



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

**Give to AgEcon Search**

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

The Stata Journal (2017)  
17, Number 4, pp. 803–833

# Identification and estimation of treatment effects in the presence of (correlated) neighborhood interactions: Model and Stata implementation via `ntreatreg`

Giovanni Cerulli  
CNR-IRCrES  
National Research Council of Italy  
Research Institute on Sustainable Economic Growth  
Rome, Italy  
giovanni.cerulli@ircres.cnr.it

**Abstract.** In this article, I present a counterfactual model identifying average treatment effects by conditional mean independence when considering peer- or neighborhood-correlated effects, and I provide a new command, `ntreatreg`, that implements such models in practical applications. The model and its accompanying command provide an estimation of average treatment effects when the stable unit treatment-value assumption is relaxed under specific conditions. I present two instructional applications: the first is a simulation exercise that shows both model implementation and `ntreatreg` correctness; the second is an application to real data, aimed at measuring the effect of housing location on crime in the presence of social interactions. In the second application, results are compared with a no-interaction setting.

**Keywords:** st0499, `ntreatreg`, ATEs, Rubin’s causal model, SUTVA, neighborhood effects

## 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 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 to consistently estimate average treatment effects (ATEs). It is thus clear that, if interference (or interaction) among units is not properly accounted for, traditional program evaluation methods such as regression adjustment,

selection models, matching, or reweighting are bound to be biased estimations of the actual treatment effect (TE).<sup>1</sup>

Rubin (1978) calls this important assumption a stable unit treatment-value assumption (SUTVA), whereas Manski (2013) calls it “individualistic treatment response” to emphasize that it restricts the form of the treatment response function that the analyst considers. SUTVA (or individualistic treatment response) implies that the treatment applied to a specific individual affects only the outcome of that individual. This means that potential “externality effects” flowing, for instance, from treated to untreated subjects are strictly ruled out.

This article is an attempt to partially relax this assumption; by excluding the alternative, it operationalizes the estimation of ATEs when peer effects are assumed to flow from treated to untreated units. This restriction, reasonable in specific contexts, allows for straightforward identification and estimation of TEs simply by invoking conditional mean independence (CMI). Although demanding, this restriction seems a valuable attempt to weaken SUTVA, even though its complete removal would require a more general approach.

Some epidemiological studies have addressed the interference topic by restricting the analysis to experimental settings with randomized treatment (see, for instance, Rosenbaum [2007]; Hudgens and Halloran [2008]; Tchetgen Tchetgen and VanderWeele [2010]; and Robins, Hernán, and Brumback [2000]). However, in this article, I move along the line traced by econometric studies, normally dealing with nonexperimental settings with sample selection (that is, no random assignment to treatment is assumed). Thus, an ex-post evaluation is envisaged (Sobel 2006). In particular, I work within the binary POM that I attempt to partially generalize to account for the presence of neighborhood effects. My theoretical reference draws upon previous works dealing with TE identification in the presence of peer effects, particularly the works by Manski (1993, 2013).

I provide a new community-contributed command, `ntreatreg`, to operationalize the estimation of the suggested model in practical applications. Stata provides a powerful package, `teffects`, for estimating TEs for observational data. `teffects` can also estimate many valuable community-contributed TE routines available for similar and more advanced purposes. However, neither official nor community-contributed commands have been provided so far to incorporate peer effects in TE estimation. The `ntreatreg` command is a first attempt of this incorporation. Therefore, it is a valuable tool for estimating ATEs for observational data when the SUTVA is relaxed according to specific conditions. Such conditions often characterize some biomedical and socioeconomic contexts of application.

---

1. The applied literature on the socioeconomics of peer effect is rather vast; here we focus on that related to peer (or neighborhood) effects within Rubin’s potential outcome model (POM). 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. His article also provides a brief survey of the literature on the subject.

This article is organized as follows: Section 2 presents some related literature and positions my approach within it. Section 3 sets out the model, its assumptions, and its propositions. Section 4 presents the model's estimation procedure. Section 5 puts forward the software implementation of the model via the community-contributed command `ntreatreg` and provides a simulative illustration by setting out the model's data-generating process (DGP); section 5.7 presents a utility command, `mkomega`, for when users want to compute a similarity matrix based on a list of covariates to be then inserted into `ntreatreg`. Section 6 illustrates an application of `ntreatreg` to real data by investigating the effect of housing location on crime; here a comparison with a no-interaction setting is also performed by using the companion routine `ivtreatreg` (Cerulli 2014). Section 7 concludes the article. Appendix A then sets out the proof of each proposition.

## 2 Related literature

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

Rosenbaum (2007) was among the first scholars to pave the way to generalize the standard randomization statistical approach for comparing different treatments with the case of units' interference. He presented a statistical model in which a 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 prespecified groups and interference between groups is ruled out.

In the same vein, Sobel (2006) provided a definition, identification, and estimation strategy for traditional ATE estimators when interference between units is allowed by using the Moving to Opportunity for Fair Housing randomized social experiment as an example. In his article, he interchangeably uses the terms "interference" and "spillover" to account for the presence of such externalities. Interestingly, he shows that a potential bias can arise when no interference is erroneously assumed, and he defines a series of direct and indirect TEs that may be identified under reasonable assumptions. Additionally, he shows some interesting links between the form of his estimators under interference and the local ATE estimator provided by Imbens and Angrist (1994), thus showing that—under interference—TEs can be identified only on specific subpopulations.

The article by Hudgens and Halloran (2008) is probably the most relevant of this literature, because these authors develop a rather general and rigorous modeling of the statistical treatment setting under randomization when interference is potentially present. Furthermore, their approach also paves the way for extensions to observational settings. Starting from the same two-stage randomization approach of Rosenbaum (2007), these authors manage to go substantially further by providing a precise characterization of the causal effects with interference in randomized trials also encompassing Sobel's approach. They define direct, indirect, total, and overall causal effects, showing the relation be-

tween these measures and providing an unbiased estimator of the upper bound of their variance.

Tchetgen Tchetgen and VanderWeele's (2010) article follows in the footsteps traced by the approach of Hudgens and Halloran (2008), providing 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 estimator when interference is present.

Aronow and Samii (2013) generalize 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 ( $w_{-i}$ ), that is,

$$y_i(w_i; w_{-i}) \quad (1)$$

which entails that—with  $N$  units and a binary treatment, for instance—a number of  $2^N$  potential outcomes may arise. Nevertheless, an alternative way of modeling unit  $i$ 's potential outcome is assuming

$$y_i(w_i; y_{-i}) \quad (2)$$

where  $y_{-i}$  is the  $(N - 1) \times 1$  vector of other units' potential outcomes, excluding unit  $i$ 's potential outcome.<sup>2</sup> The notion of interference entailed by expression (2) is different from that implied by expression (1). The latter, however, is consistent with the notion of "endogenous" neighborhood effects provided by Manski (1993, 532–533). Manski, in fact, identifies three types of effects corresponding to three arguments of the individual (potential) outcome equation incorporating social effects:<sup>3</sup>

1. *Endogenous effects.* These effects entail that the outcome of an individual depends on the outcomes of other individuals belonging to the same neighborhood.
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 the same neighborhood.

---

2. A combined regression model, including both individual treatments and outcomes, may be expressed as

$$y_i = f(w_i; y_{-i}; w_{-i})$$

Arduini, Patacchini, and Rainone (2014) provide a first attempt to modeling such a regression on individuals eligible for treatment, showing that the coefficient of  $w_i$  (that is, 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 POM. I instead provide a POM-consistent approach generalized to the case of possible interaction among units.

3. The literature is not homogeneous in singling out a unique name of such effects; dependent on context, authors interchangeably refer to peer, neighborhood, social, club, interference, or interaction effects.

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

Contextual and correlated effects are exogenous because they clearly depend on predetermined characteristics of the individuals in the neighborhood (case 2) or of the neighborhood itself (case 3). Endogenous effects are of broader interest because they are affected by the behavior (measured as “outcome”) of other individuals involved in the same neighborhood. This means that endogenous effects comprise both direct and indirect effects linked to a given external intervention on individuals.

The model proposed in this article assumes that the potential untreated outcome depends on treated units’ potential outcomes. However, because the latter are assumed to depend on a set of observable exogenous confounders in the presence of uncorrelated unobservables, this model fits only “correlated effects” as defined in Manski’s taxonomy.

To concisely position this article within the literature, I will say that previous contributions assume the following:

- i) The unit potential outcome depends on its own treatment and other units’ treatment.
- ii) The assignment is randomized or conditionally unconfounded.
- iii) The treatment is multiple.
- iv) Potential outcomes have a nonparametric form.

In this article, I instead assume the following:

- i) The unit potential outcome depends on its own treatment and other units’ potential outcome.
- ii) The assignment is mean conditionally unconfounded.
- iii) The treatment is binary.
- iv) Potential outcomes have a parametric form.

The model developed here is part of a broader class of regression-adjustment TE models. These models are suitable for observational studies, though it should be recognized that regression adjustment does not provide consistent estimates if a “killing” unobservable confounder is at work. The conditional unconfoundedness assumption (ii, above) upon which the estimator relies is shared by other wellknown estimators, such as matching and (inverse-probability) reweighting. In this sense, assuming absence of correlation between the treatment variables and the one measuring neighborhood is essential to the identification of this type of model. Therefore, in this article, I suggest a simple but workable way to relax SUTVA, one that seems easy to implement in many biomedical and socioeconomic contexts of application.

### 3 A binary treatment model with “correlated” neighborhood effects

This section presents a model for fitting the ATEs of a policy program (or a treatment) in a nonexperimental setting in the presence of “correlated” neighborhood (or externality) interactions. We consider a binary treatment variable  $w$ —taking a value of 1 for treated and 0 for untreated units—that is 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;  $N_1$  is the number of treated units;  $N_0$  is the number of untreated units;  $w_i$  is the treatment variable assuming a value of 1 if unit  $i$  is treated and 0 if it is untreated;  $y_{1i}$  is the outcome of unit  $i$  when the individual is treated;  $y_{0i}$  is the outcome of unit  $i$  when the individual is untreated;  $\mathbf{x}_i = (x_{1i}, x_{2i}, x_{3i}, \dots, x_{Mi})$  is a row vector of  $M$  exogenous observable characteristics for unit  $i = 1, \dots, N$ .

To begin with, as usual in this literature, we define unit  $i$ ’s TE as

$$\text{TE}_i = y_{1i} - y_{0i} \quad (3)$$

$\text{TE}_i$  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 ( $y_0$ ). Because  $\text{TE}_i$  refers to the same individual at the same time, the analyst can observe just one of the two quantities feeding into (3), never both. For instance, it might be the case that we can observe the investment behavior 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 to reliably recover 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 across all units  $i$ , and both  $y_{1i}$  and  $y_{0i}$  are generally explained by a structural part depending on observable factors and a nonstructural part depending on an unobservable (error) term. Nevertheless, recovering the entire distributions of  $y_{1i}$  and  $y_{0i}$  (and, consequently, the distribution of  $\text{TE}_i$ ) may be too demanding without very strong assumptions. Therefore, the literature has focused on estimating specific moments of these distributions, particularly the “mean”, thus defining the so-called population ATE and ATE conditional on  $\mathbf{x}_i$  [that is,  $\text{ATE}(\mathbf{x}_i)$ ] of a policy intervention as

$$\text{ATE} = E(y_{i1} - y_{i0}) \quad (4)$$

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

where  $E(\cdot)$  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,  $\text{ATE} = E_{\mathbf{x}}\{\text{ATE}(\mathbf{x})\}$ .

Given the definitions of the unconditional and conditional ATE in (4) and (5), respectively, one can define the same parameters in the subpopulation of treated (ATET) and untreated (ATENT) units, that is,

$$\begin{aligned}\text{ATET} &= E(y_{i1} - y_{i0} | w_i = 1) \\ \text{ATET}(\mathbf{x}_i) &= E(y_{i1} - y_{i0} | \mathbf{x}_i, w_i = 1)\end{aligned}$$

and

$$\begin{aligned}\text{ATENT} &= E(y_{i1} - y_{i0} | w_i = 0) \\ \text{ATENT}(\mathbf{x}_i) &= E(y_{i1} - y_{i0} | \mathbf{x}_i, w_i = 0)\end{aligned}$$

In this article, I aim to provide consistent parametric estimation of all previous quantities (ATEs) in the presence of neighborhood effects.

To that end, let us start with what is observable to the analyst in such a setting, that is, the actual status of the unit  $i$ , which can be obtained as

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

Equation (6) is Rubin's POM, which is the fundamental relation linking the unobservable to the observable outcome. Given (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  $\{y_{0i}, y_{1i}, w_i, \mathbf{x}_i\}$  as defined above, the following equalities hold:

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

Hence, throughout this article, we will assume unconfoundedness (that is, CMI) to hold. As we will see, CMI is a sufficient condition for identifying ATEs also when neighborhood effects are considered.

Once CMI has been assumed, we then need to properly model the potential outcomes  $y_{0i}$  and  $y_{1i}$  to get a representation of the ATEs (that is, ATE, ATET, and ATENT) while accounting for the presence of correlated externality effects. In this article, we will simplify our analysis further 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

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



Assumption 2 poses two important restrictions to the form given to the potential outcomes: i) it makes them dependent on some unknown parameters  $\theta$  (that is, the parametric form), and ii) it entails that the externality effect occurs only in one direction, that is, from the treated individuals to the untreated, while excluding the alternative.<sup>4</sup>

- *Assumption 3. Linearity and weighting matrix.* We assume that the potential outcomes are linear in the parameters and that an  $N \times N$  weighting matrix  $\Omega$  of exogenous constant numbers is known.

Under assumptions 1, 2, and 3, the model takes on the form

$$\begin{cases} y_{1i} = \mu_1 + \mathbf{x}_i\beta_1 + e_{1i} \\ y_{0i} = \mu_0 + \mathbf{x}_i\beta_0 + \gamma s_i + e_{0i} \\ s_i = \sum_{j=1}^{N_1} \omega_{ij} y_{1j}, \quad \text{with} \quad \sum_{j=1}^{N_1} \omega_{ij} = 1 \\ y_i = y_{0i} + w(y_{1i} - y_{0i}) \\ \text{CMI holds} \end{cases} \quad (7)$$

where  $i = 1, \dots, N$  and  $j = 1, \dots, N_1$ ;  $\mu_1$  and  $\mu_0$  are scalars;  $\beta_0$  and  $\beta_1$  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 an unconditional mean of 0 and a constant variance; and  $s_i$  represents the unit  $i$ th neighborhood effect because of the treatment administered to unit  $j$  ( $j = 1, \dots, N_1$ ). Observe that, by linearity,<sup>5</sup> we have

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

where the parameter  $\omega_{ij}$  is the generic element of the weighting matrix  $\Omega$  expressing some form of distance between unit  $i$  and unit  $j$ . Although not strictly required for consistency, we also assume that these weights sum to 1, that is,  $\sum_{j=1}^{N_1} \omega_{ij} = 1$ . In short, previous assumptions say that unit  $i$ 's neighborhood 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

$$y_{0i} = \mu_0 + \mathbf{x}_i\beta_0 + \gamma \sum_{j=1}^{N_1} \omega_{ij} y_{1j} + e_{0i}$$

4. In the more general case in which peer effects take place from treated to untreated units and vice versa, identifying ATEs consistent with the POM becomes trickier because various feedback terms do arise. Using a spatial regression approach, [Arduini, Patacchini, and Rainone \(2014\)](#) estimate a TE reduced form that also includes feedback terms. However, their model is not directly derived using the POM, unlike the model in this article.

5. The linearity of the spillover effect is an assumption needed to simplify the subsequent regression analysis. However, nonlinear forms might also be used. Some sensitivity analysis could show how results change according to different mathematical forms of the spillover effect.

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

Let us now consider four propositions implied by the previous assumptions.

- *Proposition 1. Formula of ATE with neighborhood interactions.* Given assumptions 2 and 3 and the implied equations established in (7), the ATE with neighborhood interactions takes on the form

$$\begin{aligned} \text{ATE} &= E(y_{1i} - y_{0i}) = \mu + E \left\{ \mathbf{x}_i \boldsymbol{\delta} - \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \gamma \boldsymbol{\beta}_1 - e_i \right\} \\ &= \mu + \bar{\mathbf{x}} \boldsymbol{\delta} - \bar{\mathbf{v}} \boldsymbol{\lambda} \end{aligned} \quad (9)$$

where  $\boldsymbol{\lambda} = \gamma \boldsymbol{\beta}_1$ ,  $\bar{\mathbf{x}} = E(\mathbf{x}_i)$ ,  $\bar{\mathbf{v}} = E(\underbrace{\sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j}_{\mathbf{v}_i})$  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.1.

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

$$\begin{aligned} \text{ATE} &= E(y_{1i} - y_{0i}) \\ &= E \left\{ (\mu_1 + \mathbf{x}_i \boldsymbol{\beta}_1 + e_{1i}) - \left( \mu_0 + \mathbf{x}_i \boldsymbol{\beta}_0 + \gamma \sum_{j=1}^{N_1} \omega_{ij} y_{1j} + e_{0i} \right) \right\} \end{aligned} \quad (10)$$

where

$$\begin{aligned} \sum_{j=1}^{N_1} \omega_{ij} y_{1j} &= \sum_{j=1}^{N_1} \omega_{ij} (\mu_1 + \mathbf{x}_j \boldsymbol{\beta}_1 + e_{1j}) \\ &= \mu_1 \sum_{j=1}^{N_1} \omega_{ij} + \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \boldsymbol{\beta}_1 + \sum_{j=1}^{N_1} \omega_{ij} e_{1j} \\ &= \mu_1 + \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \boldsymbol{\beta}_1 + \sum_{j=1}^{N_1} \omega_{ij} e_{1j} \end{aligned} \quad (11)$$

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

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

$$\text{ATE}(\mathbf{x}_i) = E(y_{1i} - y_{0i} | \mathbf{x}_i) = \text{ATE} + (\mathbf{x}_i - \bar{\mathbf{x}}) \boldsymbol{\delta} + (\bar{\mathbf{v}} - \mathbf{v}_i) \boldsymbol{\lambda} \quad (12)$$

where it is now easy to see that  $\text{ATE} = E_{\mathbf{x}}\{\text{ATE}(\mathbf{x})\}$ . The proof is in appendix A.2.

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

$$y_i = \eta + w_i \times \text{ATE} + \mathbf{x}_i \boldsymbol{\beta}_0 + w_i(\mathbf{x}_i - \bar{\mathbf{x}}) \boldsymbol{\delta} + \mathbf{z}_i \boldsymbol{\lambda} + e_i \quad (13)$$

where  $\mathbf{z}_i = \mathbf{v}_i + w_i(\bar{\mathbf{v}} - \mathbf{v}_i)$ ,  $\mathbf{v}_i = \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j$ ,  $\bar{\mathbf{v}} = 1/N \sum_{i=1}^N (\sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j)$ ,  $\boldsymbol{\lambda} = \gamma \boldsymbol{\beta}_1$ ,  $\eta = \mu_0 + \gamma \mu_1$ , and  $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$ . The proof is in appendix A.3.

- *Proposition 4. Ordinary least-squares (OLS) consistency.* Under assumptions 1 (CMI), 2, and 3, the error term of regression (12) has a mean of 0 conditional on  $(w_i, \mathbf{x}_i)$ , that is,

$$\begin{aligned} E(e_i | w_i, \mathbf{x}_i) &= E \left\{ \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} + e_{0i} + w_i(e_{1i} - e_{0i}) - w_i \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} | w_i, \mathbf{x}_i \right\} \\ &= 0 \end{aligned}$$

thus implying that (12) is a regression model with parameters that can be consistently estimated by OLS. The proof is in appendix A.4.

Once a consistent estimation of the parameters of (12) is obtained, we can estimate the ATE directly from the regression, and we can estimate  $\text{ATE}(\mathbf{x}_i)$  by plugging the estimated parameters into (11). This is because the plug-in estimator of  $\text{ATE}(\mathbf{x}_i)$  becomes a function of consistent estimates and thus becomes consistent itself:

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

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

Once the formulas for  $\widehat{\text{ATE}}$  and  $\widehat{\text{ATE}}(\mathbf{x}_i)$  are available, it is also possible to recover the  $\widehat{\text{ATE}}_{\text{T}}$  and the  $\widehat{\text{ATE}}_{\text{NT}}$  as

$$\widehat{\text{ATE}}_{\text{T}} = \widehat{\text{ATE}} + \frac{1}{\sum_{i=1}^N w_i} \sum_{i=1}^N w_i \left\{ (\mathbf{x}_i - \bar{\mathbf{x}}) \hat{\boldsymbol{\delta}} + (\bar{\mathbf{v}} - \mathbf{v}_i) \hat{\boldsymbol{\lambda}} \right\}$$

and

$$\widehat{\text{ATE}}_{\text{NT}} = \widehat{\text{ATE}} + \frac{1}{\sum_{i=1}^N (1 - w_i)} \sum_{i=1}^N (1 - w_i) \left\{ (\mathbf{x}_i - \bar{\mathbf{x}}) \hat{\boldsymbol{\delta}} + (\bar{\mathbf{v}} - \mathbf{v}_i) \hat{\boldsymbol{\lambda}} \right\}$$

These quantities are functions of observable components and parameters consistently estimated by OLS (see the next section). Once these estimates are available, standard errors for  $\widehat{\text{ATE}}_{\text{T}}$  and  $\widehat{\text{ATE}}_{\text{NT}}$  can be correctly obtained via bootstrapping (see Wooldridge [2010, 911–919]).

## 4 Estimation

Starting from the previous section's results, I suggest a simple protocol for estimating ATEs. Given an independent and identically distributed sample of observed variables for each individual  $i$ ,

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

1. Provide a weighting matrix  $\mathbf{\Omega} = [\omega_{ij}]$  measuring some type of distance between the generic unit  $i$  (untreated) and unit  $j$  (treated);
2. Using OLS, fit 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  $\{\hat{\beta}_0, \hat{\delta}, \hat{\gamma}, \hat{\beta}_1\}$  and put them into the formulas of  $\widehat{\text{ATEs}}$ .

By comparing the formulas of the ATE with ( $\gamma \neq 0$ ) and without ( $\gamma = 0$ ) the neighborhood effect, we define the estimated neighborhood bias as

$$\begin{aligned} \text{Bias} &= |\text{ATE}_{\text{without}} - \text{ATE}_{\text{with}}| = |\gamma\mu_1 + \bar{\mathbf{v}}\boldsymbol{\lambda}| \\ &= \left| \gamma\mu_1 + \left\{ \frac{1}{N} \sum_{i=1}^N \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \right\} \boldsymbol{\lambda} \right| \end{aligned}$$

This is the bias arising when one neglects peer effect in assessing TE in observational studies: it depends on the weights employed, the average of the observable confounders considered in  $\mathbf{x}$ , and the magnitude of the coefficients  $\gamma$  and  $\beta_1$ . Such bias may be positive or negative.

Furthermore, by defining

$$\gamma\beta_1 = \boldsymbol{\lambda}$$

it is also possible to determine whether this bias is statistically significant by simply testing the following null hypothesis:

$$H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_M = 0$$

If this hypothesis is rejected, we cannot exclude that neighborhood effects are in place, thus significantly affecting the estimation of the causal parameters' ATEs. In a similar way, we can also obtain an estimation of the neighborhood bias for ATET and ATENT.

## 5 Implementation via `ntreatreg`

The previous model can easily be fit using the new command `ntreatreg`, which has the syntax given below.

## 5.1 Syntax

```
ntreatreg outcome treatment [varlist] [if] [in] [weight], spill(matrix)
    [hetero(varlist_h) conf(#) graphic save_graph(filename) vce(robust)
    const(noconstant) head(noheader) beta]
```

`fweights`, `iweights`, and `pweights` are allowed; see [U] 11.1.6 **weight**.

## 5.2 Description

`ntreatreg` estimates ATEs under CMI when neighborhood interactions may be present. It incorporates such externalities within the traditional Rubin POM. As such, it provides an attempt to relax the SUTVA generally assumed in observational studies.

## 5.3 Options

`spill(matrix)` specifies the adjacent (weighted) matrix used to define the presence and strength of the units' relationships. It could be a distance matrix, with distance loosely defined as either vector or spatial. `spill()` is required.

`hetero(varlist_h)` specifies the variables over which one calculates the idiosyncratic  $ATE(\mathbf{x})$ ,  $ATET(\mathbf{x})$ , and  $ATENT(\mathbf{x})$ , where  $\mathbf{x} = \text{varlist}_h$ . The default is that the command fits the specified model without a heterogeneous average effect. `varlist_h` should be the same set or a subset of the variables specified in `varlist`.

`conf(#)` sets the confidence level equal to the number specified by the user. The default is `conf(95)`.

`graphic` requests a graphical representation of the density distributions of  $ATE(\mathbf{x})$ ,  $ATET(\mathbf{x})$ , and  $ATENT(\mathbf{x})$ . It gives an outcome only if variables are specified in `hetero()`.

`save_graph(filename)` saves in `filename` the graph obtained with the option `graphic`.

`vce(robust)` allows for robust regression standard errors.

`const(noconstant)` suppresses the regression constant term.

`head(noheader)` suppresses the header.

`beta` reports standardized beta coefficients.

## 5.4 Remarks

`ntreatreg` creates the following variables:

`_ws_varname_h` is an additional regressor used in the model's regression when the option `hetero()` is specified.

`_z_varname_h` is a spillover additional regressor.

`_v_varname_h` is the first spillover component of `_z_varname_h`.

`_ws_v_varname_h` is the second spillover component of `_z_varname_h`.

`ATE_x` is an estimate of the idiosyncratic ATE given  $\mathbf{x}$ .

`ATET_x` is an estimate of the idiosyncratic ATET given  $\mathbf{x}$ .

`ATENT_x` is an estimate of the idiosyncratic ATENT given  $\mathbf{x}$ .

## 5.5 Stored results

`ntreatreg` stores the following in `e()`:

Scalars	
<code>e(N_tot)</code>	total number of (used) observations
<code>e(N_treat)</code>	number of (used) treated units
<code>e(N_untreat)</code>	number of (used) untreated units
<code>e(ate)</code>	value of the ATE
<code>e(atet)</code>	value of the ATET
<code>e(atent)</code>	value of the ATENT

## 5.6 Simulation exercise

To provide an operational estimation of our model, we first perform an illustrative simulation exercise based on the DGP underlying the model and illustrated in (7). This step is useful, both to show that the model has a relatively simple computational counterpart and to test the syntactic and semantic correctness of `ntreatreg` as a command.

The code to properly reproduce equations in system (7) appears below. For illustrative purposes, and with no loss of generality, we consider a random treatment:

```
. ***** START SIMULATION *****
. *****
. * Step 1. Generate the matrix omega
. *****
. * Generate the matrix omega
. clear all
. set matsize 1000
. set obs 200
number of observations (_N) was 0, now 200
. set seed 10101
. generate w=rbinomial(1,0.5)
```

```

. gsort - w
. count if w==1
100
. global N1=r(N)
. global NO=_N-$N1
. matrix def M=J(_N,_N,0)
. global N=_N
. forvalues i=1/$N {
2. forvalues j=1/$N1 {
3. matrix M[`i`,`j']=runiform()
4. }
5. }
. matrix define SUM=J(_N,1,0)
. forvalues i=1/$N {
2. forvalues j=1/$N1 {
3. matrix SUM[`i',1] = SUM[`i',1] + M[`i`,`j']
4. }
5. }
. forvalues i=1/$N {
2. forvalues j=1/$N1 {
3. matrix M[`i`,`j']=M[`i`,`j']/SUM[`i',1]
4. }
5. }
. matrix omega=M
. *****
. * Step 2. Define the models data generating process (DGP)
. *****
. * Declare a series of parameters
. scalar mu1=2
. scalar b11=5
. scalar b12=3
. scalar b13=9
. scalar e1=rnormal()
. scalar mu0=5
. scalar b01=7
. scalar b02=1
. scalar b03=6
. scalar e0=rnormal()
. generate x1=rnormal()
. generate x2=rnormal()
. generate x3=5+3*rnormal()
. scalar gamma=0.8
. * Sort the treatment so to have the ones first
. gsort - w
. * Generate y1
. generate y1 = mu1 + x1*b11 + x2*b12 + e1
. generate y1_obs=w*y1
. mkmat y1_obs, mat(y1_obs)
. * Generate s
. matrix s = omega*y1_obs

```

```

. svmat s
. * Generate y0 and finally y
. generate y0 = mu0 + x1*b01 + x2*b02 + gamma*s1 + e0
. generate y = y0 + w*(y1-y0)
. * Generate the treatment effect te
. generate te=y1-y0
. summarize te

```

Variable	Obs	Mean	Std. Dev.	Min	Max
te	200	-3.140086	2.849946	-12.89648	4.528738

```

. * Put the ATE into a scalar
. scalar ATE=r(mean)
. display ATE
-3.1400858
. *****
. * Step 3. Fit the model using ntreatreg
. *****
. * y: dependent variable
. * w: treatment
. * x: [x1; x2] are the covariates
. * Matrix of spillovers: OMEGA
. * Fit the model using ntreatreg
. set more off
. ntreatreg y w x1 x2, hetero(x1) spill(omega) graphic

```

Source	SS	df	MS	Number of obs	=	200
Model	8479.76569	6	1413.29428	F(6, 193)	=	1382.01
Residual	197.36846	193	1.02263451	Prob > F	=	0.0000
				R-squared	=	0.9773
				Adj R-squared	=	0.9765
Total	8677.13415	199	43.6036892	Root MSE	=	1.0113

```


```

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
w	-3.133691	.1453545	-21.56	0.000	-3.420379 -2.847004
x1	6.869514	.1046721	65.63	0.000	6.663066 7.075962
x2	2.052399	.0718827	28.55	0.000	1.910622 2.194175
_ws_x1	-1.955304	.146996	-13.30	0.000	-2.245228 -1.665379
_z_x1	7.800202	2.114743	3.69	0.000	3.629227 11.97118
_z_x2	-1.066327	1.683491	-0.63	0.527	-4.386729 2.254075
_cons	3.107643	.4468559	6.95	0.000	2.226295 3.988991

```

(200 real changes made)
(200 real changes made)
(200 real changes made)
(100 missing values generated)
(100 missing values generated)
. scalar ate_neigh = _b[w] // put ATE into a scalar
. display ate_neigh
-3.1336913

```



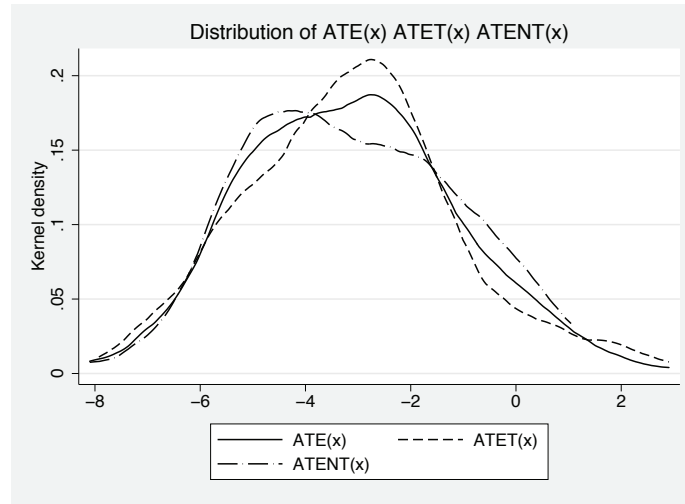


Figure 1. Resulting graph from the `ntreatreg` command with the `graphic` option

This simulation code deserves some comments:

- Step 1 provides a  $200 \times 200$  matrix,  $\Omega$ . This matrix is built by first generating a matrix  $\mathbf{M}$  of the same size as  $\Omega$  from a  $[0 - 1]$  uniform distribution and then dividing  $\mathbf{M}$  by its column sums. This last step is necessary to allow  $\Omega$  to become a matrix of weights, as entailed by the third line of system (7).
- Step 2 reproduces the model DGP as defined in system (7) by giving the potential outcomes  $y_1$  and  $y_0$  a linear form. We first need to generate  $y_1$  and, given this, the spillover variable  $s$ , which serves in turn as an explanatory variable for generating (with  $x_1$  and  $x_2$ ) the potential outcome of untreated units as entailed by the second line of system (7). Finally, by applying the potential outcome equation, we are able to generate the observable outcome of this process,  $y$ , and thus the TE for each unit (that is, the variable `te`). Given this, the “true” DGP’s ATE is obtained as the mean of `te`, which in this case is equal to  $-3.1401$ .
- Step 3 fits the model generated by the previous DGP by using `ntreatreg`. If it is correct, we expect `ntreatreg` to provide a value of ATE close to the “true” one, that is,  $-3.1401$ . We immediately see that `ntreatreg` estimates a statistically significant ATE equal to  $-3.1337$ , which is strictly close to the one provided by our simulated DGP. By running the code several times, we obtain similar outcomes (not reported). We can conclude that `ntreatreg` is reliable both syntactically and semantically. Moreover, the `graphic` option of `ntreatreg` allows one to draw the distribution of  $\text{ATE}(\mathbf{x})$ ,  $\text{ATET}(\mathbf{x})$ , and  $\text{ATENT}(\mathbf{x})$  when SUTVA is relaxed according to the assumptions underlying the model. Finally, given the random nature of the treatment assumed in the simulated DGP, it is not surprising we discovered a similar shape for the distribution of  $\text{ATE}(\mathbf{x})$ ,  $\text{ATET}(\mathbf{x})$ , and  $\text{ATENT}(\mathbf{x})$ .

## 5.7 The `mkomega` command

In the previous example, we provided the code to generate the neighborhood matrix  $\Omega$  by simulation. This was useful to understand the correct form of  $\Omega$  to insert into `ntreatreg`. However, a dedicated command to generate such a similarity matrix based on a set of variables (that is, covariates' vector distance) comes in handy. The command `mkomega` does this task, because it computes units' similarity matrices using the variables declared in *varlist*. Two types of similarity matrices are optionally computed by this routine: the correlation matrix and the inverse Euclidean distance matrix.<sup>6</sup> The syntax of `mkomega` is set out below.

### Syntax

```
mkomega treatment varlist [if] [in], sim_measure(type) out(outcome)
```

*treatment* is a binary variable taking a value of 1 for treated units and 0 for untreated ones. It is the same treatment variable that the user is going to specify in `ntreatreg`.

*varlist* is a list of numeric variables on which to build the distance measure. These variables should be of numeric significance, not categorical. Some of these variables might be specified as confounders in `ntreatreg`.

### Description

`mkomega` computes a unit's similarity matrix using the variables declared in *varlist* to be later used in the command `ntreatreg`. Two types of similarity matrices are optionally allowed by this command: the correlation matrix and the inverse Euclidean distance matrix.

### Options

`sim_measure(type)` specifies the similarity matrix to use. `sim_measure()` is required. *type* may be `corr`, for the correlation matrix, or `L2`, for the inverse Euclidean distance matrix.

`out(outcome)` specifies the outcome variable one is going to use in `ntreatreg`. `out()` is required.

---

6. Geographical distance can sometimes more suitably catch TE transmission. For this reason, the `mkomega` command provides the option `sim_measure(L2)` to calculate an (inverse) Euclidean distance matrix based on *varlist*. Indeed, if *varlist* is made of two variables identifying geographical coordinates (for instance, `X1` = latitude and `X2` = longitude), then using option `sim_measure(L2)` allows for an estimation of ATE adjusted for spillovers based on (pure) geographical distances.

### Stored results

`mkomega` stores the following in `r()`:

Scalars	
<code>r(N1)</code>	number of treated units
<code>r(N0)</code>	number of untreated units
Matrices	
<code>r(M)</code>	similarity matrix

## 6 Application to real data: The effect of housing location on crime

In this application, we consider the dataset `spatial.columbus.dta` provided by [Anselin \(1988\)](#) containing information on property crimes in 49 neighborhoods in Columbus, Ohio, in 1980.

The aim of this illustrative application is to determine the effect of housing location on crimes, that is, the causal effect of the variable `cp`—taking a value of 1 if the neighborhood is located in the “core” of the city and 0 if it is located in the “periphery”—on the number of residential burglaries and vehicle thefts per thousand households (that is, the variable `crime`).

Several conditioning (or confounding) observable factors are included in the dataset. Here we consider only two of them: household income in \$1,000s (`inc`) and housing value in \$1,000s (`hval`).

We are interested in detecting the effect of housing location on the number of crimes in such a setting, by taking into account possible interactions among neighborhoods. Our research presumption is that the number of burglaries in a specific peripheral neighborhood is not only affected by the neighborhood’s idiosyncratic characteristics (the variables `inc` and `hval`) but also by the number of burglaries that occur in the core neighborhoods. The conjecture behind this statement is that a saturation effect may take place, with thieves more prone to move toward peripheral areas of the city when core areas become saturated. This may be due to, for instance, new installations of security systems in the core whenever a certain amount of burglaries occur.

To build the matrix  $\Omega$ , we use the command `mkomega` with option type L2, which calculates the Euclidean distance matrix associated with the covariates `x` and `y`, representing respectively the longitude and latitude of the neighborhood. This matrix presents values that are greater than 0. The assumption here is that geographical distance properly catches interactions among different neighborhoods.

After running `mkomega` to generate  $\Omega$  as in system (7), we consider the relation between `crime` and `cp` using `inc` and `hval` as regressors subject to observable heterogeneity. We then estimate the model presented here by comparing the estimates obtained using `ntreatreg`, which takes into account peer effects, with those obtained

using `ivtreatreg` (Cerulli 2014), which estimates the same model but without accounting for peer effects.

The coefficient of the treatment variable, `cp`, is equal to 14.6 in the regression incorporating peer effects and equal to 13.6 in the one not incorporating them. Both are significant. The adjusted  $R$ -squared is rather high and similar in the two regressions. The percentage bias is around 7%. By performing a test to see whether the coefficients of the peer effects are jointly 0 (that is,  $H_0 : \gamma\beta_0 = 0$ ), we reject this null hypothesis, getting an  $F$  test equal to 5.78, which is highly significant (because the  $p$ -value is equal to 0.0061). This means that we cannot reject that peer effects are present in this example. We can also graphically compare the distribution of  $ATE(\mathbf{x})$ ,  $ATET(\mathbf{x})$ , and  $ATENT(\mathbf{x})$  with and without neighborhood interaction.

The whole Stata code showing all the steps needed to perform such estimates is reported below. Such code can be used as a general template for using `ntreatreg` to correctly apply the model presented in this article.

```
. ***** START IMPLEMENTATION *****
. *****
. * Step 1. Model inputs
. *****
. set scheme sj
. use "spatial_columbus", clear
. global y "crime"
. global w "cp"
. global xvars "inc hoval"
. global hvars "inc hoval"
. global dvars "x y"
. *****
. * Step 2. Deal with missing values
. *****
. * Eliminate common missing values
. quietly regress $y $w $xvars $hvars $dvars
. * Consider a dataset made of only nonmissing values
. keep if e(sample)
(0 observations deleted)
. * Sort the treatment (treated first)
. gsort - $w
. global N = r(N)
. count if $w==1
. 24
. global N1 = r(N)
. global N0 = $N-$N1
. *****
. * Step 3. Run mkomega to generate the matrix "omega"
. *****
. mkomega $w $dvars, sim_measure(L2) out($y)
(0 observations deleted)
. matrix omega=r(M)
```

```

. *****
. * Step 4. Fit the model using ntreatreg (to get ATE with neighborhood
> interactions) (resulting graph shown in figure 2)
. *****
. ntreatreg $y $w $xvars, hetero($hvars) spill(omega) graphic

```

Source	SS	df	MS	Number of obs	=	49
Model	10269.266	7	1467.038	F(7, 41)	=	18.98
Residual	3168.95358	41	77.2915508	Prob > F	=	0.0000
				R-squared	=	0.7642
				Adj R-squared	=	0.7239
Total	13438.2195	48	279.962907	Root MSE	=	8.7916

crime	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cp	14.5955	3.75345	3.89	0.000	7.015257 22.17575
inc	-.936559	.3619498	-2.59	0.013	-1.667531 -.2055866
hoval	-.1753827	.0961938	-1.82	0.076	-.3696501 .0188847
_ws_inc	-1.157042	.9291237	-1.25	0.220	-3.033445 .7193614
_ws_hoval	.1890178	.2091914	0.90	0.372	-.2334528 .6114885
_z_inc	-10.99322	7.124302	-1.54	0.131	-25.38104 3.394596
_z_hoval	-7.99784	2.5437	-3.14	0.003	-13.13495 -2.860734
_cons	400.355	111.1496	3.60	0.001	175.8839 624.8262

```

(49 real changes made)
(49 real changes made)
(49 real changes made)
(49 real changes made)
(25 missing values generated)
(24 missing values generated)

. * Store the estimates, put the ATE into a scalar, and rename variables
. estimates store REG_peer
. scalar ate_neigh = _b[$w]
. rename ATE_x _ATE_x_spill
. rename ATET_x _ATET_x_spill
. rename ATENT_x _ATENT_x_spill
. display ate_neigh
14.595504

. *****
. * Step 5. Test if the coefficients of the peer effect are jointly zero
. *****
. * If we accept the null Ho: gamma*beta0=0, the peer effect is negligible
. * If we do not accept the null, the peer effect is at work
. test _z_inc = _z_hoval = 0
(1) _z_inc - _z_hoval = 0
(2) _z_inc = 0
      F( 2, 41) = 5.78
      Prob > F = 0.0061

. * We reject that peer effects are negligible

```

```

. *****
. * Step 6. Fit the model using ivtreatreg (to get ATE without
> neighborhood interactions)
. *****
. ivtreatreg $y $w $xvars, hetero($hvars) model(cf-ols)

```

Source	SS	df	MS	Number of obs	=	49
Model	9375.05895	5	1875.01179	F(5, 43)	=	19.84
Residual	4063.1606	43	94.4921069	Prob > F	=	0.0000
				R-squared	=	0.6976
				Adj R-squared	=	0.6625
Total	13438.2195	48	279.962907	Root MSE	=	9.7207

```


```

crime	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cp	13.59008	4.119155	3.30	0.002	5.283016 21.89715
inc	-.8335211	.3384488	-2.46	0.018	-1.516068 -.1509741
hoval	-.1885477	.1036879	-1.82	0.076	-.3976543 .0205588
_ws_inc	-1.26008	1.004873	-1.25	0.217	-3.286599 .7664396
_ws_hoval	.2021829	.2300834	0.88	0.384	-.2618246 .6661904
_cons	46.52524	6.948544	6.70	0.000	32.51217 60.53832

```

(49 real changes made)
(49 real changes made)
(25 missing values generated)
(24 missing values generated)

. * Store the estimates and put the ATE into a scalar
. estimates store REG_no_peer
. scalar ate_no_neigh = _b[$w]

. *****
. * Step 7. Calculate the magnitude of the neighborhood-interactions bias
. *****
. * Bias in level
. scalar bias= ate_no_neigh - ate_neigh
. * Bias in percentage
. scalar bias_perc=(bias/ate_no_neigh)*100
. display bias_perc
-7.3981897

. ***** END IMPLEMENTATION *****

```

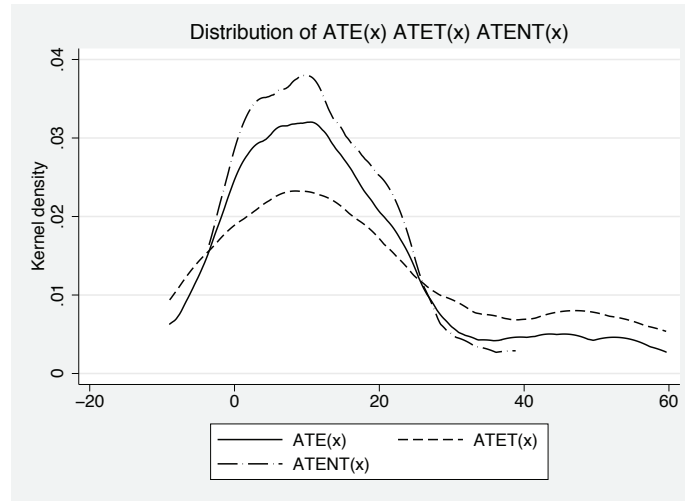


Figure 2. Graphical representation of the model with neighborhood effects

In conclusion, if the analyst does not properly consider neighborhood effects, then the actual effect of housing location on crime will be underestimated, although such underestimation seems not too large in this example. Moreover, the test shows that peer effects are relevant in such a context, because the regression coefficients of the peer component are jointly significant.

## 7 Conclusion

In this article, I presented a counterfactual model (embedded into the traditional Rubin POM) identifying ATEs by CMI when correlated peer (or neighborhood) effects are considered. I provided the command `ntreatreg` to implement such models in practical applications. Moreover, as a by-product of and designed to be used in combination with `ntreatreg`, I provided the `mkomega` command for generating a similarity matrix based on a set of covariates.

The model and its accompanying command estimate ATEs when the SUTVA is relaxed under specific conditions. I set out two instructional applications: i) a simulation exercise, useful to show both model implementation and `ntreatreg` correctness, and ii) an application to real data, aimed at measuring the effect of housing location on crime. In this second application, results are also compared with a no-interaction setting.

This model has various limitations. In what follows, I suggest some potential developments that other Stata developers can account for. Indeed, the model might be improved by

- allowing for treated units to be affected by other treated units' outcomes and untreated units to be affected by other untreated ones' outcomes;
- extending the model to "multiple" or "continuous" treatment, when treatment may be multivalued or fractional, for instance, by still holding CMI;
- allowing for unit potential outcome to depend on other units' treatment;
- identifying ATEs with neighborhood interactions when the treatment is endogenous (that is, relaxing CMI) by implementing a generalized method of moments instrumental-variables estimation procedure;
- going beyond the potential outcomes' parametric form, thus relying on a nonparametric or semiparametric specification.<sup>7</sup>

Finally, an issue that deserves further inquiry is the assumption of exogeneity of the weighting matrix  $\Omega$ . Indeed, a challenging question is: what happens if individuals strategically modify their behavior to better take advantage of others' treatment outcome? It is clear that if there exists some correlation between unobservable confounders affecting outcome and neighborhood choice, the weights may become endogenous, thus yielding further identification problems for previous causal effects. Future studies should tackle situations in which this possibility may occur.

## 8 References

- Angrist, J. D. 2014. The perils of peer effects. *Labour Economics* 30: 98–108.
- Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- Arduini, T., E. Patacchini, and E. Rainone. 2014. Identification and estimation of outcome response with heterogeneous treatment externalities. CPR Working Paper 167. <https://www.maxwell.syr.edu/uploadedFiles/cpr/publications/working-papers2/wp167.pdf>.
- Aronow, P. M., and C. Samii. 2013. Estimating average causal effects under general interference, with application to a social network experiment. ArXiv Working Paper No. arXiv:1305.6156. <https://arxiv.org/abs/1305.6156>.
- Cerulli, G. 2014. `ivtreatreg`: A command for fitting binary treatment models with heterogeneous response to treatment and unobservable selection. *Stata Journal* 14: 453–480.

---

7. For instance, one possibility would be that of relaxing the linear dependence of  $y$  on  $x$ , by using a partially linear form for the potential outcomes.



- 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: 945–960.
- Hudgens, M. G., and M. E. Halloran. 2008. Toward causal inference with interference. *Journal of the American Statistical Association* 103: 832–842.
- Imbens, G. W., and J. D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62: 467–475.
- Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60: 531–542.
- . 2013. Identification of treatment response with social interactions. *Econometrics Journal* 16: S1–S23.
- Robins, J. M., M. A. Hernán, and B. Brumback. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11: 550–560.
- Rosenbaum, P. R. 2007. Interference between units in randomized experiments. *Journal of the American Statistical Association* 102: 191–200.
- Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701.
- . 1977. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics* 2: 1–26.
- . 1978. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6: 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: 1398–1407.
- Tchetgen Tchetgen, E. J., and T. J. VanderWeele. 2010. On causal inference in the presence of interference. *Statistical Methods in Medical Research* 21: 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: 129–133.
- . 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

#### **About the author**

Giovanni Cerulli is a researcher at CNR-IRCrES, National Research Council of Italy, Institute for Research on Sustainable Economic Growth. He received a degree in statistics and a PhD in economic sciences from Sapienza University of Rome and is editor-in-chief of the *International Journal of Computational Economics and Econometrics*. His main research interest is applied

microeconometrics with a focus on counterfactual treatment-effects models for program evaluation. He is the author of the book *Econometric Evaluation of Socio-Economic Programs: Theory and Applications* (Springer, 2015). He has published articles in high-quality, refereed economics journals.

## A Appendix A

This appendix shows how to obtain the formulas of ATE and ATE( $\mathbf{x}$ ) set out in (3) and (3), and then shows how regression (12) can be obtained. Finally, the appendix details proof that assumption 1 is sufficient for consistently estimating the parameters of regression (12) by OLS.

### A.1 Formula of ATE with neighborhood interactions

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

$$\begin{aligned}
 y_{1i} &= \mu_1 + \mathbf{x}_i \beta_1 + e_{1i} \\
 y_{0i} &= \mu_0 + \mathbf{x}_i \beta_0 + \gamma s_i + e_{0i} \\
 s_i &= \sum_{j=1}^{N_1} \omega_{ij} y_{1j} \\
 \text{ATE} &= E(y_{1i} - y_{0i}) \\
 &= E \left( (\mu_1 + \mathbf{x}_i \beta_1 + e_{1i}) \right. \\
 &\quad \left. - \left[ \mu_0 + \mathbf{x}_i \beta_0 + \gamma \left\{ \mu_1 + \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \beta_1 + \sum_{j=1}^{N_1} \omega_{ij} e_{1j} \right\} + e_{0i} \right] \right) \\
 &= E \left[ \mu_1 + \mathbf{x}_i \beta_1 + e_{1i} \right. \\
 &\quad \left. - \left\{ \mu_0 + \mathbf{x}_i \beta_0 + \gamma \mu_1 + \gamma \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \beta_1 + \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} + e_{0i} \right\} \right] \\
 &= E \left\{ \mu_1 + \mathbf{x}_i \beta_1 + e_{1i} - \mu_0 - \mathbf{x}_i \beta_0 - \gamma \mu_1 - \gamma \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \beta_1 \right. \\
 &\quad \left. - \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} - e_{0i} \right\}
 \end{aligned}$$

$$\begin{aligned}
&= E \left\{ \mu_1 - \gamma \mu_1 - \mu_0 + \mathbf{x}_i \boldsymbol{\beta}_1 - \mathbf{x}_i \boldsymbol{\beta}_0 - \gamma \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \boldsymbol{\beta}_1 \right. \\
&\quad \left. - \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} + e_{1i} - e_{0i} \right\} \\
&= E \left\{ \mu_1 (1 - \gamma) - \mu_0 + \mathbf{x}_i (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) - \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \gamma \boldsymbol{\beta}_1 \right. \\
&\quad \left. - \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} + e_{1i} - e_{0i} \right\} \\
&= E \left\{ \mu_1 (1 - \gamma) - \mu_0 + \mathbf{x}_i \boldsymbol{\delta} - \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \gamma \boldsymbol{\beta}_1 \right. \\
&\quad \left. - \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} + e_{1i} - e_{0i} \right\} \\
&= \mu + E \left\{ \mathbf{x}_i \boldsymbol{\delta} - \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \gamma \boldsymbol{\beta}_1 - e_i \right\} \\
&= \mu + E(\mathbf{x}_i) \boldsymbol{\delta} - \gamma E \left( \underbrace{\sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j}_{\mathbf{v}_i} \right) \boldsymbol{\beta}_1
\end{aligned}$$

This implies that  $\text{ATE} = E(y_{1i} - y_{0i}) = \mu + E(\mathbf{x}_i) \boldsymbol{\delta} - \gamma E(\mathbf{v}_i) \boldsymbol{\beta}_1$ , which has a sample equivalent of

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

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  $\mathbf{\Omega}$  is organized as follows:

$$\begin{array}{cc} & \begin{array}{cc} \text{T} & \text{C} \end{array} \\ \begin{array}{c} \text{T} \\ \text{C} \end{array} & \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\ \omega_{21} & \omega_{22} & \omega_{23} & \omega_{24} \\ \omega_{31} & \omega_{32} & \omega_{33} & \omega_{34} \\ \omega_{41} & \omega_{42} & \omega_{43} & \omega_{44} \end{bmatrix} \end{array}$$

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

$$\begin{aligned} \widehat{\text{ATE}} &= \hat{\mu} + \frac{1}{4} \left( \sum_{i=1}^4 x_i \right) \hat{\delta} - \hat{\gamma} \times \hat{\beta}_1 \frac{1}{4} \left\{ \sum_{i=1}^4 \left( \sum_{j=1}^2 \omega_{ij} x_j \right) \right\} \\ &= \hat{\mu} + \frac{1}{4} \left( \sum_{i=1}^4 x_i \right) \hat{\delta} - \hat{\gamma} \times \hat{\beta}_1 \frac{1}{4} \left\{ \sum_{i=1}^4 (\omega_{i1} x_1 + \omega_{i2} x_2) \right\} \\ &= \hat{\mu} + \bar{x} \hat{\delta} - \hat{\gamma} \times \hat{\beta}_1 \frac{1}{4} \left\{ \underbrace{\sum_{i=1}^4 (\omega_{i1} x_1 + \omega_{i2} x_2)}_{v_1} \right\} = \hat{\mu} + \bar{x} \hat{\delta} - \hat{\gamma} \times \hat{\beta}_1 \bar{v} \end{aligned}$$

Observe that

$$\begin{aligned} v_1 &= \omega_{11} \times 1 + \omega_{12} \times 2 \\ v_2 &= \omega_{21} \times 1 + \omega_{22} \times 2 \\ v_3 &= \omega_{31} \times 1 + \omega_{32} \times 2 \\ v_4 &= \omega_{41} \times 1 + \omega_{42} \times 2 \end{aligned}$$

implying

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

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 comes only from treated to untreated units, thus excluding other types of feedbacks, it is equivalent to consider only the first two columns of  $\mathbf{\Omega}$  in the calculation of the externality component, those refereeing to the treated units. That is,

$$\begin{array}{cc}
& \begin{array}{cc} \text{T} & \text{C} \end{array} \\
\begin{array}{c} \text{T} \\ \text{C} \end{array} & \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\ \omega_{21} & \omega_{22} & \omega_{23} & \omega_{24} \\ \omega_{31} & \omega_{32} & \omega_{33} & \omega_{34} \\ \omega_{41} & \omega_{42} & \omega_{43} & \omega_{44} \end{bmatrix}
\end{array}$$

where neither of the two columns refers to the control group.

## A.2 Formula of $\text{ATE}(\mathbf{x}_i)$ with neighborhood interactions

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

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

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

### A.3 Obtaining regression (12)

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

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

$$\begin{aligned}
&= (\mu_0 + \gamma\mu_1) + \left( \gamma \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \beta_1 + w_i (\mu_1 - \mu_0 - \gamma\mu_1) + \mathbf{x}_i \beta_0 + w_i \mathbf{x}_i (\beta_1 - \beta_0) \\
&\quad - w_i \gamma \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \beta_1 + e_i \\
&= (\mu_0 + \gamma\mu_1) + \left( \gamma \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \beta_1 + w_i (\mu_1 - \mu_0 - \gamma\mu_1) + \mathbf{x}_i \beta_0 + w_i \mathbf{x}_i \delta \\
&\quad - w_i \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \gamma \beta_1 + e_i \\
&= (\mu_0 + \gamma\mu_1) + \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \gamma \beta_1 + w_i (\mu_1 - \mu_0 - \gamma\mu_1) + \mathbf{x}_i \beta_0 + w_i \mathbf{x}_i \delta \\
&\quad - w_i \left( \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j \right) \gamma \beta_1 + e_i + (w_i \bar{\mathbf{x}} \delta - w_i \bar{\mathbf{x}} \delta) \\
&\quad + \left\{ w_i E \left( \underbrace{\sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j}_{\mathbf{v}_i} \right) \gamma \beta_1 - w_i E \left( \underbrace{\sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j}_{\mathbf{v}_i} \right) \gamma \beta_1 \right\} \\
&= (\mu_0 + \gamma\mu_1) + w_i (\mu + \bar{\mathbf{x}} \delta - \bar{\mathbf{v}} \boldsymbol{\lambda}) + \mathbf{x}_i \beta_0 + w_i (\mathbf{x}_i - \bar{\mathbf{x}}) \delta + \mathbf{v}_i \boldsymbol{\lambda} + w_i \bar{\mathbf{v}} \boldsymbol{\lambda} - w_i \mathbf{v}_i \boldsymbol{\lambda} + e_i \\
&= \eta + w_i \times \text{ATE} + \mathbf{x}_i \beta_0 + w_i (\mathbf{x}_i - \bar{\mathbf{x}}) \delta + \{ \mathbf{v}_i + w_i (\bar{\mathbf{v}} - \mathbf{v}_i) \} \boldsymbol{\lambda} + e_i
\end{aligned}$$

Therefore, we can conclude that

$$y_i = \eta + w_i \times \text{ATE} + \mathbf{x}_i \beta_0 + w_i (\mathbf{x}_i - \bar{\mathbf{x}}) \delta + \mathbf{z}_i \boldsymbol{\lambda} + e_i$$

where  $\mathbf{z}_i = \mathbf{v}_i + w_i (\bar{\mathbf{v}} - \mathbf{v}_i)$ ,  $\mathbf{v}_i = \sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j$ ,  $\bar{\mathbf{v}} = 1/N \sum_{i=1}^N (\sum_{j=1}^{N_1} \omega_{ij} \mathbf{x}_j)$ ,  $\boldsymbol{\lambda} = \gamma \beta_1$ ,  $\eta = \mu_0 + \gamma\mu_1$ , and  $\delta = \beta_1 - \beta_0$ .

#### A.4 OLS consistency

Under assumption 1 (CMI), the parameters of regression (12) can be consistently estimated by OLS. Indeed, we immediately see that the mean of  $e_i$  conditional on  $(w_i; \mathbf{x}_i)$  is equal to 0:

$$\begin{aligned}
 & E \left\{ \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} + e_{0i} + w_i(e_{1i} - e_{0i}) - w_i \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} | w_i, \mathbf{x}_i \right\} \\
 &= E \left\{ \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} | w_i, \mathbf{x}_i \right\} + E \{ e_{0i} | w_i, \mathbf{x}_i \} + E \{ w_i(e_{1i} - e_{0i}) | w_i, \mathbf{x}_i \} \\
 &\quad - E \left( w_i \gamma \sum_{j=1}^{N_1} \omega_{ij} e_{1j} | w_i, \mathbf{x}_i \right) \\
 &= \gamma \sum_{j=1}^{N_1} \omega_{ij} E(e_{1j} | \mathbf{x}_i) + E(e_{0i} | \mathbf{x}_i) + w_i E \{ (e_{1i} - e_{0i}) | \mathbf{x}_i \} - w_i \gamma \sum_{j=1}^{N_1} \omega_{ij} E(e_{1j} | \mathbf{x}_i) \\
 &= 0
 \end{aligned}$$

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