Identification and Estimation of Treatment Effects in the Presence of Neighbourhood Interactions

Giovanni Cerulli
CNR - IRCrES
National Research Council of Italy
Research Institute on Sustainable Economic Growth
Via dei Taurini 19, 00185, Roma, Italy
E-mail: giovanni.cerulli@ircres.cnr.it
Tel.: +39.06.4993.7867

Structured Abstract

This paper presents a counter-factual model identifying average treatment effects (ATEs) by conditional mean independence (CMI) when externality (or neighbourhood) effects are considered. The paper aims at providing a consistent estimation of ATEs when the SUTVA (Stable Unit Treatment Value Assumption) is relaxed. Under some restrictions on SUTVA relaxation (i.e., the outcome of treated units can affect the outcome of the untreated ones, but the other way round is excluded), this model develops within a linear regression-adjustment where ATEs is identified by conditional mean independence. The paper sets out two instructional applications: the first is applied on real data, and aims at measuring the effect of education on fertility in the presence of social interactions; The second, reported in appendix, is a simulation exercise using the software package Stata. Results are compared with a no-interaction setting.

Keywords: ATEs, Rubin's causal model, SUTVA, neighbourhood effects, Stata command.

JEL classification: C21, C31, C87

June 10, 2015

Note: previous versions of this paper contained a mistake. Please consider this one as the final correct and updated version.

1. Introduction

In observational program evaluation studies, aimed at estimating the effect of an intervention on the outcome of a set of targeted individuals, it is generally assumed that *"the treatment received by one unit does not affect other units' outcome"* (Cox, 1958). Along with other fundamental assumptions - such as, for instance, the conditional independence assumption, the exclusion restriction provided by instrumental-variables estimation, or the existence of a "forcing" variable in regression discontinuity design - the no-interference assumption is additionally invoked in order to consistently estimating average treatment effects (ATEs). It is thus clear that, if interference (or interaction) among units is not properly taken into account, traditional program evaluation methods such as regression adjustment, selection models, matching or reweighting are bound to be biased estimations of the actual treatment effect[1].

Rubin (1978) calls this important assumption as Stable-Unit-Treatment-Value-Assumption (SUTVA), whereas Manski (2013) refers to Individualistic-Treatment-Response (ITR) to emphasize that this poses a restriction in the form of the treatment response function that the analyst considers. SUTVA (or ITR) implies that the treatment applied to a specific individual affects only the outcome of that individual, so that potential "externality effects" flowing from treated to untreated subjects are strictly ruled out.

This paper is an attempt to "partially" relax this assumption; it operationalizes the estimation of ATEs when peer effects are assumed to flow from treated to untreated units, by however excluding the other way round. Such restriction, reasonable in specific contexts, allows for a straightforward identification and estimation of treatment effects simply by invoking conditional mean independence. Although demanding, this restriction seems a valuable attempt to weaken SUTVA, although a complete removal of it would require more sophisticated approaches.

Epidemiological studies have addressed the "interference" topic, by however restricting the analysis to experimental settings with treatment randomization (see, for instance: Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen-Tchetgen and VanderWeele, 2010; Robins et al., 2000). Differently, this paper moves along the line traced by econometric studies, normally dealing with non-experimental settings with sample selection (i.e., no random draw is assumed), and an ex-post evaluation is thus envisaged (Sobel, 2006). In particular, we work within the binary potential outcome model we partially attempt to generalize in order to take into account for the presence of neighbourhood effects. Our theoretical reference draws upon previous works dealing with treatment

[1] The applied literature on the socio-economics of peer effect is rather vast; here we focus on that related to peer (or neighbourhood) effect within the Rubin's potential outcome model (POM). Very recently, however, Angrist (2014) has provided a comprehensive critical review of problems arising in measuring the causal effect of a peer regressor on individual performance. Such article also provides a brief survey of the literature on the subject.

effect identification in the presence of peer effects, and in particular the papers by Manski (1993; 2013).

The paper is organized as follows: section 2 presents some related literature and positions our approach within the Manski's notion of "endogenous" neighbourhood effects; section 3 sets out the model, its assumptions and propositions; section 4 presents the model's estimation procedure; section 5 puts forward an application of the model to real data, by exploring the effect of education on fertility. Two appendices conclude the paper: appendix A sets out the proof of each proposition, while appendix B illustrates a Stata (simulative) implementation of the proposed model, via the user-written command `ntreatreg` (Cerulli, 2015).

2. Related literature

The literature on the estimation of treatment effects under potential interference among units is a recent and challenging field of statistical and econometric study. So far, however, only few papers have dealt formally with this relevant topic (Angrist, 2014).

Rosenbaum (2007) was among the first scholars paving the way to generalize the standard randomization statistical approach for comparing different treatments to the case of units' interference. He presented a statistical model in which unit's response depends not only on the treatment individually received, but also on the treatment received by other units, thus showing how it is possible to test the null-hypothesis of no interference in a random assignment setting where randomization occurs within pre-specified groups and interference between groups is ruled out.

On the same vein, Sobel (2006) provided a definition, identification and estimation strategy for traditional average treatment effect estimators when interference between units is allowed, by taking as example the "Moving To Opportunity" (MTO) randomized social experiment. In his paper, he uses interchangeably the term interference and spillover to account for the presence of such a kind of externality. Interestingly, he shows that a potential bias can arise when no-interference is erroneously assumed, and defines a series of direct and indirect treatment effects that may be identified under reasonable assumptions. Moreover, this author shows some interesting links between the form of his estimators under interference and the Local Average Treatment Effect (LATE) estimator provided by Imbens and Angrist (1994), thus showing that – under interference – treatment effects can be identified only on specific sub-populations.

The paper by Hudgens and Halloran (2008) is probably the most relevant of this literature, as these authors develop a rather general and rigorous modelling of the statistical treatment setting under randomization when interference is potentially present. Furthermore, their approach paves the way also for extensions to observational settings. Starting from the same two-stage randomization

approach of Rosenbaum (2007), these authors manage to go substantially farther by providing a precise characterization of the causal effects with interference in randomized trials encompassing also the Sobel's approach. They define *direct*, *indirect*, *total* and *overall* causal effects showing the relation between these measures and providing an unbiased estimator of the upper bound of their variance.

Tchetgen-Tchetgen and VanderWeele (2010)'s paper follows in the footsteps traced by the approach of Hudgens and Halloran (2008) and provides a formal framework for statistical inference on population average causal effects in a finite sample setting with interference when the outcome variable is binary. Interestingly, they also present an original inferential approach for observational studies based on a generalization of the Inverse Probability Weighting (IPW) estimator when interference is present. Unfortunately, they do not provide the asymptotic variances for such estimators.

Aronow and Samii (2013) finally generalizes the approach proposed by Hudgens and Halloran (2008) going beyond the hierarchical experiment setting and providing a general variance estimation including covariates adjustment.

Previous literature assumes that the potential outcome y of unit i is a function of the treatment received by such a unit (w_i) and the treatment received by all the other units ($\mathbf{w}_{-i}$), that is:

$$y_i(w_i; \mathbf{w}_{-i}) \tag{1}$$

entailing that – with N units and a binary treatment for instance – a number of 2^N potential outcomes may arise. Nevertheless, an alternative way of modelling unit i's potential outcome may be that of assuming:

$$y_i(w_i; \mathbf{y}_{-i}) \tag{2}$$

where $\mathbf{y}_{-i}$ is the (N-1)x1 vector of other units' potential outcomes excluding unit i's potential outcome. The notion of interference entailed by expression (2) is different from that implied by expression (1). The latter, however, is well consistent with the notion of "endogenous" neighbourhood effects provided by Manski (1993, pp. 532-533). Manski, in fact, identifies three types of effects corresponding to three arguments of the individual (potential) outcome equation incorporating social effects[2]:

[2] The literature is not homogeneous in singling out a unique name of such effects: although context-dependent, authors interchangeably refer to peer, neighbourhood, social, club, interference or interaction effects.

1. *Endogenous effects*. Such effects entail that the outcome of an individual depends on the outcomes of other individuals belonging to his neighbourhood.

2. *Exogenous (or contextual) effects*. These effects concern the possibility that the outcome of an individual is affected by the exogenous idiosyncratic characteristics of the individuals belonging to his neighbourhood.

3. *Correlated effects*. They are effects due to belonging to a specific group and thus sharing some institutional/normative condition (that one can loosely define as "environment").

Contextual and correlated effects are to be assumed as exogenous, as they clearly depend on pre-determined characteristics of the individuals in the neighbourhood (case 2) or of the neighbourhood itself (case 3). Endogenous effects are on the contrary of broader interest, as they are affected by the behaviour (measured as "outcome") of other individuals involved in the same neighbourhood. This means that endogenous effects both comprise direct and indirect effects linked to a given external intervention on individuals. The model proposed in this paper incorporates the presence of endogenous neighbourhood effects as defined by Manski within a traditional binary counterfactual model and provides both an identification and an estimation procedure for the Average Treatment Effects (ATEs) in this specific case[3].

How can we position this paper within the literature? Very concisely, previous literature assumes that: (i) unit potential outcome depends on own treatment and other units' treatment; (ii) assignment is randomized or conditionally unconfounded; (iii) treatment is multiple; (iii) potential outcomes have a non-parametric form. This paper, instead, assumes that: (i) unit potential outcome depends on own treatment and other units' potential outcome; (ii) assignment is mean conditionally unconfounded; (iii) treatment is binary; (iv) potential outcomes have a parametric form.

As such, this paper suggests a simple but workable way to relax SUTVA, one that seems rather easy to implement in many socio-economic contexts of application.

[3] A combined regression model including both individual treatments and outcomes may be expressed as:

$$y_i = \mathrm{f}\,(w_i; \mathbf{y}_{-i}; \mathbf{w}_{-i})$$

Arduini, Patacchini and Rainone (2014) provides a first attempt to modelling such a regression on individuals eligible for treatment, showing that the coefficient of w_i (i.e., their measure of ATE) combines both treatments' and outcomes' direct and indirect effects on y. However, such a model is not embedded within the classical Rubin' potential outcome model (POM). Differently, the paper proposed here provides a POM-consistent approach, generalized to the case of possible interaction among units.

3. A binary treatment model with "endogenous" neighbourhood effects

This section presents a model for estimating the average treatment effects (ATEs) of a policy program (or a treatment) in a non-experimental setting in the presence of "endogenous" neighbourhood (or externality) interactions. We consider a binary treatment variable w - taking value 1 for treated and 0 for untreated units - assumed to affect an outcome (or target) variable y that can take a variety of forms.

Some notation can help in understanding the setting: N is the number of units involved in the experiment; N_1, the number of treated units; N_0 the number of untreated units; w_i the treatment variable assuming value "1" if unit i is treated and "0" if untreated; y_{1i} is the outcome of unit i when she is treated; y_{0i} is the outcome of unit i when she is untreated; $\mathbf{x}_i = (x_{1i}, x_{2i}, x_{3i}, \ldots, x_{Mi})$ is a row vector of M exogenous observable characteristics for unit $i = 1, \ldots, N$.

To begin with, as usual in this literature, we define the unit i's *Treatment Effect* (TE) as:

$$\mathrm{TE}_i = y_{1i} - y_{0i} \tag{3}$$

TE_i is equal to the difference between the value of the target variable when the individual is treated (y_1), and the value assumed by this variable when the same individual is untreated (y_0). Since TE_i refers to the same individual at the same time, the analyst can observe just one of the two quantities feeding into (3) but never both. For instance, it might be the case that we can observe the investment behaviour of a supported company, but we cannot know what the investment of this company would have been, had it not been supported, and vice versa. The analyst faces a fundamental missing observation problem (Holland, 1986) that needs to be tackled econometrically in order to recover reliably the causal effect via some specific imputation technique (Rubin, 1974; 1977).

The random pair (y_{1i}, y_{0i}) is assumed to be independent and identically distributed (i.i.d.) across all units i, and both y_{1i} and y_{0i} are generally explained by a structural part depending on observable factors, and a non-structural one depending on an unobservable (error) term. Nevertheless, recovering the entire distributions of y_{1i} and y_{0i} (and, consequently, the distribution of the TE_i) may be too demanding without very strong assumptions, so that the literature has focused on estimating specific moments of these distributions and in particular the "mean", thus defining the so-called population Average Treatment Effect (hereinafter ATE), and ATE conditional on $\mathbf{x}_i$ (i.e., $\mathrm{ATE}(\mathbf{x}_i)$) of a policy intervention as:

$$\mathrm{ATE} = \mathrm{E}(y_{i1} - y_{i0}) \tag{4}$$

$$\text{ATE}(\mathbf{x}_i) = \text{E}(y_{i1} - y_{i0} \mid \mathbf{x}_i) \tag{5}$$

where E(·) is the mean operator. ATE is equal to the difference between the average of the target variable when the individual is treated (y_1), and the average of the target variable when the same individual is untreated (y_0). Observe that, by the law of iterated expectations, $\text{ATE} = \text{E}_{\mathbf{x}}\{\text{ATE}(\mathbf{x})\}$.

Given the definition of the unconditional and conditional average treatment effect in (4) and (5) respectively, it is immediate to define the same parameters in the sub-population of treated (ATET) and untreated (ATENT) units, i.e.:

$$\text{ATET} = \text{E}(y_{i1}-y_{i0} \mid w_i=1)$$
$$\text{ATET}(\mathbf{x}_i) = \text{E}(y_{i1} - y_{i0} \mid \mathbf{x}_i, w_i=1)$$

and

$$\text{ATENT} = \text{E}(y_{i1}-y_{i0} \mid w_i=0)$$
$$\text{ATENT}(\mathbf{x}_i) = \text{E}(y_{i1} - y_{i0} \mid \mathbf{x}_i, w_i=0)$$

The aim of this paper is to provide consistent parametric estimation of all previous quantities (we refer to as ATEs) in the presence of neighbourhood effects.

To that end, we start with what is observable to the analyst in such a setting, i.e. the actual status of the unit *i*, that can be obtained as:

$$y_i = y_{0i} + w_i\,(y_{1i} - y_{0i}) \tag{6}$$

Equation (6) is the Rubin's potential outcome model (POM), and it is the fundamental relation linking the unobservable to the observable outcome. Given Eq. (6), we first set out all the assumptions behind the next development of the proposed model.

Assumption 1. *Unconfoundedness* (or CMI). Given the set of random variables $\{y_{1i}, y_{1i}, w_i, \mathbf{x}_i\}$ as defined above, the following equalities hold:

$$\text{E}(y_{ig} \mid w_i, \mathbf{x}_i) = \text{E}(y_{ig} \mid \mathbf{x}_i) \quad \text{with} \quad g = \{0,1\}$$

Hence, throughout this paper, we will assume unconfoundedness, i.e. Conditional Mean Independence (CMI) to hold. As we will show, CMI is a sufficient condition for identifying ATEs also when neighbourhood effects are considered.

Once CMI has been assumed, we then need to model the potential outcomes y_{0i} and y_{1i} in a proper way so to get a representation of the ATEs (i.e., ATE, ATET and ATENT) taking into account the presence of endogenous externality effects. In this paper, we simplify further our analysis by assuming some restrictions in the form of the potential outcomes.

Assumption 2. *Restrictions on the form of the potential outcomes*. Consider the general form of the potential outcome as expressed in (2), and assume this relation to depend parametrically on a vector of real numbers $\boldsymbol{\theta} = (\boldsymbol{\theta}_0; \boldsymbol{\theta}_1)$. We assume that:

$$y_{1i}(w_i; \mathbf{x}_i; \boldsymbol{\theta}_1)$$

and

$$y_{0i}(w_i; \mathbf{x}_i; \mathbf{y}_{1,-i}; \boldsymbol{\theta}_0)$$

Assumption 2 poses two important restrictions to the form given to the potential outcomes: (i) it makes them dependent on some unknown parameters $\boldsymbol{\theta}$ (i.e., parametric form); (ii) it entails that the externality effect occurs only in one direction, i.e. from the treated individuals to the untreated units, while excluding the other way round[4].

Assumption 3. *Linearity and weighting-matrix*. We assume that the potential outcomes are linear in the parameters, and that a NxN weighting-matrix $\boldsymbol{\Omega}$ of exogenous constant numbers is known.

Under Assumptions 1, 2 and 3, the model takes on this form:

[4] In the more general case in which peer effects take place from treated to untreated units and vice versa, an identification of ATEs consistent with the POM becomes trickier, as various feedback terms do arise. Using a spatial regression approach, Arduini, Patacchini and Rainone (2014) estimate a treatment effect reduced-form including also feedback terms. Their model is not however directly derived by using the POM, as this paper attempts to do.

$$\begin{cases} y_{1i} = \mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1 + e_{1i} \\ y_{0i} = \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma s_i + e_{0i} \\ s_i = \sum_{j=1}^{N_1} \omega_{ij} y_{1j}, \quad \text{with} \ \sum_{j=1}^{N_1} \omega_{ij} = 1 \\ y_i = y_{0i} + w(y_{1i} - y_{0i}) \\ \text{CMI holds} \end{cases} \tag{7}$$

where i = 1, ..., N and j = 1, ..., N_1, μ_1 and μ_0 are scalars, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are two unknown vector parameters defining the different response of unit i to the vector of covariates $\mathbf{x}$, e_0 and e_1 are two random errors with zero unconditional mean and constant variance, and s_i represents unit i-th neighbourhood effect due to the treatment administrated to units j (j = 1, ..., N_1). Observe that, by linearity[5], we have:

$$s_i = \begin{cases} \sum_{j=1}^{N_1} \omega_{ij} y_{1j} & \text{if } \ i \in \{w = 0\} \\ 0 & \text{if } \ i \in \{w = 1\} \end{cases} \tag{8}$$

where the parameter ω_{ij} is the generic element of the weighting matrix $\boldsymbol{\Omega}$ expressing some form of *distance* between unit i and unit j. Although not strictly required for consistency, we also assume that these weights add to one, i.e. $\sum_{j=1}^{N_1} \omega_{ij} = 1$. In short, previous assumptions say that units i neighbourhood effect takes the form of a weighted-mean of the outcomes of treated units and that this "social" effect has an impact only on unit i's outcome when this unit is untreated. As a consequence, by substitution of (8) into (7), we get that:

$$y_{0i} = \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma \sum_{j=1}^{N_1} \omega_{ij} y_{1j} + e_{0i} \tag{9}$$

making it clear that untreated unit's i outcome is a function of its own idiosyncratic characteristics ($\mathbf{x}_i$), the weighted outcomes of treated units multiplied by a sensitivity parameter γ, and a standard error term.

We state now a series of propositions implied by previous assumptions.

[5] The linearity of the spillover effect is an assumption needed to simplify the subsequent regression analysis. It is however clear that non-linear forms might also be used.

Proposition 1. *Formula of* ATE *with neighbourhood interactions*. Given assumptions 2 and 3 and the implied equations established in (7), the average treatment effect (ATE) with neighbourhood interactions takes on this form:

$$\begin{aligned}&\text{ATE} = \mathrm{E}(y_{1i} - y_{0i}) = \mu + \mathrm{E}\left[\mathbf{x}_i\boldsymbol{\delta} - \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1 - e_i\right] = \\ &\mu + \bar{\mathbf{x}}\boldsymbol{\delta} - \bar{\mathbf{v}}\boldsymbol{\lambda}\end{aligned} \tag{10}$$

Where $\boldsymbol{\lambda} = \gamma\boldsymbol{\beta}_1$, $\bar{\mathbf{x}} = \mathrm{E}(\mathbf{x}_i)$, $\bar{\mathbf{v}} = \mathrm{E}\left(\underbrace{\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j}_{\mathbf{v}_i}\right)$ is the unconditional mean of the vector $\mathbf{x}_i$, and $\mu = \mu_1 - \mu_0 - \gamma\mu_1$. The proof is in appendix A. See A1.

Indeed, by the definition of ATE as given in (4) and by (7), we can immediately show that for such a model:

$$\text{ATE} = \mathrm{E}(y_{1i} - y_{0i}) = \mathrm{E}\left[\left(\mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1 + e_{1i}\right) - \left(\mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\sum_{j=1}^{N_1}\omega_{ij}y_{1j} + e_{0i}\right)\right] \tag{11}$$

where:

$$\begin{aligned}&\sum_{j=1}^{N_1}\omega_{ij}y_{1j} = \sum_{j=1}^{N_1}\omega_{ij}\left(\mu_1 + \mathbf{x}_j\boldsymbol{\beta}_1 + e_{1j}\right) = \\ &\mu_1\sum_{j=1}^{N_1}\omega_{ij} + \sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\boldsymbol{\beta}_1 + \sum_{j=1}^{N_1}\omega_{ij}e_{1j} = \\ &\mu_1 + \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\boldsymbol{\beta}_1 + \sum_{j=1}^{N_1}\omega_{ij}e_{1j}\end{aligned} \tag{12}$$

and by developing ATE further using Eq. (11), we finally get the result in (10).

Proposition 2. *Formula of* ATE($\mathbf{x}_i$) *with neighbourhood interactions*. Given assumptions 2 and 3 and the result in proposition 1, we have that:

$$\mathrm{ATE}(\mathbf{x}_i) = \mathrm{E}(y_{1i} - y_{0i} \mid \mathbf{x}_i) = \mathrm{ATE} + (\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + (\bar{\mathbf{v}} - \mathbf{v}_i)\boldsymbol{\lambda} \tag{13}$$

where it is now easy to see that ATE $=\mathrm{E}_{\mathbf{x}}\{\mathrm{ATE}(\mathbf{x})\}$. The proof is in appendix A. See A2.

Proposition 3. *Baseline random-coefficient regression*. By substitution of equations (7) into the POM of Eq. (6), we obtain the following random-coefficient regression model (Wooldridge, 1997):

$$y_i = \eta + w_i \cdot \mathrm{ATE} + \mathbf{x}_i\boldsymbol{\beta}_0 + w_i(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + \mathbf{z}_i\boldsymbol{\lambda} + e_i \tag{14}$$

where: $\mathbf{z}_i = \mathbf{v}_i + w_i(\bar{\mathbf{v}} - \mathbf{v}_i)$, $\mathbf{v}_i = \sum_{j=1}^{N_1} \omega_{ij}\mathbf{x}_j$, $\bar{\mathbf{v}} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N_1} \omega_{ij}\mathbf{x}_j\right)$, $\boldsymbol{\lambda}=\gamma\boldsymbol{\beta}_1$, $\eta = \mu_0 + \gamma\mu_1$, and $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$. The proof is in appendix A. See A3.

Proposition 4. *Ordinary Least Squares (OLS) consistency*. Under assumption 1 (CMI), 2 and 3, the error tem of regression (14) has zero mean conditional on (w_i, $\mathbf{x}_i$), i.e.:

$$\mathrm{E}\left(e_i \,\middle|\, w_i, \mathbf{x}_i\right) = \mathrm{E}\left(\gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} + e_{0i} + w_i(e_{1i} - e_{0i}) - w_i\gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} \,\middle|\, w_i, \mathbf{x}_i\right) = 0 \tag{15}$$

thus implying that Eq. (14) is a regression model whose parameters can be *consistently* estimated by Ordinary Least Squares (OLS). The proof is in appendix A. See A4.

Once a consistent estimation of the parameters of (14) is obtained, we can estimate ATE directly from the regression, and ATE($\mathbf{x}_i$) by plugging the estimated parameters into formula (11). This is because ATE($\mathbf{x}_i$) becomes a function of consistent estimates, and thus consistent itself:

$$\operatorname{plim} \widehat{\mathrm{ATE}}(\mathbf{x}_i) = \mathrm{ATE}(\mathbf{x}_i)$$

where $\widehat{\mathrm{ATE}}(\mathbf{x}_i)$ is the plug-in estimator of ATE($\mathbf{x}_i$). Observe, however, that the (exogenous) weighting matrix $\boldsymbol{\Omega}=[\omega_{ij}]$ needs to be provided in advance.

Once the formulas for ATE and ATE($\mathbf{x}_i$) are available, it is also possible to recover the Average Treatment Effect on Treated (ATET) and on non-Treated (ATENT) as:

$$\text{ATET} = \text{ATE} + \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \left[(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + (\bar{\mathbf{v}} - \mathbf{v}_i)\boldsymbol{\lambda} \right] \tag{16}$$

and:

$$\text{ATENT} = \text{ATE} + \frac{1}{\sum_{i=1}^{N} (1 - w_i)} \sum_{i=1}^{N} (1 - w_i) \left[(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + (\bar{\mathbf{v}} - \mathbf{v}_i)\boldsymbol{\lambda} \right] \tag{17}$$

These quantities are functions of observable components and parameters consistently estimated by OLS (see next section). Once these estimates are available, standard errors for ATET and ATENT can be correctly obtained via bootstrapping (see Wooldridge, 2010, pp. 911-919).

4. Estimation

Starting from previous section's results, a simple protocol for estimating ATEs can be suggested. Given an i.i.d. sample of observed variables for each individual *i*:

$$\{y_i, w_i, \mathbf{x}_i\} \text{ with } i = 1, \ldots, N$$

1. provide a weighting matrix $\Omega=[\omega_{ij}]$ measuring some type of distance between the generic unit *i* (untreated) and unit *j* (treated);

2. estimate by an OLS a regression model of:

$$y_i \quad \text{on} \quad \{1, \; w_i, \; \mathbf{x}_i, \; w_i(\mathbf{x}_i - \bar{\mathbf{x}}), \; \mathbf{z}_i\}$$

3. obtain $\left\{\hat{\boldsymbol{\beta}}_0, \; \hat{\boldsymbol{\delta}}, \; \hat{\gamma}, \; \hat{\boldsymbol{\beta}}_1\right\}$ and put them into the formulas of ATEs.

By comparing the formula of ATE *with* ($\gamma \neq 0$) and *without* ($\gamma = 0$) neighbourhood effect, we get the *neighbourhood-bias* defined as:

$$\begin{aligned} \text{Bias} &= \left|\text{ATE}_{\text{no-neigh}} - \text{ATE}_{\text{with-neigh}}\right| = \left|\gamma\mu_1 + \bar{\mathbf{v}}\boldsymbol{\lambda}\right| = \\ &\left|\gamma\mu_1 + \left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\right]\boldsymbol{\lambda}\right| \end{aligned} \tag{18}$$

This can also be seen as the *externality effect* produced by the evaluated policy: it depends on the weights employed, on the average of the observable confounders considered into **x**, and on the magnitude of the coefficients γ and $\boldsymbol{\beta}_1$. Observe that such bias may be positive as well as negative. Furthermore, by defining:

$$\gamma\boldsymbol{\beta}_1 = \boldsymbol{\lambda} \tag{19}$$

it is also possible to test whether this bias is or is not statistically significant by simply testing the following null-hypothesis:

$$\mathrm{H}_0: \ \lambda_1 = \lambda_2 = ... = \lambda_M = 0$$

If this hypothesis is rejected, we cannot exclude that neighbourhood effects are pervasive, thus affecting significantly the estimation of the causal parameters ATEs. In a similar way, we can also get an estimation of the neighbourhood-bias for ATET and ATENT.

For operationalization purposes, appendix B provides the reader with a Stata implementation of our model, via the user-written command `ntreatreg` (Cerulli, 2015). This appendix sets out an illustrative simulation example along with its related Stata code.

5. Application: the effect of *education* on *fertility*

In this application, we consider the dataset "FERTIL2" accompanying the manual "*Introductory Econometrics: A Modern Approach*" by Wooldridge (2000), where we consider only *N*=152 (out of 4,361) randomly drawn women in childbearing age in Botswana.

The aim of this application is that of evaluating the impact of *education* on *fertility*, i.e. the causal effect of the variable "educ7" - taking value 1 if a woman has more than or exactly seven years of education and 0 otherwise - on the number of family children (the variable "children").

Several conditioning (or confounding) observable factors are included in the dataset, such as the age of the woman ("age"), whether or not the family owns a TV ("tv"), whether or not the woman lives in a city ("urban"), and so forth. Table 1 illustrates a description of the variables included in FERTIL2.

We are particularly interested in detecting the effect of education on fertility in such a setting, by taking into account possible *peer-interactions* among women.

Table 1. Description of the dataset "FERTIL2_reduced".

Variable name	Variable description
mnthborn	month woman born
yearborn	year woman born
age	age in years
electric	=1 if has electricity
radio	=1 if has radio
tv	=1 if has tv
bicycle	=1 if has bicycle
educ	years of education
ceb	children ever born
agefbrth	age at first birth
children	number of living children
knowmeth	=1 if know about birth control
usemeth	=1 if ever use birth control
monthfm	month of first marriage
yearfm	year of first marriage
agefm	age at first marriage
idlnchld	'ideal' number of children
heduc	husband's years of education
agesq	age^2
urban	=1 if live in urban area
urb_educ	urban*educ
spirit	=1 if religion = spirit
protest	=1 if religion = protestant
catholic	=1 if religion = catholic
frsthalf	=1 if mnthborn <= 6
educ0	=1 if educ == 0
evermarr	=1 if ever married
educ7	=1 if educ >= 7

Our research presumption is that in choosing their 'desired' number of children, less educated women (the untreated ones) are not only affected by their own (idiosyncratic) characteristics (the variables **x**), but also by the number of children chosen by more educated women. The conjecture behind this statement is that less educated women might want to be as like as possible to more educated ones as a way to avoid some form of social stigma.

In order to build the (network) matrix omega, we calculate the module of the correlation matrix associated to the covariates: *radio*, *tv*, *bicycle*, *usemeth*, *protest*, and *frsthalf*. Such matrix presents values that are only equal or greater than zero. The assumption here is that women more correlated in terms of the previous variables are also those assumed to be more linked.

Table 2. Comparison between the baseline regression results with and without peer effects.

	REG_peer	*REG_no_peer*
educ7	-0.76***	-0.53**
	(0.25)	(0.26)
age	0.33***	0.52***
	(0.12)	(0.12)
agesq	-0.00	-0.01***
	(0.00)	(0.00)
evermarr	0.94**	1.02**
	(0.43)	(0.45)
electric	-0.75**	-0.53
	(0.35)	(0.37)
tv	-0.01	-0.15
	(0.47)	(0.50)
_ws_age	0.05	-0.14
	(0.18)	(0.19)
_ws_agesq	-0.00	0.00
	(0.00)	(0.00)
_ws_evermarr	0.27	0.15
	(0.57)	(0.60)
_z_age	8.35***	
	(2.66)	
_z_agesq	-0.15***	
	(0.05)	
_z_evermarr	-7.83*	
	(4.52)	
_z_electric	22.72**	
	(10.02)	
_z_tv	-3.59	
	(7.59)	
N	152	152
adj. R^2	0.642	0.583
F-test	20.33***	24.44***

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

After standardizing this matrix to comply with the requirements of matrix omega, we consider the relation between *children* and *educ7* using *age*, *agesq*, *evermarr*, *electric*, and *tv* as controls, and *age*, *agesq*, and *evermarr* as regressors subject to observable heterogeneity. We then estimate the model presented in the previous sections by using the Stata routine "ntreatreg" (Cerulli, 2015) freely available on the Stata Statistical Software Components (SSC) archive.

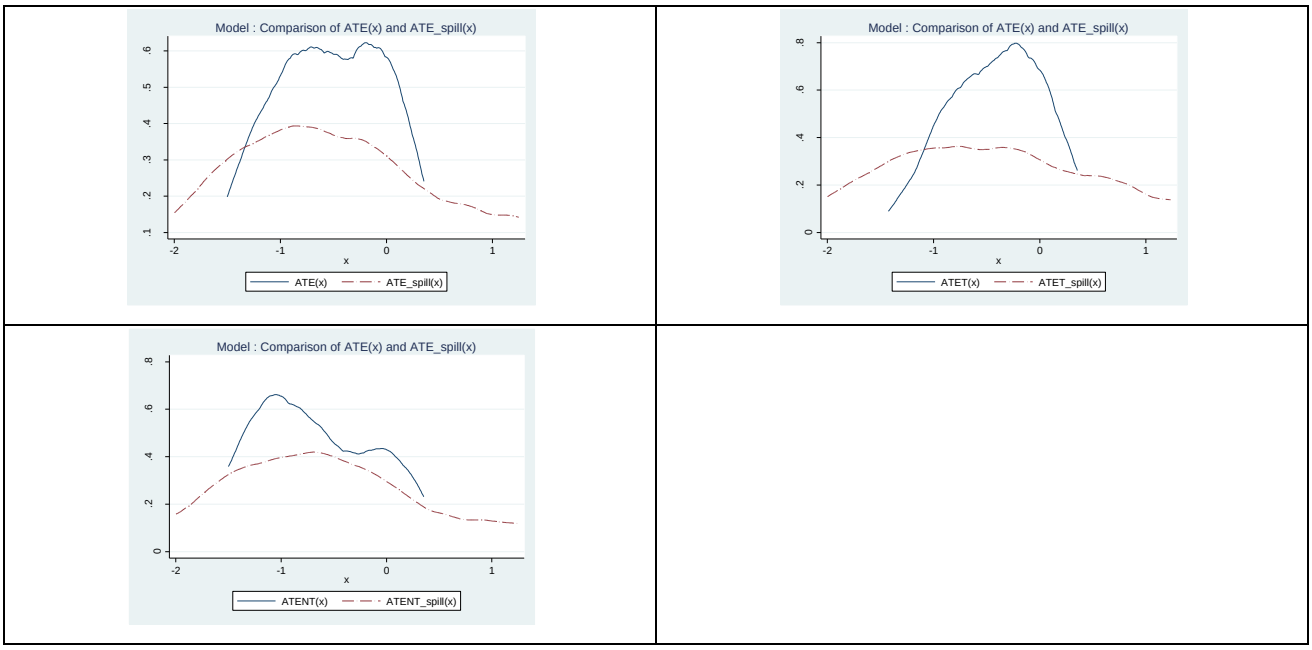
Results are set out in Table 2. The coefficient of the treatment variable, "educ7", is significantly equal to -0.76 in the regression incorporating peer effects (*REG_no_peer*), significantly equal to -0.53 in the one not incorporating them (*REG_peer*). The adjusted R-squared is higher in the first

than in the second regression, thus meaning that the fit is better when treatment peer effects are considered. The absolute bias is around 43 %, which is of a considerable magnitude.

By performing a test to see whether the coefficients of the peer-effects are jointly zero (i.e., H_0: $\gamma\boldsymbol{\beta}_0 = 0$), we reject this hypothesis getting an F-test equal 5.68 which is highly significant (as the p-value is equal to 0.0001). This means that we cannot reject that peer effects are present in this example.

We can also compare graphically the distribution of ATE(**x**), ATET(**x**) and ATENT(**x**) with and without neighbourhood-interaction. Figure 1 shows the results.

Figure 1. Comparison between ATE(**x**), ATET(**x**) and ATENT(**x**) with and without peer-effect.



As a conclusion, we can say that if the analyst does not consider neighbourhood effects, she will "down-estimate" the actual effect of education on fertility of around a 43 %. Furthermore, the test shows that the neighbourhood effect is relevant, as the regression coefficients of the neighbourhood component are jointly significant.

How can we interpret such a result? A possible argument can be that there is a peer-effect on women in choosing how many children to have. Indeed, as said before, the "desired" number of children for a woman does not depend only on her individual determinants ("age", for instance), but also on "what friends do". In our sample, the existence of such a "social interaction" seems to increase the "effect of education on fertility" (from 0.53 to 0.76 in absolute values), by showing that fertility behaviour of more educated women (unconditionally less fertile) produce a *spillover* on less educated ones, by pushing them to increase the number of children to have. This might be a

response behaviour on the part of less educated women trying to identify themselves as characterized by a different "social mission" (i.e. raising more children) compared to more educated ones. Therefore, "education" has an effect on "fertility", not only because schooling can delay the time to have a child, but also because education "triggers" some form of peer-effects.

7. Conclusion

This paper has presented a possible solution to incorporate *externality* (or *neighbourhood*) *effects* within the traditional Rubin's potential outcome model under conditional mean independence. As such, it attempts to account for treatment peer effects within the traditional regression-adjustment model within program evaluation settings when SUTVA should be relaxed.

As anticipated in previous sections, this approach presents various limitations, and in what follows we list some of its potential developments. Indeed, the model might be improved by:

- allowing also for treated units to be affected by other treated units' outcome, and untreated units to affect untreated ones;
- extending the model to "multiple" or "continuous" treatment, when treatment may be multi-valued or fractional for instance, by still holding CMI;
- Allowing for the unit potential outcome to depend on other units' treatment;
- identifying ATEs with neighbourhood interactions when w may be endogenous (i.e., relaxing CMI) by implementing GMM-IV estimation;
- trying to go beyond the potential outcomes' parametric form, thus relying on a nonparametric or semi-parametric specification.

Finally, an issue that should deserve further inquiry regards the assumption of exogeneity of the weighting matrix Ω. Indeed, a challenging question is: what does it happen if individuals strategically modify their weighting weights to better take advantage of others' treatment outcome? It is clear that weights do become endogenous, thus yielding further identification problems for previous causal effects. Future studies should tackle situations in which this possibility may occur.

References

Angrist, J.D. (2014), The Perils of Peer Effects, *Labour Economics*, forthcoming.

Arduini, T., Patacchini, E. and Rainone, E. (2014), Identification and Estimation of Outcome Response with Heterogeneous Treatment Externalities, CPR Working Paper No 167.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.

Aronow, P.M. and Cyrus, Samii C. (2013), Estimating average causal effects under interference between units. Unpublished manuscript, May 28.

Cerulli, G. (2015), NTREATREG: Stata module for estimation of treatment effects in the presence of neighbourhood interactions, *Statistical Software Components S457961*, Boston College, Department of Economics.

Cox, D.R. (1958), *Planning of Experiments*, New York, Wiley.

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

Hudgens, M. G. and Halloran, M.E. (2008), Toward causal inference with interference, *Journal of the American Statistical Association*, 103, 482, 832–842.

Imbens, W.G. and Angrist, J.D. (1994), Identification and estimation of local average treatment effects, *Econometrica*, 62, 2, 467–475.

Manski, C.F.(1993), Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies*, 60, 3, 531–542.

Manski, C.F. (2013), Identification of treatment response with social interactions, *The Econometrics Journal*, 16, 1, S1–S23.

Robins, J. M., Hernan, M.A. and Brumback, B. (2000), Marginal structural models and causal inference in epidemiology, *Epidemiology*, 11, 5, 550–560.

Rosenbaum, P.R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102, 477, 191–200.

Rubin, D.B. (1974), Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology*, 66, 5, 688–701.

Rubin, D.B. (1977), Assignment to treatment group on the basis of a covariate, *Journal of Educational Statistics*, 2, 1, 1–26.

Rubin, D.B. (1978), Bayesian inference for causal effects: The role of randomization, *Annals of Statistics*, 6, 1, 34–58.

Sobel, M.E. (2006), What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference, *Journal of the American Statistical Association*, 101, 476, 1398–1407.

Tchetgen-Tchetgen, E. J. and VanderWeele, T.J. (2010), On causal inference in the presence of interference, *Statistical Methods in Medical Research*, 21, 1, 55-75.

Wooldridge, J.M. (1997), On two stage least squares estimation of the average treatment effect in a random coefficient model, *Economics Letters*, 56, 2, 129-133.

Wooldridge, J.M. (2010), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press.

Appendix A

In this appendix, we show how to obtain the formulas of ATE and ATE(**x**) set out in (12) and (13). Then, we show how regression (14) can be obtained and, finally, we prove that Assumption 1 is sufficient for consistently estimating the parameters of regression (14) by OLS.

A1. *Formula of* ATE *with neighbourhood interactions.*

Given assumptions 2 and 3, and the implied equations in (7), we get that:

$$y_{1i} = \mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1 + e_{1i}$$
$$y_{0i} = \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma s_i + e_{0i}$$
$$s_i = \sum_{j=1}^{N_1}\omega_{ij}y_{1j}$$

$$\text{ATE} = \mathrm{E}(y_{1i} - y_{0i}) = \mathrm{E}\left[\left(\mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1 + e_{1i}\right) - \left(\mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\left[\mu_1 + \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\boldsymbol{\beta}_1 + \sum_{j=1}^{N_1}\omega_{ij}e_{1j}\right] + e_{0i}\right)\right] =$$

$$= \mathrm{E}\left[\mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1 + e_{1i} - \left(\mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\mu_1 + \gamma\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\boldsymbol{\beta}_1 + \gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} + e_{0i}\right)\right] =$$

$$= \mathrm{E}\left[\mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1 + e_{1i} - \mu_0 - \mathbf{x}_i\boldsymbol{\beta}_0 - \gamma\mu_1 - \gamma\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\boldsymbol{\beta}_1 - \gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} - e_{0i}\right] =$$

$$= \mathrm{E}\left[\mu_1 - \gamma\mu_1 - \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_1 - \mathbf{x}_i\boldsymbol{\beta}_0 - \gamma\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\boldsymbol{\beta}_1 - \gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} + e_{1i} - e_{0i}\right] =$$

$$= \mathrm{E}\left[\mu_1(1-\gamma) - \mu_0 + \mathbf{x}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) - \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1 - \gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} + e_{1i} - e_{0i}\right] =$$

$$= \mathrm{E}\left[\mu_1(1-\gamma) - \mu_0 + \mathbf{x}_i\boldsymbol{\delta} - \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1 - \gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} + e_{1i} - e_{0i}\right] =$$

$$= \mu + \mathrm{E}\left[\mathbf{x}_i\boldsymbol{\delta} - \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1 - e_i\right] = \mu + \mathrm{E}(\mathbf{x}_i)\boldsymbol{\delta} - \gamma\mathrm{E}\left(\underbrace{\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j}_{\mathbf{v}_i}\right)\boldsymbol{\beta}_1$$

This implies that $\text{ATE} = \mathrm{E}(y_{1i} - y_{0i}) = \mu + \mathrm{E}(\mathbf{x}_i)\boldsymbol{\delta} - \gamma\mathrm{E}(\mathbf{v}_i)\boldsymbol{\beta}_1$ whose sample equivalent is:

$$\text{ATE} = \hat{\mu} + \frac{1}{N}\left(\sum_{i=1}^{N}\mathbf{x}_i\right)\hat{\boldsymbol{\delta}} - \hat{\gamma}\frac{1}{N}\left(\sum_{i=1}^{N}\mathbf{v}_i\right)\hat{\boldsymbol{\beta}}_1 = \mu + \frac{1}{N}\left(\sum_{i=1}^{N}\mathbf{x}_i\right)\hat{\boldsymbol{\delta}} - \hat{\gamma}\frac{1}{N}\left(\sum_{i=1}^{N}\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\right)\hat{\boldsymbol{\beta}}_1$$

where $\mu = \mu_1(1-\gamma) - \mu_0$, and $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$. ■

As an example, consider the case in which N=4, and N_1=N_0=2. Suppose that the matrix $\boldsymbol{\Omega}$ is organized as follows:

	T	C
T	$\begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$	$\begin{bmatrix} \omega_{13} & \omega_{14} \\ \omega_{23} & \omega_{24} \end{bmatrix}$
C	$\begin{bmatrix} \omega_{31} & \omega_{32} \\ \omega_{41} & \omega_{42} \end{bmatrix}$	$\begin{bmatrix} \omega_{33} & \omega_{34} \\ \omega_{34} & \omega_{44} \end{bmatrix}$

Suppose to have just one confounder x. In this case, we have:

$$\text{ATE} = \hat{\mu} + \frac{1}{4}\left(\sum_{i=1}^{4} x_i\right)\hat{\delta} - \hat{\gamma}\cdot\hat{\beta}_1\frac{1}{4}\left(\sum_{i=1}^{4}\left(\sum_{j=1}^{2}\omega_{ij}x_j\right)\right) = \hat{\mu} + \frac{1}{4}\left(\sum_{i=1}^{4} x_i\right)\hat{\delta} - \hat{\gamma}\cdot\hat{\beta}_1\frac{1}{4}\left(\sum_{i=1}^{4}\left[\omega_{i1}x_1 + \omega_{i2}x_2\right]\right) =$$

$$\hat{\mu} + \bar{x}\hat{\delta} - \hat{\gamma}\cdot\hat{\beta}_1\frac{1}{4}\left(\sum_{i=1}^{4}\underbrace{\left[\omega_{i1}x_1 + \omega_{i2}x_2\right]}_{v_i}\right) = \hat{\mu} + \bar{x}\hat{\delta} - \hat{\gamma}\cdot\hat{\beta}_1\bar{v}$$

Observe that:

$$\begin{aligned} v_1 &= \omega_{11}x_1 + \omega_{12}x_2 \\ v_1 &= \omega_{21}x_1 + \omega_{22}x_2 \\ v_3 &= \omega_{31}x_1 + \omega_{32}x_2 \\ v_4 &= \omega_{41}x_1 + \omega_{42}x_2 \end{aligned}$$

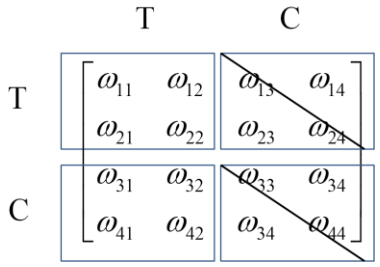
implying:

$$\text{ATE} = \hat{\mu} + \bar{x}\hat{\delta} - \hat{\gamma}\cdot\hat{\beta}_1\frac{1}{4}\left(\sum_{i=1}^{4}\underbrace{\left[\omega_{i1}x_1 + \omega_{i2}x_2\right]}_{v_i}\right) = \hat{\mu} + \bar{x}\hat{\delta} - \hat{\gamma}\cdot\hat{\beta}_1\left[\bar{\omega}_{\cdot 1}x_1 + \bar{\omega}_{\cdot 2}x_2\right]$$

where:

$$\bar{\omega}_{\cdot 1} = \frac{1}{4}\sum_{i=1}^{4}\omega_{i1} \quad \text{and} \quad \bar{\omega}_{\cdot 2} = \frac{1}{4}\sum_{i=1}^{4}\omega_{i2}$$

This means that, by assuming that the externality effect only comes from treated to untreated units thus excluding other types of feedbacks, is equivalent to consider *only* the first two columns of **Ω** in the calculation of the externality component, those refereeing to the treated units, i.e.:



where no use of the two columns referring to the control group occurs.

A2. *Formula of* ATE($\mathbf{x}_i$) *with neighbourhood interactions.*

Given assumptions 2 and 3, and the result in A1, we get:

$$\mathrm{ATE}(\mathbf{x}_i) = \mathrm{E}(y_{1i} - y_{0i} \mid \mathbf{x}_i) = \mu + \mathrm{E}\left[\mathbf{x}_i\boldsymbol{\delta} - \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1 - e_i \mid \mathbf{x}_i\right] = \mu + \mathbf{x}_i\boldsymbol{\delta} - \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1 +$$

$$+[\bar{\mathbf{x}}\boldsymbol{\delta} - \bar{\mathbf{x}}\boldsymbol{\delta}] + \left[\mathrm{E}\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1 - \mathrm{E}\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1\right] =$$

$$\left(\mu + \bar{\mathbf{x}}\boldsymbol{\delta} - \mathrm{E}\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\gamma\boldsymbol{\beta}_1\right) + (\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + \left[\mathrm{E}\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right) - \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\right)\right]\gamma\boldsymbol{\beta}_1 =$$

$$\mathrm{ATE} + (\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + (\bar{\mathbf{v}} - \mathbf{v}_i)\boldsymbol{\lambda}$$

where $\boldsymbol{\lambda}=\gamma\boldsymbol{\beta}_1$.

■

A3. *Obtaining regression (14).*

By substitution of the potential outcome as in (7) into the potential outcome model, we get that:

$$y_i = \left(\mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\sum_{j=1}^{N_1}\omega_{ij}y_{1j} + e_{0i}\right) + w_i\left[\left(\mu_1 + \mathbf{x}_i\boldsymbol{\beta}_1 + e_{1i}\right) - \left(\mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\sum_{j=1}^{N_1}\omega_{ij}y_{1j} + e_{0i}\right)\right] =$$

$$= \left(\mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\sum_{j=1}^{N_1}\omega_{ij}y_{1j} + e_{0i} \right) + w_i(\mu_1 - \mu_0) + w_i\mathbf{x}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + w_i(e_{1i} - e_{0i}) - w_i\gamma\sum_{j=1}^{N_1}\omega_{ij}y_{1j} =$$

$$= \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\mu_1 + \left(\gamma\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 + \gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} + e_{0i} + w_i(\mu_1 - \mu_0) + w_i\mathbf{x}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + w_i(e_{1i} - e_{0i}) -$$

$$w_i\gamma\mu_1 - w_i\gamma\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j\boldsymbol{\beta}_1 - w_i\gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} =$$

$$= \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\mu_1 + \left(\gamma\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 + \underbrace{[\gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j} + e_{0i} + w_i(e_{1i} - e_{0i}) - w_i\gamma\sum_{j=1}^{N_1}\omega_{ij}e_{1j}]}_{e_i} + w_i(\mu_1 - \mu_0) +$$

$$w_i\mathbf{x}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) - w_i\gamma\mu_1 - w_i\gamma\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 =$$

$$= \mu_0 + \mathbf{x}_i\boldsymbol{\beta}_0 + \gamma\mu_1 + \left(\gamma\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 + w_i(\mu_1 - \mu_0) + w_i\mathbf{x}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) - w_i\gamma\mu_1 - w_i\gamma\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 + e_i =$$

$$= (\mu_0 + \gamma\mu_1) + \left(\gamma\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 + w_i(\mu_1 - \mu_0 - \gamma\mu_1) + \mathbf{x}_i\boldsymbol{\beta}_0 + w_i\mathbf{x}_i(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) - w_i\gamma\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 + e_i =$$

$$= (\mu_0 + \gamma\mu_1) + \left(\gamma\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\boldsymbol{\beta}_1 + w_i(\mu_1 - \mu_0 - \gamma\mu_1) + \mathbf{x}_i\boldsymbol{\beta}_0 + w_i\mathbf{x}_i\boldsymbol{\delta} - w_i\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\gamma\boldsymbol{\beta}_1 + e_i =$$

$$= (\mu_0 + \gamma\mu_1) + \left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\gamma\boldsymbol{\beta}_1 + w_i(\mu_1 - \mu_0 - \gamma\mu_1) + \mathbf{x}_i\boldsymbol{\beta}_0 + w_i\mathbf{x}_i\boldsymbol{\delta} - w_i\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)\gamma\boldsymbol{\beta}_1 + e_i +$$

$$+\left[w_i\bar{\mathbf{x}}\boldsymbol{\delta} - w_i\bar{\mathbf{x}}\boldsymbol{\delta}\right] + \left[w_i\mathrm{E}\left(\underbrace{\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j}_{\mathbf{v}_i} \right)\gamma\boldsymbol{\beta}_1 - w_i\mathrm{E}\left(\underbrace{\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j}_{\mathbf{v}_i} \right)\gamma\boldsymbol{\beta}_1 \right] =$$

$$= (\mu_0 + \gamma\mu_1) + w_i(\mu + \bar{\mathbf{x}}\boldsymbol{\delta} - \bar{\mathbf{v}}\boldsymbol{\lambda}) + \mathbf{x}_i\boldsymbol{\beta}_0 + w_i(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + \mathbf{v}_i\boldsymbol{\lambda} + w_i\bar{\mathbf{v}}\boldsymbol{\lambda} - w_i\mathbf{v}_i\boldsymbol{\lambda} + e_i =$$

$$= \eta + w_i \cdot \mathrm{ATE} + \mathbf{x}_i\boldsymbol{\beta}_0 + w_i(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + \left[\mathbf{v}_i + w_i(\bar{\mathbf{v}} - \mathbf{v}_i)\right]\boldsymbol{\lambda} + e_i$$

Therefore, we can conclude that:

$$y_i = \eta + w_i \cdot \mathrm{ATE} + \mathbf{x}_i\boldsymbol{\beta}_0 + w_i(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + \mathbf{z}_i\boldsymbol{\lambda} + e_i$$

where: $\mathbf{z}_i = \mathbf{v}_i + w_i(\bar{\mathbf{v}} - \mathbf{v}_i)$, $\mathbf{v}_i = \sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j$, $\bar{\mathbf{v}} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N_1}\omega_{ij}\mathbf{x}_j \right)$, $\boldsymbol{\lambda}=\gamma\boldsymbol{\beta}_1$, $\eta = \mu_0 + \gamma\mu_1$, and $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$

■

A4. *Ordinary Least Squares (OLS) consistency*.

Under Assumption 1 (CMI), the parameters of regression (14) can be consistently estimated by OLS. Indeed, it is immediate to see that the mean of e_i conditional on (w_i; $\mathbf{x}_i$) is equal to zero:

$$\mathrm{E}\left[\gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} + e_{0i} + w_i (e_{1i} - e_{0i}) - w_i \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} \,\middle|\, w_i, \mathbf{x}_i \right] =$$

$$\mathrm{E}\left[\gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} \,\middle|\, w_i, \mathbf{x}_i \right] + \mathrm{E}\left[e_{0i} \,\middle|\, w_i, \mathbf{x}_i \right] + \mathrm{E}\left[w_i (e_{1i} - e_{0i}) \,\middle|\, w_i, \mathbf{x}_i \right] - \mathrm{E}\left[w_i \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} \,\middle|\, w_i, \mathbf{x}_i \right] =$$

$$\gamma \sum_{j=1}^{N_1} \omega_{ij} \mathrm{E}\left[e_{1j} \,\middle|\, \mathbf{x}_i \right] + \mathrm{E}\left[e_{0i} \,\middle|\, \mathbf{x}_i \right] + w_i \mathrm{E}\left[(e_{1i} - e_{0i}) \,\middle|\, \mathbf{x}_i \right] - w_i \gamma \sum_{j=1}^{N_1} \omega_{ij} \mathrm{E}\left[e_{1j} \,\middle|\, \mathbf{x}_i \right] = 0$$

where $\eta = \mu_0 + \gamma \mu_1$. ■

Appendix B.

This appendix provides a Stata implementation of the model via the user-written command `ntreatreg` (Cerulli, 2015). The Stat syntax of `ntreatreg` takes on this form:

ntreatreg *outcome treatment varlist* **, hetero(***varlist_h***)**
spill(*matrix***) graphic**

where:

outcome: is the *y* of the previous model, i.e. the target variable of the policy considered.

treatment: is the *w* of the previous model, i.e. the binary policy (treatment) indicator.

varlist: is the **x** of the previous model, i.e. the vector of observable unit characteristics.

hetero(*varlist_h***)**: is an optional subset of **x** to allow for observable heterogeneity.

spill(*matrix***)**: is the weighting-matrix **Ω,** to be provided by the user.

graphic: returns a graph of the distribution of ATE(**x**), ATET(**x**) and ATENT(**x**).

In order to provide the interested reader with an operational estimation of our model, we perform an illustrative simulation exercise based on the data generating process (DGP) underlying the model presented in the paper, and fitted by `ntreatreg`. The Stata code is reported below where, for illustrative purposes, we consider a random treatment:

```
*************************************
* 1. Generate the matrix "omega"
*************************************
* Generate the matrix "omega"
. clear
. set matsize 1000 , permanently
. set obs 200
. set seed 10101
. gen w=rbinomial(1,0.5)
. gsort - w
. count if w==1
. global N1=r(N)
. global N0=_N-$N1
. mat def M=J(_N,_N,0)
. global N=_N

* Generate a matrix M from a Uniform distribution
forvalues i=1/$N{
forvalues j=1/$N1{
mat M[`i',`j']=runiform()
}
}

* Generate a vector SUM containing the column sum of M
mat def SUM=J(_N,1,0)
forvalues i=1/$N{
forvalues j=1/$N1{
```

```
mat SUM[`i',1] = SUM[`i',1] + M[`i',`j']
}
}

* Generate the matrix omega as defined in figure #
forvalues i=1/$N{
forvalues j=1/$N1{
mat M[`i',`j']=M[`i',`j']/SUM[`i',1]
}
}
mat omega=M

*****************************************************
* 2. Define the model's data generating process (DGP)
*****************************************************
* Declare a series of parameters
scalar mu1=2
scalar b11=5
scalar b12=3
scalar e1=rnormal()
scalar mu0=5
scalar b01=7
scalar b02=1
scalar e0=rnormal()
gen x1=rnormal()
gen x2=rnormal()
scalar gamma=0.8

* Sort the treatment so to have the "ones" first
gsort - w

* Generate "y1"
gen y1 = mu1 + x1*b11 + x2*b12 + e1
gen y1_obs=w*y1
mkmat y1_obs , mat(y1_obs)

* Generate "s"
mat s = omega*y1_obs
mat list s
svmat s

* Generate "y0" and finally "y"
gen y0 = mu0 + x1*b01 + x2*b02 + gamma*s1 + e0
gen y = y0 + w*(y1-y0)

* Generate the treatment effect "te"
gen te=y1-y0
sum te

* Put the ATE into a scalar
scalar ATE=r(mean)
di ATE

************************************
* 3. Estimate the model using ntreatreg
************************************
* y: dependent variable
* w: treatment
* x: [x1; x2] are the covariates
* Matrix of spillovers: OMEGA

* Estimate the model using "NTREATREG" ///
```

```
set more off
xi: ntreatreg y w x1 x2  , ///
hetero(x1 x2) spill(omega) graphic
scalar ate_neigh = _b[w]  // put ATE into a scalar
di ate_neigh

* END OF THE SIMULATION
```

Previous Stata code: (i) starts by providing the matrix **Ω**; (ii) forms the model DGP as defined in (7); (iii) estimates the model by `ntreatreg` using the DGP simulated data.

By running this code, we get a value of ATE as predicted by `ntreatreg` equal to -3.854401, which is equal to the DGP value of the ATE, which is set to -3.8544013. We can run many simulations getting similar results.