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

where 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 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, we have that:

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

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 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.

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:

$$\text{ATE} = \text{E}(y_{1i} - y_{0i}) = \mu + \bar{\mathbf{x}}_i\boldsymbol{\delta} - \left(\sum_{j=1}^{N_1} \omega_{ij} \bar{\mathbf{x}}_j\right)\gamma\boldsymbol{\beta}_1 \tag{10}$$

where $\bar{\mathbf{x}}_i = \text{E}(\mathbf{x}_i)$ is the unconditional mean of the vector $\mathbf{x}_i$, and $\mu = \mu_1 - \mu_0 - \gamma\mu_1$. The proof is in Appendix. See A1.

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

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