 (such that $E[Y_0(1,0)|Z=2,\mathcal{T}=c]=E[Y_0(0,0)|Z=2,\mathcal{T}=c]$), the third from subtracting and adding $E[Y_1(0,0)|Z=1,\mathcal{T}=c]$, and the fourth from Assumptions 13 and 16.

Acknowledgments

We thank the editor Alfonso Flores-Lagunes and three anonymous referees for valuable suggestions. We have benefited from comments by seminar participants in Konstanz, as well as conference participants at CRC Ohlstadt 2018, Challenges in Evaluating Regional and Urban Policy 2018, European Association of Labor Economists Annual Meetings 2017, LISER Workshop on Causal Inference, Program Evaluation, and External Validity 2017, and the 2017 Annual Congress of the Swiss Society of Economics and Statistics. We are very grateful to Josef Zweimüller for his support and advice concerning the empirical application. Joachim Winter provided helpful comments on the draft. Pavel Obraztcov provided excellent research assistance.

Funding

Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged.

References

Abadie, A. (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263. [8]

——— (2005), "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72, 1–19. [12]

Angelucci, M., and Giorgi, G. D. (2009), "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?," *American Economic Review*, 99, 486–508. [2,4,5]

Angelucci, M., and Maro, V. D. (2016), "Programme Evaluation and Spillover Effects," *Journal of Development Effectiveness*, 8, 22–43. [2]

Angelucci, M., Prina, S., Royer, H., and Samek, A. (2018), "When Incentives Backfire: Spillover Effects in Food Choice," *American Economic Journal: Economic Policy* (forthcoming). [2]

Angrist, J., Imbens, G., and Rubin, D. (1996), "Identification of Causal Effects Using Instrumental Variables" (with discussion), *Journal of American Statistical Association*, 91, 444–472. [3]

Baird, S., Bohren, A., McIntosh, C., and Ozler, B. (2014), "Designing Experiments to Measure Spillover Effects," PIER Working Paper Archive 14-032, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania. [2,3,4,5]

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013), "Do Labor Market Policies Have Displacement Effects? Evidence From a Clustered Randomized Experiment," *Quarterly Journal of Economics*, 128, 531–580. [2,4,5]

Dahl, G. B., Løken, K. V., and Mogstad, M. (2014), "Peer Effects in Program Participation," *American Economic Review*, 104, 2049–2074. [3]

Deaton, A., and Cartwright, N. (2016), "Understanding and Misunderstanding Randomized Controlled Trials," NBER Working Paper 22595. [1]

Deuchert, E., Huber, M., and Schelker, M. (2018), "Direct and Indirect Effects Based on Difference-in-differences With an Application to Political Preferences Following the Vietnam Draft Lottery," *Journal of Business and Economic Statistics* (forthcoming). [8]

Ferracci, M., Jolivet, G., and van den Berg, G. J. (2014), "Evidence of Treatment Spillovers Within Markets," *The Review of Economics and Statistics*, 96, 812–823. [2,4,5]

Flores, C. A., and Flores-Lagunes, A. (2009), "Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment Under Unconfoundedness," IZA Dicussion Paper 4237. [7]

Forastiere, L., Mealli, F., and VanderWeele, T. J. (2016), "Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification," *Journal of the American Statistical Association*, 111, 510–525. [2,6]

Frangakis, C., and Rubin, D. (2002), "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29. [2,3]

Frölich, M., and Michaelowa, K. (2011), "Peer Effects and Textbooks in African Primary Education," *Labour Economics*, 18, 474–486. [3]

Graham, B. (2008), "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica*, 76, 643–660. [2]

Heckman, J., Lochner, L., and Taber, C. (1998), "General Equilibrium Treatment Effects: A Study of Tution Policy," NBER Working Paper 6426. [1]

Hirano, K., Imbens, G. W., and Ridder, G. (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189. [8]

Hong, G. (2010), "Ratio of Mediator Probability Weighting for Estimating Natural Direct and Indirect Effects," in *Proceedings of the American Statistical Association, Biometrics Section*, Alexandria, VA: American Statistical Association, pp. 2401–2415. [2,7]

Hong, G., and Raudenbush, S. W. (2006), "Evaluating Kindergarten Retention Policy," *Journal of the American Statistical Association*, 101, 901–910. [1,3]

Huber, M. (2014), "Identifying Causal Mechanisms (Primarily) Based on Inverse Probability Weighting," *Journal of Applied Econometrics*, 29, 920–943. [2,7]

Hudgens, M. G., and Halloran, M. E. (2008) "Toward Causal Inference With Interference," *Journal of the American Statistical Association*, 103, 832–842. [2]

Imai, K., Keele, L., and Yamamoto, T. (2010), "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25, 51–71. [2,7,8]

Imai, K., Tingley, D., and Yamamoto, T. (2013), "Experimental Designs for Identifying Causal Mechanisms," *Journal of the Royal Statistical Society*, Series A, 176, 5–51. [7]

Imbens, G. W. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics*, 86, 4–29. [7]

Lalive, R., and Cattaneo, M. A. (2009), "Social Interactions and Schooling Decisions," *Review of Economics and Statistics*, 91, 457–477. [4,5]

Lalive, R., Landais, C., and Zweimüller, J. (2015), "Market Externalities of Large Unemployment Insurance Extension Programs," *American Economic Review*, 105, 3564–3596. [1,2,4,5,6,8,10,12,13]

Lechner, M. (2009), "Sequential Causal Models for the Evaluation of Labor Market Programs," *Journal of Business and Economic Statistics*, 27, 71–83. [2]

Lechner, M., and Miquel, R. (2010), "Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions," *Empirical Economics*, 39, 111–137. [2]

Manski, C. F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542. [2]

Miguel, E., and Kremer, M. (2004), "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72, 159–217. [3]

Moffitt, R. (2001), "Policy Interventions, Low-Level Equilibria and Social Interactions," in *Social Dynamics*, eds. S. Durlauf and H. Young, Cambridge, MA: MIT Press. [2,3]

Nimczik, J. (2018), "Job Mobility Networks and Endogenous Labor Markets," Mimeo. [13]

Pearl, J. (2001), "Direct and Indirect Effects," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 411–420, San Francisco: Morgan Kaufman. [2,5,7]

Petersen, M. L., Sinisi, S. E., and van der Laan, M. J. (2006), "Estimation of Direct Causal Effects," *Epidemiology*, 17, 276–284. [2]

Robins, J. M. (1986), "A New Approach to Causal Inference in Mortality Studies With Sustained Exposure Periods—Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, 7, 1393–1512. [2]

——— (1989), "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in *Health Service Research Methodology: A Focus on AIDS*, eds. L. Sechrest, H. Freeman, and A. Mulley, Washington, DC: U.S. Public Health Service, pp. 113–159. [2]

——— (2003), "Semantics of Causal DAG Models and the Identification of Direct and Indirect Effects," in *Highly Structured Stochastic Systems*, eds. P. Green, N. Hjort, and S. Richardson, Oxford: Oxford University Press, pp. 70–81. [2,3]

Robins, J. M., and Greenland, S. (1992), "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143–155. [2]

Robins, J. M., Hernan, M. A., and Brumback, B. (2000), "Marginal Structural Models and Causal Inference in Epidemiology," *Epidemiology*, 11, 550–560. [2]

Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701. [3]

——— (1990), "Formal Mode of Statistical Inference for Causal Effects," *Journal of Statistical Planning and Inference*, 25, 279–292. [1]

——— (2004), "Direct and Indirect Causal Effects via Potential Outcomes," *Scandinavian Journal of Statistics*, 31, 161–170. [2]

Sobel, M. E. (2006), "What Do Randomized Studies of Housing Mobility Demonstrate?," *Journal of the American Statistical Association*, 101, 1398–1407. [1]

Tchetgen Tchetgen, E. J., and Shpitser, I. (2012), "Semiparametric Theory for Causal Mediation Analysis: Efficiency Bounds, Multiple Robustness, and Sensitivity Analysis," *The Annals of Statistics*, 40, 1816–1845. [7]

VanderWeele, T. J. (2008), "Simple Relations Between Principal Stratification and Direct and Indirect Effects," *Statistics & Probability Letters*, 78, 2957–2962. [2]

——— (2009), "Marginal Structural Models for the Estimation of Direct and Indirect Effects," *Epidemiology*, 20, 18–26. [2]

——— (2010), "Direct and Indirect Effects for Neighborhood-Based Clustered and Longitudinal Data," *Sociological Methods & Research*, 38, 515–544. [6,7]

——— (2012), "Comments: Should Principal Stratification Be Used to Study Mediational Processes?," *Journal of Research on Educational Effectiveness*, 5, 245–249. [2]

VanderWeele, T. J., Hong, G., Jones, S. M., and Brown, J. L. (2013), "Mediation and Spillover Effects in Group-Randomized Trials: A Case Study of the 4Rs Educational Intervention," *Journal of the American Statistical Association*, 108, 469–482. [7]

Vansteelandt, S., Bekaert, M., and Lange, T. (2012), "Imputation Strategies for the Estimation of Natural Direct and Indirect Effects," *Epidemiologic Methods*, 1, 129–158. [7]

Winter-Ebmer, R. (2003), "Benefit Duration and Unemployment Entry: A Quasi-Experiment in Austria," *European Economic Review*, 47, 259–273. [11]