Estimation and inference in the presence of neighborhood unobservables

Tiziano Arduini[a], Federico Belotti[b,*], Edoardo Di Porto[c]

[a]*Department of Economics and Finance, University of Rome Tor Vergata.*
[b]*Department of Economics and Finance, University of Rome Tor Vergata.*
[c]*University of Naples Federico II, CSEF and UCFS Uppsala University.*

Abstract

Neglecting neighborhood unobservables or shocks may hinder the identification of causal effects. However, if unobservables are smooth over space and units are paired based on proximity, a neighborhood data transformation can effectively eliminate their influence. This paper studies neighborhood differencing and within-neighborhood estimation strategies in a finite population framework. We establish their asymptotic distribution and provide guidance on standard errors adjustment for within and between neighborhood correlations. We also develop a test for smooth fixed effects, allowing practitioners to select the optimal threshold for data transformation. We examine the behavior of the proposed tools through simulations, showing good finite sample properties. We illustrate the usefulness of our approach using geocoded data from a clustered randomized experiment.

Keywords: Neighborhood data transformation, Smooth fixed effects, Hausman-like test, Finite population
JEL: C12, C21, I25

[*]We thank Debopam Bhattacharya, Franco Peracchi, Sergio Pastorello, Eric Auerbach, Robert Chirinko, and Edoardo Rainone; Jerome Simons and the participants of the Econometrics seminar at the Faculty of Economics, University of Cambridge; Giovanni Mellace, Federico Crudu, Michaela Kesina, Tom Boot, Lingwei Kong, Tan Chang, and the participants of the EEF seminar series at the Department of Economics, Econometrics, and Finance, University of Groningen; Elena Dal Torrione, Faranak Alikhah, and the participants of the Tor Vergata Microeconometrics Reading Group; as well as the participants of IAAE 2024, the XV World Conference of the Spatial Econometric Association, and the internal seminar series of the Department of Economics and Finance, Tor Vergata University of Rome, for their valuable discussions. Corresponding author: Federico Belotti, Department of Economics and Finance, University of Rome Tor Vergata, via Columbia 2, 00133 Rome, Italy. E-mail: federico.belotti@uniroma2.it, phone/fax: +39 06 7259 5635. The authors acknowledge financial support from the Tor Vergata University of Rome Research Grant 2021/E83C22002540005.
[1]This version: July 1^{th}, 2025.

1. Introduction

Economic agents are influenced by the characteristics and behaviors of their neighbors. When these factors are not observed, they might confound the causal effect of an explanatory variable on the outcome of interest. For instance, better schools tend to be located in better neighborhoods, which complicates the estimation of the causal effect of school quality on house prices. If the researcher does not observe neighborhood characteristics, then the impact of school quality may be overestimated (Black, 1999). Similarly, the impact of local taxation on firms' performance may be confounded by unobserved site-specific characteristics which affect firms' location choices (Duranton et al., 2011). In the context of health-related outcomes, counties in low-regulation states bordering high-regulation states may exhibit higher cancer death rates due to increased industrial activity and pollution exposure (Kahn, 2004). Estimating the causal impact of local regulations on health outcomes is challenging, as unobserved factors, such as cross-border pollution, can confound the observed effects.

A popular solution to deal with these issues is to pair units according to their proximity. If neighborhood unobservables are additively-separable and "sufficiently smooth", then an appropriate linear transformation can be used to rule them out. This strategy was pioneered by Holmes (1998), who exploited geo-coded data to apply a spatial "differencing" transformation.[2] Despite its widespread use, the existing literature neither rigorously investigates the properties of the differencing estimator nor provides guidance on how to properly select and pair neighboring units. Moreover, when the sample represents only a small fraction of the population, inference primarily relies on sampling uncertainty. However, as sampling rates approach zero, such as in surveys or small-sample administrative data, the number of observed neighbors in the sample may be very limited. Conversely, when the sample effectively constitutes the entire population, as is often the case with spatial data, the availability of observed neighbors is no longer a concern, and sampling uncertainty becomes negligible.

This paper studies neighborhood differencing (ND) and within-neighborhood (NW) estimators in a finite population framework. We establish their asymptotic distribution and develop a strategy to test for the presence of smooth fixed effects and select the optimal threshold for data transformation.

We follow the inferential framework proposed by Abadie et al. (2020) and Xu and Wooldridge (2022) to account for design uncertainty and model the spatial dependence. In particular, we adopt a finite population viewpoint by introducing spatial correlation within the population before sampling. We derive sufficient conditions to apply the central limit theorem (CLT) of Xu and Wooldridge (2022) under spatial Near Epoch Dependence (NED, Jenish and Prucha, 2012), which allows for zero sampling probability in the limit. The neighborhood data transformations (NDT) remove the spatial dependence induced by the smooth fixed effects, but introduce a mechanical correlation. We find that there is no need to correct standard errors for this correlation if we are sampling from a superpopulation. Additionally, when the sampling probability is positive, the finite population asymptotic variance is smaller than that of the

[2] Applications of this strategy are many and include studies on the effects of school quality on house prices (Black, 1999, Fack and Gret, 2010, Gibbons et al., 2013, Harjunen et al., 2018), the effect of local taxes on firm performance (Belotti et al., 2021, Duranton et al., 2011), the effects of tax policies on county level outcomes (Chirinko and Wilson, 2008), the evaluation of placed-based policy (Einiö and Overman, 2020) and the effect of pollution havens on mortality (Kahn, 2004).

superpopulation counterpart. This result extends the findings on the conservativeness of finite population variance estimation (Neyman, 1923, Abadie et al., 2020, Xu and Wooldridge, 2022) in the context of neighborhood unobservables.

This paper extends the linear regression framework in Abadie et al. (2020) accounting for spatial dependence. In doing so, it exploits lower-level conditions of the M-estimation theory in Xu and Wooldridge (2022). Our approach is related to, but different from, the pairwise differencing strategies in Auerbach (2022) and Druckenmiller and Hsiang (2018). Auerbach (2022) proposes a partially linear model for network data in which an unknown function of social unobservables drives the linking behavior. The author considers a nonparametric pairwise differencing estimator that removes the unobserved heterogeneity for units with similar linking behavior. Differently from Auerbach (2022), we impose a higher level condition, i.e., the existence of a neighborhood specification allowing the elimination of the fixed effects, which enters the model linearly and additively. However, we do not assume any specific neighborhood formation model. Our framework nests Druckenmiller and Hsiang (2018) who study identification and estimation of spatial first differences regression models where the spatial difference transformation is only applied to first-order contiguous neighbors.

We develop a fixed-sequence testing strategy allowing us to detect smooth fixed effects and, if the data support their presence, to select the optimal proximity threshold, i.e., the best neighborhood specification, to transform the data.

Specifically, to test the null hypothesis of non-smooth fixed effects, we propose using the Mundlak (1978) approach. This method is particularly well-suited to our setting because, regardless of the selected distance threshold, neighborhood averages fail to capture fixed effects when they are not spatially smooth. However, when fixed effects are smooth, these averages effectively detect their presence. Once smooth fixed effects have been detected, we test for the optimal distance by leveraging the contrast between the ND and NW estimators. Unlike the standard version of the Hausman test (Hausman, 1978), we compare estimators that are both inconsistent under the alternative. As pointed out by Ruud (1984), what matters for a specification test to have power is that it is based on estimators that diverge under the alternative and that the sampling variance of their difference is sufficiently small. We show that, since our alternative estimators rely on different functions of the data, they generally converge in probability to distinct points in the parameter space as the distance increases. Notably, the asymptotic biases of the ND and NW estimators converge to the OLS bias when the transformation is applied using the maximum observed distance between units. Clearly, in cases where both estimators exhibit the same degree of inconsistency, our test lacks power.

We illustrate the usefulness of our approach using geocoded data from the seminal work by Miguel and Kremer (2004), which investigates the effect of a deworming medical treatment on school absenteeism and health status in Kenya. The authors exploit a specific type of field experiment, where the randomization occurs at the school level, to disentangle the direct average treatment effect from the indirect cross-school externalities. We exploit our testing strategy to detect the presence of smooth fixed effects and to determine whether they extend beyond a specific radius. We detect smooth fixed effects within a six-kilometer radius - three kilometers beyond the distance selected by the authors. This finding supports adopting a longer and more granular specification of cross-school externality terms, allowing for the estimation of a larger and more precise average cross-school externality effect.

The article is organized as follows. Section 2 introduces our statistical model. Section 3 describes the proposed estimators and studies their asymptotic properties. Section 4 presents a testing strategy to detect the presence of smooth fixed effects and select the optimal distance for data transformation. Section 5 investigates the finite sample properties of the proposed estimators and test statistics using Monte Carlo simulations. In Section 6, we illustrate the proposed methodology exploiting geocoded data from a clustered randomized medical treatment program. Section 7 concludes.

2. The statistical model

Following Abadie et al. (2020) and Xu and Wooldridge (2022), we consider a sequence of finite populations subset D_n, where D is a potentially irregular lattice in $\mathbb{R}^d$ ($d \geq 1$), and n indexes a sequence of finite populations. The location $\ell : 1, \ldots, n \to D_n \subset D$ is a mapping of individual i to its location $\ell(i) \in D_n$. All locations are located at a minimum distance greater than 0. Let $\{\boldsymbol{t}_{\ell(i),n}, \ell(i) \in D_n,\ n \geq 1\}$ and $\{\theta_{\ell(i),n}, \ell(i) \in D_n,\ n \geq 1\}$ be triangular arrays of random fields defined on a probability space $(\Omega, \mathcal{F}, P)$. For simplicity, we use the notation $\{\boldsymbol{t}_i, i \in D_n,\ n \geq 1\}$ to indicate $\{\boldsymbol{t}_{\ell(i),n}, \ell(i) \in D_n,\ n \geq 1\}$. Same for the other variables.

Unit i in the population is characterized by a potential outcome function, $y_i(\cdot)$, which maps the treatment assignment column vector $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{iK})'$ into outcomes. Realized outcomes are denoted by y_i. While potential outcomes and attributes remain fixed through repeated sampling, the outcomes and the assignment may change. For each population, we have an associated sample, $R_i = 1$ if unit i in population n is sampled, and 0 otherwise, so that $N = \sum_i R_i$. We assume the following linear model for the potential outcomes

$$y_i(\boldsymbol{t}_i) = \boldsymbol{t}_i'\boldsymbol{\beta}_i + \theta_i + \epsilon_i, \tag{1}$$

where θ_i is a stochastic unobserved attribute arbitrarily correlated with $\boldsymbol{t}_i$, ϵ_i is an unobserved non-stochastic attribute, and $\boldsymbol{\beta}_i$ is a unit-level column vector of treatment responses. The researcher observes $(y_i, \boldsymbol{t}_i', \boldsymbol{z}_i', d_{ij})$ for each unit in the sample, where $\boldsymbol{z}_i$ is a column vector of non-stochastic attributes, and d_{ij} is a general proximity measure, i.e., social, economic or geographical distance between units. The researcher specifies/selects a neighborhood B_i^d based on a distance threshold d. Thus, economic and social indicators should be carefully chosen to minimize potential confounding.

Let $n_i^d = |B_i^d|$ be the cardinality of B_i^d, i.e., unit i has n_i^d neighbors including i itself. We postulate the existence of a neighborhood specification $B_i^{d^*}$ for each unit i with d^* arbitrary small such that $\theta_i \approx \theta_j$, $\forall j \in B_i^{d^*}$, i.e., there exists with probability one a small area in D where neighboring units share similar unobservables. This is true if, for example, θ_i changes smoothly across space. Figure 1 shows two examples of neighborhood selections. Neighborhoods B_1^d and B_2^d (on the left) are based on the proximity threshold d and $B_1^{d^*}$ and $B_2^{d^*}$ (on the right) are specified using $d^* < d$. The unit-specific unobservables θ_i, $i = 1, \ldots, 4$ are assumed to be smooth over the space considered. This implies that there exists a neighborhood specification $B_1^{d^*}$ and $B_2^{d^*}$ where unobservables are approximately equal. Observe that, $n_1^d = n_4^d = 3$ while $n_1^{d^*} = n_4^{d^*} = 2$. Further, we allow for overlapping neighborhoods, i.e., $B_i^d \cap B_k^d \neq \emptyset$ for all $i \neq k$ and d.

3. Estimation based on neighborhood data transformations

Let Δ^d be the neighborhood-difference operator. The neighborhood (pairwise) difference transformation takes the difference between the observations of unit i and each $j \in B_i^d \setminus i$, i.e. $\Delta_{ij}^d \boldsymbol{x}_i = \boldsymbol{x}_i - \boldsymbol{x}_j$. We denote the sample counterpart of the neighborhood-difference operator as $\tilde{\Delta}_{ij}^d = R_i \boldsymbol{x}_i - R_j \boldsymbol{x}_j = R_{ij} \Delta_{ij}^d \boldsymbol{x}_i$, where $R_{ij} = 1$ if the pair (i, j) is sampled and 0 otherwise. Let $\sum_{i<j}$ be shorthand for $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n}$. In the population, the total number of pairs can be calculated as $n_p = \frac{\sum_i (n_i^d - 1)}{2}$, while in the sample, it is denoted as $N_p = \sum_{i<j, j \in B_i \setminus i} R_{ij} = \frac{\sum_i (N_i^d - 1)}{2}$, where N_i^d represents the number of sampled units in the neighborhood of unit i. Alternatively, the within-neighborhood transformation takes the difference between the unit i and the population average $\bar{\boldsymbol{x}}_i = \frac{1}{n_i^d} \sum_j x_j$ computed considering all units $j \in B_i^d$, i.e. $\Delta_{i,n_i}^d \boldsymbol{x}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}_i$. We denote the sample counterpart of the within-neighborhood operator as $\tilde{\Delta}_{i,n_i}^d \boldsymbol{x}_i = R_i \boldsymbol{x}_i - \frac{1}{N_i^d} \sum_j R_j \boldsymbol{x}_j$.

To simplify the derivation of the properties of the proposed estimators, we adopt the same approach as presented in Abadie et al. (2020) and use the following transformations

$$\Delta_{ij} \boldsymbol{x}_i = \Delta_{ij} \boldsymbol{t}_i - \boldsymbol{\lambda} \Delta_{ij} \boldsymbol{z}_i, \;\; where\ \boldsymbol{\lambda} = \left(\sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij} \boldsymbol{t}_i) \Delta_{ij} \boldsymbol{z}_i' \right) \left(\sum_{i<j, j \in B_i \setminus i} \Delta_{ij} \boldsymbol{z}_i \Delta_{ij} \boldsymbol{z}_i' \right)^{-1}, \tag{2}$$

and

$$\Delta_{i,n_i} \boldsymbol{x}_i = \Delta_{i,n_i} \boldsymbol{t}_i - \boldsymbol{\lambda}_{NW} \Delta_{i,n_i} \boldsymbol{z}_i, \;\; where\ \boldsymbol{\lambda}_{NW} = \left(\sum_i E(\Delta_{i,n_i} \boldsymbol{t}_i) \Delta_{i,n_i} \boldsymbol{z}_i' \right) \left(\sum_i \Delta_{i,n_i} \boldsymbol{z}_i \Delta_{i,n_i} \boldsymbol{z}_i' \right)^{-1},$$

which remove the correlation with the attributes, assuming that $\sum_{i<j, j \in B_i \setminus i} \Delta_{ij} \boldsymbol{z}_i \Delta_{ij} \boldsymbol{z}_i'$, and $\sum_i \Delta_{i,n_i} \boldsymbol{z}_i \Delta_{i,n_i} \boldsymbol{z}_i'$ are full rank.

Then, the class of least squares ND estimators indexed by the distance threshold d can be defined as

$$(\hat{\boldsymbol{\beta}}_{ND}^d, \hat{\boldsymbol{\gamma}}_{ND}^d) = \underset{(\boldsymbol{\beta}_{ND}^d, \boldsymbol{\gamma}_{ND}^d)}{\operatorname{argmin}} \sum_{i<j, j \in B_i \setminus i} (\tilde{\Delta}_{ij}^d y_i - \tilde{\Delta}_{ij}^d \boldsymbol{x}_i' \boldsymbol{\beta}_{ND}^d - \tilde{\Delta}_{ij}^d \boldsymbol{z}_i' \boldsymbol{\gamma}_{ND}^d)^2.$$

Similarly, we can define the class of NW estimators as

$$(\hat{\boldsymbol{\beta}}_{NW}^d, \hat{\boldsymbol{\gamma}}_{NW}^d) = \underset{(\boldsymbol{\beta}_{NW}^d, \boldsymbol{\gamma}_{NW}^d)}{\operatorname{argmin}} \sum_{i=1}^{n} (\tilde{\Delta}_{i,n_i}^d y_i - \tilde{\Delta}_{i,n_i}^d \boldsymbol{x}_i' \boldsymbol{\beta}_{NW}^d - \tilde{\Delta}_{i,n_i}^d \boldsymbol{z}_i' \boldsymbol{\gamma}_{NW}^d)^2.$$

To ease the notation, in what follows, we remove the d superscript unless we want to stress the dependence on d. We now introduce the causal estimands of interest. Let us define the following matrices

$$\Delta_{ij}\boldsymbol{W}_n = \frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}\boldsymbol{w}_i,$$

$$\Delta_{ij}\boldsymbol{\Omega}_n = \frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}E\Delta_{ij}\boldsymbol{w}_i,$$

where the expectation is taken over the distribution of $\boldsymbol{x}$ and $\Delta_{ij}\boldsymbol{w}_i = \begin{pmatrix}\Delta_{ij}y_i\\ \Delta_{ij}\boldsymbol{x}_i\\ \Delta_{ij}\boldsymbol{z}_i\end{pmatrix}\begin{pmatrix}\Delta_{ij}y_i\\ \Delta_{ij}\boldsymbol{x}_i\\ \Delta_{ij}\boldsymbol{z}_i\end{pmatrix}'$. Consider the sample counterpart of $\Delta_{ij}\boldsymbol{W}_n$

$$\tilde{\Delta}_{ij}\boldsymbol{W}_n = \frac{1}{N_p}\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{w}_i,$$

where $\tilde{\Delta}_{ij}\boldsymbol{w}_i = \begin{pmatrix}\tilde{\Delta}_{ij}y_i\\ \tilde{\Delta}_{ij}\boldsymbol{x}_i\\ \tilde{\Delta}_{ij}\boldsymbol{z}_i\end{pmatrix}\begin{pmatrix}\tilde{\Delta}_{ij}y_i\\ \tilde{\Delta}_{ij}\boldsymbol{x}_i\\ \tilde{\Delta}_{ij}\boldsymbol{z}_i\end{pmatrix}'$. The same notation applies for the NW transformation by substituting Δ_{ij} with Δ_{i,n_i}. Below, we define our causal estimands

$$\begin{pmatrix}\boldsymbol{\beta}^c_{ND}\\ \boldsymbol{\gamma}^c_{ND}\end{pmatrix} = \begin{pmatrix}\Delta_{ij}\boldsymbol{\Omega}^{xx}_n & \Delta_{ij}\boldsymbol{\Omega}^{xz}_n\\ \Delta_{ij}\boldsymbol{\Omega}^{zx}_n & \Delta_{ij}\boldsymbol{\Omega}^{zz}_n\end{pmatrix}^{-1}\begin{pmatrix}\Delta_{ij}\boldsymbol{\Omega}^{xy}_n\\ \Delta_{ij}\boldsymbol{\Omega}^{zy}_n\end{pmatrix},$$

$$\begin{pmatrix}\boldsymbol{\beta}^c_{NW}\\ \boldsymbol{\gamma}^c_{NW}\end{pmatrix} = \begin{pmatrix}\Delta_{i,n_i}\boldsymbol{\Omega}^{xx}_n & \Delta_{i,n_i}\boldsymbol{\Omega}^{xz}_n\\ \Delta_{i,n_i}\boldsymbol{\Omega}^{zx}_n & \Delta_{i,n_i}\boldsymbol{\Omega}^{zz}_n\end{pmatrix}^{-1}\begin{pmatrix}\Delta_{i,n_i}\boldsymbol{\Omega}^{xy}_n\\ \Delta_{i,n_i}\boldsymbol{\Omega}^{zy}_n\end{pmatrix},$$

where the superscripts used for the blocks denote the partition of interest. Following Abadie et al. (2020), we could also define causal-sample estimands. However, to simplify the exposition, we focus exclusively on causal estimands.[3]

The ND and NW estimators are defined accordingly

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}^c_{ND}\\ \hat{\boldsymbol{\gamma}}^c_{ND}\end{pmatrix} = \begin{pmatrix}\tilde{\Delta}_{ij}\boldsymbol{W}^{xx}_n & \tilde{\Delta}_{ij}\boldsymbol{W}^{xz}_n\\ \tilde{\Delta}_{ij}\boldsymbol{W}^{zx}_n & \tilde{\Delta}_{ij}\boldsymbol{W}^{zz}_n\end{pmatrix}^{-1}\begin{pmatrix}\tilde{\Delta}_{ij}\boldsymbol{W}^{xy}_n\\ \tilde{\Delta}_{ij}\boldsymbol{W}^{zy}_n\end{pmatrix}, \tag{3}$$

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}^c_{NW}\\ \hat{\boldsymbol{\gamma}}^c_{NW}\end{pmatrix} = \begin{pmatrix}\tilde{\Delta}^d_{i,n_i}\boldsymbol{W}^{xx}_n & \tilde{\Delta}^d_{i,n_i}\boldsymbol{W}^{xz}_n\\ \tilde{\Delta}^d_{i,n_i}\boldsymbol{W}^{zx}_n & \tilde{\Delta}^d_{i,n_i}\boldsymbol{W}^{zz}_n\end{pmatrix}^{-1}\begin{pmatrix}\tilde{\Delta}^d_{i,n_i}\boldsymbol{W}^{xy}_n\\ \tilde{\Delta}^d_{i,n_i}\boldsymbol{W}^{zy}_n\end{pmatrix}. \tag{4}$$

Let us now state the main assumptions needed to study the behavior of the proposed estimators as $n \to \infty$.

Assumption 1 (Increasing domain). *Suppose D_n is a sequence of finite subsets of D such that $|D_n| \to \infty$ an $n \to \infty$, where the lattice $D \subset \mathbb{R}^{d_0}$, with $d_0 \geq 1$, is infinitely countable. The location l :*

[3]For a formal definition of these estimands in the standard OLS framework see Abadie et al. (2020).

$\{1,\ldots,n\} \to D_n$ is a mapping of individual i to its location $l(i) \in D_n$. All locations are located at a minimum distance, $\bar{d}_0$, greater than 0, i.e, $d_{ij} \geq \bar{d}_0$.

This condition imposes a minimum distance requirement and implies that, for any distance threshold d, there are at most kd^{d_0} points in B_i^d and at most kd^{d_0-1} points in the space $B_i^{d+1} \setminus B_i^d$, where $k > 0$ is a constant (see Lemma A.1 in Jenish and Prucha, 2009).[4] This implies that, given a small threshold d, the number of neighbors n_i is constant. Additionally, considering the space as fixed results in a constant N_i with $N_i \to n_i$ as $n \to \infty$.

Assumption 2 (Removing fixed effects). *There exists at least one small distance $d^* \geq \bar{d}_0$ such that, $\lim_{d\to d^*} E(\Delta_{ij}^d \theta_i)^2 = 0$ $\left(\lim_{d\to d^*} E(\Delta_{i,n_i}^d \theta_i)^2 = 0\right)$ uniformly in $i = 1,\ldots,n$, and $j \in B_i^d \setminus i$ as $n \to \infty$.*

Assumption 2 is crucial for the consistency of the estimators defined in (3) and (4). The key ingredient is indeed the specification of a shrinking neighborhood such that the resulting data transformation can approximately rule out fixed effects for an arbitrary small d^*. This is true if θ_i changes smoothly across space, i.e., for each $\delta > 0$ there exists an arbitrarily small distance d^* such that $|\Delta_{ij}^{d^*}\theta_i| < \delta$ (or $|\Delta_{i,n_i}^{d^*}\theta_i| < \delta$) for each unit. Duranton et al. (2011) impose this continuity condition in the specific case of geographical proximity. Assumption 2 extends this deterministic condition to expectations.

Assumption 3 (Assignment Mechanism). *The assignments $\boldsymbol{t}_1 \ldots \boldsymbol{t}_n$ are independent of the sampling indicators.*

Assumption 4 (Random Sampling). *a) There is a sequence of sampling probabilities, ρ_n, such that*

$$Pr(\boldsymbol{R} = \boldsymbol{r}) = \rho_n^{\sum_i r_i}(1-\rho_n)^{n-\sum_i r_i},$$

for all vectors $\mathbf{r}$ with element $r_i \in [0,1]$.

b) The sequence of sampling rates, ρ_n, satisfies $n\rho_n \to \infty$ and $\rho_n \to \rho \in (0,1]$.

Note that $\rho = 0$ is excluded from the range of admissible values for the sampling rate, since in that case the sample would contain no neighbors, making spatial transformations inapplicable. Assumption 3 presupposes the independence between assignments and sampling indicators. However, we allow the (expected) assigments to be spatially correlated (Xu and Wooldridge, 2022).[5] Additionally, Assumption (4) (a) and (b) are standard random sampling assumptions for a sequence of finite populations and allow for the expected sample size to grow with n as outlined in Abadie et al. (2020). Observe that this implies that for a given d there is a sequence of paired sampling probabilities, ρ_n^2, such that

$$Pr(\boldsymbol{R}_p = \boldsymbol{r}_p) = \rho_n^{2\sum_{i<j,j\in B_i\setminus i} r_{ij}}(1-\rho_n^2)^{(n_p - \sum_{i<j,j\in B_i\setminus i} r_{ij})},$$

for all vectors $\boldsymbol{r}_{ij}$ with element $r_{ij} \in [0,1]$.

[4] Additional details about set cardinalities for irregular lattices can be found in Lemma 8, Appendix A.3.
[5] See Assumption 6 below.

Assumption 5 (Moments). *a) There exist some* $\delta > 0$ *such that the sequences*
$\frac{1}{n}\sum_{i=1}^{n} E(|y_i|^{4+\delta})$, $\frac{1}{n}\sum_{i=1}^{n} E(||\boldsymbol{x}_i||^{4+\delta})$, $\frac{1}{n}\sum_{i=1}^{n}(||\boldsymbol{z}_i||^{4+\delta})$,
are uniformly bounded

b) $\Delta_{ij}\boldsymbol{\Omega}_n \to \Delta_{ij}\boldsymbol{\Omega}$*, which is full rank and* $\Delta_{i,n_i}\boldsymbol{\Omega}_n \to \Delta_{i,n_i}\boldsymbol{\Omega}$*, which is full rank.*

Assumption 5 *a)* imposes a regularity condition that constrains moments within defined bounds. Observe that by Cauchy-Schwarz inequality also the average of expectations of cross products, i.e., $\frac{1}{n}\sum_i \sum_{j\neq i} E(y_i y_j)$ are bounded. Since our causal estimands are functions of population-level averages of second moments or the expected product of random variables, Assumption 5 *b)* ensures the convergence of the expected values of these sequences in the population. The next Lemma shows the convergence of $\tilde{\Delta}_{ij}\boldsymbol{W}_n$ and $\tilde{\Delta}_{i,n_i}\boldsymbol{W}_n$.

Lemma 1. *Under Assumptions 1- 5 a),* $\tilde{\Delta}_{ij}\boldsymbol{W}_n - \Delta_{ij}\boldsymbol{\Omega}_n \xrightarrow{p} \boldsymbol{0}$*,*
b) $\tilde{\Delta}_{i,n_i}\boldsymbol{W}_n - \Delta_{i,n_i}\boldsymbol{\Omega}_n \xrightarrow{p} \boldsymbol{0}$*.*

In contrast to conventional exogeneity or unconfoundedness conditions, where residuals are assumed to be mean-independent of the regressors, the following assumption relaxes the strict randomness of the assignment mechanism.

Assumption 6 (Expected Assignment). *a) There exists a sequence of functions* $f_n(\boldsymbol{z}_i)$ *such that*

$$E(\boldsymbol{t}_i|\theta_i) = f_n(\boldsymbol{z}_i, \theta_i)$$

b) there exists a sequence of matrices $\boldsymbol{\phi}_n$ *and* q_n*, such that, for all z, as* $n \to \infty$

$$f_n(z, \theta) = \boldsymbol{\phi}_n z + q_n\theta.$$

Assumption 6 implies that the expected value of the neighborhood-transformed assignment depends only on the transformed attributes when the population is large, the distance threshold is sufficiently small, and the unobservables are "smooth." This assumption, together with the transformation in (2), implies that $\Delta_{ij}\boldsymbol{\Omega}_n^{zx}$ and $\Delta_{ij}\boldsymbol{\Omega}_n^{xz}$ (or equivalently, $\Delta_{i,n_i}\boldsymbol{\Omega}_n^{zx}$ and $\Delta_{i,n_i}\boldsymbol{\Omega}_n^{xz}$) are matrices with entries that are approximately zero for large populations.

Observe that, under these assumptions, our causal estimands $\boldsymbol{\beta}_{ND}^c$ and $\boldsymbol{\beta}_{NW}^c$ can be interpreted as weighted averages of unit-level treatment responses, as in Abadie et al. (2020). This interpretation holds if the sums of the assignment cross-moments are bounded. Further details are provided in Lemma 3 in Appendix A.2.

3.1. Asymptotic distributions

We begin by defining the population residuals relative to the population causal estimands, $\Delta_{ij}\varepsilon_i = \Delta_{ij}y_i - \Delta_{ij}\boldsymbol{x}_i'\boldsymbol{\beta}_{ND}^c - \Delta_{ij}\boldsymbol{z}_i'\boldsymbol{\gamma}_{ND}^c$ and $\Delta_{i,n}\varepsilon_i = \Delta_{i,n}y_i - \Delta_{i,n}\boldsymbol{x}_i'\boldsymbol{\beta}_{NW}^c - \Delta_{i,n}\boldsymbol{z}_i'\boldsymbol{\gamma}_{NW}^c$. We are agnostic about the dependence structure of θ_i. However, we need to impose some restrictions on the spatial dependence to derive the asymptotic distribution. We apply a Central Limit Theorem (CLT) as derived in Xu and Wooldridge (2022), which is a variation of the CLT introduced in Jenish and Prucha (2012). Notably, this CLT takes into account the finite population framework, enhancing its applicability. We adopt the definition of near-epoch dependent (NED) random fields in Jenish and Prucha (2012). Detailed definitions, along with the mixing and NED conditions required for the application of the CLT, are provided in the Appendix A.1.

To derive the asymptotic distribution of the ND and NW estimators, we first need to strengthen the identification conditions in Assumption 2.

Assumption 7. $\forall \epsilon > 0$, $\lim_{d \to d^*, n \to \infty} \sqrt{N_p} E|\Delta_{ij}^d \theta_i| = 0$, *(or* $\lim_{d \to d^*, n \to \infty} \sqrt{N} E|\Delta_{i,n_i}^{d^*} \theta_i|) = 0$ *uniformly in* $i = 1, \ldots, n$*, and* $j \in B_i^{d^*} \setminus i$*.*[6]

Assumption 7 reinforces the smoothness requirement for θ_i so that the ND and NW estimators' asymptotic distributions are centered. We provide the proofs for all the propositions in Section Appendix A.2.

3.1.1. The neighborhood-difference estimator

We begin by postulating the existence of limits for the components of the asymptotic variance of the key statistics, $\sum_{i<j, j \in B_i \setminus i} \Delta_{ij} \boldsymbol{x}_i \Delta_{ij} \varepsilon_i$.

Assumption 8.

$$\boldsymbol{B}^{cond} = \lim_{n \to \infty} \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} var(\Delta_{ij} \boldsymbol{x}_i \Delta_{ij} \varepsilon_i),$$

$$\boldsymbol{B}^{cov} = \lim_{n \to \infty} \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} \sum_{k \neq i < l, l \in B_k \setminus k} cov(\Delta_{ij} \boldsymbol{x}_i \Delta_{ij} \varepsilon_i \Delta_{kl} \boldsymbol{x}_k' \Delta_{kl} \varepsilon_l),$$

and

$$\boldsymbol{B}^{ehw} = \boldsymbol{B}^{u} + \boldsymbol{B}^{cond} = \lim_{n \to \infty} \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij} \varepsilon_i^2 \Delta_{ij} \boldsymbol{x}_i \Delta_{ij} \boldsymbol{x}_i'),$$

with

$$\boldsymbol{B}^{u} = \lim_{n \to \infty} \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij} \boldsymbol{x}_i \Delta_{ij} \varepsilon_i) E(\Delta_{ij} \boldsymbol{x}_i \Delta_{ij} \varepsilon_i)',$$

exist and are positive definite.

We can now derive the asymptotic distribution of the ND estimator viewed as an estimator of the causal estimand $\boldsymbol{\beta}_{ND}^c$.

Proposition 1. a) *Under Assumptions, 1-8, and Assumptions 10, 11, and 12 a) in Appendix A.1*

$$\sqrt{N_p}(\hat{\boldsymbol{\beta}}_{ND} - \boldsymbol{\beta}_{ND}^c) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{A}^{-1} \boldsymbol{B} \boldsymbol{A}^{-1}), \tag{5}$$

where $\boldsymbol{A} = \lim_{n \to \infty} \Delta_{ij} \boldsymbol{\Omega}_n^{xx}$ *and* $\boldsymbol{B} = \rho^2(\boldsymbol{B}^{cond} + \boldsymbol{B}^{cov}) + (1 - \rho^2) \boldsymbol{B}^{ehw}$.

[6] We thank Eric Auerbach for his assistance in providing sufficient conditions for centering the asymptotic distribution of the estimators.

Given the inherent "mechanical" dependence induced by the NDT, the variance of $\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\varepsilon_i$ will be a linear combination of the sum of the variances of the key statistics $\boldsymbol{B}^{cond}$, a covariance term $\boldsymbol{B}^{cov}$ and the expected outer product, $\boldsymbol{B}^{ehw}$, all of which are weighted by functions of the paired sampling rates. Although this case is explicitly ruled out by Assumption 4, when $\rho = 0$, representing the case of a small sample drawn from a large population, the asymptotic variance reduces to the standard Eicker-Huber-White (EHW) variance. This case bears resemblance to the result presented in Xu and Wooldridge (2022). The underlying intuition is that, when sampling only a negligible portion of the population, most of the unit's neighbors remain unobserved. Consequently, spatial transformations are generally not applicable and there is no need to correct for any induced spatial dependence. On the contrary, when all units in the population are observed $\mathbf{B} = \mathbf{B}^{cond} + \mathbf{B}^{cov}$, the researcher must take into account the dependence induced by the transformation. Let us further decompose the covariance matrix as $\boldsymbol{B}^{cov} = \boldsymbol{B}^{ecov} - \boldsymbol{B}^{ucov}$, where

$$\boldsymbol{B}^{ecov} = \lim_{n\to\infty}\frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\ \sum_{k\neq i<l,l\in B_k\setminus k} E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\varepsilon_i\Delta_{kl}\boldsymbol{x}_k\Delta_{kl}\varepsilon_l), \tag{6}$$

$$\boldsymbol{B}^{ucov} = \lim_{n\to\infty}\frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\ \sum_{k\neq i<l,l\in B_k\setminus k} E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\varepsilon_i)E(\Delta_{kl}\boldsymbol{x}_k\Delta_{kl}\varepsilon_l)'.$$

Observe that when $\rho > 0$, the difference between the "superpopulation" part of the variance $\boldsymbol{B}^{ehw} + \rho^2\boldsymbol{B}^{ecov}$ and the finite population asymptotic variance $\mathbf{B}$ is positive semidefinite and equals $\rho^2(\boldsymbol{B}^{u} + \boldsymbol{B}^{ucov})$. This result extends the findings on the conservativeness of finite population variance estimation (Neyman, 1923, Abadie et al., 2020, Xu and Wooldridge, 2022) in the context of neighborhood unobservables.

In matrix notation, we can write the ND estimator as

$$\hat{\boldsymbol{\beta}}_{ND} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{D}}'\tilde{\boldsymbol{D}}\tilde{\boldsymbol{X}})^{-1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{D}}'\tilde{\boldsymbol{D}}\tilde{\boldsymbol{y}}) \tag{7}$$

where $\tilde{\boldsymbol{D}}$ is the $N_p \times N$ neighborhood differencing matrix. Hence, if the neighborhood specification forms the sampled pairs $(1,3)$, $(1,5)$, $(2,3)$, and $(2,4)$, among others, the first rows of the neighborhood differencing matrix will be

$$\tilde{\boldsymbol{D}} = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & -1 & \dots \\ 0 & 1 & -1 & 0 & 0 & \dots \\ 0 & 1 & 0 & -1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

3.2. The within-neighborhood estimator

For the variance of the NW estimator's key statistic, expressed as $\sum_i \Delta_{i,ni}\boldsymbol{x}_i\Delta_{i,ni}\varepsilon_i$, there are numerous components. Similar to the ND case, we must assume the existence of their limits.

Assumption 9. $\boldsymbol{B}_{NW}^{cond} = \lim_{n\to\infty}\frac{1}{n}\sum_i(1-\frac{2}{n_i})Var(\boldsymbol{x}_i\varepsilon_i)$,

$\boldsymbol{B}_{NW}^{neigh} = \lim_{n\to\infty} \frac{1}{n}\sum_i \frac{1}{n_i^2}\sum_{j\in B_i} Var(\boldsymbol{x}_j\varepsilon_j)$,

$\boldsymbol{B}_{NW}^{ehw} = \boldsymbol{B}_{NW}^{u} + \boldsymbol{B}_{NW}^{cond} = \lim_{n\to\infty} \frac{1}{n}\sum_i (1-\frac{2}{n_i})E(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$,
$\boldsymbol{B}_{NW}^{ehwneigh} = \boldsymbol{B}_{NW}^{uneigh} + \boldsymbol{B}_{NW}^{neigh} = \lim_{n\to\infty} \frac{1}{n}\sum_i \frac{1}{n_i^2}\sum_{j\in B_i} E(\varepsilon_j^2\boldsymbol{x}_j\boldsymbol{x}_j')$,

with $\boldsymbol{B}_{NW}^{u} = \lim_{n\to\infty} \frac{1}{n}\sum_i (1-\frac{2}{n_i})E(\boldsymbol{x}_i\varepsilon_i)E(\boldsymbol{x}_i\varepsilon_i)'$,
$\boldsymbol{B}_{NW}^{uneigh} = \lim_{n\to\infty} \frac{1}{n}\sum_i \frac{1}{n_i^2}\sum_{j\in B_i} E(\boldsymbol{x}_j\varepsilon_j)E(\boldsymbol{x}_j\varepsilon_j)'$, *and*

$\boldsymbol{B}_{NW}^{B_i} = \lim_{n\to\infty} \frac{1}{n}\sum_i \frac{1}{n_i^2}\sum_{j\in B_i}\sum_{k\neq j\in B_i} cov\Big((\boldsymbol{x}_j\varepsilon_j),(\boldsymbol{x}_k\varepsilon_k)\Big)$,
$\boldsymbol{B}_{NW}^{iB_i} = \lim_{n\to\infty} \frac{1}{n}\sum_i \frac{1}{n_i}\sum_{j\neq i\in B_i} cov\Big((\boldsymbol{x}_i\varepsilon_i),(\boldsymbol{x}_j\varepsilon_j)\Big)$,
$\boldsymbol{B}_{NW}^{\Delta B_i B_k} = \lim_{n\to\infty} \frac{1}{n}\sum_i \sum_{k\neq i} cov\Big((\Delta_{i,n_i}\boldsymbol{x}_i, \Delta_{i,n_i}\varepsilon_i)(\Delta_{k,n_k}\boldsymbol{x}_k, \Delta_{k,n_k}\varepsilon_k)\Big)$,
exist and are positive definite.

Let $\boldsymbol{B}_{NW}^{cov} = \boldsymbol{B}_{NW}^{B_i} - 2\boldsymbol{B}_{NW}^{iB_i} + \boldsymbol{B}_{NW}^{\Delta B_i B_k}$. The covariance can be decomposed in the usual way, as shown in equation (6), for the ND case. We can now derive the asymptotic distribution of the NW estimator viewed as an estimator of the causal estimand $\boldsymbol{\beta}_{NW}^{c}$.

Proposition 2. *Under Assumptions, 1- 7, 9, and Assumptions 10, 11, and 12 b) in Appendix A.1*

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{NW} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{A}_{NW}^{-1}\boldsymbol{B}_{NW}\boldsymbol{A}_{NW}^{-1}), \tag{8}$$

where $\boldsymbol{A}_{NW} = \lim_{n\to\infty} \Delta_{i,n_i}\boldsymbol{\Omega}_n^{xx}$ *and*

$\boldsymbol{B}_{NW} = \rho(\boldsymbol{B}_{ND}^{cond} + \boldsymbol{B}_{NW}^{neigh}) + (1-\rho)(\boldsymbol{B}_{NW}^{ehw} + \boldsymbol{B}_{NW}^{ehwneigh}) + \rho\boldsymbol{B}_{NW}^{cov}$.

The same logic as the ND asymptotic variance applies here. Nevertheless, because it's not possible to factor out the sampling indicators, we are compelled to break down the variance by distinguishing between the individual and neighborhood components of the within-transformation.

The NW transformation can be implemented using a within-neighborhood matrix $\boldsymbol{G}_n$ defined as $\boldsymbol{G}_n = \boldsymbol{I}_n - \boldsymbol{C}_n$, where $\boldsymbol{C}_n$ is a $n \times n$ matrix obtained in two steps: *i*) create a (binary) interaction matrix according to the neighborhood specification; *ii*) substitute the main diagonal with ones and row-normalize. In matrix notation, we have

$$\hat{\boldsymbol{\beta}}_{NW} = (\boldsymbol{X}'\boldsymbol{G}_n'\boldsymbol{G}_n\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{G}_n'\boldsymbol{G}_n\boldsymbol{y}). \tag{9}$$

Remark 1 (Non-overlapping neighborhoods). *If neighborhoods do not overlap, $\boldsymbol{G}_n$ is a projector and we have a formal equivalence with panel/clustered data econometrics. In this case, if the neighborhoods have the same size, then the ND and NW estimators are numerically equivalent. Furthermore, the NW estimator is more efficient. A formal proof applies the GLS device to show that the NW estimator can be written as a GLS ND estimator (see, among the others, Arellano, 2013).*

3.3. Estimation of the asymptotic variance

Now, we shift our focus to the challenge of estimating the asymptotic variance of our causal estimands. Observe that $\boldsymbol{B}^{cond}$ and $\boldsymbol{B}^{cov}$ for the ND estimator, and $\boldsymbol{B}^{cond}_{NW}$, $\boldsymbol{B}^{neigh}_{NW}$ and $\boldsymbol{B}^{cov}_{NW}$ for NW estimator are challenging to estimate because the researcher cannot observe different potential outcomes for the same unit (Neyman, 1923). For this reason, we focus on estimating the EHW and the expectation of the cross-product component of the asymptotic variance, namely $\boldsymbol{B}^{ehw}$, $\boldsymbol{B}^{ecov}$, $\boldsymbol{B}^{ehw}_{NW}$, $\boldsymbol{B}^{ehwneigh}_{NW}$, and $\boldsymbol{B}^{ecov}_{NW}$. We propose the following estimator for the asymptotic variance of the ND estimator

$$\hat{\boldsymbol{V}}_{ND} = \hat{\mathbf{A}}^{-1}\hat{\boldsymbol{B}}\hat{\mathbf{A}}^{-1}, \tag{10}$$

where

$$\begin{aligned}
\hat{\boldsymbol{B}} &= \sum_{i<j, j\in B_i\setminus i}\ \sum_{k<l, l\in B_k\setminus k} \mathbf{1}_{(ij=kl|k\in B_i|i\in B_k|k=j|i=l)} \tilde{\Delta}_{ij}\hat{\varepsilon}_i \tilde{\Delta}_{kl}\hat{\varepsilon}_k \tilde{\Delta}_{ij}\hat{\boldsymbol{x}}_i \tilde{\Delta}_{kl}\hat{\boldsymbol{x}}'_k \\
\hat{\mathbf{A}} &= \sum_{i<j, j\in B_i\setminus i} \tilde{\Delta}_{ij}\hat{\boldsymbol{x}}_i \tilde{\Delta}_{ij}\hat{\boldsymbol{x}}'_i = \sum_{i<j, j\in B_i\setminus i} (\tilde{\Delta}_{ij}\boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}\tilde{\Delta}_{ij}\boldsymbol{z}_i)(\tilde{\Delta}_{ij}\boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}\tilde{\Delta}_{ij}\boldsymbol{z}_i)', \ \ where \\
\hat{\boldsymbol{\lambda}} &= \left(\sum_{i<j, j\in B_i\setminus i} \tilde{\Delta}_{ij}\boldsymbol{t}_i \tilde{\Delta}_{ij}\boldsymbol{z}'_i\right)\left(\sum_{i<j, j\in B_i\setminus i} \tilde{\Delta}_{ij}\boldsymbol{z}_i \tilde{\Delta}_{ij}\boldsymbol{z}'_i\right)^{-1},
\end{aligned} \tag{11}$$

$\tilde{\Delta}_{ij}\hat{\varepsilon}_i = \tilde{\Delta}_{ij}y_i - (\tilde{\Delta}_{ij}\boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}\tilde{\Delta}_{ij}\boldsymbol{z}_i)'\hat{\boldsymbol{\beta}}^c_{ND} - \tilde{\Delta}_{ij}\boldsymbol{z}'_i\hat{\boldsymbol{\gamma}}^c_{ND}$, and $1_{(ij=kl|k\in B_i|i\in B_k|k=j|i=l)}$ is used to select pairs for computing variances, and pairs sharing at least one common node. This estimator is a special case of the general dyadic covariance estimator proposed by Tabord-Meehan (2019). Under our set of Assumptions, we can show that (10) is consistent following Lemma 2 in Abadie et al. (2020). The finite sample properties of this estimator have been studied in Belotti et al. (2018) who show, through Monte Carlo simulations, that (10) outperforms other robust competitors proposed in the related literature (see, e.g. Fack and Gret, 2010). Furthermore, this estimator is similar to the Spatial HAC (Heteroskedasticity and Autocorrelation Consistent) estimator suggested by Xu and Wooldridge (2022), where weights are set to one if $d_{ij} \leq d^*$, and zero otherwise. Observe, however, that this weight definition leads to an omission of units that overlap between neighborhoods in the summation terms.

Similarly, we can write the NW asymptotic variance estimator as

$$\hat{\boldsymbol{V}}_{NW} = \hat{\mathbf{A}}^{-1}_{NW}\hat{\boldsymbol{B}}_{NW}\hat{\mathbf{A}}^{-1}_{NW}, \tag{12}$$

where

$$\hat{\boldsymbol{B}}_{NW} = \sum_i \sum_k \mathbf{1}_{(B_i \cap B_k \neq \emptyset)} \tilde{\Delta}_{i,n_i} \hat{\varepsilon}_i \tilde{\Delta}_{kn_k} \hat{\varepsilon}_k \tilde{\Delta}_{i,n_i} \hat{\boldsymbol{x}}_i \tilde{\Delta}_{kn_k} \hat{\boldsymbol{x}}'_k.$$

$$\hat{\mathbf{A}}_{NW} = \sum_{i=1}^{n} \tilde{\Delta}_{i,n_i} \hat{\boldsymbol{x}}_i \tilde{\Delta}_{i,n_i} \hat{\boldsymbol{x}}'_i = \sum_i (\tilde{\Delta}_{i,n_i} \boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}_{NW} \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i)(\tilde{\Delta}_{i,n_i} \boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}_{NW} \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i)', \ \ where$$

$$\hat{\boldsymbol{\lambda}}_{NW} = \left(\sum_i \tilde{\Delta}_{i,n_i} \boldsymbol{t}_i \tilde{\Delta}_{i,n_i} \boldsymbol{z}'_i \right) \left(\sum_i \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i \tilde{\Delta}_{i,n_i} \boldsymbol{z}'_i \right)^{-1}, \tag{13}$$

where $\tilde{\Delta}_{i,n_i}\hat{\varepsilon}_i = \tilde{\Delta}_{i,n_i} y_i - (\tilde{\Delta}_{i,n_i}\boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}_{NW}\tilde{\Delta}_{i,n_i}\boldsymbol{z}_i)'\hat{\boldsymbol{\beta}}^c_{NW} - \tilde{\Delta}_{i,n_i}\boldsymbol{z}'_i\hat{\boldsymbol{\gamma}}^c_{NW}$, and $1_{(B_i \cap B_k \neq \emptyset)}$ is the indicator of the event $B_i \cap B_k \neq \emptyset$, i.e., it selects units that belong to the same neighborhood, as well as those shared between two neighborhoods. Let us define $\boldsymbol{V}^{ehwcov}_{ND} = \boldsymbol{A}^{-1}(\boldsymbol{B}^{ehw} + \rho^2\boldsymbol{B}^{ecov})\boldsymbol{A}^{-1}$, and $\boldsymbol{V}^{ehwcov}_{NW} = \boldsymbol{A}^{-1}_{NW}(\boldsymbol{B}^{ehw}_{NW} + \boldsymbol{B}^{ehwneigh}_{NW} + \rho\boldsymbol{B}^{ecov}_{NW})\boldsymbol{A}^{-1}_{NW}$. The next Lemma shows that the proposed asymptotic variance estimators are conservative for the finite population asymptotic variances.

Lemma 2. *a) Under Assumptions, 1-8, and Assumption 12 a) in Appendix A.2,* $\hat{\boldsymbol{V}}_{ND} \xrightarrow{p} \boldsymbol{V}^{ehwcov}_{ND}$.
b) Under Assumptions, 1- 7, 9, and Assumption 12 b) in Appendix A.2, $\hat{\boldsymbol{V}}_{NW} \xrightarrow{p} \boldsymbol{V}^{ehwcov}_{NW}$.

We conjecture that we can use attributes information to improve the variance estimator, similar to the approach in Abadie et al. (2020).

4. Testing for smooth fixed effects and optimal distance

In what follows, we propose a sequential testing strategy allowing to test for non smooth fixed effects, and conditionally on smoothness, test for the optimal specification of the neighborhood. For clarity and simplicity, we assume that the attributes are uncorrelated with the assignments, i.e. $\boldsymbol{t}_i = \boldsymbol{x}_i$. By considering constant treatment effects, we can formally represent the linear regression function describing the potential outcomes as

$$y_i(\boldsymbol{x}_i) = \boldsymbol{x}'_i\boldsymbol{\beta} + \theta_i + \epsilon_i, \tag{14}$$

where ϵ_i is assumed to have zero mean and unit variance.

As outlined in the previous section, the proposed estimators outperform OLS only when the fixed effect θ_i exhibits spatial smoothness. Conditional on spatial smoothness, identifying the optimal distance threshold becomes crucial to satisfy Assumption (2). This natural ordering suggests the application of a fixed-sequence testing strategy (Westfall and Krishen, 2001). The latter is based on a stepwise procedure in which, for each hypothesis, testing is conditional upon rejecting all hypotheses earlier in the sequence. As long as significant results are observed in all the preceding steps, this allows to controls the family-wise error rate without the need for a multiplicity adjustment (Dmitrienko and Tamhane, 2010). In our case, we have two sequential null hypotheses. Under the null hypothesis of non-smooth fixed effects, both the proposed estimators and OLS are inconsistent, rendering the determination of an optimal distance threshold irrelevant. Conversely, if the null hypothesis is rejected in favor of smooth fixed effects, testing for an optimal distance threshold becomes relevant and appropriate.

In order to test for the null of non-smooth fixed effects, we propose to leverage the Mundlak (1978) device.[7] Indeed, as in the case of panel data, in our context, assuming that $\theta_i = \alpha + \bar{\boldsymbol{x}}_i'\boldsymbol{\xi} + e_i$, where e_i has zero mean and is assumed to be uncorrelated with $\boldsymbol{x}_i$, and $\bar{\boldsymbol{x}}_i = \frac{1}{n_i^d}\sum_j \boldsymbol{x}_j$ for all units $j \in B_i^d$, provides an alternative way to account for smooth fixed effects. Plugging-in for θ_i in equation (14), gives the expanded model

$$y_i(\boldsymbol{x}_i) = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + \bar{\boldsymbol{x}}_i'\boldsymbol{\xi} + v_i. \tag{15}$$

The OLS is a consistent estimator of α, $\boldsymbol{\beta}$, and $\boldsymbol{\xi}$ in equation (15) when the threshold d for the NDT is optimal. The quasi-Mundlak (QM) estimator is particularly useful in this context because, regardless of the distance threshold, neighborhood averages fail to account for fixed effects when they are not spatially smooth. However, when fixed effects are smooth, neighborhood averages effectively capture their influence. This enables us to construct a powerful test for the presence of smooth fixed effects by simply examining the joint statistical significance of the parameters $\boldsymbol{\xi}$ in model (15). It is worth noting that the equivalence between our QM and the NW estimator only holds when neighborhoods do not overlap, forming clusters. Section 5 presents the finite-sample performance of the QM estimator in the presence of overlapping neighborhoods, demonstrating a behavior similar to that of NW.

If the null hypothesis of non-smooth fixed effects is rejected, the next step in our sequential testing strategy can be undertaken. To identify d^*, the optimal threshold that satisfies Assumption (2), we propose a test leveraging the contrast between the ND and NW estimators. A similar strategy has been proposed by Bartolucci et al. (2015) to test for time-invariant (versus time-varying) unit effects in generalized linear models for panel data. In our case, if the ND/NW transformations rule out smooth fixed effects, then both estimators are consistent for $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\beta}}_{ND} - \hat{\boldsymbol{\beta}}_{NW} \xrightarrow{p} 0$. On the other hand, when the transformation is not able to get rid of the smooth fixed effects, both estimators are inconsistent but converge in probability to different points in the parameter space, provided that $d << d_{max} = \sup_{ij} d_{ij}$. Under the null hypothesis of optimal distance,

$$\tau_n \begin{pmatrix} \hat{\boldsymbol{\beta}}_{ND} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\beta}}_{NW} - \boldsymbol{\beta} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{ND} & \boldsymbol{C}_{ND,NW} \\ \boldsymbol{C}'_{ND,NW} & \boldsymbol{V}_{NW} \end{pmatrix}\right],$$

where $\tau_n = \sqrt{N}$ is the rate of convergence for the NW estimator, the slower one. This implies that the asymptotic null distribution of $\tau_n\hat{\boldsymbol{\delta}} = \tau_n(\hat{\boldsymbol{\beta}}_{ND} - \hat{\boldsymbol{\beta}}_{NW})$ is Gaussian with mean zero and variance $\boldsymbol{V}_\delta = \boldsymbol{V}_{ND} + \boldsymbol{V}_{NW} - \boldsymbol{C}_{ND,NW} - \boldsymbol{C}'_{ND,NW}$. Consistent estimators of $\boldsymbol{V}_{ND}$ and $\boldsymbol{V}_{NW}$ are provided in Appendix A.4 under the homoskedasticity assumption. A consistent estimator of $\boldsymbol{C}_{ND,NW}$ is $\widehat{\boldsymbol{C}}_{ND,NW} = \widehat{\boldsymbol{A}}_{NW}^{-1}\widehat{\boldsymbol{B}}_{ND,NW}\widehat{\boldsymbol{A}}_{ND}^{-1}$, with

$$\widehat{\boldsymbol{B}}_{ND,NW} = \hat{\sigma}^2\boldsymbol{X}'\boldsymbol{D}\boldsymbol{D}'\boldsymbol{G}_n\boldsymbol{G}_n'\boldsymbol{X},$$

[7] We thank an anonymous referee for suggesting the possibility of constructing an "almost exact" Mundlak (1978) estimator, which we call the quasi-Mundlak estimator.

where $\hat{\sigma}^2$ can be estimated either with (A.16) or (A.18), and $\widehat{\boldsymbol{A}}^{-1}$ is provided in Appendix A.4 for both estimators.[8]

Therefore, our test statistic is

$$\hat{\xi} = \tau_n \hat{\boldsymbol{\delta}}' \widehat{\boldsymbol{V}}_{\delta}^{-} \hat{\boldsymbol{\delta}}, \tag{16}$$

where $\widehat{\boldsymbol{V}}_{\delta} = \widehat{\boldsymbol{V}}_{ND} + \widehat{\boldsymbol{V}}_{NW} - \widehat{\boldsymbol{C}}_{ND,NW} - \widehat{\boldsymbol{C}}'_{ND,NW}$, and $\widehat{\boldsymbol{V}}_{\delta}^{-}$ denotes a generalized inverse of $\widehat{\boldsymbol{V}}_{\delta}$.[9] By construction, $\widehat{\boldsymbol{V}}_{\delta}$ is guaranteed to be non-negative definite. The asymptotic null distribution of $\hat{\xi}$ as $n \to \infty$ is χ^2 with a number of degrees of freedom equal to the rank of $\boldsymbol{V}_{\delta}$.

Similarly to the panel data case, the proposed test lacks power when the ND and NW estimators are algebraically equivalent – that is, when each unit has the same number of neighbors, and all neighborhoods are non-overlapping (see Remark 1). The power of the test depends on the degree of divergence between the ND and NW estimators.

In the following, we examine the inconsistency of the ND and NW estimators in the presence of spatially smooth unobserved heterogeneity to assess the power of the proposed test. Let us define $x_i = \phi\theta_i + (1-\phi^2)^{1/2}\omega_i$, and focus, for simplicity, on a single assignment assuming that $E(x_i) = 0$. Here, ω_i represents the random component of the treatment assignment, and it is considered to be uncorrelated to θ_i and has zero mean and unit variance. Given a distance threshold $d \neq d^*$, the ND and NW estimation errors can then be written as

$$\hat{\beta}_{NW} - \beta = \frac{\sum_{i=1}^{n} \tilde{x}_i \tilde{u}_i}{\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i}, \quad \hat{\beta}_{ND} - \beta = \frac{\sum_{i<j, j \in B_i \setminus i} \Delta x_{ij} \Delta u_{ij}}{\sum_{i<j, j \in B_i \setminus i} \Delta x_{ij} \Delta x_{ij}},$$

where, $\tilde{x}_i = \Delta_{i,n_i} x_i$, $\Delta x_{ij} = \Delta_{ij} x_i$ and $u_i = \theta_i + \epsilon_i$. The error u_i is now stochastic, as the transformation is applied at a suboptimal distance and the randomness of θ_i is not fully eliminated.
As $n \to \infty$, the ND and NW asymptotic biases are[10]

$$\hat{\beta}_{NW} - \beta \xrightarrow{p} \frac{E(\tilde{x}_i \tilde{u}_i)}{E(\tilde{x}_i \tilde{x}_i)}, \quad \hat{\beta}_{ND} - \beta \xrightarrow{p} \frac{E(\Delta x_{ij} \Delta u_{ij})}{E(\Delta x_{ij} \Delta x_{ij})}.$$

We can show that $E(\tilde{x}_i \tilde{u}_i) = \phi\tilde{\tau}$, $E(\tilde{x}_i \tilde{x}_i) = \phi^2\tilde{\tau} + (1-\phi^2)$, $E(\Delta x_{ij} \Delta u_{ij}) = \phi\Delta\tau$ and $E(\Delta x_{ij} \Delta x_{ij}) = \phi^2\Delta\tau + (1-\phi^2)$, where $\tilde{\tau} = E(\theta_i - \bar{\theta}_i)^2$ with $\bar{\theta}_i = n_i^{-1}\sum_{j \in B_i} \theta_j$, and $\Delta\tau = E(\theta_i - \theta_j)^2, \forall\, j \in B_i$. The estimator asymptotic biases become

$$\hat{\beta}_{NW} - \beta \xrightarrow{p} \frac{\phi\tilde{\tau}}{\phi^2\tilde{\tau} + (1-\phi^2)}, \quad \hat{\beta}_{ND} - \beta \xrightarrow{p} \frac{\phi\Delta\tau}{\phi^2\Delta\tau + 2(1-\phi^2)}. \tag{17}$$

This shows that our test has no power when $\phi = 0$ because, in this case, both estimators are consistent

[8] It is important to note that the differing unit-wise and pair-wise structures of the ND and NW estimators make the derivation of a robust version of $\widehat{\boldsymbol{B}}_{ND,NW}$ nontrivial. Developing a heteroskedasticity-robust version of the test remains an avenue for future research.

[9] Generalized inverses are not unique, but Holly and Monfort (1986) show that test statistics of the form (16) are invariant to the choice of generalized inverse.

[10] Convergence in probability of the numerator can be proved following the same argument used for the convergence of the denominators in Lemma 1 in Appendix A.2.

regardless of the distance threshold used for the NDT, nor when $\phi = \pm 1$, because both converge to $\beta \pm 1$. For a sufficiently large n, as d increases, the test loses power because the two estimators' asymptotic biases converge to the OLS estimator's asymptotic bias. Specifically, when d reaches its maximum value, $d_{\max}$, all units belong to the same neighborhood, causing the asymptotic biases of both estimators to coincide with that of the OLS estimator. In all the other cases, the fact that $\hat{\boldsymbol{\beta}}_{ND}$ and $\hat{\boldsymbol{\beta}}_{NW}$ behave differently as functions of d represent the source of the increasing power of our test, provided that $d << d_{max}$. We show the behavior of the asymptotic biases in (17) using a DGP that resembles our assumptions (see Section 5 for additional details). Figure 2 shows the behavior of ND and NW asymptotic biases (top panel) and the corresponding $plim(\hat{\beta}_{ND} - \hat{\beta}_{NW})$ (bottom panel) over the support of d. As expected, the biases are approximately equal to zero when $d = d^*$ is the optimal threshold. Then, they increase differently as d gets larger and converge to the OLS asymptotic bias when d is close to the maximum observed distance between units. $plim(\hat{\beta}_{ND} - \hat{\beta}_{NW})$ mimics the behavior of the asymptotic power of the test, i.e., rapidly increases when d moves away from the optimal threshold, reaches its maximum when the divergence between the two asymptotic biases is the largest and goes to zero when $d = d_{max}$.

Observe that the class of hypotheses is indexed by the threshold d. Given the sequential nature of the Hausman-like tests, we suggest starting from a small enough distance threshold and sequentially increasing it until the null hypothesis of optimal distance cannot be rejected. In doing so, Rosenbaum (2008) shows that this procedure allows to bound the probability of rejecting the null hypothesis when it is true at the nominal size α. In Appendix Appendix A.5, we formally link our testing strategy with the *testing hypotheses in order* framework in Rosenbaum (2008).

5. Monte Carlo evidence

We now present Monte Carlo evidence on the following aspects: *i*) the behavior of the variance estimators in Section 3.3; and *ii*) the size and power of the two-step testing strategy proposed in Section 4. For all the analyses presented in this Section, we generate units located on a regular lattice using the following potential outcomes data generating process (DGP)

$$y_i(x_i) = \theta_i + x_i\beta_i + \epsilon_i,$$

where θ_i is simulated as $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{1}, \mathbf{V})$, where the covariance matrix is given by $\mathbf{V} = \exp\left[-\frac{\mathbf{M}^2}{2s^2}\right]$. Here, $\mathbf{M}$ represents the Euclidean distance matrix for the regular lattice, and $s = 2$ is a smoothing parameter that governs fixed effect smoothness. Alternatively, when fixed effects are assumed to be i.i.d., θ_i is drawn from a uniform distribution, $\theta_i \sim \mathcal{U}[0, 3]$.[11]

Figure 3 shows the distribution of θ_i over the lattice for different sample sizes, comparing the cases with and without smoothness. The population values of ϵ_i and β_i are set as $\epsilon_i \sim \mathcal{N}(0, 1)$, and $\beta_i = \mathcal{N}(z_i, 1)$, with $z_i \sim \mathcal{N}(2, 4)$. We follow Abadie et al. (2020) by keeping fixed θ_i, β_i, and ϵ_i

[11]For the non-smooth case, we also simulated $\theta_i \sim \mathcal{N}(1, \sigma^2)$. The results of these alternative simulation experiments remain qualitatively similar, but the differing support of θ_i makes the graph less clear-cut.

over simulation repetitions. Thus, the treatment is the only stochastic component of the DGP, and is generated in each simulation replication. It is correlated with the smooth unobservables according to $x_i = \phi\theta_i + (1-\phi^2)^{1/2}\omega_i$, where $\omega_i \sim \mathcal{N}(0,1)$ is the random part of the assignment and ϕ is the correlation parameter. All experiments are conducted with $10,000$ simulation repetitions.

To study the behavior of the variance estimators in equations (10) and (12), we set $s = 2$ and $\phi = 0.5$, generating a population of size $n = 10,000$ units. In each simulation repetition, we randomly sample units with probability ρ from this population. We consider four designs with different sampling probabilities, $\rho = (0.1, 0.5, 0.9, 1)$. Consequently, the sample size N is random for the first three designs with $E[N] = n\rho$. For each design and repetition, we apply the ND, NW and QM estimators (see 7, 9 and 15) and the corresponding proposed robust variance estimators (see 10 and 12) using the optimal distance threshold.[12] Table 1 presents the results. Specifically, we report the standard deviations of $(\hat{\beta}_{ND} - \beta^c_{ND})$, $(\hat{\beta}_{NW} - \beta^c_{NW})$, and $(\hat{\beta}_{QM} - \beta^c_{QM})$ across repetitions, along with the average standard errors from the proposed estimators and their corresponding 95% confidence interval coverage rates. For the first design where $\rho = 0.1$, the proposed variance estimators provide accurate estimates of the standard deviations of $(\hat{\beta}_{ND} - \beta^c_{ND})$ and $(\hat{\beta}_{NW} - \beta^c_{NW})$, while the EHW estimator is already conservative for the average causal effect in the case of the QM estimator. When ρ gets larger, all the considered variance estimators are conservative, and their coverage rate is almost equal to the nominal coverage. This evidence extends the findings on the conservativeness of finite population variance estimation (Neyman, 1923, Abadie et al., 2020, Xu and Wooldridge, 2022) to regression models with smooth neighborhood unobservables, where spatial analogs of first-differencing and within-group transformations are required.

As for the size and power analysis of the two-step testing strategy proposed in Section 4, we use the same DGP as in the previous exercise and set $\phi = 0.5$. However, rather than sampling from a larger finite population, we generate smaller populations of varying sizes, $n \in 400, 1600, 3600$, allowing for randomness in the error term.[13]

First, we examine the properties of the test for detecting smooth fixed effects. We generate i.i.d. fixed effects to assess the test size, while power is analyzed using spatially smooth fixed effects. For each simulation repetition, we apply the QM estimator for a set of distance thresholds excluding the optimal distance and test whether the coefficient associated with the neighborhood averages ($\bar{x}_i$) is equal to zero using a heteroskedasticity-robust F statistic. Table 2 reports the size (panel a) and power (panel b) of our test by sample size, along with the mean and standard deviation (SD) of the test statistic. The table provides two clear takeaways. First, the proposed test exhibits overall small size distortions, especially as the population increases. However, size distortions tend to rise with larger distance thresholds and remain moderate when $n = 400$. Second, the test shows excellent power regardless of the distance threshold or population size.

Second, we analyze the properties of the Hausman-like specification test, once again setting $s = 2$

[12]For the QM estimator, being the OLS estimator of the parameters in model (15), we use the standard EHW sandwich covariance matrix.

[13]This approach is equivalent to drawing a small random sample from a large population. The simulation results remain quantitatively unchanged when non-stochastic errors are used.

for generating smooth fixed effects. Figure 4 illustrates the power of the test as a function of population size and the distance threshold used to define neighborhoods, starting from the optimal distance. The red dotted line represents the nominal size $\alpha = 0.05$. Consistent with the asymptotic behavior of the ND and NW estimators (see Figure 2), both estimators converge to the true average causal effect when the distance is optimal. In this case, under the null hypothesis, the test exhibits a size that closely matches the expected nominal level. As the distance threshold increases, the two estimators begin to diverge, and the test becomes highly powerful even for minor deviations from the optimal distance. As expected, as the population size gets larger, the power of the test significantly improves. However, as the distance threshold further increases, the ND and NW estimators gradually converge toward the OLS estimator (see also Figure 2), resulting in a progressive decline in test power.

6. Empirical illustration: Miguel and Kremer (2004)

For illustrative purposes, we use data from the primary school deworming project conducted in western Kenya (Miguel and Kremer, 2004, MK).[14] Worm infection rates were relatively high in this area, especially among school-age children. Indeed, 37% of interviewed children reported having at least one moderate-to-heavy helminth infection.

An essential feature of the deworming program is that the randomization takes place at the school level, allowing the identification of the effect of deworming even in the presence of externalities. As pointed out by MK, school-level randomization naturally generates local variation in the density of treatment that can be exploited to disentangle the direct treatment effect from the indirect effects, i.e., externalities across schools.

The authors consider the following spatial externality model

$$y_{is} = \alpha + \beta T_s + \underbrace{\sum_d (\gamma_d N_{ds}^T)}_{\text{observed externalities}} + \sum_d (\phi_d N_{ds}) + \boldsymbol{x}_{is}\boldsymbol{\delta} + \epsilon_{is} \tag{18}$$

where y_{is} is an indicator of school attendance or health status, s refers to the school, i to the student, T_{is} is treatment status, N_{ds} the total number of pupils in primary schools at geographical distance d from school s, and N_{ds}^T is the number of these pupils randomly assigned to the treatment, while $\boldsymbol{x}_{is}$ are school and pupil characteristics. Thus, the overall average deworming effect is: $\beta + \sum_d (\gamma_d \bar{N}_{ds}^T)$, where $\bar{N}_{ds}^T$ is the average number of treated school pupils located at distance d from school s. The sample includes 2,328 pupils and 49 schools, 25 assigned to the treatment. The terms denoted by the curly bracket below in equation (18) capture cross-school externalities that, in this context, are likely to diffuse smoothly over space.

Table 3, Columns 3 and 6, replicates the results for moderate-to-heavy helminth infections reported

[14] Here, we use the updated data available in the Miguel and Kremer (2004)'s replication package, downloadable here. A detailed description of the deworming program can be found in Miguel and Kremer (2004). The replication code for the analysis in this Section is available from the authors upon request.

in Miguel and Kremer (2014, Table B.1) using a weighted linear probability model.[15] In particular, Column 6 presents the LPM estimates of model (18), where both N^T_{ds} and N_{ds} are included for pupils located within 0-3 and 3-6 kilometers.[16] As observed, there is no evidence of cross-school externalities at a distance of 3-6 kilometers. Furthermore, including this term in the regression renders the average cross-school externality effect statistically insignificant. Based on this evidence, MK argue that incorporating the 3-6 kilometer effect is inappropriate and select the specification that excludes this term as the preferred model. This specification, replicated in Column 6, yields a statistically significant average cross-school externality effect and provides a more precise estimate of the overall deworming effect.

While a primary objective in MK is to estimate the overall deworming effect by distinguishing between direct effects and cross-school externalities, our focus here is on testing for the presence of local smooth fixed effects. Such effects may arise from imperfect randomization, selective school absenteeism, treatment non-compliance, or infection transmission dynamics, potentially introducing bias in the estimation of the overall deworming effect (Miguel and Kremer, 2004). Importantly, they may also capture observed, potentially smooth, externalities.

To address this, we first use the QM-type regression in equation (15), which can be interpreted as a homogeneous version of the MK model in equation (18), to test the null hypothesis of non-smooth fixed effects. Specifically, we estimate the following model using OLS

$$y_{is} = \alpha + \beta T_s + \gamma_d N^T_{ds} + \boldsymbol{x}_{is}\boldsymbol{\delta} + \epsilon_{is}. \tag{19}$$

Notably, our test does not require including N_{ds} in the model, as in equation (18), since this term is essential for capturing cross-school externalities but redundant when testing for smooth fixed effects. If smooth fixed effects are detected, the next step is to apply the Hausman-like specification test in (16) to determine the optimal distance at which they no longer pose a problem for inference. Given the structural similarity between QM- and MK-type regressions, this distance allows us to specify an MK-type model that effectively captures cross-school externalities while minimizing potential bias from locally smooth, correlated unobservables.

Table 4, Column 5, shows that the null hypothesis of non-smooth fixed effects can be rejected at all distances except 2 kilometers, indicating the presence of smooth, correlated unobservables. Additionally, Column 6 reveals that 4 kilometers is the first distance threshold at which we fail to reject the null hypothesis that ND and NW estimators are equal.[17] Interestingly, at 5 kilometers, the null hypothesis is rejected again, but from 6 kilometers onward, it is no longer rejected. We interpret this as evidence that some smooth fixed effects may persist beyond 4 kilometers, suggesting that the MK-type model should account for a finer specification of externalities up to 6 kilometers. This specification helps control for

[15] We specifically refer to Columns 2 and 3 of Table B.1 in the Miguel and Kremer (2014, replication manual), which updates the results in Table 7 of the published paper. We thank the authors for kindly providing us with the matrix reporting the geographical distance between schools.

[16] It is important to note that, while MK present average treatment effects from a probit model, our LPM replication closely approximate their results.

[17] The test statistic is computed by contrasting the $\hat{\beta}$ coefficient in model (18), after partialling out all other regressors.

unobserved local heterogeneity while improving the identification of cross-school externalities.

Guided by this evidence, we also estimated model (18) using a more granular specification of cross-school externality terms. Table 3, Columns 1 and 2, present these estimates. We find that specifying externality terms in 2-kilometer steps captures larger and more precisely estimated average cross-school externality effects, which are statistically different at 5% (χ^2-stat = 5.85 for the difference between the two effects[18]) from those selected by MK, and reported in Column 6. This result also underscores that including the 4-6 kilometers externality effect is justified from a mean squared error perspective, as advocated by the authors (Miguel and Kremer, 2014, Replication Manual, Section 4.9).

7. Concluding remarks

Nowadays, there is a growing availability of larger datasets incorporating geo-coded information. If inference relies on sampling uncertainty, the researcher may find that only a few neighbors are observed in the sample when the sampling rate is negligible. Conversely, sampling uncertainty approaches zero when the sample effectively represents the entire population, as is often the case with spatial data.

This article proposes estimation strategies based on neighborhood data transformations (NDT) for models with additively-separable smooth unobserved heterogeneity. The proposed framework integrates design-based with sampling-based uncertainty. We study the asymptotic properties of the neighborhood difference (ND) and within neighborhood (NW) class of estimators. Following the inferential framework outlined by Abadie et al. (2020) and Xu and Wooldridge (2022) to account for design uncertainty and model the spatial dependence, we adopt a finite population viewpoint by introducing spatial correlation within the population before sampling. We propose a testing strategy to assess whether fixed effects are spatially smooth by leveraging a quasi-Mundlak approach and selecting the optimal distance based on the contrast between ND and NW estimators. Through Monte Carlo simulations, we validate the theoretical predictions about the finite population asymptotic variance and evaluate the performance of the proposed variance estimators. Further, we study the size and power of our testing strategy across varying threshold distances. We find that, as suggested by theory, ND and NW estimators approach consistency when the fixed effects are smooth and the distance threshold used for transformation is optimal. When the distance threshold increases, the estimators start diverging, making powerful the test based on their contrast.

Finally, we illustrate the usefulness of our approach using data from the seminal study by Miguel and Kremer (2004) on a health program in Kenya. While the authors leverage school-level randomization to identify direct and indirect average treatment effects, we focus on testing for local smooth fixed effects and determining the optimal distance threshold. Applying an NDT at different thresholds reveals significant smooth fixed effects up to six kilometers, supporting an MK-type specification that includes all cross-school externality terms, rather than only those up to three kilometers, as originally suggested by the authors. Notably, we find that the average cross-school externality effect is larger and more precisely estimated. This suggests that expanding similar interventions across a wider geographic area could

[18]We run the test using standard errors clustered at school level and taking into account the covariance between the two average cross-school externality effects.

amplify their overall impact, maximizing health benefits at a lower marginal cost.

Future extensions of our framework naturally include partially linear and nonlinear models, along with the formal derivation of the quasi-Mundlak estimator's properties.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1):265–296.

Arellano, M. (2013). *Panel Data Econometrics*. Oxford University Press.

Auerbach, E. (2022). Identification and estimation of a partially linear regression model using network data. *Econometrica*, 90(1):347–365.

Bartolucci, F., Belotti, F., and Peracchi, F. (2015). Testing for time-invariant unobserved heterogeneity in generalized linear models for panel data. *Journal of Econometrics*, 184(1):111–123.

Belotti, F., Di Porto, E., and Santoni, G. (2018). Spatial differencing: estimation and inference. *CESifo Economic Studies*, 64(2):241–254.

Belotti, F., Di Porto, E., and Santoni, G. (2021). The effect of local taxes on firm performance: Evidence from geo-referenced data. *Journal of Regional Science*, 61(2):492–510.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation Of Elementary Education. *The Quarterly Journal of Economics*, 114(2):577–599.

Bradley, R. C. and Tone, C. (2017). A central limit theorem for non-stationary strongly mixing random fields. *Journal of Theoretical Probability*, 30(2):655–674.

Chirinko, R. S. and Wilson, D. J. (2008). State investment tax incentives: A zero-sum game? *Journal of Public Economics*, 92(12):2362 – 2384. New Directions in Fiscal Federalism.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Dmitrienko, A. and Tamhane, A. C. (2010). Gatekeeping procedures in clinical trials. In *Multiple Testing Problems in Pharmaceutical Statistics*, pages 165–192. CRC Press.

Druckenmiller, H. and Hsiang, S. (2018). Accounting for unobservable heterogeneity in cross section using spatial first differences. Working Paper 25177, National Bureau of Economic Research.

Duranton, G., Gobillon, L., and Overman, H. G. (2011). Assessing the effects of local taxation using microgeographic data. *Economic Journal*, 121(555):1017–1046.

Einiö, E. and Overman, H. G. (2020). The effects of supporting local business: Evidence from the UK. *Regional Science and Urban Economics*, 83:1–14.

Fack, G. and Gret, J. (2010). When do better schools raise housing prices? Evidence from Paris public and private schools. *Journal of Public Economics*, 94:59 – 77.

Gibbons, S., Machin, S., and Silva, O. (2013). Valuing school quality using boundary discontinuities. *Journal of Urban Economics*, 75:15 – 28.

Harjunen, O., Kortelainen, M., and Saarimaa, T. (2018). Best education mo can buy? Capitalization of school quality in finland. *CESifo Economic Studies*, 64(2):150–175.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Holmes, T. J. (1998). The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of political Economy*, 106(4):667–705.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150(1):86–98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of econometrics*, 170(1):178–190.

Kahn, M. E. (2004). Domestic pollution havens: Evidence from cancer deaths in border counties. *Journal of Urban Economics*, 56(1):51 – 69.

Miguel, E. and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Miguel, E. and Kremer, M. (2014). Replication data for: Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, pages 465–472. Reprinted in 1990.

Rosenbaum, P. R. (2008). Testing hypotheses in order. *Biometrika*, 95(1):248–252.

Ruud, P. A. (1984). Tests of specification in econometrics. *Econometric Reviews*, 3(2):211–242.

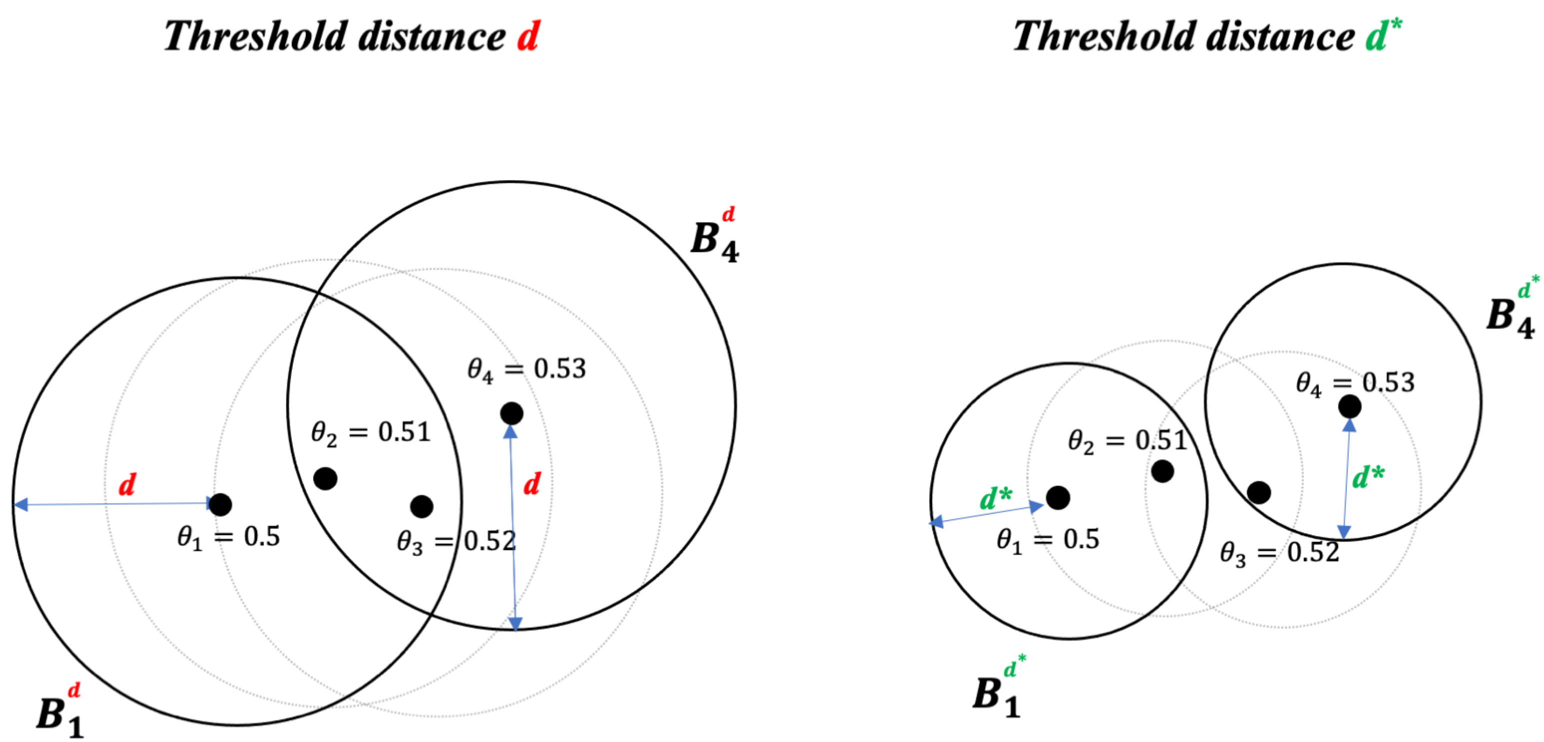
Sales, A. C. (2017). Sequential specification tests to choose a model: A change-point approach. *arXiv preprint arXiv:1708.00907*.

Tabord-Meehan, M. (2019). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business & Economic Statistics*, 37(4):671–680.

Westfall, P. H. and Krishen, A. (2001). Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *Journal of Statistical Planning and Inference*, 99(1):25–40.

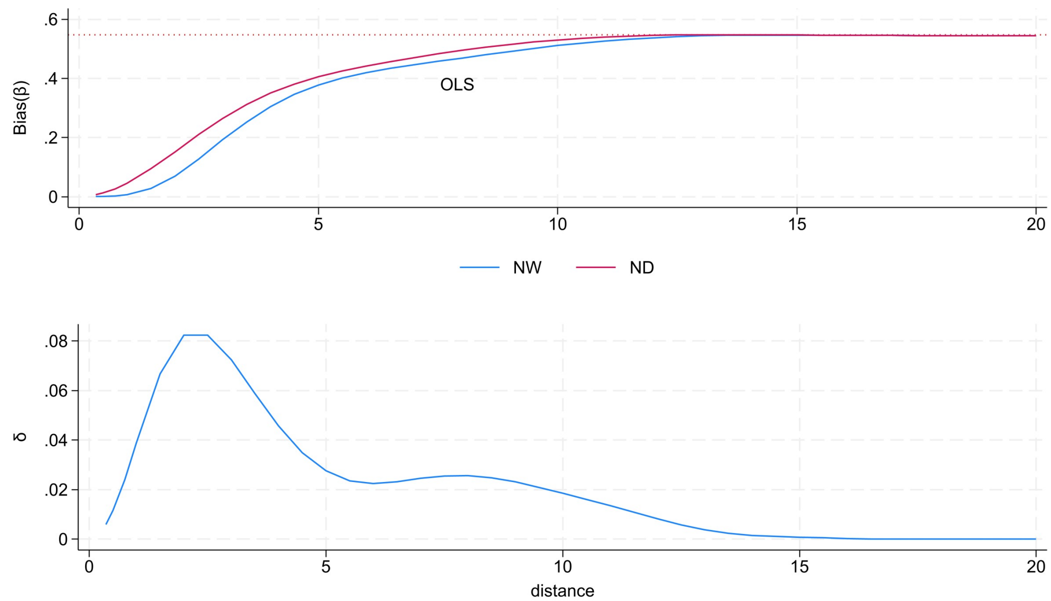
Xu, R. and Wooldridge, J. M. (2022). A design-based approach to spatial correlation.

Figure 1: Neighborhood Selection by Threshold Distances



Notes. The graph reports neighborhood specifications based on two different proximity thresholds. The four neighborhoods are centered towards the units with radius d (left) and d^* (right). d^* is assumed to be the threshold defining neighborhoods where units are more similar in unobservables. The example assumes that the four unit-specific unobservables are smooth over the space considered.

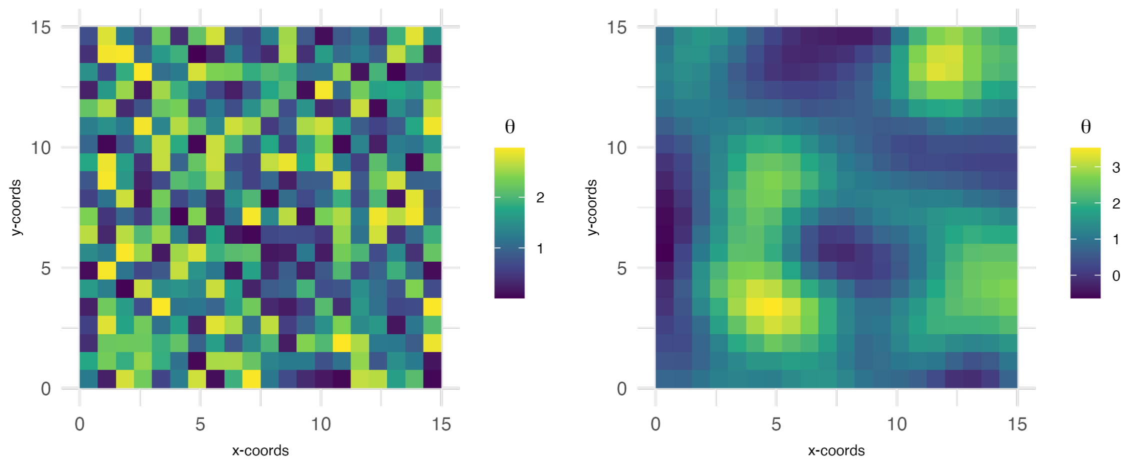
Figure 2: Asymptotic biases of ND and NW.



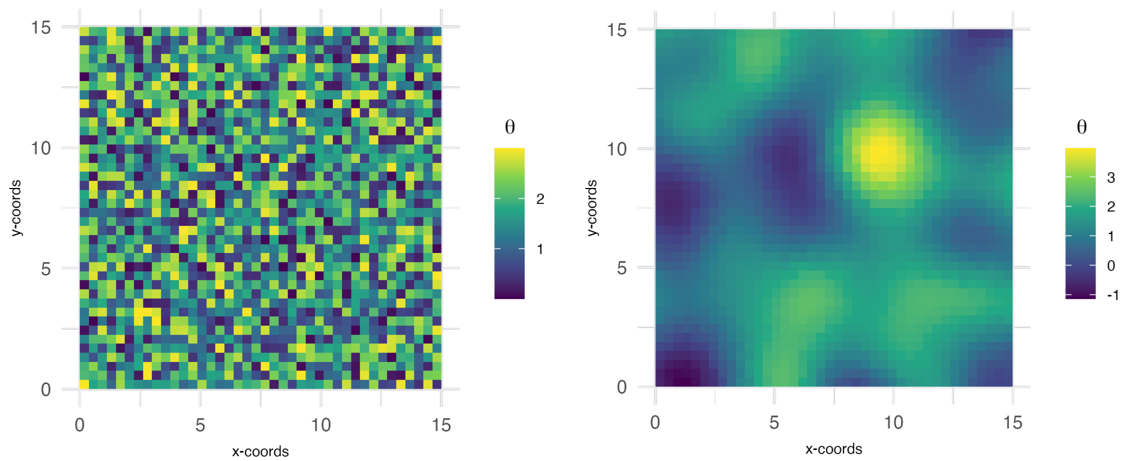
Notes. The top graph reports the asymptotic biases in (17) at different distance thresholds $d \in [d^*, \sup_{ij} d_{ij}]$. The dotted red line indicates the bias of the OLS estimator. The bottom graph shows the computation of $plim(\hat{\beta}_{ND} - \hat{\beta}_{NW})$ based on the difference of the ND and NW asymptotic biases in 17. θ_i is generated according to the data generating process described in Section 5, setting $\phi = 0.5$ and $n = 3,600$.

Figure 3: Distribution of θ.

(a) $n = 20$



(b) $n = 40$



(c) $n = 60$

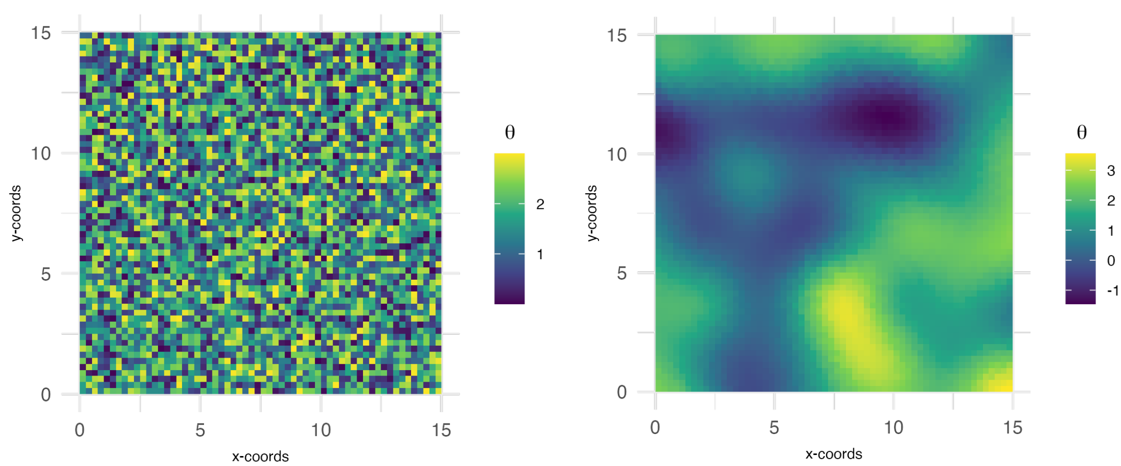


Table 1: Monte Carlo evidence: Standard errors and coverage for nominal 95% confidence intervals

ND				
	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1$
sd$(\hat{\beta} - \beta^c)$	0.248	0.072	0.048	0.046
Coverage based on sd$(\hat{\beta} - \beta^c)$	0.948	0.936	0.954	0.946
Average $\hat{se}$	0.249	0.076	0.053	0.050
Coverage based on $\hat{se}$	0.943	0.949	0.967	0.961
NW				
	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1$
sd$(\hat{\beta} - \beta^c)$	0.246	0.073	0.049	0.048
Coverage based on sd$(\hat{\beta} - \beta^c)$	0.951	0.942	0.952	0.949
Average $\hat{se}$	0.248	0.076	0.053	0.050
Coverage based on $\hat{se}$	0.950	0.949	0.967	0.953
Quasi-Mundlak				
	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1$
sd$(\hat{\beta} - \beta^c)$	0.238	0.069	0.047	0.046
Coverage based on sd$(\hat{\beta} - \beta^c)$	0.946	0.941	0.945	0.950
Average $\hat{se}$	0.243	0.073	0.052	0.049
Coverage based on $\hat{se}$	0.948	0.953	0.968	0.957

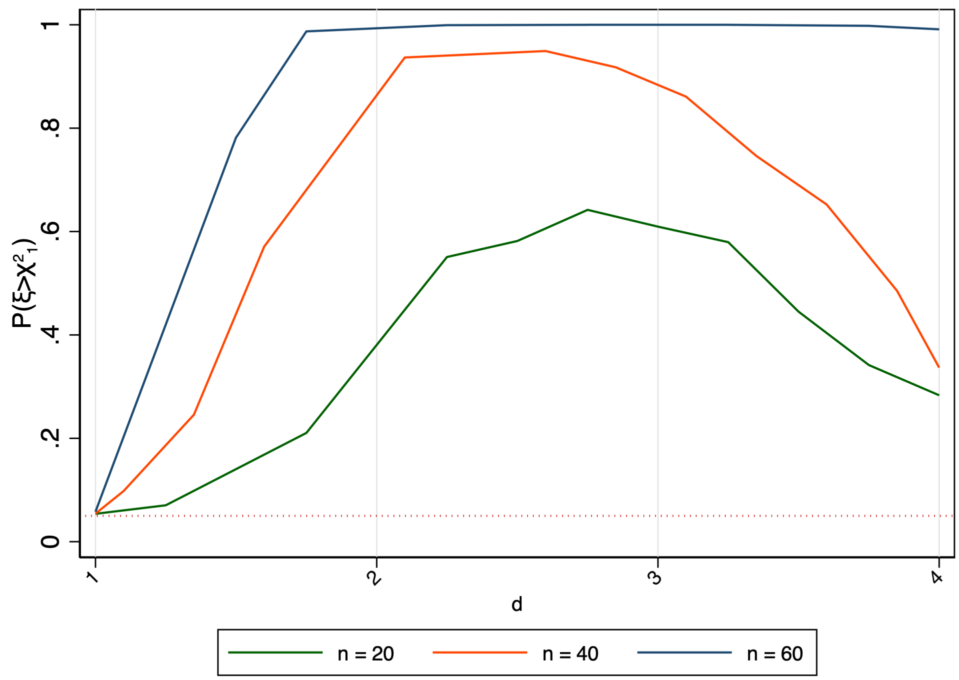
Notes: The table reports the standard deviation of $(\hat{\beta} - \beta^c)$ over replication, the coverage for nominal 95% intervals for different sampling rates: $\rho = (0.1, 0.5, 0.9, 1)$ for a population of $n = 10,000$ units with $\beta_i = \mathcal{N}(z_i, 1)$, with $z_i \sim \mathcal{N}(2, 4)$. We employ $10,000$ simulation repetitions.

Table 2: Size and power analysis of the test for smooth fixed effects by sample size (Step 1)

(a) Non-smooth				(b) Smooth			
$n = 400$				$n = 400$			
d	Mean	SD	Size	d	Mean	SD	Power
0.75	0.8	1.12	0.026	0.75	11.4	6.40	0.911
1	0.8	1.22	0.035	1	21.9	8.97	0.997
1.25	0.8	1.13	0.029	1.25	26.4	9.67	1.000
1.5	0.9	1.30	0.039	1.5	29.1	10.21	1.000
1.75	0.8	1.10	0.026	1.75	29.3	10.05	1.000
2	0.8	1.11	0.028	2	30.1	10.30	1.000
2.25	0.8	1.17	0.029	2.25	29.8	10.39	1.000
2.5	0.8	1.08	0.023	2.5	27.2	9.51	1.000
2.75	0.7	0.98	0.019	2.75	25.3	9.17	1.000
3	0.7	1.06	0.024	3	23.5	8.79	0.999
3.25	0.7	0.96	0.015	3.25	19.2	7.88	0.997
3.5	0.7	0.97	0.017	3.5	16.0	7.08	0.986
$n = 1600$				$n = 1600$			
d	Mean	SD	Size	d	Mean	SD	Power
0.75	0.9	1.36	0.046	0.75	78.9	16.80	1.000
1	0.9	1.30	0.039	1	94.4	18.20	1.000
1.25	1.0	1.36	0.046	1.25	100.4	18.74	1.000
1.5	1.0	1.32	0.042	1.5	102.9	18.84	1.000
1.75	1.0	1.48	0.056	1.75	103.9	18.98	1.000
2	1.0	1.44	0.054	2	101.1	18.81	1.000
2.25	1.1	1.48	0.057	2.25	97.2	18.15	1.000
2.5	1.1	1.60	0.066	2.5	92.6	17.68	1.000
2.75	1.2	1.62	0.071	2.75	85.5	16.64	1.000
3	1.2	1.71	0.076	3	77.2	15.84	1.000
3.25	1.3	1.74	0.083	3.25	68.6	14.84	1.000
3.5	1.4	1.84	0.090	3.5	58.2	13.42	1.000
$n = 3600$				$n = 3600$			
d	Mean	SD	Size	d	Mean	SD	Power
0.75	1.1	1.50	0.058	0.75	368.7	39.31	1.000
1	0.9	1.33	0.043	1	397.6	40.89	1.000
1.25	0.9	1.25	0.039	1.25	410.5	41.40	1.000
1.5	0.8	1.18	0.031	1.5	414.9	41.58	1.000
1.75	0.9	1.24	0.036	1.75	411.9	40.74	1.000
2	0.9	1.22	0.035	2	398.3	40.00	1.000
2.25	0.9	1.24	0.038	2.25	380.9	38.56	1.000
2.5	0.9	1.22	0.035	2.5	357.2	36.97	1.000
2.75	0.9	1.22	0.036	2.75	331.1	34.90	1.000
3	1.0	1.30	0.042	3	302.4	33.11	1.000
3.25	1.0	1.29	0.042	3.25	275.9	30.53	1.000
3.5	0.9	1.26	0.039	3.5	248.2	28.40	1.000

Notes: The table reports the size (panel a) and power (panel b) of our test by sample size, along with the mean and standard deviation (SD) of the test statistic. We employ $10,000$ simulation repetitions.

Figure 4: Power curve of the test for optimal distance by sample size (step 2).



Notes. The optimal distances are $d^* = 0.75$ ($n = 20$), $d^* = 0.40$ ($n = 40$) , and $d^* = 0.25$ ($n = 60$). The distances reported on the x-axis has been shifted so that all three lines originate from the same starting point ($d^* = 1$). At the optimal distance, the test operates under the null hypothesis, with the plot displaying the size of the test. As d increases, the lines depict the power of the test (10,000 Monte Carlo replications).

Table 3: Replication and new evidence on the findings of Miguel and Kremer (2014) (Any moderate-heavy helminth infection, 1999)

	1km step (long)	2km step (long)	3km step (long)	1km step (short)	2km step (short)	3km step (short)
Indicator for Group 1 (1998 Treatment) School	-0.199***	-0.240***	-0.297***	-0.256***	-0.272***	-0.318***
	(0.0555)	(0.0531)	(0.0548)	(0.0495)	(0.0496)	(0.0501)
Group 1 pupils within 1 km (per 1000 pupils)	0.627*			0.608*		
	(0.3466)			(0.3578)		
Group 1 pupils within 1-2 km (per 1000 pupils)	-0.061			-0.144		
	(0.1349)			(0.1418)		
Group 1 pupils within 2-3 km (per 1000 pupils)	-0.325***			-0.296***		
	(0.0929)			(0.0983)		
Group 1 pupils within 3-4 km (per 1000 pupils)	-0.302**			-0.068		
	(0.1168)			(0.1022)		
Group 1 pupils within 4-5 km (per 1000 pupils)	-0.061					
	(0.0931)					
Group 1 pupils within 5-6 km (per 1000 pupils)	-0.072					
	(0.1076)					
Group 1 pupils within 2 km (per 1000 pupils)		0.025			-0.093	
		(0.1517)			(0.1492)	
Group 1 pupils within 2-4 km (per 1000 pupils)		-0.280***			-0.217***	
		(0.0773)			(0.0773)	
Group 1 pupils within 4-6 km (per 1000 pupils)		-0.083				
		(0.0562)				
Group 1 pupils within 3 km (per 1000 pupils)			-0.186**			-0.205**
			(0.0873)			(0.0814)
Group 1 pupils within 3-6 km (per 1000 pupils)			-0.047			
			(0.0687)			
F-stat	5.804	4.496	2.297	8.022	4.820	6.344
(p-value)	0.000	0.007	0.112	0.000	0.012	0.015
$\bar{R}^2$	0.519	0.508	0.505	0.510	0.503	0.502
Avg overall cross-school externality effect	-0.217	-0.218	-0.127	-0.100	-0.135	-0.084
(t-stat)	-2.879	-2.997	-1.339	-2.320	-3.098	-2.519
Overall deworming effect	-0.416	-0.458	-0.424	-0.355	-0.407	-0.402
(t-stat)	-5.975	-7.410	-5.484	-6.743	-7.842	-8.080

Notes: Grade 3-8 pupils. Linear probability model estimation, cluster robust standard errors at school level in parentheses. Observations are weighted by total school population. Significantly different from zero at 99% (***), 95% (**), and 90% (*) confidence levels. The 1999 parasitological survey data are for Group 1 and Group 2 schools. The pupil population data is from the 1998 School questionnaire. The specification includes the same set of control variables used in Miguel and Kremer (2014): grade indicators, school assistance controls, district exam scores and the total number of children attending primary school within a certain distance from the school. The F-stat reported in the bottom panel is testing the joint statistical significance of the cross-school externality terms included in the model. $\bar{R}^2$ is the adjusted R^2.

Table 4: Testing for non-smooth fixed effects and optimal distance.

Distance	# pupils	avg neighbors	# pairs	H_0: non-smooth fixed-effects	H_0: Optimal distance
2 Km	1,502	69.14	25,962	0.38 (0.542)	14.16 (0.000)
3 Km	2,088	133.02	69,438	4.65 (0.036)	7.23 (0.007)
4 Km	2,328	190.97	111,143	4.59 (0.037)	1.45 (0.228)
5 Km	2,328	291.96	169,923	5.25 (0.026)	4.40 (0.036)
6 Km	2,328	389.44	226,651	5.20 (0.027)	1.08 (0.299)

Notes: The table reports test statistics and p-values (in parentheses) for the two-step testing strategy by distance thresholds for the sample of Grade 3-8 pupils. For additional details see Section 4.

Appendix A. Appendix

Appendix A.1. NED conditions and CLT

We first introduce useful definitions based on the NED and mixing concepts in Jenish and Prucha (2012) and Xu and Wooldridge (2022).

Definition 1. *For any random vector Y, $\|Y\|_p = (E|Y|^p)^{1/p}$ denotes its L_p-norm. Denote $\mathcal{F}_{i,n}(s)$ as a σ-field generated by the random vectors θ's located within the neighborhood B_i^s, which is a ball centered at the location $l(i)$ with a radius s in d^0-dimensional Euclidean space D.*

Definition 2. *Let $\boldsymbol{M} = \{M_{in}, i \in D_n, m \geq 1\}$ be a random field, and let $\boldsymbol{\theta} = \{\theta_i, i \in T_n, n \geq 1\}$ be another random field, where $|T_n| \to \infty$ as $n \to \infty$. Let $\boldsymbol{\nu} = \{\nu_{in}, i \in D, n \geq 1\}$ be an array of finite positive constants. Then the random field $\boldsymbol{M}$ is said to be $L_p(d)$-near-epoch dependent on the random field $\boldsymbol{\theta}$ if:*

$$\|M_{in} - E(M_{in}|\mathcal{F}_{in}(s))\|_p \leq \nu_{in}\psi(s)$$

for some sequence $\psi(s) \geq 0$ with $\lim\limits_{s\to\infty} \psi(s) = 0$. The $\psi(s)$ are called the NED coefficients, and the ν_{in} are called the NED scaling factors. $\boldsymbol{M}$ is said to be L_p-NED on $\boldsymbol{\theta}$ of size $-\boldsymbol{\lambda}$ if $\psi(s) = O(s^{-\boldsymbol{\mu}})$ for some $\boldsymbol{\mu} > \boldsymbol{\lambda} > 0$.

To derive the limiting distributions of our estimators we need to assume that both y_i, and $\boldsymbol{x}_i$ are NED on θ_i that is a mixing process. This is true for example if θ_i is an infinite moving average random field (so that it can be assumed to be smooth over the space) under some conditions, i.e. $\theta_i = \sum_{j\in\mathbb{Z}^d} b_{ij}\zeta_j$, with $\lim\limits_{s\to\infty} \sup_{n,i\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d, d_{ij}>s} |b_{ij,n}| < 0$, and $\sup_{n,i\in\mathbb{Z}^d} ||\zeta_{in}||_p$ is finite with $p > 1$ (Jenish and Prucha, 2012). The NED property is preserved when performing summation, multiplication, and Lipschitz transformations, thus the NED properties will be preserved under NDT. In accordance with the approach described in Xu and Wooldridge (2022), we adhere to the mixing definition presented in Bradley and Tone (2017). The distance between any subsets K, $V \in D$ is defined as $d(K, V) = \inf\{d_{ij} : i \in K, \, and \, j \in V\}$.

Definition 3. *Let $\mathcal{A}$ and $\mathcal{B}$ be two sub-σ-algebras of $\mathcal{F}$, and let*

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\left(|P(AB) - P(A)P(B)|, A \in \mathcal{A}, B \in \mathcal{B}\right)$$

and

$$\rho(\mathcal{A}, \mathcal{B}) = \sup|corr(f, g)|, f \in L^2_{real}(\mathcal{A}), g \in L^2_{real}(\mathcal{B}).$$

For $K \subseteq D_n$ and $V \subseteq D_n$, let $\sigma_n(K) = \sigma(\theta_i, i \in K)$ and $\alpha_n(K, V) = \alpha(\sigma_n(K), \sigma_n(V))$. Then, the α-mixing coefficient for the random field θ is defined as:

$$\overline{\alpha}(r) = \sup_n \sup_{K,V} \left(\alpha_n(K, V), d(K, V) \geq r\right).$$

The maximal correlation coefficient is defined as:

$$\overline{\rho}(r) = \sup_n \sup_{K,V} \left(\rho_n(K, V), d(K, V) \geq r\right).$$

Based on Lemma B.4 from Xu and Wooldridge (2022), we employ a modified Central Limit Theorem (CLT) for Non-Equidistantly Dependent (NED) processes to establish convergence in distribution.

Theorem 1 (CLT for NED processes). *Let $\{q_{ni}, 0 \leq i \leq D_n, , n \geq 1\}$ be a real valued zero-mean random field that is L_2-NED on $\boldsymbol{\theta} = \{\theta_{i,n}, i \in D_n,\ n \geq 1\}$, with the scaling factors ν_{in} and the NED coefficients $\psi(s)$. Let us define, $Q_n = \sum_{i \in D_n} \frac{R_i}{\sqrt{\rho_n}} q_{in}$ and $\sigma_n = Var(Q_n)$. Let us assume*

a) $\sup_{n, i \in D_n} E(|q_{ni}/c_{in}|)^{2+\delta} < \infty$ *for some $\delta > 0$, where $c_{in} > 0$ is a sequence of constants.*

b) for any fixed $s > 0$, there exist a positive constatnt C such that for any n and every nonempty set $K \subseteq D_n$, $E(\sum_{i \in K} \frac{R_i}{\sqrt{\rho_n}} E(q_{ni}/j_n|\mathcal{F}(s)))^2 \geq C \sum_{i \in K} E(\frac{R_i}{\sqrt{\rho_n}} E(q_{ni}/j_n|\mathcal{F}(s)))^2$, where $j_n = \max_{i \in D_n}\{c_{in}, \nu_{in}\}$,

c) $\inf_n |D_n|^{-1} j_n^{-2} \sigma_n^2 > 0$,

d) NED coefficients $\psi(s)$ is of size $-\boldsymbol{\nu}$,

e) NED scaling factors satisfy $\sup_{n, i \in D_n} c_{in}^{-1} \nu_{in} \leq C < \infty$,

then, under Assumptions 1, 3, 4, and 10

$$Q_n \sigma_n^{-1} \xrightarrow{d} N(0,1).$$

To establish the asymptotic normality of $\boldsymbol{\beta}_{ND}$ and $\boldsymbol{\beta}_{NW}$, we need to introduce additional technical conditions for the application of Theorem 1. These supplementary assumptions may not be straightforward to interpret, as exemplified in prior work (see, for example, Jenish and Prucha, 2012). You can refer to Lemma 4 and 6 for a detailed examination of how these conditions are applied. We provide a list of these conditions below.

Assumption 10 (Mixing condition). *For the input random field θ: (i) $\overline{\alpha}(r) \to 0$ as $r \to \infty$; (ii)* $\lim_{r \to \infty} \overline{\rho}(r) < 1$.

Assumption 11 (NED condition). *The random field $\boldsymbol{M} = (y, \boldsymbol{x})$ is L_2-NED on $\theta = \{\theta_{in}, i \in T_n, n \geq 1\}$ with the scaling factors $\boldsymbol{\nu}$ and the NED coefficients $\psi(s)$ of size $-\boldsymbol{\nu}$.*

Assumption 12. *a) ND estimator: let $\Delta_{ij}\varepsilon_i = \Delta_{ij} y_i - \Delta_{ij} \boldsymbol{x}_i \boldsymbol{\beta}^{cs} - \Delta_{ij} \boldsymbol{z}_i \boldsymbol{\gamma}^{cs}$ and $\Delta_{ij} V_{ni} = \boldsymbol{a}' \Delta_{ij} \boldsymbol{x}_i \Delta_{ij} \varepsilon_i$ for any conformable vector $\boldsymbol{a}$. We assume, that, $\inf_{n_p} n_p^{-1} \sigma_n^2 > 0$, where $\sigma_n^2 = Var(\sum_{i<j \in B_i \setminus i} \frac{R_{ij}}{\sqrt{\rho_n^2}} (\Delta_{ij} V_{ni} - \mu_{ij}))$, and $\sup_{n, i \in D_n} \nu_{in} < \infty$. Furthermore, there exists a constan C such that,*

$$\sum_{i<j \in B_i \setminus i} \sum_{k \neq i < l \in B_k \setminus k} E\Bigg(E\Bigg(\Delta_{ij} V_i \Bigg| \mathcal{F}_{in}(s)\Bigg) E\Bigg(\Delta_{kl} V_k \Bigg| \mathcal{F}_{kn}(s)\Bigg)\Bigg) \geq C.$$

b) NW estimator: let $\Delta_{i,n_i}\varepsilon_i = \Delta_{i,n_i} y_i - \Delta_{i,n_i} \boldsymbol{x}_i \boldsymbol{\beta}^c - \Delta_{i,n_i} \boldsymbol{z}_i \boldsymbol{\gamma}^c$ and $\Delta_{i,n_i} V_{ni} = \boldsymbol{a}' \Delta_{i,n_i} \boldsymbol{x}_i \Delta_{i,n_i} \varepsilon_i$ for any conformable vector $\boldsymbol{a}$. We assume that $\inf_n n^{-1} \sigma_n^2 > 0$, where $\sigma_n^2 = Var(\sum_i \frac{1}{\sqrt{\rho_n}} (R_i V_i - \frac{1}{n_i} \sum_j R_j V_j - \rho_n(\mu_i - \frac{1}{n_i} \sum_j \mu_j))$, and $\sup_{n, i \in D_n} \nu_{in} < \infty$. Furthermore, there exist a constant C such that

$$\sum_i \sum_{k \neq i} E\Bigg(E\Bigg(V_i \Bigg| \mathcal{F}_{in}(s)\Bigg) E\Bigg(V_k \Bigg| \mathcal{F}_{kn}(s)\Bigg)\Bigg) \geq C,$$

$$\sum_i \sum_{k \neq i} E\Bigg(E\Bigg(\frac{1}{n_i} \sum_{j \in B_i} V_j \Bigg| \mathcal{F}_{in}(s)\Bigg) E\Bigg(\frac{1}{n_k} \sum_l V_l \Bigg| \mathcal{F}_{kn}(s)\Bigg)\Bigg) \geq C.$$

Appendix A.2. Proofs of propositions

Under Assumption 2, it's important to note that all expectations and probabilities are implicitly conditioned on $d = d^*$. Throughout the subsequent proofs, we frequently rely on Assumption 3 (the Assignment mechanism) without explicit reference.

Proof of Lemma 1 *a*).

The proof follows the same argument as in Abadie et al. (2020). Let

$$\Delta_{ij}\boldsymbol{W}_n = \frac{1}{n_p}\sum_{i<j, j\in B_i\setminus i}\Delta_{ij}\boldsymbol{w}_i$$

$$\Delta_{ij}\boldsymbol{\Omega}_n = \frac{1}{n_p}\sum_{i<j, j\in B_i\setminus i}E\Delta_{ij}\boldsymbol{w}_i,$$

where the expectation is taken over the distribution of $\boldsymbol{x}$ and $\Delta_{ij}\boldsymbol{w}_i = \begin{pmatrix}\Delta_{ij}y_i\\ \Delta_{ij}\boldsymbol{x}_i\\ \Delta_{ij}\boldsymbol{z}_i\end{pmatrix}\begin{pmatrix}\Delta_{ij}y_i\\ \Delta_{ij}\boldsymbol{x}_i\\ \Delta_{ij}\boldsymbol{z}_i\end{pmatrix}'$. Consider the sample counterpart of $\Delta_{ij}\boldsymbol{W}_n$

$$\tilde{\Delta}_{ij}\boldsymbol{W}_n\frac{1}{N_p}\sum_{i<j, j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{w}_i,$$

where $\tilde{\Delta}_{ij}\boldsymbol{w}_i = \begin{pmatrix}\tilde{\Delta}_{ij}y_i\\ \tilde{\Delta}_{ij}\boldsymbol{x}_i\\ \tilde{\Delta}_{ij}\boldsymbol{z}_i\end{pmatrix}\begin{pmatrix}\tilde{\Delta}_{ij}y_i\\ \tilde{\Delta}_{ij}\boldsymbol{x}_i\\ \tilde{\Delta}_{ij}\boldsymbol{z}_i\end{pmatrix}'$.

Let $\Delta_{ij}W_i^{(k,l)}$ be the (k,l) element of the matrix $\Delta_{ij}\boldsymbol{W}_n$, the same apply for the other matrices. By Assumption 4 for any fixed $0 < \epsilon < 1$, there is $n_{p,\epsilon}$ such that $n_p > n_{p,\epsilon}$, we have $n_p\rho_n^2 > -log(\epsilon)$. So, for $n_p > n_{p,\epsilon}$ we have

$$Pr(N_p = 0) = \left(1 - \frac{n_p\rho_n^2}{n_p}\right)^{n_p} < \left(1 + \frac{log(\epsilon)}{n_p}\right)^{n_p} < e^{log(\epsilon)} = \epsilon.$$

This implies that $Pr(N_p = 0) \to 0$ and

$$E\Big((\tilde{\Delta}_{ij}W_i^{(k,l)} - \Delta_{ij}\Omega_i^{(k,l)})^2|N_p = 0\Big)Pr(N_p = 0) = (\Delta_{ij}\Omega_i^{(k,l)})^2 Pr(N_p = 0) \to 0$$

by Assumption 5 *a)*, Holder's inequality, and w.l.o.g. assuming that the elements of $\tilde{\Delta}_{ij}\boldsymbol{W}_n = \tilde{\Delta}_{ij}\boldsymbol{\Omega}_n = \boldsymbol{0}$ when $N_p = 0$. For any integer $1 \leq n_1 \leq n_p$, observe that

$$E\left(\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)} - E(\Delta_{ij}w_i^{(k,l)})\middle| N_p = n_1\right) = \frac{n_p}{n_1}E(R_{ij}(w_i^{(k,l)} - w_j^{(k,l)})|N_p = n_1) - E(\Delta_{ij}w_i^{(k,l)}) = 0 \tag{A.1}$$

given that $E(R_{ij}|N_p = n_1) = \frac{n_1}{n_p}$.

Now, to show the convergence in probability of the terms of interest we compute the convergence in (conditional) quadratic mean. We start with the following term

$$E\Bigg(\Bigg(\frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)}-E(\Delta_{ij}w_i^{(k,l)})\Bigg)^2\Bigg|N_p=n_1\Bigg).$$

We can rewrite this as:

$$\frac{1}{n_p^2}\sum_{i<j,j\in B_i\setminus i}\Bigg(E\Bigg(\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)}-E(\Delta_{ij}w_i^{(k,l)})\Bigg)^2\Bigg|N_p=n_1\Bigg) \tag{A.2}$$
$$+\frac{1}{n_p^2}\Bigg(\sum_{i<j,j\in B_i\setminus i}\sum_{f\neq i<g,g\in B_f\setminus f}E\Bigg(\Bigg(\frac{n_p}{n_1}\tilde{\Delta}_{ij}w_i^{(k,l)}-E(\Delta_{ij}w_i^{(k,l)})\Bigg)\Bigg(\frac{n_p}{n_1}\tilde{\Delta}_{fg}w_f^{(k,l)}-E(\Delta_{fg}w_f^{(k,l)})\Bigg)\Bigg|N_p=n_1\Bigg).$$

Thus, the first term of (A.2) becomes

$$\frac{1}{n_p^2}\Bigg(\sum_{i<j,j\in B_i\setminus i}E\Bigg(\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)}-E(\Delta_{ij}w_i^{(k,l)})\Bigg)^2\Bigg|N_p=n_1\Bigg)\leq\frac{1}{n_p^2}\Bigg(\sum_{i<j,j\in B_i\setminus i}E\Bigg(\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)}\Bigg)^2\Bigg|N_p=n_1\Bigg).$$

Then,

$$\frac{1}{n_p^2}\Bigg(\sum_{i<j,j\in B_i\setminus i}E\Bigg(\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)}\Bigg)^2\Bigg|N_p=n_1\Bigg)=$$
$$=\frac{1}{n_p^2}\Bigg(\sum_{i<j,j\in B_i\setminus i}\frac{n_p^2}{n_1^2}E(R_{ij}(w_i^{(k,l)}-w_j^{(k,l)})|N_p=n_1)^2\Bigg)=$$
$$=\frac{1}{n_p^2}\Bigg(\sum_{i<j,j\in B_i\setminus i}\frac{n_p^2}{n_1^2}E(R_{ij}^2|N_p=n_1)E((\Delta_{ij}w_i^{(k,l)})^2)\Bigg).$$

So, conditional on $N_p=n_1$, we can factor for $E(R_{ij})=\frac{n_1}{n_p}$

$$=\frac{1}{n_p^2}\Bigg(\sum_{i<j,j\in B_i\setminus i}\frac{n_p}{n_1}E((\Delta_{ij}w_i^{(k,l)})^2)\Bigg)\leq\frac{C}{n_1},$$

where the last inequality holds by Assumption 5 *a)* for large n_p.

Let

$$\tilde{\Xi}_{n_1}=\begin{cases}\Bigg(\frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)}-E(\Delta_{ij}w_i^{(k,l)})\Bigg)^2 & if\ \ n_1>0\\ 0 & if\ \ n_1=0\end{cases}$$

and

$$\Xi_{n_1} = \begin{cases} C/n_1 & if \;\; n_1 > 0 \\ 0 & if \;\; n_1 = 0 \end{cases}$$

$$E\left(\left(\frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)} - E(\Delta_{ij}w_i^{(k,l)})\right)^2 \middle| N_p > 0\right)Pr(N_p > 0) = E(\tilde{\Xi}_{N_p}) \leq E(\Xi_{N_p}).$$

Now, we want to study the behavior of Ξ_{N_p} when n_p goes to infinity. Observe that, for any $\epsilon > 0$, $Pr(\Xi_{N_p} > \epsilon) \leq Pr(0 < N_p < C/\epsilon) < Pr(N_p < C/\epsilon)$ because Ξ_{N_p} can be also zero and the last interval includes the previous ones. Now for any $\epsilon > 0$ if we apply the Chernoff's bounds to a sum of bernoully rvs we have:

$$Pr(N_p < C/\epsilon) = Pr(N_p < (1-\delta)n_p\rho_n^2) \leq exp(-\delta^2 n_p \rho_n^2/2) \to 0,$$

where $\delta = \frac{(n_p\rho_n^2 - C/\epsilon)}{n_p\rho^2}$. Thus, $\Xi_{N_p} = o_p(1)$ and given that it is bounded also $E(\Xi_{N_p}) = o(1)$. Let us denote $E(\Delta_{ij}w_i^{(k,l)}) = \mu_{ij}$. The second term of (A.2) becomes

$$\frac{1}{n_p^2}\left(\sum_{i<j,j\in B_i\setminus i}\;\sum_{f\neq i<g,g\in B_f\setminus f} E\left(\left(\frac{n_p}{n_1}\tilde{\Delta}_{ij}w_i^{(k,l)} - E(\Delta_{ij}w_i^{(k,l)})\right)\left(\frac{n_p}{n_1}\tilde{\Delta}_{fg}w_f^{(k,l)} - E(\Delta_{fg}w_f^{(k,l)})\right)\middle| N_p = n_1\right) = \tag{A.3}$$

$$= \frac{1}{n_p^2}\left(\sum_{i<j,j\in B_i\setminus i}\;\sum_{f\neq i<g,g\in B_f\setminus f}\left(E(\Delta_{ij}w_i^{(k,l)}\Delta_{fg}w_f^{(k,l)}) - 2\mu_{fg}\mu_{ij} + \mu_{ij}\mu_{fg}\right)\right)$$

$$= \frac{1}{n_p^2}\left(\sum_{i<j,j\in B_i\setminus i}\;\sum_{f\neq i<g,g\in B_f\setminus f}\left(E(\Delta_{ij}w_i^{(k,l)}\Delta_{fg}w_f^{(k,l)}) - \mu_{fg}\mu_{ij}\right)\right) \to 0,$$

given that $E(R_{ij}|N_p = n_1) = \frac{n_1}{n_p}$ in the second equality, by Assumption 5 *a)*, and n_p^2 that is dominating the convergence. Therefore,

$$E\left(\left(\frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\frac{n_p}{N_p}\tilde{\Delta}_{ij}w_i^{(k,l)} - E(\Delta_{ij}w_i^{(k,l)})\right)^2\right) = o(1).$$

■

Proof of Lemma 1 *b*).

Let

$$\Delta_{i,n_i}\boldsymbol{W}_n = \frac{1}{n}\sum_i \Delta_{i,n_i}\boldsymbol{w}_i$$

$$\Delta_{i,n_i}\boldsymbol{\Omega}_n = \frac{1}{n}\sum_i E\Delta_{ij}\boldsymbol{w}_i,$$

where the expectation is taken over the distribution of $\boldsymbol{x}$ and $\Delta_{i,n_i}\boldsymbol{w}_i = \begin{pmatrix} \Delta_{i,n_i}y_i \\ \Delta_{i,n_i}\boldsymbol{x}_i \\ \Delta_{i,n_i}\boldsymbol{z}_i \end{pmatrix} \begin{pmatrix} \Delta_{i,n_i}y_i \\ \Delta_{i,n_i}\boldsymbol{x}_i \\ \Delta_{i,n_i}\boldsymbol{z}_i \end{pmatrix}'$.

Consider the sample counterpart of $\Delta_{i,n_i}\boldsymbol{W}_n$

$$\tilde{\Delta}_{i,n_i}\boldsymbol{W}_n \frac{1}{N}\sum_i \tilde{\Delta}_{i,n_i}\boldsymbol{w}_i,$$

where $\tilde{\Delta}_{i,n_i}\boldsymbol{w}_i = \begin{pmatrix} \tilde{\Delta}_{i,n_i}y_i \\ \tilde{\Delta}_{i,n_i}\boldsymbol{x}_i \\ \tilde{\Delta}_{i,n_i}\boldsymbol{z}_i \end{pmatrix} \begin{pmatrix} \tilde{\Delta}_{i,n_i}y_i \\ \tilde{\Delta}_{i,n_i}\boldsymbol{x}_i \\ \tilde{\Delta}_{i,n_i}\boldsymbol{z}_i \end{pmatrix}'$.

Let $\Delta_{i,n_i}W_i^{(k,l)}$ be the (k,l) element of the matrix $\Delta_{i,n_i}\boldsymbol{W}_n$, the same apply for the other matrices. By Assumption 4 for any fixed $0 < \epsilon < 1$, there is n_ϵ such that $n > n_\epsilon$, we have $n\rho_n > -log(\epsilon)$. So, for $n > n_\epsilon$ we have

$$Pr(N = 0) = \left(1 - \frac{n\rho_n}{n}\right)^n < \left(1 + \frac{log(\epsilon)}{n}\right)^n < e^{log(\epsilon)} = \epsilon.$$

This implies that $Pr(N = 0) \to 0$ and

$$E\left((\tilde{\Delta}_{i,n_i}W_i^{(k,l)} - \Delta_{i,n_i}\Omega_i^{(k,l)})^2|N = 0\right)Pr(N = 0) = (\Delta_{i,n_i}\Omega_i^{(k,l)})^2 Pr(N = 0) \to 0$$

by Assumption 5 *a)*, Holder's inequality, and assuming that the elements of $\tilde{\Delta}_{i,n_i}\boldsymbol{W}_n = \tilde{\Delta}_{i,n_i}\boldsymbol{\Omega}_n = \boldsymbol{0}$ when $N = 0$. For any integer $1 \leq n_1 \leq n$, observe that

$$E\left(\frac{n}{N}\tilde{\Delta}_{i,n_i}w_i^{(k,l)} - E(\Delta_{i,n_i}w_i^{(k,l)})\middle| N = n_1\right) = \frac{n}{n_1}E(R_i w_i - \frac{1}{N_i}\sum_j R_j w_j | N = n_1) - E(\Delta_{i,n_i}w_i^{(k,l)}) =$$

$$\frac{n}{n_1}\left(\frac{n_1}{n}E(w_i) - \frac{1}{N_i}\frac{n_1}{n}\sum_j E(w_j)\right) - \left(E(w_i) - \frac{1}{n_i}\sum_j E(w_j)\right) \to \boldsymbol{0}, \tag{A.4}$$

given that $E(R_i|N = n_1) = E(R_j|N = n_1) = \frac{n_1}{n}$, and $N_i \to n_i$ by Assumption 1.

Now, to show the convergence in probability of the terms of interest we compute the convergence in (conditional) quadratic mean. We start with the following term

$$E\left(\left(\frac{1}{n}\sum_i \frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)} - E(\Delta_{i,n_i} w_i^{(k,l)})\right)^2 \Big| N = n_1\right) = \tag{A.5}$$

$$= \frac{1}{n^2}\left(\sum_i E\left(\frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)} - E(\Delta_{i,n_i} w_i^{(k,l)})\right)^2 \Big| N = n_1\right) +$$

$$\frac{1}{n^2}\left(\sum_i \sum_{j\neq i} E\left(\left(\frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)} - E(\Delta_{i,n_i} w_i^{(k,l)})\right)\left(\frac{n}{N}\tilde{\Delta}_{j,n_j} w_j^{(k,l)} - E(\Delta_{j,n_j} w_j^{(k,l)})\right) \Big| N = n_1\right).$$

Let us focus on the first term

$$\frac{1}{n^2}\left(\sum_i E\left(\frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)} - E(\Delta_{i,n_i} w_i^{(k,l)})\right)^2 \Big| N = n_1\right) \leq \frac{1}{n^2}\left(\sum_i E\left(\frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)}\right)^2 \Big| N = n_1\right).$$

Then,

$$\frac{1}{n^2}\left(\sum_i E\left(\frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)}\right)^2 \Big| N = n_1\right) = \tag{A.6}$$

$$= \frac{1}{n^2}\left(\sum_i \frac{n^2}{n_1^2} E(R_i w_i^{(k,l)} - \frac{1}{N_i}\sum_j R_j w_j^{(k,l)} | N = n_1)^2\right) =$$

$$= \frac{1}{n^2}\left(\sum_i \frac{n^2}{n_1^2}\left(E(R_i^2 | N = n_1) E(w_i^{(k,l)2})\left(1 - \frac{2}{N_i} + \frac{1}{N_i^2}\right)\right.\right.$$

$$- 2E(R_i | N = n_1) E(w_i^{(k,l)}) \frac{1}{N_i}\sum_{j\neq i} E(R_j | N = n_1) E(w_j^{(k,l)})$$

$$\left.\left. + \frac{1}{N_i^2}\sum_{j\neq i} E(R_j | N = n_1) E(w_j^{(k,l)2})\right)\right) =$$

$$= \frac{1}{n}\left(\sum_i \frac{1}{n_1}\left(E(w_i^{(k,l)2})\left(1 - \frac{2}{N_i} + \frac{1}{N_i^2}\right) - 2\frac{n_1}{n} E(w_i^{(k,l)}) \frac{1}{N_i}\sum_{j\neq i} E(w_j^{(k,l)}) + \frac{1}{N_i^2}\sum_{j\neq i} E(w_j^{(k,l)2})\right)\right),$$

where we used that $E(R_i|N = n_1) = E(R_j|N = n_1) = E(R_i^2|N = n_1) = \frac{n_1}{n}$. by Assumption 5 *a)* for large n, the first term

$$\frac{1}{n_1}\frac{1}{n}\sum_i E(w_i^{(k,l)2})\left(1 - \frac{2}{N_i} + \frac{1}{N_i^2}\right) \leq \frac{C}{n_1},$$

the second

$$-\frac{2}{n^2}\sum_i \left(E(w_i^{(k,l)}) \frac{1}{N_i}\sum_{j\neq i} E(w_j^{(k,l)})\right) \leq O(n^{-1}) = o(1), \;\; and$$

the third

$$= \frac{1}{n_1}\frac{1}{n}\sum_i \left(\frac{1}{N_i^2}\sum_{j\neq i} E(w_j^{(k,l)2})\right) \leq \frac{C}{n_1}.$$

Let

$$\tilde{\Xi}_{n_1} = \begin{cases} \left(\frac{1}{n}\sum_i \frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)} - E(\Delta_{i,n_i} w_i^{(k,l)})\right)^2 & if \;\; n_1 > 0 \\ 0 & if \;\; n_1 = 0 \end{cases}$$

and

$$\Xi_{n_1} = \begin{cases} 2C/n_1 & if \;\; n_1 > 0 \\ 0 & if \;\; n_1 = 0 \end{cases}$$

$$E\left(\left(\frac{1}{n}\sum_i \frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)} - E(\Delta_{i,n_i} w_i^{(k,l)})\right)^2 \middle| N > 0\right) Pr(N > 0) = E(\tilde{\Xi}_N) \leq E(\Xi_N).$$

Next, we want to study the behavior of Ξ_N when $n \to \infty$. Observe that, for any $\epsilon > 0$, $Pr(\Xi_N > \epsilon) \leq Pr(0 < N < 2C/\epsilon) < Pr(N < 2C/\epsilon)$ because Ξ_N can be also zero and the last interval includes the previous ones. Now for any $\epsilon > 0$ if we apply the Chernoff's bounds to a sum of bernoully rvs we have:

$$Pr(N < 2C/\epsilon) = Pr(N < (1-\delta)n\rho_n) \leq exp(-\delta^2 n\rho_n/2) \to 0,$$

where $\delta = \frac{(n\rho_n - 2C/\epsilon)}{n\rho_n}$. Thus, $\Xi_N = o_p(1)$ and given that it is bounded also $E(\Xi_N) = o(1)$. For the second term in (A.5), let us take the first product

$$= \frac{1}{n^2}\left(\sum_i\sum_{j\neq i} E\left(\left(\frac{n}{N}R_i w_i^{(k,l)} - \frac{n}{N}\frac{1}{N_i}\sum_j R_j w_j^{(k,l)}\right)\left(\frac{n}{N}R_f w_f^{(k,l)} - \frac{n}{N}\frac{1}{N_f}\sum_g R_g w_g^{(k,l)}\right)\middle| N = n_1\right) = \tag{A.7}$$

$$= \frac{1}{n^2}\left(\sum_i\sum_{j\neq i} E(w_i^{(k,l)} w_f^{(k,l)}) - E(w_i^{(k,l)}\frac{1}{n_f}\sum_g w_g^{(k,l)}) - \right.$$
$$\left. - E(w_f^{(k,l)}\frac{1}{n_i}\sum_j w_j^{(k,l)}) + E(\frac{1}{n_f n_i}\sum_g w_g^{(k,l)}\sum_j w_j^{(k,l)})\right) \leq C/n^2 \to 0,$$

by Assumption 5 *a)*. Similar to Lemma 1 a), this holds for all the terms of the products in (A.5). Therefore,

$$E\left(\left(\frac{1}{n}\sum_i \frac{n}{N}\tilde{\Delta}_{i,n_i} w_i^{(k,l)} - E(\Delta_{i,n_i} w_i^{(k,l)})\right)^2\right) = o(1).$$

■

Lemma 3. *Under Assumptions 1- 6, as* $n \to \infty$, *if* $\sum_{i<j, j\in B_i\setminus i} E(\boldsymbol{x}_i \boldsymbol{x}_j') < C$, *where C is a positive constant, then*

$$\boldsymbol{\beta}^c_{ND} = \boldsymbol{\beta}^c_{NW} = \sum_i E(\boldsymbol{x}_i\boldsymbol{x}_i)^{-1} \sum_i E(\boldsymbol{x}_i\boldsymbol{x}_i')\boldsymbol{\beta}_i + o(1).$$

Proof of Lemma 3.

As $n \to \infty$, $\sum_{i<j, j \in B_i \setminus i} \Delta_{ij}\boldsymbol{z}_i \Delta_{ij}\boldsymbol{z}_i'$ is full rank and $\boldsymbol{\lambda}$, exists, so that $\Delta_{ij}\boldsymbol{\Omega}^{xz}_n = \boldsymbol{0}$, and

$$\boldsymbol{\beta}^c_{ND} = \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i(\Delta_{ij}\boldsymbol{x}_i'))^{-1} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i(\Delta_{ij}\boldsymbol{y}_i)).$$

Moreover, as n get large, by Assumption 6 $\boldsymbol{\lambda} = \boldsymbol{\phi}_n$, which implies $E(\Delta_{ij}\boldsymbol{x}_i) = \boldsymbol{0}$.
Then, given (1) and Assumption 2, we have that

$$E(\Delta_{ij} y_i) = E(\Delta_{ij}\boldsymbol{t}_i\boldsymbol{\beta}_i) + \Delta_{ij}\epsilon_i + E(\Delta_{ij}\theta_i).$$

Thus,

$$E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}y_i) = E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\boldsymbol{t}_i\boldsymbol{\beta}_i) + E(\Delta_{ij}\boldsymbol{x}_i)\Delta_{ij}\epsilon_i + E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\theta_i) = E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\boldsymbol{x}_i\boldsymbol{\beta}_i) + E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\theta_i).$$

Then,

$$\boldsymbol{\beta}^c_{ND} = \left(\frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E\Delta_{ij}\boldsymbol{x}_i(\Delta_{ij}\boldsymbol{x}_i')\right)^{-1} \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\boldsymbol{x}_i\boldsymbol{\beta}_i) + \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\theta_i).$$

W.l.o.g. let us focus on the component-wise convergence of $\frac{1}{n_p}\sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\theta_i)$. Therefore,

$$\frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}x_i\Delta_{ij}\theta_i) \leq \max_{ij} E(\Delta_{ij}x_i\Delta_{ij}\theta_i) \leq \max_{ij} E|\Delta_{ij}x_i\Delta_{ij}\theta_i| \leq$$

$$\max_{ij} E(|\Delta_{ij}x_i|) \max_{ij} E(|\Delta_{ij}\theta_i|) = O(1) \max_{ij} E|(\Delta_{ij}\theta_i)| = O(1)o(1) = o(1),$$

by Assumptions 2 and 5 *a)* and

$$\boldsymbol{\beta}^c_{ND} = \left(\frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E\Delta_{ij}\boldsymbol{x}_i(\Delta_{ij}\boldsymbol{x}_i')\right)^{-1} \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i(\boldsymbol{x}_i\boldsymbol{\beta}_i - \boldsymbol{x}_j\boldsymbol{\beta}_j)) + o(1).$$

Now, let us focus on $\frac{1}{n_p}\sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i(\boldsymbol{x}_i\boldsymbol{\beta}_i - \boldsymbol{x}_j\boldsymbol{\beta}_j))$. We have,

$$\frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\Delta_{ij}\boldsymbol{x}_i(\boldsymbol{x}_i\boldsymbol{\beta}_i - \boldsymbol{x}_j\boldsymbol{\beta}_j)) =$$

$$= \frac{1}{n_p} \sum_i \left((n_i - 1)E(\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\beta}_i) + \sum_{j \in B_i} E(\boldsymbol{x}_j\boldsymbol{x}_j'\boldsymbol{\beta}_j)\right) - \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} \left(E(\boldsymbol{x}_i\boldsymbol{x}_j'\boldsymbol{\beta}_j) - E(\boldsymbol{x}_j\boldsymbol{x}_i'\boldsymbol{\beta}_i)\right).$$

Given Assumption 5, the cross moments terms are bounded. Also $\frac{1}{n_p}\sum_{j \in B_i} E(\boldsymbol{x}_j\boldsymbol{x}_j'\boldsymbol{\beta}_j) \to 0$ given that

$n_i < \infty$, by Assumption 1. Finally,

$$\frac{1}{\sum_i (n_i - 1)} \sum_i (n_i - 1) E(\boldsymbol{x}_i \boldsymbol{x}_i') \boldsymbol{\beta}_i - \frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} E(\boldsymbol{x}_i \boldsymbol{x}_j')(\boldsymbol{\beta}_i + \boldsymbol{\beta}_j) \rightarrow$$

$$\rightarrow \frac{1}{n} \sum_i E(\boldsymbol{x}_i \boldsymbol{x}_i') \beta_i + o(1).$$

We obtain our result by employing the same strategy to the denominator of $\boldsymbol{\beta}^c_{ND}$. The same arguments can be applied to $\boldsymbol{\beta}^c_{NW}$.

■

Lemma 4. *Let* $\{V_{ni}, 0 \leq i \leq D_n, n \geq 1\}$ *be a real valued random field that is* L_2-*NED on* $\{\theta_{i,n}, i \in D_n, n \geq 1\}$, *with the scaling factors* ν_{in} *and the NED coefficients* $\psi(s)$ *of size* $-\nu$, $\mu_{nij} = E(\Delta_{ij} V_{ni})$, *and* $\sigma_n^2 = Var(\sum_{i<j \in B_i \setminus i} \frac{R_{ij}}{\sqrt{\rho_n^2}} (\Delta_{ij} V_{ni} - \mu_{ij}))$. *Suppose that* $R_{ij}, \ldots R_{n_p}$ *are independent of* $\Delta_{ij} V_{ni}$, *Assumption 1, 4, and 10 hold,* $\sup_{n, i \in D_n} E|(V_{ni} - \mu_i)|^{2+\delta} < \infty$, *and* $\sup_{n, i \in D_n} \nu_{in} < \infty$ *for some* $\delta > 0$, $\inf_n n_p^{-1} \sigma_n^2 > 0$,

$$\sum_{i<j, j \in B_i \setminus i} \mu_{nij} = 0,$$

$$\frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} Var(\Delta_{ij} V_{ni}) \rightarrow \sigma^2,$$

$$\frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} \mu_{ij}^2 \rightarrow \kappa^2, \ \textit{and}$$

$$\frac{1}{n_p} \sum_{i<j, j \in B_i \setminus i} \sum_{k \neq i < l, l \in B_k \setminus k} cov(V_{ij}, V_{kl}) \rightarrow \sigma_{ijkl}.$$

Furthermore, there exists a constan C *such that,*

$$\sum_{i<j \in B_i \setminus i} \sum_{k \neq i < l \in B_k \setminus k} E\bigg(E\bigg(\Delta_{ij} V_i - \mu_{ij} \bigg| \mathcal{F}_{in}(s)\bigg) E\bigg(\Delta_{kl} V_k - \mu_{kl} \bigg| \mathcal{F}_{kn}(s)\bigg) \geq C$$

where $\sigma^2 + \rho^2 \sigma_{ijkl} + (1 - \rho^2)\kappa^2 > 0$ *and* n_p *denotes the number of pairs in the population. Then,*

$$\frac{1}{\sqrt{N_p}} \sum_{i<j, j \in B_i \setminus i} \bigg(R_{nij} \Delta_{ij} V_i\bigg) \xrightarrow{d} \mathcal{N}(0, \sigma^2 + 2\sigma_{ijkl} + (1 - \rho^2)\kappa^2),$$

where N_p *is the number of sampled pairs.*

Proof. To easy the notation let us drop the subscript n from the array. As in Abadie et al. (2020) we have that given d, $N_p \sim Bin(n_p, \rho_n^2)$ with $E(N_p) = n_p \rho_n^2$.

$$Var\bigg(\frac{N_p}{\rho_n^2 n_p}\bigg) = \frac{\rho_n^2 (1 - \rho_n^2) n_p}{n_p^2 \rho_n^4} \rightarrow 0.$$

Then by the continuous mapping theorem, we have

$$\left(\frac{n_p\rho_n^2}{N_p}\right)^{1/2} \xrightarrow{p} 1.$$

As a consequence, it suffices to prove that

$$\frac{1}{\sqrt{n_p}} \sum_{i<j, j\in B_i\setminus i} \left(\frac{R_{nij}\Delta_{ij}V_i}{\sqrt{\rho_n^2}}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2 + \rho^2\sigma_{ijkl} + (1-\rho_n^2)\kappa^2).$$

Now define

$$s_n^2 = \frac{1}{n_p} Var\left(\sum_{i<j, j\in B_i\setminus i} (R_{ij}\Delta_{ij}V_i - \rho_n^2\mu_{ij})\right).$$

Observe that,

$$E\left(\frac{R_{ij}\Delta_{ij}V_i - \rho_n^2\mu_{ij}}{s_n\sqrt{n_p\rho_n^2}}\right) = 0,$$

for n large enough and $s_n^2 > 0$. Let us now focus on the variance,

$$Var(R_{ij}\Delta_{ij}V_i - \rho_n^2\mu_{ij}).$$

Focusing on the variance of the first demeaned term we have

$$Var(R_{ij}\Delta_{ij}V_i - \rho_n^2\mu_{ij}) = Var(R_{ij}\Delta_{ij}V_i) = \rho_n^2 E(\Delta_{ij}V_i^2) - \rho_n\mu_{ij}^2 = \rho_n^2(Var(\Delta_{ij}V_i) + (1-\rho_n^2)\mu_{ij}^2).$$

We need now to compute the $Var\left(\sum_{i<j, j\in B_i\setminus i}(R_{ij}\Delta_{ij}V_i - (\rho_n^2\mu_{ij}))\right)$. This is equal to the sum of the variances plus an extra covariance term that is equal to the covariances of the pairs which have one unit in common. Formally,

$$\begin{aligned}
&Var\left(\sum_{i<j, j\in B_i\setminus i}(R_{ij}\Delta_{ij}V_i - (\rho_n^2\mu_{ij}))\right) = \\
&= \sum_{i<j, j\in B_i\setminus i}\left(\rho_n^2(Var(\Delta_{ij}V_i) + (1-\rho_n^2)\mu_{ij}^2)\right) \\
&+ \rho_n^4 \sum_{i<j, j\in B_i\setminus i} \sum_{(k\neq i)<l\in B_k^d\setminus k} cov(\Delta_{ij}V_i, \Delta_{kl}V_k).
\end{aligned}$$

Let us now define $q_{ni} = \Delta_{ij}V_i - \mu_{ij}$. From the NED definition,

$$\begin{aligned}
&\|q_{in} - E(q_{in}|\mathcal{F}_{in}(s))\|_p \leq \\
&\|(V_i - \mu_i) - (V_j - \mu_j)\|_p + \|E((V_i - \mu_i)|\mathcal{F}_{in}(s)) - E((V_j - \mu_j)|\mathcal{F}_{jn}(s))\|_p \leq \\
&\nu_{in}\psi(s) + \nu_{jn}\psi(s) \leq \nu_{ijn}2\psi(s),
\end{aligned}$$

where $\nu_{ijn} = \max\{\nu_i, \nu_j\}$. Thus, q_{in} is NED on θ_i of the same size, see also Theorem 17.8 pag. 267 in Davidson (1994). Given that θ_i and q_{ni} have uniformly bounded moments, we can set $\nu_{in} = c_{in} = 1$ (Jenish and Prucha, 2012). First, observe that given n, R_{ij} is i.i.d. and so is also m-dependent (See Defintion 3 and Lemma B.4 in Xu and Wooldridge, 2022). We check conditions *a)-d)* in Theorem 1. Let us start with *a)*

$$\sup_{n,i\in D_n} E(|q_{ni}/c_{in}|)^{2+\delta} < \infty$$

for some $\delta > 0$, where $c_{in} > 0$ is a sequence of constants. Setting, $c_{in} = 1$, allows us to satisfy the requirement by Assumption. For condition *b)*, we need that for any fixed $s > 0$, there exist a positive constatnt C such that for any n and every nonempty set $K \subseteq D_n$, $E(\sum_{i\in K} \frac{R_i}{\sqrt{\rho_n}} E(q_{ni}/j_n|\mathcal{F}(s)))^2 \geq C\sum_{i\in K} E(\frac{R_i}{\sqrt{\rho_n}} E(q_{ni}/j_n|\mathcal{F}(s)))^2$, where $j_n = \max_{i\in D_n}\{c_{in}, \nu_{in}\}$. Thus, given $j_n = 1$,

$$E\bigg(\sum_{i<j\in B_i\setminus i} \frac{R_{ij}}{\sqrt{\rho_n^2}} E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)\bigg)^2 \geq CE\sum_{i<j\in B_i\setminus i}\bigg(\frac{R_{ij}}{\sqrt{\rho_n^2}} E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)\bigg)^2.$$

Focusing on the LHS of the inequality, we have

$$\begin{aligned}
&E\bigg(\sum_{i<j\in B_i\setminus i} \frac{R_{ij}}{\sqrt{\rho_n^2}} E\Big(\frac{\Delta_{ij}V_i - \mu_{ij}}{j_n}\Big|\mathcal{F}_{in}(s)\Big)\bigg)^2 = \\
&\frac{E(R_{ij}^2)}{\rho_n^2}\sum_{i<j\in B_i\setminus i} E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)^2 + \\
&+\sum_{i<j\in B_i\setminus i}\ \sum_{k\neq i<l\in B_k\setminus k} E\bigg(E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)E\Big(\Delta_{kl}V_k - \mu_{kl}\Big|\mathcal{F}_{kn}(s)\Big)\bigg) = \\
&\sum_{i<j\in B_i\setminus i} E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)^2 + \sum_{i<j\in B_i\setminus i}\ \sum_{k\neq i<l\in B_k\setminus k} E\bigg(E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)E\Big(\Delta_{kl}V_k - \mu_{kl}\Big|\mathcal{F}_{kn}(s)\Big)\bigg).
\end{aligned} \tag{A.8}$$

The RHS can be written as

$$\begin{aligned}
&CE\sum_{i<j\in B_i\setminus i}\bigg(\frac{R_{ij}}{\sqrt{\rho_n^2}} E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)\bigg)^2 = \\
&\frac{E(R_{ij}^2)}{\rho_n^2}\sum_{i<j\in B_i\setminus i} E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)^2 = \\
&\sum_{i<j\in B_i\setminus i} E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)^2.
\end{aligned} \tag{A.9}$$

Thus, using (A.8)-(A.9), we have

$$\sum_{i<j\in B_i\setminus i}\ \sum_{k\neq i<l\in B_k\setminus k} E\bigg(E\Big(\Delta_{ij}V_i - \mu_{ij}\Big|\mathcal{F}_{in}(s)\Big)E\Big(\Delta_{kl}V_k - \mu_{kl}\Big|\mathcal{F}_{kn}(s)\Big)\bigg) \geq C.$$

This is sufficient for condition *b)* to hold within our framework. While the NED coefficients $\psi(s)$ are assumed to have a magnitude of size $-\boldsymbol{\nu}$ (d)), condition *e)* (stating that the NED scaling factors must satisfy $\sup_{n,i\in D_n} c_{in}^{-1}\nu_{in} \leq C < \infty$) is trivially satisfied.

Therefore,

$$\sum_{i<j,j\in B_i\backslash i}\left(\frac{R_{ij}\Delta_{ij}V_i - \mu_{ij}\rho_n^2}{(s_n\sqrt{\rho_n^2 n_p})}\right)\xrightarrow{d}\mathcal{N}(0,1),$$

given that $\frac{s_n}{\sqrt{\sigma^2+\rho^2\sigma_{ijkl}+(1-\rho^2)\kappa^2)}} \to 1.$

■

Lemma 5. *Under Assumptions, 1-8, and Assumption 12 a) in Appendix A.2. Let* $\boldsymbol{B}^u = \boldsymbol{B}^{ehw} - \boldsymbol{B}^{cond}$, *and* $\tilde{\Delta}_{ij}\tilde{\varepsilon}_i = \tilde{\Delta}_{ij}y_i - \tilde{\Delta}_{ij}\boldsymbol{x}_i\boldsymbol{\beta}^{cs} - \tilde{\Delta}_{ij}\boldsymbol{z}_i\boldsymbol{\gamma}^{cs}$*. Then,*

$$\sum_{i<j,j\in B_i\backslash i}\frac{1}{\sqrt{N_p}}\tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{ij}\varepsilon_i\xrightarrow{d}\mathcal{N}(\boldsymbol{0},\boldsymbol{B}^{cond}+\rho^2\boldsymbol{B}^{cov}+(1-\rho^2)\boldsymbol{B}^u)$$

Proof. Let us start with a). We first study the convergence of this term

$$\frac{1}{\sqrt{n_p}}\sum_{i<j,j\in B_i\backslash i}E\bigg(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\theta_i\bigg).$$

Without loss of generality, let us focus on the component wise convergence. We have that

$$\frac{1}{\sqrt{n_p}}\sum_{i<j,j\in B_i\backslash i}E(\Delta_{ij}x_i\Delta_{ij}\theta_i)\leq\sqrt{n_p}\max_{ij}E(\Delta_{ij}x_i\Delta_{ij}\theta_i)\leq\sqrt{n_p}\max_{ij}E|\Delta_{ij}x_i\Delta_{ij}\theta_i|\leq$$

$$\sqrt{n_p}\max_{ij}E(|\Delta_{ij}x_i|)\max_{ij}E(|\Delta_{ij}\theta_i|)=O(1)\sqrt{n_p}\max_{ij}E|(\Delta_{ij}\theta_i)|=O(1)o(1)=o(1),$$

where the second inequality follows by triangular inequality, the third inequality by Cauchy-Schwartz inequality, the first equality by Assumption 5 a) and Holder's inequality, and the second equality by Assumption 7. Consider $\Delta_{ij}V_{ni} = a'\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\varepsilon_i$. Let us verify the conditions of Lemma 4. We start with $\frac{1}{n}\sum_i E(|V_{ni}|^{2+\delta})$ is bounded by a positive constant for some $\delta > 0$.

$$\frac{1}{n_p}\sum_{i<j,j\in B_i\backslash i}E\left(\left|\left|\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\varepsilon_i\right|\right|^{2+\delta}\right)\leq\frac{||a||^{2+\delta}}{n_p}\sum_{i<j,j\in B_i\backslash i}E\bigg(||\boldsymbol{x}_i||^{2+\delta}(|y_i|+||\boldsymbol{x}_i||||\boldsymbol{\beta}||+||\boldsymbol{z}_i||||\boldsymbol{\gamma}||)^{2+\delta}\bigg)+$$
$$E\bigg(||\boldsymbol{x}_j||^{2+\delta}(|y_j|+||\boldsymbol{x}_j||||\boldsymbol{\beta}||+||\boldsymbol{z}_j||||\boldsymbol{\gamma}||)^{2+\delta}\bigg)\leq C,$$

by Minkowski's inequality and Assumption 5 a). Furthermore,

$$\sum_{i<j,j\in B_i\backslash i}\mu_{nij}=a'\sum_{i<j,j\in B_i\backslash i}E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\varepsilon_i)=0,$$

by defining the coefficient β^c_{ND} as orthogonality condition.

Let $a \neq 0$. We have that,

$$\frac{1}{n_p} Var\Big(\sum_{i<j, j\in B_i\setminus i} \Delta_{ij} V_i \Big) = \frac{1}{n_p} a' Var\Big(\sum_{i<j, j\in B_i\setminus i} \Delta_{ij} \boldsymbol{x}_i \Delta_{ij}\varepsilon_i \Big) a \to a'(\boldsymbol{B}^{cond} + \boldsymbol{B}^{cov})a,$$

$$\frac{1}{n_p} \sum_{i<j, j\in B_i\setminus i} \mu_{ij}^2 = \frac{1}{n_p} a' \Big(\sum_{i<j, j\in B_i\setminus i} E\Big(\Delta_{ij} \boldsymbol{x}_i \Delta_{ij}\varepsilon_i\Big) E\Big(\Delta_{ij} \boldsymbol{x}_i \Delta_{ij}\varepsilon_i\Big)' \Big) a \to a'(\boldsymbol{B}^{u})a.$$

This implies

$$a' \Big(\sum_{i<j, j\in B_i\setminus i} \frac{1}{\sqrt{N_p}} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij}\varepsilon_i \Big) \xrightarrow{d} \mathcal{N}(\mathbf{0}, a'(\boldsymbol{B}^{cond} + \rho^2 \boldsymbol{B}^{cov} + (1-\rho^2)\boldsymbol{B}^{u})a).$$

Using the Cramer-Wold device, the result follows

$$\sum_{i<j, j\in B_i\setminus i} \frac{1}{\sqrt{N_p}} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij}\varepsilon_i \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{B}^{cond} + \rho^2 \boldsymbol{B}^{cov} + (1-\rho^2)\boldsymbol{B}^{u}).$$

The proof of *b)* follows the same arguments.

■

Proof of Proposition 1 .

To prove a),

We can write the ND estimator as

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{ND} \\ \hat{\boldsymbol{\gamma}}_{ND} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^c_{ND} \\ \boldsymbol{\gamma}^c_{ND} \end{pmatrix} + \Big(\sum_{i<j, j\in B_i\setminus i} \begin{pmatrix} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} \boldsymbol{x}_i' & \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} \boldsymbol{z}_i' \\ \tilde{\Delta}_{ij} \mathbf{z}_i \tilde{\Delta}_{ij} \mathbf{x}_i' & \tilde{\Delta}_{ij} \mathbf{z}_i \tilde{\Delta}_{ij} \mathbf{z}_i' \end{pmatrix} \Big)^{-1} \sum_{i<j, j\in B_i\setminus i} \begin{pmatrix} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} (\varepsilon_i + \theta_i) \\ \tilde{\Delta}_{ij} \boldsymbol{z}_i \tilde{\Delta}_{ij} (\varepsilon_i + \theta_i). \end{pmatrix}$$

Thus as $n \to \infty$ and by Lemma 5 with Assumption 7 ,

$$\sqrt{N_p} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{ND} - \boldsymbol{\beta}^c_{ND} \\ \hat{\boldsymbol{\gamma}}_{ND} - \boldsymbol{\gamma}^c_{ND} \end{pmatrix} = \Big(\frac{1}{N_p} \sum_{i<j, j\in B_i\setminus i} \begin{pmatrix} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} \boldsymbol{x}_i' & \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} \boldsymbol{z}_i' \\ \tilde{\Delta}_{ij} \mathbf{z}_i \tilde{\Delta}_{ij} \mathbf{x}_i' & \tilde{\Delta}_{ij} \mathbf{z}_i \tilde{\Delta}_{ij} \mathbf{z}_i' \end{pmatrix} \Big)^{-1} \frac{1}{\sqrt{N_p}} \sum_{i<j, j\in B_i\setminus i} \begin{pmatrix} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} \varepsilon_i \\ \tilde{\Delta}_{ij} \boldsymbol{z}_i \tilde{\Delta}_{ij} \varepsilon_i . \end{pmatrix} + o_p(1)$$

$$= \begin{pmatrix} \Delta_{ij} \boldsymbol{\Omega}_n^{xx} & \Delta_{ij} \boldsymbol{\Omega}_n^{xz} \\ \Delta_{ij} \boldsymbol{\Omega}_n^{zx} & \Delta_{ij} \boldsymbol{\Omega}_n^{zz} \end{pmatrix}^{-1} \frac{1}{\sqrt{N_p}} \sum_{i<j, j\in B_i\setminus i} \begin{pmatrix} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} \varepsilon_i \\ \tilde{\Delta}_{ij} \boldsymbol{z}_i \tilde{\Delta}_{ij} \varepsilon_i \end{pmatrix} + r_n,$$

where

$$r_n = \left[\begin{pmatrix} \tilde{\Delta}_{ij} \boldsymbol{W}_n^{xx} & \tilde{\Delta}_{ij} \boldsymbol{W}_n^{xz} \\ \tilde{\Delta}_{ij} \boldsymbol{W}_n^{zx} & \tilde{\Delta}_{ij} \boldsymbol{W}_n^{zz} \end{pmatrix}^{-1} - \begin{pmatrix} \Delta_{ij} \boldsymbol{\Omega}_n^{xx} & \Delta_{ij} \boldsymbol{\Omega}_n^{xz} \\ \Delta_{ij} \boldsymbol{\Omega}_n^{zx} & \Delta_{ij} \boldsymbol{\Omega}_n^{zz} \end{pmatrix}^{-1} \right] \frac{1}{\sqrt{N_p}} \sum_{i<j, j\in B_i\setminus i} \begin{pmatrix} \tilde{\Delta}_{ij} \boldsymbol{x}_i \tilde{\Delta}_{ij} \varepsilon_i \\ \tilde{\Delta}_{ij} \boldsymbol{z}_i \tilde{\Delta}_{ij} \varepsilon_i \end{pmatrix} + o_p(1).$$

Let us focus on

$$\frac{1}{\sqrt{N_p}}\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{z}_i\tilde{\Delta}_{ij}\varepsilon_i.$$

Rewrite the term as

$$\sqrt{\frac{n_p\rho_n^2}{N_p}}n_p^{-1/2}\sum_{i<j,j\in B_i\setminus i}\frac{1}{\sqrt{\rho_n^2}}\tilde{\Delta}_{ij}\boldsymbol{z}_i\tilde{\Delta}_{ij}\varepsilon_i.$$

As shown in Lemma 4, $\sqrt{\frac{n_p\rho_n^2}{N_p}}\xrightarrow{p}1$ and

$$E\bigg(\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{z}_i\tilde{\Delta}_{ij}\varepsilon_i\bigg)=\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{z}_iE(\tilde{\Delta}_{ij}\varepsilon_i)=\mathbf{0}.$$

Now if we show that the element-wise variance is bounded, then this implies that

$$\frac{1}{\sqrt{N_p}}\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{z}_i\tilde{\Delta}_{ij}\varepsilon_i=O_p(1),$$

by Chebyshev's inequality. Thus,

$$\begin{aligned}&Var\bigg(n_p^{-1/2}\sum_{i<j,j\in B_i\setminus i}\frac{1}{\sqrt{\rho_n^2}}\tilde{\Delta}_{ij}z_i\tilde{\Delta}_{ij}\varepsilon_i\bigg)=\\&n_p^{-1}\sum_{i<j,j\in B_i\setminus i}\big(Var(\Delta_{ij}z_i\Delta_{ij}\varepsilon_i)+(1-\rho_n^2)(E(\Delta_{ij}z_i\Delta_{ij}\varepsilon_i))^2\big)\\&+\rho_n^2\sum_{i<j,j\in B_i\setminus i}\sum_{(k\neq i)<l\in B_k^d\setminus k}cov(\Delta_{ij}z_i\Delta_{ij}\varepsilon_i,\Delta_{kl}z_k\Delta_{kl}\varepsilon_k).\end{aligned}$$

Let us focus on the three terms separately,

$$n_p^{-1}\sum_{i<j,j\in B_i\setminus i}(Var(\Delta_{ij}z_i\Delta_{ij}\varepsilon_i)=n_p^{-1}\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}z_i^2Var(\Delta_{ij}\varepsilon_i)\leq n_p^{-1}\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}z_i^2E(\Delta_{ij}\varepsilon_i^2),$$

$$n_p^{-1}(1-\rho_n^2)(E(\Delta_{ij}z_i\Delta_{ij}\varepsilon_i))^2=(1-\rho_n)n_p^{-1}\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}z_i^2(E(\Delta_{ij}\varepsilon_i))^2,\ and$$

$$\begin{aligned}&n^{-1}\rho_n^2\sum_{i<j,j\in B_i\setminus i}\sum_{(k\neq i)<l\in B_k^d\setminus k}cov(\Delta_{ij}z_i\Delta_{ij}\varepsilon_i,\Delta_{kl}z_k\Delta_{kl}\varepsilon_k)\\&\leq n^{-1}\sum_{i<j,j\in B_i\setminus i}\sum_{(k\neq i)<l\in B_k^d\setminus k}\Delta_{ij}z_i\Delta_{kl}z_k(E(\Delta_{ij}\varepsilon_iz_k\Delta_{kl}\varepsilon_k)),\end{aligned}$$

which are bounded by Assumption 5.
Therefore given that $\Delta_{ij}\boldsymbol{\Omega}_n^{xz}=\mathbf{0}$ and $\frac{1}{\sqrt{N_p}}\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{z}_i\tilde{\Delta}_{ij}\varepsilon_i=O_p(1)$, we have

$$\sqrt{N_p}(\hat{\boldsymbol{\beta}}_{ND} - \boldsymbol{\beta}_{ND}^c) = (\Delta_{ij}\boldsymbol{\Omega}_n^{xx})^{-1} + \frac{1}{\sqrt{N_p}} \sum_{i<j, j\in B_i\setminus i} \tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{ij}\varepsilon_i + o_p(1).$$

The result follows by Lemma 5 and Slutsky's theorem.

■

Lemma 6. *Let $\{V_{ni}, 0 \le i \le D_n, , n \ge 1\}$ be a real valued random field that is L_2-NED on $\{\theta_{i,n}, i \in D_n, n \ge 1\}$, with the scaling factors ν_{in} and the NED coefficients $\psi(s)$ of size $-\nu$, $\mu_{ni} = E(V_{ni})$, and $\sigma_n^2 = Var(\sum_i \frac{1}{\sqrt{\rho_n}}(R_iV_i - \frac{1}{n_i}\sum_j R_jV_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_j \mu_j))$. Suppose that $R_i, \dots R_n$ are independent of $\Delta_{i,n_i}V_{ni}$, Assumption 1, 4, and 10 hold, $\sup_{n,i\in D_n} E|(V_{ni} - \mu_i)|^{2+\delta} < \infty$, $\sup_{n,i\in D_n} \nu_{in} < \infty$ for some $\delta > 0$, $\inf_n n^{-1}\sigma_n^2 > 0$, and*
$\sum_i \mu_{ni} - \frac{1}{n_i}\sum_j \mu_j = 0$,

$\frac{1}{n}\sum_{i=1}^n Var(V_{ni})\left(1 - \frac{2}{n_i}\right) \to \sigma^2$, $\frac{1}{n}\sum_{i=1}^n \mu_i^2\left(1 - \frac{2}{n_i}\right) \to \kappa^2$,

$\frac{1}{n}\sum_{i=1}^n \frac{1}{n_i^2}\sum_{j\in B_i} Var(V_{nj}) \to \sigma_{neigh}^2$, $\frac{1}{n}\sum_{i=1}^n \frac{1}{n_i^2}\sum_{j\in B_i} \mu_j^2 \to \kappa_{neigh}^2$,

$\frac{1}{n}\sum_i \frac{1}{n_i^2}\sum_{j,k\in B_i} cov(V_jV_k) \to \sigma_{B_i}$, $\frac{1}{n}\sum_i \frac{1}{n_i}\sum_{j\neq i\in B_i} cov(V_i, V_j) \to \sigma_{i,B_i}$,

$\frac{1}{n}\sum_i \sum_{k\neq i} cov\left(\left(V_i - \frac{1}{n_i}\sum_{j\in B_i} V_j\right), \left(V_k - \frac{1}{n_k}\sum_{l\in B_k} V_l\right)\right) \to \sigma_{\Delta B_i, B_k}$,
Furthermore, there exist a constant C such that

$$\sum_i \sum_{k\neq i} E\left(E\left(V_i - \mu_i \middle| \mathcal{F}_{in}(s)\right)E\left(V_k - \mu_k \middle| \mathcal{F}_{kn}(s)\right)\right) \ge C,$$

$$\sum_i \sum_{k\neq i} E\left(E\left(\frac{1}{n_i}\sum_{j\in B_i}(V_j - \mu_j) \middle| \mathcal{F}_{in}(s)\right)E\left(\frac{1}{n_k}\sum_{l\in B_k}(V_l - \mu_l) \middle| \mathcal{F}_{kn}(s)\right)\right) \ge C,$$

where $(\sigma^2 + \sigma_{neigh}^2 + (1-\rho)(\kappa^2 + \kappa_{neigh}^2) + \rho(\sigma_{B_i} - \sigma_{i,B_i} + \sigma_{\Delta B_i, B_k})) > 0$. Then,

$$\frac{1}{\sqrt{N}}\sum_i \left(R_{ni}V_i - \frac{1}{N_i}\sum_j R_{nj}V_j\right) \xrightarrow{d}$$
$$\mathcal{N}\left(0, (\sigma^2 + \sigma_{neigh}^2 + (1-\rho)(\kappa^2 + \kappa_{neigh}^2) + \rho\sigma_{cov})\right).$$

where $\sigma_{cov} = \sigma_{B_i} - \sigma_{i,B_i} + \sigma_{\Delta B_i, B_k}$, and N is the number of sampled units.

Proof. To easy the notation let us drop the subscript n from the array. As in Abadie et al. (2020) we have that $N \sim Bin(n, \rho_n)$.

$$E\left[\frac{N}{n\rho_n}\right] = 1, \;\; Var\left(\frac{N}{n\rho_n}\right) = \frac{n\rho_n(1-\rho_n)}{(n\rho_n)^2} \to 0.$$

Then by the continuous mapping theorem, we have

$$\left(\frac{n\rho_n}{N}\right)^{1/2} \xrightarrow{p} 1.$$

As a consequence, and given that $N_i \to n_i$ by Assumption 1, it suffices to prove

$$\frac{1}{\sqrt{n}}\sum_i \left(\frac{R_iV_i - \frac{1}{n_i}\sum_j R_jV_j}{\sqrt{\rho_n}}\right) \xrightarrow{d}$$
$$\mathcal{N}\Bigg(0, (\sigma^2 + \sigma^2_{neigh} + (1-\rho)(\kappa^2 + \kappa^2_{neigh}) + \rho\sigma_{cov})\Bigg),$$

where $\sigma^2_{cov} = \sigma_{B_i} - \sigma_{i,B_i} + \sigma_{\Delta B_i,B_k}$. Now define

$$s_n^2 = \frac{1}{n}Var\left(\sum_i (R_iV_i - \rho_n\mu_i - \frac{1}{N_i}\sum_j R_jV_j + \rho_n\frac{1}{n_i}\sum_j \mu_j)\right).$$

Observe that,

$$E\Bigg(\frac{R_iV_i - \frac{1}{n_i}\sum_j R_jV_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_j \mu_j)}{s_n\sqrt{n\rho_n}}\Bigg) = 0,$$

for n large enough and $s_n^2 > 0$. Let us now focus on the variance,

$$Var(R_iV_i - \frac{1}{n_i}\sum_j R_jV_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_j \mu_j)).$$

Let us fix i and focus on the variance of $\frac{1}{n_i}\sum_{j\in B_i} R_jV_j$,

$$Var\Bigg(\frac{1}{n_i}\sum_{j\in B_i} R_jV_j - \frac{1}{n_i}\sum_{j\in B_i}\rho_n\mu_j\Bigg) = \rho_n\Bigg(\frac{1}{n_i^2}\sum_{j\in B_i}\Bigg(Var(V_j) + (1-\rho_n)\mu_j^2\Bigg) + \frac{1}{n_i^2}\rho_n\sum_{j\in B_i}\sum_{k\neq j\in B_i} cov(V_jV_k)\Bigg).$$

Thus,

$$Var(R_iV_i - \frac{1}{n_i}\sum_{j\in B_i} R_jV_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_{j\in B_i}\mu_j)) = \tag{A.10}$$

$$= \rho_n(Var(V_i) + (1-\rho_n)\mu_i^2) + \rho_n\frac{1}{n_i^2}\left(\sum_{j\in B_i}\left(Var(V_j) + (1-\rho_n)\mu_j^2\right) + \rho_n\sum_{j\in B_i}\sum_{k\neq j\in B_i} cov(V_jV_k)\right)$$

$$- 2cov(R_iV_i, \frac{1}{n_i}\sum_{j\in B_i} R_jV_j)$$

$$Var(R_iV_i - \frac{1}{n_i}\sum_{j\in B_i} R_jV_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_{j\in B_i}\mu_j)) =$$

$$= \rho_n(Var(V_i) + (1-\rho_n)\mu_i^2)\left(1 - \frac{2}{n_i}\right) + \rho_n\frac{1}{n_i^2}\left(\sum_{j\in B_i}\left(Var(V_j) + (1-\rho_n)\mu_j^2\right) + \rho_n\sum_{j\in B_i}\sum_{k\neq j\in B_i} cov(V_jV_k)\right)$$

$$- 2\frac{1}{n_i}\rho_n^2\sum_{j\neq i\in B_i} cov(V_i, V_j)$$

We need now to compute the

$$Var\left(\sum_i\left(R_{ni}V_i - \frac{1}{n_i}\sum_{j\in B_i} R_jV_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_{j\in B_i}\mu_j)\right)\right).$$

This is equal to the sum of the variances plus two times an extra covariance term that consider the spatial dependence induced by the mixing process θ. Formally,

$$\sum_i\sum_{k\neq i} cov\left(\left(R_iV_i - \frac{1}{n_i}\sum_{j\in B_i} R_jV_j\right)\cdot\left(R_kV_k - \frac{1}{n_k}\sum_{l\in B_k} R_lV_l\right)\right) =$$

$$\sum_i\sum_{k\neq i}\left(\rho_n^2 cov(V_i, V_k) - \rho_n^2\frac{1}{n_k}\sum_{l\in B_k} cov(V_i, V_l) - \rho_n^2\frac{1}{n_i}\sum_{j\in B_i} cov(V_k, V_j) + \frac{1}{n_i}\frac{1}{n_k}\rho_n^2\sum_{j\in B_i}\sum_{l\in B_k} cov(V_l, V_j)\right)$$

$$\sum_i\sum_{k\neq i}\rho_n^2\left(cov(V_i, V_k) - \frac{1}{n_k}\sum_{l\in B_k} cov(V_i, V_l) - \frac{1}{n_i}\sum_{j\in B_i} cov(V_k, V_j) + \frac{1}{n_i}\frac{1}{n_k}\sum_{j\in B_i}\sum_{l\in B_k} cov(V_l, V_j)\right) =$$

$$\rho_n^2\sum_i\sum_{k\neq i} cov\left(\left(V_i - \frac{1}{n_i}\sum_{j\in B_i} V_j\right), \left(V_k - \frac{1}{n_k}\sum_{l\in B_k} V_l\right)\right).$$

So plugin in this extra term into (A.10) we have

$$Var\left(\sum_i R_{ni}V_i - \frac{1}{n_i}\sum_j R_jV_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_j \mu_j)\right) =$$

$$= \sum_i \rho_n\Bigg[(Var(V_i) + (1-\rho_n)\mu_i^2)\left(1 - \frac{2}{n_i}\right) +$$

$$+ \frac{1}{n_i^2}\Bigg(\sum_{j\in B_i}\left(Var(V_j) + (1-\rho_n)\mu_j^2\right) + \rho_n \sum_{j\in B_i}\sum_{k\neq j\in B_i} cov(V_jV_k)\Bigg) - \frac{1}{n_i}\rho_n \sum_{j\neq i\in B_i} cov(V_i,V_j) +$$

$$+ \rho_n \sum_i \sum_{k\neq i} cov\Bigg(\Big(V_i - \frac{1}{n_i}\sum_{j\in B_i} V_j\Big), \Big(V_k - \frac{1}{n_k}\sum_{l\in B_k} V_l\Big)\Bigg)\Bigg].$$

Let us define $q_{ni} = \sum_i V_i - \frac{1}{n_i}\sum_j V_j - (\mu_i - \frac{1}{n_i}\sum_j \mu_j)$. From the NED definition,

$$\|q_{in} - E(q_{in}|\mathcal{F}_{in}(s))\|_p \leq$$

$$\|(V_i - \mu_i) - \frac{1}{n_i}\sum_{j\in B_i}(V_j - \mu_j)\|_p + \|E((V_i - \mu_i)|\mathcal{F}_{in}(s)) - E((\frac{1}{n_i}\sum_{j\in B_i}(V_j - \mu_j))|\mathcal{F}_{jn}(s))\|_p \leq$$

$$\nu_{in}\psi(s) + \nu_{j\in B_i}\psi(s) \leq \nu_{i,j\in B_i}2\psi(s),$$

where $\nu_{i,j\in B_i} = \max\{\nu_i, ...\nu_{n_i}\}, \forall j \in B_i$. Thus, q_{in} is NED on θ_i of the same size, see also Theorem 17.8 pag. 267 in Davidson (1994). To directly apply Theorem 1 it is convenient to decompose q_{ni} in two components: the " individual" and " neighborhood" ones. Let us start with the the " individual" component, $q_{1ni} = R_i(V_i - \mu_i)$. Given that θ_i and q_{1ni} have uniformly bounded moments, we can set $\nu_{in} = c_{in} = 1$ (Jenish and Prucha, 2012). First, observe that given n, R_{ij} is i.i.d. and so is also m-dependent (See Defintion 3 and Lemma B.4 in Xu and Wooldridge, 2022). We check conditions *a)-d)* in Theorem 1. Let us start with *a)*

$$\sup_{n,i\in D_n} E(|q_{1ni}/c_{in}|)^{2+\delta} < \infty$$

for some $\delta > 0$, where $c_{in} > 0$ is a sequence of constants. Setting, $c_{in} = 1$, allows us to satisfy the requirement by Assumption.

For condition *b)*, we need that for any fixed $s > 0$, there exist a positive constant C such that for any n and every nonempty set $K \subseteq D_n$, $E(\sum_{i\in K}\frac{R_i}{\sqrt{\rho_n}}E(q_{1ni}/j_n|\mathcal{F}(s)))^2 \geq C\sum_{i\in K}E(\frac{R_i}{\sqrt{\rho_n}}E(q_{1ni}/j_n|\mathcal{F}(s)))^2$, where $j_n = \max_{i\in D_n}\{c_{in}, \nu_{1ni}\}$. Thus, given that $j_n = 1$, let us start with the first part of

$$E\Bigg(\sum_i \frac{R_i}{\sqrt{\rho_n}}E\Big(V_i - \mu_i\Big|\mathcal{F}_{in}(s)\Big)\Bigg)^2 \geq CE\sum_i\left(\frac{R_i}{\sqrt{\rho_n}}E\Big(V_i - \mu_i\Big|\mathcal{F}_{in}(s)\Big)\right)^2.$$

Focusing on the LHS of the inequality, we have

$$E\Big(\sum_i \frac{R_i}{\sqrt{\rho_n}} E\Big(V_i - \mu_i \Big| \mathcal{F}_{in}(s)\Big)\Big)^2 = \tag{A.11}$$
$$\frac{E(R_i^2)}{\rho_n}\sum_i E\Big(V_i - \mu_i\Big|\mathcal{F}_{in}(s)\Big)^2 + \sum_i\sum_{k\neq i} E\Big(E\Big(V_i-\mu_i\Big|\mathcal{F}_{in}(s)\Big)E\Big(V_k-\mu_k\Big|\mathcal{F}_{kn}(s)\Big)\Big) =$$
$$\sum_i E\Big(V_i - \mu_i\Big|\mathcal{F}_{in}(s)\Big)^2 + \sum_i\sum_{k\neq i} E\Big(E\Big(V_i-\mu_i\Big|\mathcal{F}_{in}(s)\Big)E\Big(V_k-\mu_k\Big|\mathcal{F}_{kn}(s)\Big)\Big).$$

The RHS can be written as

$$CE\sum_i \Big(\frac{R_i}{\sqrt{\rho_n}} E\Big(V_i - \mu_i\Big|\mathcal{F}_{in}(s)\Big)\Big)^2 = \tag{A.12}$$
$$\frac{E(R_i^2)}{\rho_n}\sum_i E\Big(vV_i - \mu_i\Big|\mathcal{F}_{in}(s)\Big)^2 =$$
$$\sum_i E\Big(V_i - \mu_i\Big|\mathcal{F}_{in}(s)\Big)^2.$$

Thus, using (A.11)-(A.12), we have

$$\sum_i\sum_{k\neq i} E\Big(E\Big(V_i-\mu_i\Big|\mathcal{F}_{in}(s)\Big)E\Big(V_k-\mu_k\Big|\mathcal{F}_{kn}(s)\Big)\Big) \geq C.$$

This is sufficient for condition *b)* to hold within our framework. While the NED coefficients $\psi(s)$ are assumed to have a magnitude of size $-\boldsymbol{\nu}$ (*d*)), condition *e)* (stating that the NED scaling factors must satisfy $\sup_{n,i\in D_n} c_{1ni}^{-1}\nu_{1ni} \leq C < \infty$) is trivially satisfied. Let us define $Q_{1n} = \sum_i \frac{R_i}{\sqrt{\rho_n}} q_{1in}$ and $\sigma_{1n} = Var(Q_{1n})$. By Theorem 1, we have $Q_{1n}\sigma_{1n}^{-1} \xrightarrow{d} N(0,1)$.

Now, let us apply Theorem 1 once again to the " neighborhood" component. By setting $q_{2ni} = \frac{1}{n_i}\sum_{j\in B_i} R_j(V_j - \mu_j)$, and $c_{2in} = \nu_{2in} = 1$. Conditions *a)*, *c)*, and *e)* hold following the same reasoning as before for the " individual" part. Observe that *d)* holds with the same size because the process $\sum_{j\in B_i} V_j$ for each i has size equal to $-\min_{j\in B_i}(\nu_{jn})$. However for *b)* to hold we need a different sufficient condition,

$$\sum_i\sum_{k\neq i} E\Big(E\Big(\frac{1}{n_i}\sum_{j\in B_i}(V_j-\mu_j)\Big|\mathcal{F}_{in}(s)\Big)E\Big(\frac{1}{n_k}\sum_{l\in B_k}(V_l-\mu_l)\Big|\mathcal{F}_{kn}(s)\Big)\Big) \geq C.$$

Let us define $Q_{2n} = \sum_i \sum_{j\in B_i} \frac{1}{n_i}\frac{R_j}{\sqrt{n\rho_n}} q_{2in}$ and $\sigma_{2n} = Var(Q_{2n})$. By Theorem 1, we have $Q_{2n}\sigma_{2n}^{-1} \xrightarrow{d} N(0,1)$. Thus, both the " individual" and " neighborhood" components are asymptotically normally distributed.

Therefore,

$$\sum_i \Big(\frac{R_{ni}V_i - \frac{1}{n_i}\sum_j R_j V_j - \rho_n(\mu_i - \frac{1}{n_i}\sum_j \mu_j)}{(s_n\sqrt{\rho_n n})}\Big) \xrightarrow{d} \mathcal{N}(0,1),$$

given that $\frac{s_n}{\sqrt{\sigma^2+\sigma^2_{neigh}+(1-\rho)(\kappa^2+\kappa^2_{neigh})+\rho\sigma_{cov}}} \to 1$, and $\sum_i \mu_{ni} - \frac{1}{n_i}\sum_{j\in B_i}\mu_j = 0$.

■

Lemma 7. *Under Assumptions, 1- 7, 9, and Assumption 12 b) in Appendix A.2. Let* $\boldsymbol{B}^u_{NW} = \boldsymbol{B}^{ehw}_{NW} - \boldsymbol{B}^{cond}_{NW}$*, and* $\boldsymbol{B}^{cov}_{NW} = \boldsymbol{B}^{B_i}_{NW} - \boldsymbol{B}^{iB_i}_{NW} + \boldsymbol{B}^{\Delta B_i B_k}_{NW}$*. Then,*

$$\sum_i \frac{1}{\sqrt{N}}\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\varepsilon_i \xrightarrow{d}$$
$$\mathcal{N}\Bigg(\boldsymbol{0}, (\boldsymbol{B}^{cond}_{ND} + \boldsymbol{B}^{neigh}_{NW} + (1-\rho)(\boldsymbol{B}^u_{NW} + \boldsymbol{B}^{uneigh}_{NW}) + \rho\boldsymbol{B}^{cov}_{NW})\Bigg).$$

Proof. We first study the convergence of this term

$$\frac{1}{\sqrt{n}}\sum_i E\Bigg(\Delta_{i,n_i}\boldsymbol{x}_i\Delta_{i,n_i}\theta_i\Bigg).$$

Without loss of generality, let us focus on the component-wise convergence. We have that

$$\frac{1}{\sqrt{n}}\sum_i E(\Delta_{i,n_i}x_i\Delta_{i,n_i}\theta_i) \leq \sqrt{n}\max_i E(\Delta_{i,n_i}x_i\Delta_{i,n_i}\theta_i) \leq \sqrt{n}\max_i E|\Delta_{i,n_i}x_i\Delta_{i,n_i}\theta_i| \leq$$

$$\sqrt{n}\max_i E(|\Delta_{ij}x_i|)\max_i E(|\Delta_{ij}\theta_i|) = O(1)\sqrt{n}\max_i E|(\Delta_{ij}\theta_i)| = O(1)o(1) = o(1),$$

where the first inequality follows by triangular inequality, the third inequality by Cauchy-Schwartz inequality, the first equality by Assumption 5 a) and Holder's inequality, and the second equality by Assumption 7. Consider $\Delta_{i,n_i}V_{ni} = a'\Delta_{i,n_i}\boldsymbol{x}_i\Delta_{i,n_i}\varepsilon_i$. Let us verify the conditions of Lemma 4. We start with $\frac{1}{n}\sum_i E(|V_{ni}|^{2+\delta})$ is bounded by a positive constant for some $\delta > 0$.

$$\frac{1}{n}\sum_i E\Bigg(\Bigg|\Bigg|\Delta_{i,n_i}\boldsymbol{x}_i\Delta_{i,n_i}\varepsilon_i\Bigg|\Bigg|^{2+\delta}\Bigg) \leq \frac{||a||^{2+\delta}}{n}\sum_i E\Bigg(||\boldsymbol{x}_i||^{2+\delta}(|y_i| + ||\boldsymbol{x}_i||||\boldsymbol{\beta}|| + ||\boldsymbol{z}_i||||\boldsymbol{\gamma}||)^{2+\delta}\Bigg) +$$
$$\frac{||a||^{2+\delta}}{n_i}\sum_j E\Bigg(||\boldsymbol{x}_j||^{2+\delta}(|y_j| + ||\boldsymbol{x}_j||||\boldsymbol{\beta}|| + ||\boldsymbol{z}_j||||\boldsymbol{\gamma}||)^{2+\delta}\Bigg) \leq C,$$

by Minkowski's inequality and Assumption 5 a). Furthermore,

$$\sum_i \mu_{ni} - \frac{1}{n_i}\sum_j \mu_j = 0 = a'\sum_i E(\Delta_{i,n_i}\boldsymbol{x}_i\Delta_{i,n_i}\varepsilon_i) = 0,$$

by defining the coefficient β^c_{NW} as orthogonality condition. Let

$\frac{1}{n}\sum_{i=1}^n Var(V_{ni})\Bigg(1 - \frac{2}{n_i}\Bigg) = a'\Bigg(\frac{1}{n}\sum_i Var(\boldsymbol{x}_i\varepsilon_i)\Bigg(1 - \frac{2}{n_i}\Bigg)\Bigg)a \to a'\boldsymbol{B}^{cond}_{NW}a,$

$\frac{1}{n}\sum_{i=1}^n \mu_i^2\Bigg(1 - \frac{2}{n_i}\Bigg) = a'\Bigg(\frac{1}{n}\sum_i E(\boldsymbol{x}_i\varepsilon_i)E(\boldsymbol{x}_i\varepsilon_i)'\Bigg(1 - \frac{2}{n_i}\Bigg)\Bigg)a \to a'\boldsymbol{B}^u_{NW}a,$

$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j\in B_i}Var(V_{nj}) \ = a'\bigg(\frac{1}{n}\sum_i\frac{1}{n_i^2}\sum_{j\in B_i}Var(\boldsymbol{x}_j\varepsilon_j)\bigg)a \to a'\boldsymbol{B}_{NW}^{neigh}a,$

$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j\in B_i}\mu_j^2 = a'\bigg(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i^2}\sum_{j\in B_i}E(\boldsymbol{x}_j\varepsilon_j)E(\boldsymbol{x}_j\varepsilon_j)'\bigg)a \to a'\boldsymbol{B}_{NW}^{uneigh}a,$

$\frac{1}{n}\sum_i\frac{1}{n_i^2}\sum_{j\in B_i}\sum_{k\neq j\in B_i}cov(V_jV_k) = a'\bigg(\frac{1}{n}\sum_i\frac{1}{n_i^2}\sum_{j\in B_i}\sum_{k\neq j\in B_i}cov\Big((\boldsymbol{x}_j\varepsilon_j),(\boldsymbol{x}_k\varepsilon_k)\Big)\bigg)a \to a'\boldsymbol{B}_{NW}^{B_i}a,$

$\frac{1}{n}\sum_i\frac{1}{n_i^2}\sum_{j\neq i\in B_i}cov(V_i,V_j) = a'\bigg(\frac{1}{n}\sum_i\frac{1}{n_i}\sum_{j\neq i\in B_i}cov\Big((\boldsymbol{x}_i\varepsilon_i),(\boldsymbol{x}_j\varepsilon_j)\Big)\bigg)a \to a'\boldsymbol{B}_{NW}^{i,B_i}a,$

$\frac{1}{n}\sum_i\sum_{k\neq i}cov\Big(\Delta_{i,n_i}V_i,\Delta_{k,n_k}V_k\Big) = a'\bigg(\frac{1}{n}\sum_i\sum_{k\neq i}cov\Big((\Delta_{i,n_i}\boldsymbol{x}_i,\Delta_{i,n_i}\varepsilon_i)(\Delta_{k,n_k}\boldsymbol{x}_k,\Delta_{k,n_k}\varepsilon_k)\Big)\bigg)a \to$
$a'\boldsymbol{B}_{NW}^{\Delta B_iB_k}a,$
and $\boldsymbol{B}_{NW}^{cov} = \boldsymbol{B}_{NW}^{B_i} - \boldsymbol{B}_{NW}^{iB_i} + \boldsymbol{B}_{NW}^{\Delta B_iB_k}$. By Lemma 6, this implies

$$a'\frac{1}{\sqrt{N}}\sum_i\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\varepsilon_i \xrightarrow{d}$$
$$\mathcal{N}\Bigg(\mathbf{0}, a'\bigg(\boldsymbol{B}_{ND}^{cond} + \boldsymbol{B}_{NW}^{neigh} + (1-\rho)(\boldsymbol{B}_{NW}^{u} + \boldsymbol{B}_{NW}^{uneigh}) + \rho\boldsymbol{B}_{NW}^{cov}\bigg)a\Bigg).$$

Using the Cramer-Wold device, the result follows

$$\frac{1}{\sqrt{N}}\sum_i\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\varepsilon_i \xrightarrow{d}$$
$$\mathcal{N}\bigg(\mathbf{0}, (\boldsymbol{B}_{ND}^{cond} + \boldsymbol{B}_{NW}^{neigh} + (1-\rho)(\boldsymbol{B}_{NW}^{u} + \boldsymbol{B}_{NW}^{uneigh}) + \rho\boldsymbol{B}_{NW}^{cov})\bigg).$$

■

Proof of Proposition 2 .

We can write the NW estimator as

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}_{NW}\\ \hat{\boldsymbol{\gamma}}_{NW}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_{NW}^{c}\\ \boldsymbol{\gamma}_{NW}^{c}\end{pmatrix} + \Bigg(\sum_i\begin{pmatrix}\tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i' & \tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\boldsymbol{z}_i'\\ \tilde{\Delta}_{i,n_i}\mathbf{z}_i\tilde{\Delta}_{i,n_i}\mathbf{x}_i' & \tilde{\Delta}_{i,n_i}\mathbf{z}_i\tilde{\Delta}_{i,n_i}\mathbf{z}_i'\end{pmatrix}\Bigg)^{-1}\sum_i\begin{pmatrix}\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}(\varepsilon_i+\theta_i)\\ \tilde{\Delta}_{i,n_i}\boldsymbol{z}_i\tilde{\Delta}_{i,n_i}(\varepsilon_i+\theta_i).\end{pmatrix}$$

Thus as $n\to\infty$ and by Lemma 5 with Assumption 7 ,

$$\sqrt{N_p}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{NW}-\boldsymbol{\beta}_{NW}^{c}\\ \hat{\boldsymbol{\gamma}}_{NW}-\boldsymbol{\gamma}_{NW}^{c}\end{pmatrix} = \Bigg(\frac{1}{N}\sum_i\begin{pmatrix}\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i' & \tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\boldsymbol{z}_i'\\ \tilde{\Delta}_{i,n_i}\mathbf{z}_i\tilde{\Delta}_{i,n_i}\mathbf{x}_i' & \tilde{\Delta}_{i,n_i}\mathbf{z}_i\tilde{\Delta}_{i,n_i}\mathbf{z}_i'\end{pmatrix}\Bigg)^{-1}\frac{1}{\sqrt{N}}\sum_i\begin{pmatrix}\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\varepsilon_i\\ \tilde{\Delta}_{i,n_i}\boldsymbol{z}_i\tilde{\Delta}_{i,n_i}\varepsilon_i.\end{pmatrix}+o_p(1)$$

$$= \begin{pmatrix}\Delta_{i,n_i}\boldsymbol{\Omega}_n^{xx} & \Delta_{i,n_i}\boldsymbol{\Omega}_n^{xz}\\ \Delta_{i,n_i}\boldsymbol{\Omega}_n^{zx} & \Delta_{i,n_i}\boldsymbol{\Omega}_n^{zz}\end{pmatrix}^{-1}\frac{1}{\sqrt{N}}\sum_i\begin{pmatrix}\tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\varepsilon_i\\ \tilde{\Delta}_{i,n_i}\boldsymbol{z}_i\tilde{\Delta}_{i,n_i}\varepsilon_i\end{pmatrix} + r_n,$$

where

$$r_n = \left[\begin{pmatrix} \tilde{\Delta}_{i,n_i} \boldsymbol{W}_n^{xx} & \tilde{\Delta}_{i,n_i} \boldsymbol{W}_n^{xz} \\ \tilde{\Delta}_{i,n_i} \boldsymbol{W}_n^{zx} & \tilde{\Delta}_{i,n_i} \boldsymbol{W}_n^{zz} \end{pmatrix}^{-1} - \begin{pmatrix} \Delta_{i,n_i} \boldsymbol{\Omega}_n^{xx} & \Delta_{i,n_i} \boldsymbol{\Omega}_n^{xz} \\ \Delta_{i,n_i} \boldsymbol{\Omega}_n^{zx} & \Delta_{i,n_i} \boldsymbol{\Omega}_n^{zz} \end{pmatrix}^{-1} \right] \frac{1}{\sqrt{N}} \sum_i \begin{pmatrix} \tilde{\Delta}_{i,n_i} \boldsymbol{x}_i \tilde{\Delta}_{i,n_i} \varepsilon_i \\ \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i \tilde{\Delta}_{i,n_i} \varepsilon_i \end{pmatrix} + o_p(1).$$

Let us focus on

$$\frac{1}{\sqrt{N}} \sum_i \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i \tilde{\Delta}_{i,n_i} \varepsilon_i.$$

Rewrite the term as

$$\sqrt{\frac{n\rho_n}{N}} n^{-1/2} \sum_i \frac{1}{\sqrt{\rho_n}} \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i \tilde{\Delta}_{i,n_i} \varepsilon_i.$$

As shown in Lemma 4, $\sqrt{\frac{n_p \rho_n^2}{N}} \xrightarrow{p} 1$ and

$$E(\sum_i \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i \tilde{\Delta}_{i,n_i} \varepsilon_i) = \sum_i \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i E(\tilde{\Delta}_{i,n_i} \varepsilon_i) = \mathbf{0}.$$

Now if we show that the element-wise variance is bounded, then this implies that

$$\frac{1}{\sqrt{N}} \sum_i \tilde{\Delta}_{i,n_i} \boldsymbol{z}_i \tilde{\Delta}_{i,n_i} \varepsilon_i = O_p(1),$$

by Chebyshev's inequality. Thus,

$$\begin{aligned} Var\left(n^{-1/2} \sum_i \frac{1}{\sqrt{\rho_n}} \tilde{\Delta}_{i,n_i} z_i \tilde{\Delta}_{i,n_i} \varepsilon_i\right) &= n^{-1} Var\left(\sum_i \frac{1}{\sqrt{\rho_n}} \tilde{\Delta}_{i,n_i} z_i \tilde{\Delta}_{i,n_i} \varepsilon_i\right) = \qquad \text{(A.13)} \\ = \sum_i \Bigg[&(Var(z_i\varepsilon_i) + (1-\rho_n)(E(z_i\varepsilon_i))^2) \left(1 - \frac{2}{n_i}\right) + \\ + \frac{1}{n_i^2} \Bigg(\sum_{j \in B_i} &\left(Var(z_j\varepsilon_j) + (1-\rho_n)(E(z_j\varepsilon_j))^2\right) + \rho_n \sum_{j,k \in B_i} cov(z_j\varepsilon_j, z_k\varepsilon_k) \Bigg) - \frac{1}{n_i} \rho_n \sum_{j \neq i \in B_i} cov(z_i\varepsilon_i, z_j\varepsilon_j) + \\ + \rho_n \sum_i \sum_{k \neq i} &cov\left(\left(z_i\, \varepsilon_i - \frac{1}{n_i} \sum_{j \in B_i} z_j\varepsilon_j\right), \left(z_k\varepsilon_k - \frac{1}{n_k} \sum_{l \in B_k} z_l\varepsilon_l\right)\right) \Bigg]. \end{aligned}$$

The first terms,

$$n^{-1} \sum_i Var(z_i\varepsilon_i) = n^{-1} \sum_i z_i^2 Var(\varepsilon_i) \leq n^{-1} \sum_i z_i^2 E(\varepsilon_i^2),$$

$$(1-\rho_n) n^{-1} \sum_i (E(z_i\varepsilon_i))^2 = (1-\rho_n) n^{-1} \sum_i z_i^2 (E(\varepsilon_i))^2, \ and$$

$$n^{-1} \rho_n \sum_{j,k \in B_i} cov(z_j\varepsilon_j, z_k\varepsilon_k) \leq n^{-1} \sum_{j,k \in B_i} z_j z_k (E(z_j\varepsilon_j z_k\varepsilon_k)),$$

are bounded by Assumption 5. The same can be applied to the other terms. Therefore given that

$\Delta_{i,n_i}\boldsymbol{\Omega}_n^{xz} = \mathbf{0}$ and $\frac{1}{\sqrt{N}}\sum_i \tilde{\Delta}_{i,n_i}\boldsymbol{z}_i\tilde{\Delta}_{i,n_i}\varepsilon_i = O_p(1)$, we have

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{NW} - \boldsymbol{\beta}^c_{NW}) = (\Delta_{i,n_i}\boldsymbol{\Omega}_n^{xx})^{-1} + \frac{1}{\sqrt{N}}\sum_i \tilde{\Delta}_{i,n_i}\boldsymbol{x}_i\tilde{\Delta}_{i,n_i}\varepsilon_i + o_p(1).$$

The result follows by Lemma 7 and Slutsky's theorem.

■

Proof of Lemma 2 .

First, observe that by Lemma 1 and 6, $\boldsymbol{\lambda}$ exists, it is equal to $\boldsymbol{\phi}_n$ and

$$\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = \left(\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{ij}\boldsymbol{z}_i'\right)\left(\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\boldsymbol{z}_i\tilde{\Delta}_{ij}\boldsymbol{z}_i'\right)^{-1} \xrightarrow{p} \mathbf{0}.$$

Then using the definition of $\hat{\mathbf{A}} = \sum_{i<j,j\in B_i\setminus i}(\tilde{\Delta}_{ij}\boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}\tilde{\Delta}_{ij}\boldsymbol{z}_i)(\tilde{\Delta}_{ij}\boldsymbol{t}_i - \hat{\boldsymbol{\lambda}}\tilde{\Delta}_{ij}\boldsymbol{z}_i)'$, we have

$$\hat{\mathbf{A}} - \tilde{\Delta}_{ij}\boldsymbol{W}_n^{xx} = (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})\tilde{\Delta}_{ij}\boldsymbol{W}_n^{zz}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})' - \tilde{\Delta}_{ij}\boldsymbol{W}_n^{xz}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})' - (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})\tilde{\Delta}_{ij}\boldsymbol{W}_n^{xz} \xrightarrow{p} \mathbf{0}.$$

Thus, Lemma 1 and Assumption 5 imply that $\hat{\mathbf{A}} \xrightarrow{p} \mathbf{A}$ that is full rank. In the same way, we can show that $\hat{\mathbf{A}}_{\mathbf{NW}} \xrightarrow{p} \mathbf{A}_{\mathbf{NW}}$.

Let us define

$$\tilde{\boldsymbol{B}}_n^{ecov} = \frac{1}{N_p}\sum_{i<j,j\in B_i\setminus i}\sum_{k\neq i<l,l\in B_k\setminus k}\tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{ij}\varepsilon_i\tilde{\Delta}_{kl}\boldsymbol{x}_k'\tilde{\Delta}_{kl}\varepsilon_l,$$

$$\tilde{\boldsymbol{B}}_n^{ehw} = \frac{1}{N_p}\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\varepsilon_i^2\tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{ij}\boldsymbol{x}_i',$$

$$\bar{\boldsymbol{B}}_n^{ecov} = \frac{1}{N_p}\sum_{i<j,j\in B_i\setminus i}\sum_{k\neq i<l,l\in B_k\setminus k}\tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{ij}\hat{\varepsilon}_i\tilde{\Delta}_{kl}\boldsymbol{x}_k'\tilde{\Delta}_{kl}\hat{\varepsilon}_l,$$

$$\bar{\boldsymbol{B}}_n^{ehw} = \frac{1}{N_p}\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\hat{\varepsilon}_i^2\tilde{\Delta}_{ij}\boldsymbol{x}_i\tilde{\Delta}_{ij}\boldsymbol{x}_i'.$$

$$\boldsymbol{B}_n^{ecov} = \frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}\sum_{k\neq i<l,l\in B_k\setminus k}E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\varepsilon_i\Delta_{kl}\boldsymbol{x}_k\Delta_{kl}\varepsilon_l),$$

and

$$\boldsymbol{B}_n^{ehw} = \frac{1}{n_p}\sum_{i<j,j\in B_i\setminus i}E(\Delta_{ij}\varepsilon_i^2\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\boldsymbol{x}_i').$$

Let us define, $\hat{\boldsymbol{B}}^{ecov} = \frac{1}{N_p}\sum_{i<j,j\in B_i\setminus i}\sum_{k\neq i<l,l\in B_k\setminus k}\mathbf{1}_{(k\in B_i|i\in B_k|k=j|i=l)}\tilde{\Delta}_{ij}\hat{\epsilon}_i\tilde{\Delta}_{kl}\hat{\epsilon}_k\tilde{\Delta}_{ij}\hat{\boldsymbol{x}}_i\tilde{\Delta}_{kl}\hat{\boldsymbol{x}}_k'$, and $\hat{\boldsymbol{B}}^{ehw} = \frac{1}{N_p}\sum_{i<j,j\in B_i\setminus i}\tilde{\Delta}_{ij}\hat{\epsilon}_i^2\tilde{\Delta}_{ij}\hat{\boldsymbol{x}}_i\tilde{\Delta}_{ij}\hat{\boldsymbol{x}}_i'$ and rewrite $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{B}}^{ehw} + \hat{\boldsymbol{B}}^{ecov})$. Observe that $\lim_{n\to\infty}E(\hat{\boldsymbol{B}}) = \boldsymbol{B}^{ehw} + \rho^2\boldsymbol{B}^{ecov}$. Subsequently, employing the same rationale as presented in the proof of Lemma 2 in

Abadie et al. (2020) (page 293), we can demonstrate that $\hat{\boldsymbol{B}} - (\boldsymbol{B}^{ehw} + \rho^2 \boldsymbol{B}^{ecov}) \xrightarrow{p} \mathbf{0}$ under the conditions of Assumptions 5a) with $\delta = 4$ and 8, where $\boldsymbol{B}^{ecov} = \lim_{n\to\infty} \boldsymbol{B}_n^{ecov}$. This derivation utilizes both the moment convergences presented in in Lemma 1 *a)*, the assertion that $\hat{\boldsymbol{\beta}}_{ND} \xrightarrow{p} \boldsymbol{\beta}_{ND}$ as established in Proposition 1, and $\hat{\boldsymbol{\gamma}}_{ND} \xrightarrow{p} \boldsymbol{\gamma}_{ND}$ (using again Lemma 1 *a)*). Using the same reasoning, we can show that $\hat{\boldsymbol{B}}_{NW} - (\boldsymbol{B}_{NW}^{ehw} + \boldsymbol{B}_{NW}^{ehwneigh} + \rho \boldsymbol{B}_{NW}^{ecov}) \xrightarrow{p} \mathbf{0}$, under the conditions of Assumptions 5a) with $\delta = 4$ and 9. This derivation utilizes both the moment convergences presented in Lemma 1 *b)*, the assertion that $\hat{\boldsymbol{\beta}}_{NW} \xrightarrow{p} \boldsymbol{\beta}_{NW}$ as established in Proposition 2, and $\hat{\boldsymbol{\gamma}}_{NW} \xrightarrow{p} \boldsymbol{\gamma}_{NW}$ (using again Lemma 1 *b)*).

■

Appendix A.3. Set cardinalities for irregular lattices

We report Lemma A.1 (ii) and (iii) in Jenish and Prucha (2009).

Lemma 8. *Let $D \subset R^{d_0}$, $d_0 \geq 1$, be an infinitely countable unevenly spaced lattice. For any distance d there are at most $k_1 d^{d_0}$ points in B_i^d and $k_2 d^{d_0 - 1}$ points in B_i^d / B_i^{d-1}, where k_1 and k_2 are positive constants.*

Appendix A.4. Homoskedastic case

When errors are homoskedastic, i.e. $\mathbf{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' | \boldsymbol{X}, \boldsymbol{\theta}) = \sigma^2 \boldsymbol{I}_n$, then a consistent estimator of the asymptotic variance of the ND estimator is

$$\hat{\boldsymbol{V}}_{ND}^{hom} = \hat{\sigma}_2 \hat{\mathbf{A}}^{-\mathbf{1}} \hat{\boldsymbol{B}}^{hom} \hat{\mathbf{A}}^{-1}, \tag{A.14}$$

with

$$\hat{\boldsymbol{B}}^{hom} = \boldsymbol{X}' \boldsymbol{D}' \boldsymbol{D} \boldsymbol{D}' \boldsymbol{D} \boldsymbol{X}, \tag{A.15}$$

and an unbiased estimator of σ^2 (see also, e.g. Duranton et al., 2011) is

$$\hat{\sigma}^2 = \Big[tr(\boldsymbol{D}\boldsymbol{D}') - tr(\hat{\mathbf{A}} \hat{\boldsymbol{B}}) \Big]^{-1} \sum_{i<j, j \in B_i \setminus i} \Delta_{ij} \hat{\epsilon}_i \Delta_{ij} \hat{\epsilon}_i. \tag{A.16}$$

Similarly, for the NW estimator we have

$$\hat{\boldsymbol{V}}_{NW}^{hom} = \hat{\sigma}_2 \hat{\mathbf{A}}_W^{-1} \hat{\boldsymbol{B}}_W^{hom} \hat{\mathbf{A}}_W^{-1}, \tag{A.17}$$

where $\hat{\boldsymbol{B}}_W^{hom}$ have the same formulation of $\hat{\boldsymbol{B}}$ with $\boldsymbol{G}_n$ in place of $\boldsymbol{D}$, while an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \big[tr(\boldsymbol{G}_n \boldsymbol{G}_n') \big]^{-1} \sum_{i=1}^{n} \Delta_{i,n_i} \hat{\epsilon}_i \Delta_{i,n_i} \hat{\epsilon}_i. \tag{A.18}$$

Appendix A.5. Sequential Hausman-like tests and error rate

This Appendix shows that the sequence of Hausman-like tests performed to find the optimal threshold d^* does not result in more frequent rejection of the true hypotheses. In doing so, we follow the approach *for testing hypotheses in order* given by Rosenbaum (2008).[19] Let $\mathcal{D} = 1, 2 \ldots, \kappa$ be a totally ordered set with order $\preceq$. In our framework, $\mathcal{D}$ represents the set of threshold distances. Let H_d, $d \in \mathcal{D}$, be a class of hypotheses, indexed by the threshold d. For each hypothesis, H_d, the researcher fixes a nominal size α. If H_d is true then $pr(p_d \leq \alpha) \leq \alpha$, where p_d is the p-value for the Hausman-like test implemented at the threshold distance d. The sequence of hypotheses H_d is indexed by all the distances that satisfy the inequality $d \leq d^*$. In this context, the distance threshold $\kappa - 1$ is preferred to κ, i.e. $\kappa \preceq \kappa - 1$ if $\kappa - 1 \leq \kappa$. Further, we assume that there is some H_{d^*}, that is true and for all $d \prec d^*$, H_d is false.

Sequential Hausman-like Tests procedure. For each $d \in \mathcal{D}$, test H_d at nominal size α if and only if H_{d_1}, $d_1 \in \mathcal{D}$, has been previously tested (at nominal size α) and rejected for all $d_1 \prec d\,(d_1 < d)$; otherwise do not test H_d.

The *Sequential Hausman-like Tests procedure* falls into the Method 1. proposed in Rosenbaum (2008). The author shows that the probability of the researcher rejecting at least one true hypothesis using Method 1. is at most α (Proposition 1). In other words, under these assumptions and the ordering nature of the hypotheses, the sequentiality of this procedure does not affect the probability of type I error.

[19] Sales (2017) provides an alternative method for sequential specification tests using an illustrative example similar to ours: the selection of bandwidth for a regression discontinuity design.