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

Recent literature emphasizes the importance of accounting for local unobserved effects, which may stem from either omitted neighborhood characteristics or shocks. As in a panel data setting, we exploit neighborhood data transformations to eliminate these effects. We consider a class of estimators indexed by a proximity threshold, allowing the researcher to select the neighborhood. Gaining insights from the recent dyadic data literature, we study their asymptotic properties allowing for overlapping neighborhoods. We develop a Hausman-like testing strategy to detect the presence of correlated local unobserved effects. Our test does not require assumptions on the neighborhood unobserved heterogeneity distribution and allows practitioners to select the best neighborhood specification to transform the data. We illustrate our approach using data from Miguel and Kremer (2004). Our method allows us to recover the same direct average treatment effect by considering treatment spillovers as local unobserved effects.

Keywords: Neighborhood data transformation, Local unobserved heterogeneity, Hausman-like test
JEL: C12, C21, I25

[*] We thank Franco Peracchi, Sergio Pastorello, Robert Chirinko, Edoardo Rainone, Giovanni Mellace, Federico Crudu, Michaela Kesina, Tom Boot, Lingwei Kong, Tan Chang, and participants of the EEF seminar series at the Department of Economics, Econometrics and Finance of the University of Groningen and the participants of the XV World Conference of the Spatial Econometric Association and the internal seminar series of the Department of Economics and Finance, University of Rome Tor Vergata, for useful 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. This version: June 28, 2022.

1. Introduction

Neighboring units may share similar behaviors and characteristics, which, in turn, might confound the impact assessment of an explanatory variable, say some nonrandomized treatment, on the outcome of interest. This is especially true when the researcher does not observe these factors. 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 cannot adequately control for neighborhood characteristics, then the impact of school quality may be overestimated (Black, 1999).

A popular solution to deal with these issues is collecting geo-coded data and pairing units according to their geographical proximity. If neighborhood unobservables are "sufficiently smooth" over space, then a "spatial differencing" approach can be used to rule them out. This strategy was pioneered by Holmes (1998) in models with additively-separable unobserved effects.[1] However, despite its popularity, existing literature neither provides criteria to select the threshold distance for pairing units nor rigorously investigates the properties of the differencing estimator. Furthermore, proximity is not only geographical and can be defined along many other dimensions, e.g., economic, social, cognitive, institutional, cultural, and technological.

This paper fills this gap by developing a class of estimators indexed by a general proximity threshold, allowing the researcher to select the neighborhood. Gaining insights from the literature on linear panel data model estimation, we consider two data transformations for models with additively-separable unobserved heterogeneity: the neighborhood-difference (ND) and the within-neighborhood (NW) transformations. We draw inspiration from the inferential framework for dyadic data proposed by Tabord-Meehan (2019) to model the dependence between observations induced by the considered transformations. In particular, we model the dependence across overlapping neighborhoods using a dependency graph. Thanks to this dependence structure, we derive sufficient conditions to apply the central limit theorem (CLT) of Janson et al. (1988).[2] In the spirit of Bartolucci et al. (2015), we develop a Hausman-like testing strategy allowing to detect correlated local unobserved effects and, if the data support their presence, to select the best neighborhood specification to transform the data.

Our estimation framework 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 the linking behavior is driven by an unknown function of social unobservables. 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 to eliminate the unit-specific neighborhood unobserved heterogeneity (NUH), which enters the model linearly and additively. However, we do not assume any specific neighborhood formation model. Our framework

[1] Applications of this strategy are many and include studies on the effects of school quality on house prices (Black, 1999, Fack and Grenet, 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).

[2] For inference in the case of non-overlapping neighborhoods see Hansen and Lee (2019).

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 illustrate the usefulness of our approach using 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, i.e., the impact of deworming for students in schools near the treated ones. By exploiting the neighborhood specification suggested by our test, we show how to retrieve the same direct causal effect in Miguel and Kremer (2004) treating cross-school externalities as NUH. This is especially important when a clustered experimental design cannot be implemented, for example, due to budget constraints.

The article is organized as follows. Section 2 introduces our statistical model and presents different interpretations. Section 3 describes the proposed estimators and studies their asymptotic properties. Section 4 presents a Hausman-like test to detect the presence of smooth NUH. This is especially important when a clustered experimental setting cannot be implemented, for example, due to budget constraints Section 5 investigates the finite sample properties of the proposed estimators and test statistic using Monte Carlo simulations. In Section 6, we illustrate the proposed methodology exploiting data from a clustered randomized medical treatment program. Section 7 concludes.

2. The statistical model and interpretations

We consider a random sample $\{y_i, \boldsymbol{x}_i, \theta_i\}_{i=1}^n$ drawn from a large population. The baseline model for the outcome is the following

$$y_i = \theta_i + \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i, \tag{1}$$

where $\boldsymbol{x}_i$ is a $1 \times K$ vector of explanatory variables, $\boldsymbol{\beta}$ is K-vector of unknown parameters, θ_i denotes unit i's unobserved effect, and ϵ_i the idiosyncratic error. We treat θ_i as a fixed-effect arbitrarily correlated with one or more of the explanatory variables in $\boldsymbol{x}_i$.

The researcher observes $\{y_i, \boldsymbol{x}_i, d_{ij}\}_{i,j=1j\neq i}^n$, where d_{ij} is a general proximity measure, i.e., social, economical or geographical distance between units. The researcher specifies/selects a neighborhood B_i^d based on a distance threshold d. Suppose the proximity measure is constructed using economic or social indicators, e.g., relative GDP or trade volume. In that case, it could likely be correlated with the outcome of interest, and the resulting neighborhood specification B_i^d forms endogenously (Qu and Lee, 2015). However, our approach is not affected by this issue.

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 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. We assume 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.

Model (1) can be applied in many contexts. For instance, θ_i can be interpreted as a treatment spillover when a treatment assigned to unit j also affects the outcome of the neighboring unit i. Almost all the welfare and public health programs might generate spillovers: a cash transfer provided to the household's head may affect her and the partner's labor supply; a vaccine affects the individual's probability of getting infected and the likelihood of disease transmission within the community. In this setting, Miguel and Kremer (2004) study the effect of a deworming medical treatment on school absenteeism and health status in Kenya, assuming the following linear model for the outcome

$$y_{is} = \underbrace{\sum_{\bar{d}} \gamma_{\bar{d}} T_{s\bar{d}}^N}_{\theta_s} + \underbrace{\delta T_s + \boldsymbol{z}_{is}\boldsymbol{\beta}}_{\boldsymbol{x}_{is}} + \epsilon_{is}, \quad i = 1, \ldots, n, \;\; s = 1, \ldots, S \tag{2}$$

where T_s is a binary indicator equal to one if the deworming treatment is assigned to all pupils in school s, $T_{s\bar{d}}^N$ is the total number of pupils in treated schools within geographical distance $\bar{d}$ of school s, and $\gamma_{\bar{d}}$ are the unknown spillover effects. Hence, the latter are assumed to be homogeneous across pupils but heterogeneous across distances. According to our setting, we postulate for each school s the existence of a neighborhood, B_s^d, based on a general distance function, $d_{sj} = g(T^N, \bar{d})$ $\forall j \neq s$ and a threshold d, such that $\theta_s \approx \theta_j$, $\forall j \in B_s^d$. This distance is specified by comparing the number of treated pupils between schools weighted by their geographical distance. The idea is to pair similar schools in terms of spillover effects so that if they are smooth enough, a neighborhood data transformation exists to rule them out. See Section 6 for further details.

A different setting where model (1) can be applied is when unobservables are location or site-specific (Duranton et al., 2011, Belotti et al., 2021). In this cases, the model for the outcome can be represented as

$$y_{iza} = \theta_{iz} + \gamma_a + \boldsymbol{x}_{iza}\boldsymbol{\beta} + \epsilon_{iza}, \quad i = 1, \ldots, n, \quad a = 1, \ldots, m, \quad z = 1, \ldots, Z, \tag{3}$$

where θ_{iz} is a local effect that is assumed to change smoothly over the unobserved location z, defined at a finer spatial scale than the observed administrative location a. As before, we assume the existence of a neighborhood for each unit i defined according to a geographical distance threshold d, B_{iz}^d, where $\theta_{iz} \approx \theta_{jz}$, $\forall j \in B_{iz}^d$.

3. Estimation based on neighborhood data transformations

Let Δ^d be the neighborhood-difference operator. Let $N^d = \frac{\sum_i (n_i^d - 1)}{2}$ be the total number of pairs in the sample. 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. $\Delta_{ij}^d \boldsymbol{x}_i = \boldsymbol{x}_i - \boldsymbol{x}_j$, Alternatively, the within-neighborood transformation takes the difference between unit i and the average $\bar{\boldsymbol{x}}_i = \frac{1}{n_i^d} \sum_{j \in B_i^d} \boldsymbol{x}_j$ computed considering all units $j \in B_i^d$, i.e. $\Delta_{i,n_i}^d \boldsymbol{x}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}_i$.

Let $\sum_{i<j}$ be shorthand for $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n}$. Then, the class of ND estimators indexed by the distance threshold d can be defined as

$$\hat{\boldsymbol{\beta}}_{ND}^d = \left[\sum_{i<j, j \in B_i^d \setminus i} \Delta_{ij}^d \boldsymbol{x}_i' \Delta_{ij}^d \boldsymbol{x}_i \right]^{-1} \sum_{i<j, j \in B_i^d \setminus i} \Delta_{ij}^d \boldsymbol{x}_i' \Delta_{ij}^d y_i. \tag{4}$$

Similarly, we can define the class of NW estimators as

$$\hat{\boldsymbol{\beta}}_{NW}^d = \left[\sum_{i=1}^{n} \Delta_{i,n_i}^d \boldsymbol{x}_i' \Delta_{i,n_i}^d \boldsymbol{x}_i \right]^{-1} \sum_{i=1}^{n} \Delta_{i,n_i}^d \boldsymbol{x}_i' \Delta_{i,n_i}^d y_i. \tag{5}$$

Let us now state the main assumptions needed to study the asymptotic behavior of the proposed estimators as $n \to \infty$. To ease the notation, in what follows, we remove the d superscript unless we want to stress the dependence on d.

Assumption 1. *There exists a neighborhood $B_i^{d^*}$ such that $\Delta_{ij}^{d^*} \theta_i \xrightarrow{p} 0$ (or $\Delta_{in_i}^{d^*} \theta_i \xrightarrow{p} 0$) as $d^* \to 0$ uniformly in $i = 1, \ldots, n$, and $j \in B_i^{d^*} \setminus i$.*[3]

Assumption 2. *(i) $\{\Delta_{ij} \boldsymbol{x}_i, \Delta_{ij} \epsilon_i\}$ are identically distributed across $i = 1, \ldots, n$ and each $j \in B_i \setminus i$. $\{\Delta_{ij} \boldsymbol{x}_i, \Delta_{ij} \epsilon_i\}$ and $\{\Delta_{kl} \boldsymbol{x}_k, \Delta_{kl} \epsilon_k\}$ are independent for each j and l if $B_i \cap B_k = \emptyset$ with $j \in B_i \setminus i$ and $l \in B_k \setminus k$; (ii) $\{\Delta_{in_i} \boldsymbol{x}_i, \Delta_{in_i} \epsilon_i\}$ are identically distributed across $i = 1, \ldots, n$. $\{\Delta_{in_i} \boldsymbol{x}_i, \Delta_{in_i} \epsilon_i\}$ and $\{\Delta_{kn_k} \boldsymbol{x}_k, \Delta_{kn_k} \epsilon_k\}$ are independent for each i and k if $|B_i \cap B_k| \leq c_n$, where $c_n \geq 0$ is a data-driven constant.*

Assumption 1 is crucial for the consistency of the estimators defined in (4) and (5). The key ingredient is indeed the specification of a shrinking neighborhood such that the resulting data transformation can approximately rule out NUH for an arbitrary small d^*. This is true if, for example, 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 $|\Delta_{in_i}^{d^*} \theta_i| < \delta$) for each unit. Duranton et al. (2011) impose this continuity condition in the specific case of geographical proximity. Assumption 1 extends this deterministic condition to stochastic continuity. We define the implied dependence structure in Assumption 2. In particular, while we assume that the transformed observations are identically distributed, the ND transformed observations for which the neighborhoods overlap are dependent. For the NW transformation, observations are dependent only if the share of overlapping units is large enough. The importance of the data-driven constant c_n for the asymptotic properties of the NW estimator is clarified in Assumption 7. It is worth noting that if $B_i \cap B_k = \emptyset \; \forall \; i \neq k$, then neighborhoods do not overlap, and inference can be conducted using standard tools for clustered data (Hansen and Lee, 2019).

[3] We define $\Delta_{ij}^{d^*} \theta_i \xrightarrow{p} 0$ when $d^* \to 0$ as $\forall \epsilon > 0$, $\lim_{d^* \to 0} Pr(|\Delta_{ij}^{d^*} \theta_i| > \epsilon) = 0$.

To study the asymptotic distribution of this class of estimators we follow the inferential approach for dyadic data proposed by Tabord-Meehan (2019).[4] The author studies the asymptotic properties of the OLS estimator for dyadic data using the CLT for dependency graphs introduced by Janson et al. (1988). Intuitively, a dependency graph is defined in our framework as a family of random variables with vertex set $\{\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\epsilon_i,\, i=1,\dots n, \text{ and } j\in B_i\setminus i\}$ ($\{\Delta_{in_i}\boldsymbol{x}_i,\Delta_{in_i}\epsilon_i,\, i=1,\dots n,\}$) and edges formed by vertexes with overlapping neighborhoods (more specifically, random variables that are dependent according to Assumption 2). The dependency graph definition and further details on the Janson et al. (1988)' CLT can be found in the Appendix A. Let $\sigma_n^2 = var(\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}x_i\Delta_{ij}\epsilon_i)$ or $\sigma_n^2 = var(\sum_i \Delta_{in_i}x_i\Delta_{in_i}\epsilon_i)$. We refer to the arguments of the $var()$ operator as OLS key statistics. For the ND and NW estimator respectively, the key condition to apply the Janson et al. (1988)' CLT is

$$\frac{\left(\frac{\#\text{ obs}}{M^u}\right)^{1/l} M^u}{\sigma_n} \to 0 \;\; as\, n\to\infty \text{ for some } \; l\geq 3, \tag{6}$$

where $\#obs$ equal to N or n, and M^u is a measure of the maximum level of dependence across observations defined as

$$M^u_{ND} = \max_i |\{k:\, B_i\cap B_k\neq\emptyset\}|, \quad M^u_{SW} = \max_i |\{k:\, B_i\cap B_k > c_n\}|.$$

Therefore, to satisfy condition (6) the ratio between M^u and σ_n has to be controlled (as n gets larger). We use M^u without subscript when it is not important to refer to a specific estimator. Observe that $M^u\leq n-1$. Similarly, the minimum level of dependence across observations is

$$M^l_{ND} = \min_i |\{k:\, B_i\cap B_k\neq\emptyset\}|, \quad M^l_{SW} = \min_i |\{k:\, B_i\cap B_k > c_n\}|.$$

We consider two asymptotic approximations: (*i*) sparseness, and (*ii*) denseness. Under (*i*) the maximum number of overlapping neighborhood is bounded as $n\to\infty$, while under (*ii*) M^u is allowed to increase with n.

Assumption 3 (Sparseness). *$M^u < c$ for some constant $c>0$ as $n\to\infty$.*

Assumption 4 (Denseness). *$M^l_{ND}\geq c\,n$ for some constant $c>0$ as $n\to\infty$.*

Assumption 5. *$\{x_i,\epsilon_i\}_{i=1}^n$ have uniformly bounded support for all n.*

Assumption 3 is the same as Assumption AF1 in Tabord-Meehan (2019), and implies that the resulting dependency graph for the OLS key statistics is sparse. For the denseness case, we assume that the minimum number of overlapping neighborhoods across units grows with n. This Assumption is needed to ensure a proper rate of growth for σ_n, which then satisfies the CLT condition (A.1) in Appendix A.[5] To make the derivation of the asymptotic properties manageable, Assumption 5 imposes a bounded support

[4] See also Fafchamps and Gubert (2007), Cameron and Miller (2014), Aronow et al. (2015) for other works on dyadic-robust inference.

[5] For more details about the dense asymptotic approximation, see Section 2.3 in Tabord-Meehan (2019).

for explanatory variables and errors. This implies that also $\{\Delta_{ij}x_i, \Delta_{ij}\epsilon_i\}$ and $\{\Delta_{in_i}x_i, \Delta_{in_i}\epsilon_i\}$ have bounded support for each $i = 1, \ldots, n$.[6]

3.1. Examples

We now discuss two popular settings in which the nature of the data allows to specify neighborhoods such that the proposed estimators can be implemented.

3.1.1. Geo-coded data

With geo-coded, units are located on a (possibly) unevenly spaced lattice $D \subset \mathbb{R}^{d_0}$, $d_0 \geq 1$, infinitely countable. The neighborhood, $B^d_{\ell(i)}$, may depend on a given geographical distance threshold d, where $\ell(i)$ denotes the location of unit i. In this setting, two location units $\ell(i)$ and $\ell(j)$ are considered neighbors if their pairwise distance $d_{\ell(i)\ell(j)}$ is below the distance threshold d. Under sparseness, i.e. increasing domain asymptotics, a sufficient condition for having $n_{\ell(i)}$ bounded is given by Jenish and Prucha (2009):

Condition 1. *The lattice $D \subset \mathbb{R}^{d_0}$, with $d_0 \geq 1$, is infinitely countable. 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.*

This condition imposes a minimum distance requirement and implies that, for any distance threshold d, there are at most kd^{d_0} points in $B_{\ell(i)}$ and at most kd^{d_0-1} points in the space $B^{d+1}_{\ell(i)} \setminus B^d_{\ell(i)}$, where $k > 0$ is a constant (see Lemma A.1 in Jenish and Prucha, 2009).[7] Hence, the maximum degree of overlapping neighborhoods $M^u < c$ for some constant $c > 0$ as $n \to \infty$ and Assumption 3 holds. Consequently, the sufficient key condition 6 for the CLT to hold becomes

$$\frac{\left(\frac{\#\text{ obs}}{kd^{d_0}}\right)^{1/l} kd^{d_0}}{\sigma_n} \to 0 \;\; as\, n \to \infty \text{ for some } \; l \geq 3, \tag{7}$$

where k and d_0 are constants, $\#obs$ is equal to N or n for ND and NW, respectively, and d is the distance threshold used to characterize the neighborhoods.

3.1.2. Network data

With (social) network data, the researcher observes a set of units i, $i = 1, \ldots, n$, each of which is supposed to have a reference group, i.e. peers group, B_i. An example can be students interacting in a classroom. The degree n_i, i.e. the number of peers of unit i in a given classroom, is finite so that Assumption 3 holds. The non-stochastic interaction matrix G represents the social network within the class, where $g_{ij} = 1$ if j is friend of i, 0 otherwise. Then, M^u represents the maximum number of common friends between two students. This can be computed as $\max_i(\sum_{j \neq i} 1[g^{[2]}_{ij} > 0] + n_i)$ where $g^{[2]}_{ij}$ denotes the elements of the G^2 matrix. The elements of G^2 can be interpreted as the number of transitive

[6] Notice that Assumption 5 can be relaxed by adding some moment condition on (x_i, ϵ_i).
[7] Additional details about set cardinalities for irregular lattices can be found in Lemma 1, Appendix A.

triads between two units. In other words, both the number of first and second order connections are considered to get M^u.

In this setting, Auerbach (2022) proposes a pairwise difference estimator for the more general partially linear model. The outcome is assumed to be generated by an unknown function of unit-specific unobservables that drives the linking behavior in a network. The proposed nonparametric pairwise difference estimator removes the unobserved social influence for those units with similar linking behaviors. The author shows that, for a large class of nonparametric network formation models, the latter can be recovered from the columns of G^2. Differently from Auerbach (2022), we impose a higher level condition in Assumption 1 and that the unobservables enter model (1) linearly and additively. However, we do not impose any specific neighborhood formation model.

3.2. Asymptotic properties

Given its popularity in applied works (see Section 1), we start by investigating the properties of the ND estimator. While the main asymptotic results are a particular case of the OLS estimator for dyadic data (Tabord-Meehan, 2019), to the best of our knowledge, we are the first to investigate the asymptotic properties of the NW estimator in the presence of overlapping clusters. We report the proofs for all the propositions in Appendix A.

3.2.1. The neighborhood-difference estimator

We start by assuming the rate of growth of the key statistics' variance for the two considered asymptotic approximations.

Assumption 6. a) *Under **sparseness***

$$\boldsymbol{B} = \lim_{n\to\infty} \frac{1}{N} var\left(\sum_{i<j,j\in B_i\setminus i} \Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\epsilon_i\right) = \frac{1}{N}\lim_{n\to\infty}\sum_{i<j,j\in B_i\setminus i}\sum_{k<l\in B_k\setminus k} cov(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\epsilon_i, \Delta_{kl}\boldsymbol{x}_k\Delta_{kl}\epsilon_k)$$

is positive definite.

b) *Under **denseness***

$$\boldsymbol{B} = \lim_{n\to\infty} \frac{1}{Nn^r} var\left(\sum_{i<j,j\in B_i\setminus i} \Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\epsilon_i\right) = \frac{1}{Nn^r}\lim_{n\to\infty}\sum_{i<j,j\in B_i\setminus i}\sum_{k<l\in B_k\setminus k} cov(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\epsilon_i, \Delta_{kl}\boldsymbol{x}_k\Delta_{kl}\epsilon_k)$$

is positive definite for some $r \in [0, 1]$.

These assumptions are related to the ones made on the variance of the key statistics in the dyadic robust literature (Aronow et al., 2015, Cameron and Miller, 2014). For an exhaustive discussion about the relationship between σ_n^2 in clustered and dyadic data refer to Tabord-Meehan (2019). We are now ready to characterize the asymptotic distribution of the ND estimator under the two asymptotic approximations.

Proposition 1. a) *Under Assumptions 1, 2, 3, 5, and 6* a)*,*

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

where $\boldsymbol{A} = E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\boldsymbol{x}_i')^{-1}$ *and* $\boldsymbol{B}$ *as in Assumption 6* a).
b) *Under Assumptions 1, 2, 4, 5, and 6* b) *with* $r > 0$

$$\sqrt{\frac{N}{n^r}}(\hat{\boldsymbol{\beta}}_{ND} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}), \tag{9}$$

where $\boldsymbol{A} = E(\Delta_{ij}\boldsymbol{x}_i\Delta_{ij}\boldsymbol{x}_i')^{-1}$ *and* $\boldsymbol{B}$ *as in Assumption 6* b).

Under denseness, the smaller is r, the closer we are to the sparse asymptotic approximation characterized by less dependence between observations. In matrix notation, we can write the ND estimator as

$$\hat{\boldsymbol{\beta}}_{ND} = (\boldsymbol{X}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{y}) \tag{10}$$

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

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

3.3. The within-neighborhood estimator

The denseness assumption must be strengthened for the NW estimator to decrease dependence on the OLS key statistics. This is because the rate of growth of the variance, i.e., n^{r+1}, is the same as M^u. The sample size is n (not N), so we cannot exploit the Nn^r rate of growth assumed for σ_n^2 in sub-section 3.2.1. In particular, to satisfy condition 6, we make the following additional assumption to restrict the growth rate of M_{NW}^u.

Assumption 7 (denseness). $\frac{M_{NW}^u}{n} \to 0$ *as* $n \to \infty$.

Assumption 8. a) *Under* ***sparseness***

$$\boldsymbol{B}_W = \lim_{n\to\infty} \frac{1}{n} var(\sum_i \Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\epsilon_i) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n} cov\Big(\Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\epsilon_i, \Delta_{kn_k}\boldsymbol{x}_k\Delta_{kn_k}\epsilon_k\Big)$$

is positive definite.
b) *Under* ***denseness***

$$\boldsymbol{B}_W = \lim_{n\to\infty} \frac{1}{n\,n^r} var(\sum_i \Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\epsilon_i) = \lim_{n\to\infty} \frac{1}{n\,n^r}\sum_{i=1}^{n}\sum_{k=1}^{n} cov\Big(\Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\epsilon_i, \Delta_{kn_k}\boldsymbol{x}_k\Delta_{kn_k}\epsilon_k\Big)$$

is positive definite for some $r \in [0,1]$.

Observe that also the constant c_n in Assumption 2 must be large enough so that M_{NW}^u grows slower than n. We are now ready to characterize the asymptotic distribution of the NW estimator.

Proposition 2. a) *Under Assumptions 1, 2, 3, 5, and 8* a)*,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{NW} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{A}_W^{-1}\boldsymbol{B}_W\boldsymbol{A}_W^{-1}), \tag{11}$$

where $\boldsymbol{A}_W = E(\Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\boldsymbol{x}_i')^{-1}$ *and* $\boldsymbol{B}_W$ *as in Assumption 6* a)*.*
b) *Under Assumptions 1, 2, 7, 5, and 8* b) *with* $r \in (0, 1)$

$$\sqrt{\frac{n}{n^r}}(\hat{\boldsymbol{\beta}}_{NW} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{A}_W^{-1}\boldsymbol{B}_W\boldsymbol{A}_W^{-1}), \tag{12}$$

where $\boldsymbol{A}_W = E(\Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\boldsymbol{x}_i')^{-1}$ *and* $\boldsymbol{B}_W$ *as in Assumption 6* b)*.*

Unlike the ND estimator under denseness, we need to exclude in (12) the case in which $r = 1$; otherwise, the asymptotic distribution is degenerate. The NW transformation can be implemented using a within-neighborhood matrix $\boldsymbol{W}_n$ defined as

$$\boldsymbol{W}_n = \boldsymbol{I}_n - \boldsymbol{C}_n,$$

where $\boldsymbol{C}_n$ is a $n \times n$ matrix obtained in two steps:

1. create a (binary) interaction matrix according to the neighborhoods specification;
2. substitute the main diagonal with ones and row-normalize.

We use the subscript n to stress that, in general, $\boldsymbol{W}_n$ is a triangular array. In matrix notation we have

$$\hat{\boldsymbol{\beta}}_{NW} = (\boldsymbol{X}'\boldsymbol{W}_n'\boldsymbol{W}_n\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{W}_n'\boldsymbol{W}_n\boldsymbol{y}). \tag{13}$$

Remark 1 (Non-overlapping neighborhoods). *If neighborhoods do not overlap,* $\boldsymbol{W}_n$ *is a projector and we have a formal similarity with panel 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).*

3.4. Estimation of the variance covariance matrix

3.4.1. Homoskedastic case

When errors in model (1) are homoskedastic, i.e. $\mathbf{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\boldsymbol{I}_n$, then a consistent estimator of the asymptotic variance of the ND estimator is

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

with

$$\begin{aligned}\hat{\boldsymbol{B}} &= \boldsymbol{X}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{D}'\boldsymbol{D}\boldsymbol{X},\\ \hat{\boldsymbol{A}} &= \boldsymbol{X}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{X},\end{aligned}$$

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

$$\hat{\sigma}^2 = \left[tr(\boldsymbol{D}\boldsymbol{D}') - tr(\hat{\boldsymbol{A}}\hat{\boldsymbol{B}})\right]^{-1} \sum_{i<j,j\in B_i\setminus i} \Delta_{ij}\hat{\epsilon}_i\Delta_{ij}\hat{\epsilon}_i. \tag{15}$$

Similarly, for the NW estimator we have

$$\hat{\boldsymbol{V}}_{NW} = \hat{\sigma}_2\hat{\boldsymbol{A}}_W^{-1}\hat{\boldsymbol{B}}_W\hat{\boldsymbol{A}}_W^{-1}, \tag{16}$$

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

$$\hat{\sigma}^2 = \left[tr(\boldsymbol{W}_n\boldsymbol{W}_n')\right]^{-1} \sum_{i=1}^{n} \Delta_{i,n_i}\hat{\epsilon}_i\Delta_{i,n_i}\hat{\epsilon}_i. \tag{17}$$

3.4.2. General case

In this case, an estimator of the asymptotic variance of the ND estimator is

$$\hat{\boldsymbol{V}}_{ND} = \hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{A}}^{-1} \tag{18}$$

where

$$\begin{aligned}\hat{\boldsymbol{B}} &= \sum_i\sum_k\sum_{j\in B_i\cap B_k} \mathbf{1}_{ij,kj\neq ji,jk}\Delta_{ij}\hat{\epsilon}_i\Delta_{kj}\hat{\epsilon}_k\Delta_{ij}\boldsymbol{x}_i'\Delta_{kj}\boldsymbol{x}_k\\ \hat{\boldsymbol{A}} &= \sum_{i<j,j\in B_i\setminus i} \Delta_{ij}\boldsymbol{x}_i'\Delta_{ij}\boldsymbol{x}_i,\end{aligned}$$

and $\mathbf{1}_{ij,kj\neq ji,jk}$ is the indicator of the event $\Delta_{ij}\Delta_{kj} \neq \Delta_{ji}\Delta_{jk}$. 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 (18) is consistent following Proposition 3.3 in Tabord-Meehan (2019). The finite sample properties of this estimator has been studied in Belotti et al. (2018) who show, through Monte Carlo simulations, that (18) outperforms other robust competitors proposed in the related literature (see, e.g. Fack and Grenet, 2010).

Similarly, we can write the asymptotic variance of the NW estimator as

$$\hat{\boldsymbol{V}}_{NW} = \hat{\boldsymbol{A}}_W^{-1}\hat{\boldsymbol{B}}_W\hat{\boldsymbol{A}}_W^{-1}, \tag{19}$$

where $\hat{\boldsymbol{A}}_W = \sum_{i=1}^{n} \Delta_{in_i} \boldsymbol{x}_i' \Delta_{in_i} \boldsymbol{x}_i$. We propose the following estimator for $\hat{\boldsymbol{B}}_W$

$$\hat{\boldsymbol{B}}_W \quad = \quad \sum_i \sum_k \mathbf{1}_{[(k \in B_i) \vee (i \in B_k) \vee (|B_i \cap B_k > c_n|)]} \Delta_{in_i} \hat{\epsilon}_i \Delta_{kn_k} \hat{\epsilon}_k \Delta_{in_i} \boldsymbol{x}_i' \Delta_{kn_k} \boldsymbol{x}_k,$$

where the indicator function selects units within the same neighborhood and units for which there is a relevant degree of overlapping between neighborhoods, and c_n is a data-driven constant that defines the degree of neighborhoods overlapping. We conjecture that this estimator is consistent for $\boldsymbol{V}_{NW}$ under an appropriate selection rule for c_n. We left this issue for future research.

4. Testing the neighborhood specification

Assumption 1 is a strong requirement for the consistency of the proposed estimators. To test for the null hypothesis of smooth NUH, i.e. Assumption 1 holds, we propose a Hausman-like test based on the contrast between ND and NW estimators that are both inconsistent under the alternative. This 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. If the NUH is ruled out by the ND/NW transformations (H_{0d}), then both estimators are consistent and $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_{ND} - \hat{\boldsymbol{\beta}}_{NW} \xrightarrow{p} 0$. On the other hand, when the transformations cannot entirely remove NUH because the latter is not smooth enough or the neighborhood selection given the threshold is not optimal, the two estimators generally converge in probability to different points in the parameter space. Thus, $\sqrt{\tau_n}\hat{\boldsymbol{\delta}} \xrightarrow{d} \mathcal{N}(\boldsymbol{\delta}, \boldsymbol{V}_\delta)$, where $\boldsymbol{\delta} \neq 0$, τ_n is the rate of convergence for the NW estimator (it is slower than the ND estimator), and $\boldsymbol{V}_\delta$ may be singular. Because of the asymptotic normality of our estimators, the Hausman-like test statistic

$$\xi = \tau_n \hat{\boldsymbol{\delta}}^T \widehat{\boldsymbol{V}}_\delta^- \hat{\boldsymbol{\delta}}, \tag{20}$$

has approximately a χ^2 asymptotic distribution with number of degrees of freedom equal to the rank of $\widehat{\boldsymbol{V}}_\delta^-$. Observe that $\widehat{\boldsymbol{V}}_\delta^- = \hat{\boldsymbol{V}}_{ND} + \hat{\boldsymbol{V}}_{NW} - \hat{\boldsymbol{V}}_{ND,NW} - \hat{\boldsymbol{V}}_{ND,NW}'$, where $\hat{\boldsymbol{V}}_{ND,NW} = \hat{\boldsymbol{A}}_{NW}^{-1} \hat{\boldsymbol{B}}_{ND,NW} \hat{\boldsymbol{A}}_{ND}^{-1}$, with

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

where $\hat{\sigma}^2$ can be estimated either with (15) or (17), and $\hat{\boldsymbol{V}}^-$ is a generalized inverse. The proposed test has no power when *i*) the ND and NW estimators are algebraically equivalent, i.e. when the number of neighbors is the same for each i and all neighborhood are non-overlapping (see Remark 1); *ii*) the covariate(s) of interest is not correlated with the unobserved effects.

The power of our test depends on the divergence between ND and NW estimators. Hence, in this Section we study their asymptotic bias for a specific set of alternatives. We assume that

$$y_i = \theta_i + x_i \beta + \epsilon_i, \quad i = 1, \ldots, n, \tag{21}$$

$x_i = \phi\theta_i + (1 - \phi^2)^{1/2} \xi_i$, where the ϵ_i and the ξ_i are i.i.d. r.v. with zero mean and unit variance and

$u_i = \theta_i + \epsilon_i$.

Given a distance threshold $d \neq d^*$, the ND and NW estimation errors can then be written as

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

where $\tilde{x}_i = \Delta_{i,n_i} x_i$ and $\Delta x_{ij} = \Delta_{ij} x_i$.
As $n \to \infty$, the ND and NW asymptotic biases are[8]

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

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 $\tilde{\tau} = E(\theta_i - \bar{\theta}_i)^2$ and $\Delta\tau = E(\theta_i - \theta_j)^2, \forall\, j \in B_i$. The estimator asymptotic biases become

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

To obtain sharper results, we characterize the data generating process (DGP) of the NUH. Let us assume that the natural logarithm of the proximity between agents is a weighted sum between two components to ease the derivation. The first represents the difference between unobservables, and the second is a random dyadic shock. Formally, the distance model is the following

$$\bar{d}_{ij} = \alpha\,(\theta_i - \theta_j) + (1-\alpha^2)^{1/2} \varepsilon_{ij} \tag{22}$$

where $\bar{d}_{ij}$ is the natural logarithm of the distance, ε_{ij} is iid across dyads with mean zero, unit variance and uncorrelated with the θs. When $\alpha = 1$ the logarithm of the distance between i and j is equal to the difference between θ_i and θ_j. In this case, we are under H_{0d} if for a small distance threshold d^* the neighborhood specification is such that $\theta_i = \theta_j\ \forall i = 1, \ldots, n$, and $j \in B_i^{d^*} \setminus i$. When $\alpha = 0$ the distance between unit i and j is random and uncorrelated to their unobservables. The detailed derivation of the asymptotic biases under the distance model (22) and a different specification where we generate NUH following a treatment spillover model as in 2 can be found in Appendix A.2. Under the distance model (22), and conditioning on the event $A = 1(\bar{d}_{ij} < d)$, it can be shown that the asymptotic biases will converge to

$$\hat{\beta}_{ND} - \beta \xrightarrow{p} \frac{\phi \frac{1}{\alpha^2} \left\{ E[\bar{d}_{ij}^2 | A] - (1-\alpha^2) \right\}}{\phi^2 \frac{1}{\alpha^2} \left\{ E[\bar{d}_{ij}^2 | A] - (1-\alpha^2) \right\} + (1-\phi^2)},$$

[8] Convergence in probability of the numerator can be proved following the same argument used for the convergence of the denominators (A.3) and (A.6) in Appendix A.

$$\hat{\beta}_{NW} - \beta \xrightarrow{p} \frac{\phi \frac{1}{\alpha^2} \left\{ E\left[\left(n_i^{-1} \sum_{j \in B_i} \bar{d}_{ij} \right)^2 | A \right] - (1-\alpha^2) E\left[n_i^{-2} \sum_{j \in B_i} \varepsilon_{ij}^2 | A \right] \right\}}{\phi^2 \frac{1}{\alpha^2} \left\{ E\left[\left(n_i^{-1} \sum_{j \in B_i} \bar{d}_{ij} \right)^2 | A \right] - (1-\alpha^2) E\left[n_i^{-2} \sum_{j \in B_i} \varepsilon_{ij}^2 | A \right] \right\} + (1-\phi^2)},$$

where the conditional expectations are defined as

$$E[\bar{d}_{ij}^2 | A] = \frac{\sum_{d:\ i,j \in A} \bar{d}_{ij}^2 Pr(\bar{d}_{ij})}{Pr(\bar{d}_{ij} < d)},$$

$$E\left[\left(n_i^{-1} \sum_{j \in B_i} \bar{d}_{ij} \right)^2 \middle| A \right] = \frac{\sum_{d:\ i,j \in A} (n_i^{-1} \sum_{j \in B_i} \bar{d}_{ij})^2 Pr(\bar{d}_{ij})}{Pr(\bar{d}_{ij} < d)}.$$

Notice that if $\sup_{ij} Pr(\bar{d}_{ij} < d) = Pr(\sup_{ij} \bar{d}_{ij} < d) = 1$, then the two asymptotic biases converge to the asymptotic bias of OLS since all the units are in the same neighborhood. This occurs when we fix n, and we let d grow. See Section 5 for further details. On the other hand, if $Pr(\bar{d}_{ij} < d) = 0 \ \forall i, j$, then the conditional expectations are not defined. In general, $\Delta\tau$ increases in the conditional second moment of the logarithm of the distance (and decreases in α). However, the asymptotic bias will converge in probability, assuming that the second moment is bounded. Observe that the test has no power if $\phi = 0$ and when the conditional second moment of the distance equals zero and $\alpha = 1$. An example where the distance might not vary is when observations are clustered across space.[9] As for the NW estimator, $\tilde{\tau}$ depends on the conditional second moment of the logarithm of the average neighborhood distance, i.e., $\tilde{\tau}$ is a function of both distance and the number of neighbors.

To conclude, we suggest the following testing strategy to the practitioners.

Step 1 *Testing for correlated NUH:* we suggest to construct a test based on the contrast between OLS and NW estimators. This test resembles a standard Hausman test, which exploits the contrast between the random effects and within-group estimators in panel data setting. Indeed, If (*i*) the unobserved effects are absent, i.e., $\theta_i = 0, \forall i = 1, \ldots, n$ or (*ii*) $\theta_i \neq 0, \forall i = 1, \ldots, n$ but are uncorrelated with $\boldsymbol{x}_i$ (H_0), then both the OLS and NW estimators are consistent and $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\beta}}_{OLS} - \hat{\boldsymbol{\beta}}_{NW} \xrightarrow{p} 0$. In terms of efficiency, OLS is the most efficient estimator under (*i*) while the NW estimator is the most efficient under (*ii*). Since the OLS and NW estimators depend on different functions of the data, they generally converge in probability to different points in the parameter space in the presence of correlated NUH, unless $Pr(\sup_{ij} \bar{d}_{ij} < d) = 1$, i.e. the inconsistency of both estimators is the same. Thus, this test has power against a variety of alternatives resulting in smooth and non-smooth NUH, such as omitted neighborhood factors. Because of the asymptotic normality of our estimators, the statistic

$$\xi_{NUH} = \tau_n \hat{\boldsymbol{\lambda}}^T \widehat{\boldsymbol{V}}_{\lambda}^{-} \hat{\boldsymbol{\lambda}},$$

[9] See Section 5 for a discussion about the test size when data are generated using a clustered design.

has approximately a centered χ^2 asymptotic distribution with the number of degrees of freedom equal to the rank of $\widehat{\boldsymbol{V}}_{\lambda}^{-} = \hat{\boldsymbol{V}}_{OLS} - \hat{\boldsymbol{V}}_{NW}$.

Step 2 *Testing for neighborhood specification/smooth unobservables:* If H_0 in step 1 is rejected by the data, we suggest implementing the test in (20) to select the optimal threshold d^* for the neighborhood specification. In this case, the class of hypotheses is index by the threshold d. Given the sequential nature of the Hausman-like tests, we suggest starting from a large enough distance threshold and sequentially decreasing the threshold until the H_{0d} hypothesis cannot be rejected by the data. 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.3, we formally link our testing strategy with the *testing hypotheses in order* framework in Rosenbaum (2008).

5. Monte Carlo evidence

We now present some Monte Carlo evidence about the finite sample properties of the ND and NW estimators and the size and power properties of the test statistic in (20).[10]

5.1. Location-specific unobserved effects

We consider two DGPs. The first represents the case where local unobservables are the same within small non-overlapping neighborhoods (DGP_1). In this case, we can determine *a priori* which is the threshold distance that allows removing the local unobserved heterogeneity completely. The second represents a more general case where the unobservables are unit-specific, and the optimal distance is unknown (DGP_2). Our data generating process for the outcome variable is

$$y_{i,b} = \theta_{i,b} + x_{i,b}\beta + \epsilon_{i,b}, \quad i = 1, \dots, n, \tag{23}$$

where $b = 1, \dots, B$ denotes the neighbourhood and $\theta_{i,b} \sim \lfloor \mathcal{U}(1, 10) \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part. We then add the following model that links the distance between units to their unobservables

$$d_{ij} = \alpha \left| \theta_{i,b} - \theta_{j,b} \right| + (1 - \alpha^2)^{1/2} \varepsilon_{ij} \tag{24}$$

where $\alpha \in (0, 1)$ and $\varepsilon_{ij} \sim \mathcal{N}^{+}(0, 1)$ is i.i.d. across dyads. In the first DGP $\alpha = 1$. This implies that the distance between units i and j is equal to the euclidean distance between their θs and that $|\theta_{i,b} - \theta_{j,b}| = 0$ if units i and j are in the same neighborhood b. DGP_2 considers a design without clusters (the subscript b is removed) where the θs are unit-specific, and set $\alpha = 0.5$. To allow for dependence between the unobservables θ_i and the regressor, we generate the latter as $x_i = \phi\theta_i + (1 - \phi^2)^{1/2} z_i$, where the z_i are i.i.d. standard Gaussian. As aforementioned, with DGP_1 we control the optimal threshold d^*, i.e. the distance allowing to fully remove local unobservables; setting $\alpha = 1$ implies that $d^* = 1$.

[10] We present the results of the simulation studies in the main text using Figures 2 - 4. The corresponding tabulations are reported in the Supplementary Appendix (Tables S.1 - S.3).

For DGP_1 we consider three sample sizes n = 250, 500 and 1000, $\beta = 1$, $\phi = 0.5$ (1,000 replications per experiment). We set B = 50, 100 and 500, respectively (i.e., five units per neighborhood), $\alpha = 1$ and consider ten different distance thresholds, $d = 1, 2, \ldots, 10$ for defining the neighboroods. In DGP_2 we set n = 250, 500, 1000, and 2000, $\alpha = 0.5$ resulting in a continuous support for d_{ij}. It is worth noting that, in general, setting a small distance threshold can be effective in removing the NUH, but this may come at the cost of reducing the sample, causing estimators' efficiency losses.

Figure 2 and 3 report the results.The top panel shows the ND and NW estimates by percentiles of d_{ij}, the latter used to define the neighborhood data transformations (NDT). Here, the green dotted line is the true direct effect δ and the red dotted line is the biased OLS estimate. The bottom panel depicts the power of the Hausman-like test again by percentiles of d_{ij}. The dotted line represents the nominal size α= 0.05. In line with the asymptotic behavior of ND and NW estimators (see Sections 3.2.1), both estimators are consistent under the H_{0d}, i.e. when $d = 1$ (DGP_1) or d is between the first and the fourth percentile of d_{ij} (DGP_2). As expected, when the percentile of d_{ij} grows, the two estimators start diverging, making a test based on their difference powerful. This is confirmed by the results in the bottom panel, which show tiny size distortions even though the test exhibits a slight tendency to over-reject when n is small. On the other hand, the test is very powerful with just small deviations from the optimal distance, i.e. moving from d=1 to d=2 (DGP_1) or from d=0.5 to d=1 (DGP_2). As n gets larger, estimates start diverging much earlier, i.e., the power of the test increases for smaller percentiles as expected. When all the observations are used as neighbors (i.e., we use the last percentile of d_{ij} as the threshold for the NDT), the ND and NW estimators converge to OLS, and the power drops to zero.

5.2. Treatment spillover effects

The outcome is generated following the treatment spillover model interpretation is Section 2

$$y_i = \theta_i + T_i\delta + x_i\beta + \epsilon_i, \quad i = 1, \ldots, n, \tag{25}$$

where $x_i \sim \mathcal{N}(0, 1)$ is an exogenous explanatory variable, $T_i = 1(x_i > 0)$ is a binary treatment indicator, $\theta_i = \sum_{j \neq i \in C_i} T_j \frac{\gamma}{\bar{d}_{ij}}$ with C_i^r denoting the ball with radius r and the distance $\bar{d}_{ij} = 1 + |x_i - x_j|$ a function of the distance between i and j in terms of the explanatory variable, and $\epsilon_i \sim \mathcal{N}(0, 0.25)$ the idiosyncratic error. This setting ensures that $Corr(T_i, \theta_i) >> 0$, i.e. the spillover effects θ_i are correlated with the treatment.[11] We then define $D^r = [d_{ij}]$, a distance matrix based on the number of treated units within the r-ball with $d_{ij}^r = |N_{ri}^T - N_{rj}^T|$ and $N_{ri}^T = \sum_{j \neq i \in C_i^r} \frac{T_j}{\bar{d}_{ij}}$. We set, $r = 3$, $\delta = \beta = 0.5$, and we run three experiments with n = 2300, 3000 and 6000. See Section 6 for ND and NW estimation of model (25) using real data from Miguel and Kremer (2004).[12] [13]

Figure 4 displays our results.The top panel shows the ND and NW estimates by the percentiles of

[11] This correlation range from about 0.5 to 0.7 in our simulations.

[12] We set n=2300 following the number of observations in the empirical application, i.e., 2,328 pupils.

[13] We perform additional Monte Carlo experiments for DGP (25): we consider a generalized distance definition equivalent to the empirical application, i.e., $d_{ik}^r = |N_{ri}^T - N_{rk}^T|\bar{d}_{ik}$ with $N_{ri}^T = \sum_{j \neq i \in C_i^r} T_j$; we set the true δ to 0.25, that is the benchmark direct treatment effect of the deworming program (see Section 6 for details). Results remain qualitatively unchanged.

the distance (d_{ij}^r), which, as before, have been used to define the neighborhoods for applying the NDT. Here, the green dotted line is the actual direct effect δ, and the red dotted line is the biased OLS estimate, obtained using a model in which the spillover effects are not controlled for. The bottom panel depicts the power of the Hausman-like test again by the percentiles of the distance. The dotted line represents the nominal size α= 0.05. Figure 4 shows that the ND and NW estimates start diverging as we increase the threshold distance percentiles, i.e., the test becomes powerful. As the sample size increases, estimates start diverging earlier. The power curve follows the same pattern - the test is consistent for smaller percentiles of d_{ij}^r. When all the observations are used as neighbors (i.e., the last distance percentile is used as threshold), the ND and NW estimators converge to OLS, and the power drops to zero.

6. Empirical application: Miguel and Kremer (2004)

Health programs can generate indirect effects on untreated units in the presence of externalities. This section illustrates how practitioners can exploit NDT to estimate the direct effect of a treatment ruling out externalities. We use data from the Primary School Deworming Project conducted in western Kenya (Miguel and Kremer, 2004).[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 (Miguel and Kremer, 2004).

An essential feature of the deworming program is that the randomization takes place at the school level allowing to identify the direct effect of deworming even in the presence of externalities. As pointed out by Miguel and Kremer (2004), 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., spillovers across schools.

The authors consider the following specification

$$y_{is} = \alpha + \beta T_s + \underbrace{\sum_{\bar{d}} (\gamma_{\bar{d}} N_{\bar{d}s}^T)}_{\text{observed spillover effect}} + \sum_{\bar{d}} (\phi_{\bar{d}} N_{\bar{d}s}) + \boldsymbol{x}_{is}\boldsymbol{\delta} + \epsilon_{is} \tag{26}$$

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, $N_{\bar{d}s}$ the total number of pupils in primary schools at geographical distance $\bar{d}$ from school s, and $N_{\bar{d}s}^T$ is the number of these pupils randomly assigned to the treatment, while $\boldsymbol{x}_{is}$ are school and pupil characteristics. Thus, the overall average treatment effect is: $\beta + \sum_{\bar{d}} (\gamma_{\bar{d}} \bar{N}_{\bar{d}s}^T)$, where $\bar{N}_{\bar{d}s}^T$ is the average number of treated school pupils located at distance $\bar{d}$ from school s. The sample includes 2,328 pupils and 49 schools, 25 assigned to the treatment. In the parlance of our model, model (26) refers to equation (2).

Table 1 replicates, using a linear probability model (LPM), the results on deworming health direct

[14] Here, we used the updated data downloaded from Edward Miguel's website. See the replication manual. A detailed description of the program can be found in Miguel and Kremer (2004).

effects and externalities across schools in the case of any moderate-heavy helminth infection.[15] It shows that the fraction of students with moderate to heavy infection is around 25% lower for treated schools than for the untreated ones. It is worth noting that while Miguel and Kremer (2004) report average marginal effects from a probit model, our LPM estimates well approximate their baseline results. However, the latter can be easily compared with ND and NW estimates. While one of the main objectives in Miguel and Kremer (2004) is to identify direct and indirect effects separately, we focus on estimating the direct effects by treating the spillover effects as local unobservables. The NDT for ND and NW estimators follows closely the settings in the Monte Carlo experiments for DGP (25) in Section 5. Consider schools s and k, we first count the number of treated pupils within $\bar{d}\,km$ from school s and k, i.e. $N^T_{\bar{d}s}$ and $N^T_{\bar{d}k}$, and then implement a generalized distance as

$$d_{sk} = |N^T_{\bar{d}s} - N^T_{\bar{d}k}|\bar{d}_{sk}, \tag{27}$$

where $\bar{d}_{sk}$ is the geographical distance in km between school s and k. Thus, the geographical distance amplifies the difference between the exposure functions of schools s and k. We use $\bar{d} = 6$ based on the evidence that medical treatment spillovers are likely to occur within a 6 km from the treated school (see Miguel and Kremer, 2004). We apply the ND and NW estimators using different distance thresholds d based on deciles of d_{sk}. Figure 5 displays our main results; the top panel shows the ND and NW estimates by distance deciles. Here, the green dotted line is the "benchmark" direct effect reported in Table 1 while, the red one, is the biased OLS estimates obtained using model (26) without including the terms $\sum_{\bar{d}}(\gamma_{\bar{d}}N^T_{\bar{d}s})$ and $\sum_{\bar{d}}(\phi_{\bar{d}}N_{\bar{d}s})$. Figure 5, bottom panel, depicts the values of the test statistic by deciles of d_{sk}, the dotted line represents the critical value (nominal size α= 0.1). We find that H_{0d}, i.e., Assumption 1 holds, is rejected before the 4th decile, the latter being the first decile where the two estimators converge to the true "benchmark" value.[16] As expected, when the test rejects H_{0d}, also the NW and ND direct effect estimates diverge. After the 6th decile of d_{sk}, the test statistics increase significantly up to the 9th decile where it suddenly drops because the ND and NW estimators converge to the biased OLS estimates. Figure 6 replicates the same exercise by setting $\bar{d} = 3$. Results remain qualitatively unchanged. When the ND and NW estimates are close to the true value at the third decile of d_{sk}, H_{0d} cannot be rejected from the data. As the ND and NW estimates start diverging, the test becomes powerful.

This evidence highlights the potential for the proposed tools to remove the bias due to spillovers when the "high" level randomization exploited in Miguel and Kremer (2004) (e.g., at school, firm, or county level) cannot be implemented, for example, due to budget constraints.

[15] We specifically refer to Table A7 in the 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 the schools.

[16] The test statistic is computed on β after that all the regressors have been "partialled-out".

7. Concluding remarks

Nowadays, the growing availability of richer and larger datasets allows for a broad definition of proximity-based neighborhoods. For example, networks data allows for the specification of neighborhoods based on social distance, geo-coded data can be used to define neighborhoods exploiting geographical distance, and large-scale surveys may provide neighborhood representations based on economic or cultural distances.

This article proposes estimation strategies based on neighborhood data transformations (NDT) for models with additively-separable smooth unobserved heterogeneity. We study the asymptotic properties of the neighborhood difference (ND) and within neighborhood (NW) class of estimators. Following the inferential framework for dyadic data proposed by Tabord-Meehan (2019), we model the dependence between observations induced by the NDT. We present a Hausman-like test for the null hypothesis of smooth neighborhood unobserved heterogeneity (NUH) based on the contrast between ND and NW estimators. We describe how to exploit the test to search for the optimal distance threshold to transform the data. Using Monte Carlo simulations, we evaluate the performance of the ND and NW estimators and the test as a function of the threshold distance. The ND and NW estimators approach consistency when the NUH is smooth, and the distance threshold is small enough. When the distance threshold increases, the estimators start diverging and the test becomes powerful.

We finally demonstrate the validity of our approach using data from the seminal paper by Miguel and Kremer (2004) on a health program in Kenya. While the authors exploit the school level randomization to identify the direct and indirect average treatment effects, we specify the model assuming that the cross-school externalities are unobservable. By applying the NDT based on the distance threshold suggested by our test, we rule out the externalities and retrieve the same direct causal effect in Miguel and Kremer (2004). The extensions of our framework to partially linear and nonlinear models are natural next steps.

References

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

Aronow, P. M., Samii, C., and Assenova, V. A. (2015). Cluster–robust variance estimation for dyadic data. *Political Analysis*, pages 564–577.

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.

Cameron, A. C. and Miller, D. L. (2014). Robust inference for dyadic data. Technical report, University of California at Davis.

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.

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

Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350.

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

Hansen, B. E. and Lee, S. (2019). Asymptotic theory for clustered samples. *Journal of Econometrics*, 210(2):268–290.

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

Holmes, T. J. (1998). The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders. *Journal of Political Economy, University of Chicago Press*, 106(4):667–705.

Janson, S. et al. (1988). Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *Annals of probability*, 16(1):305–312.

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.

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.

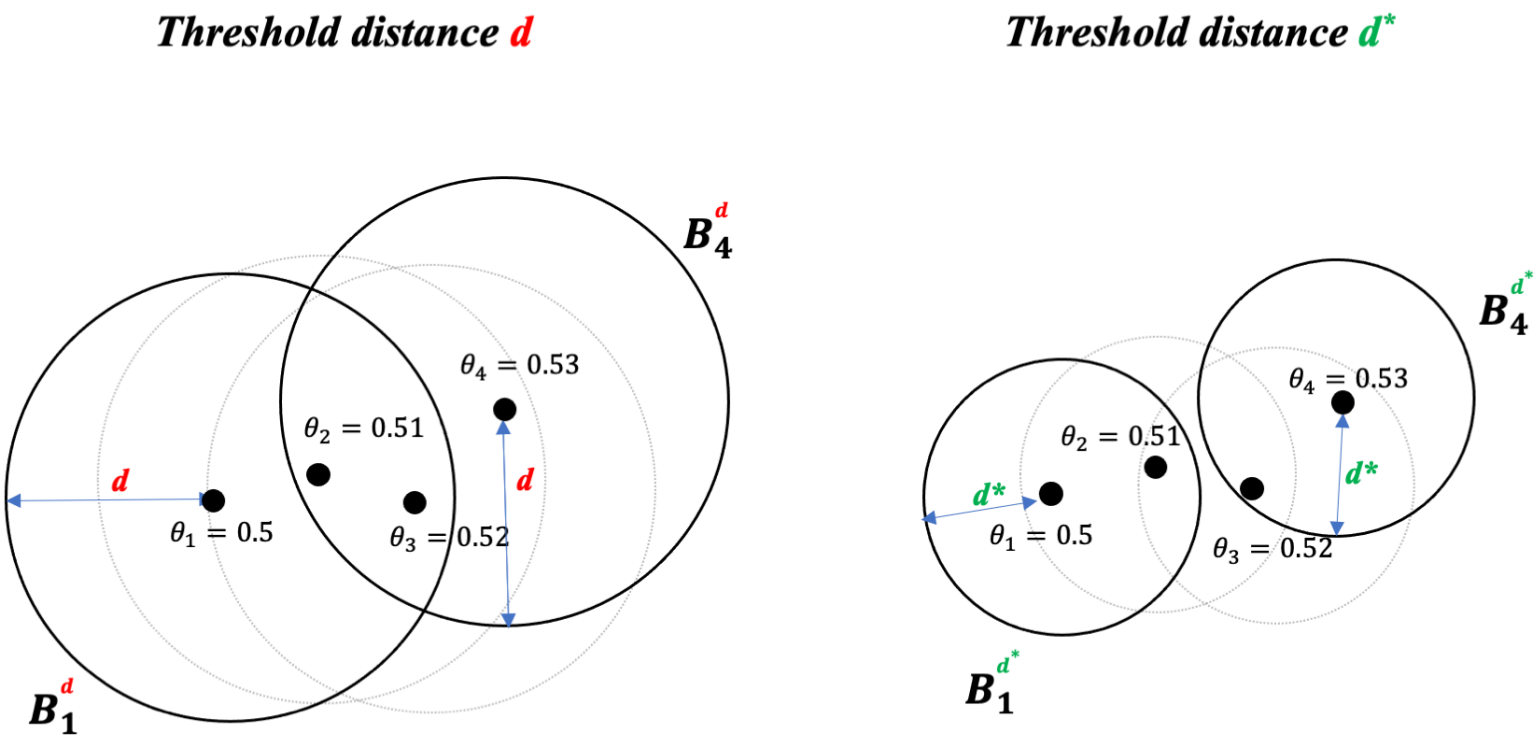
Qu, X. and Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2):209–232.

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

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

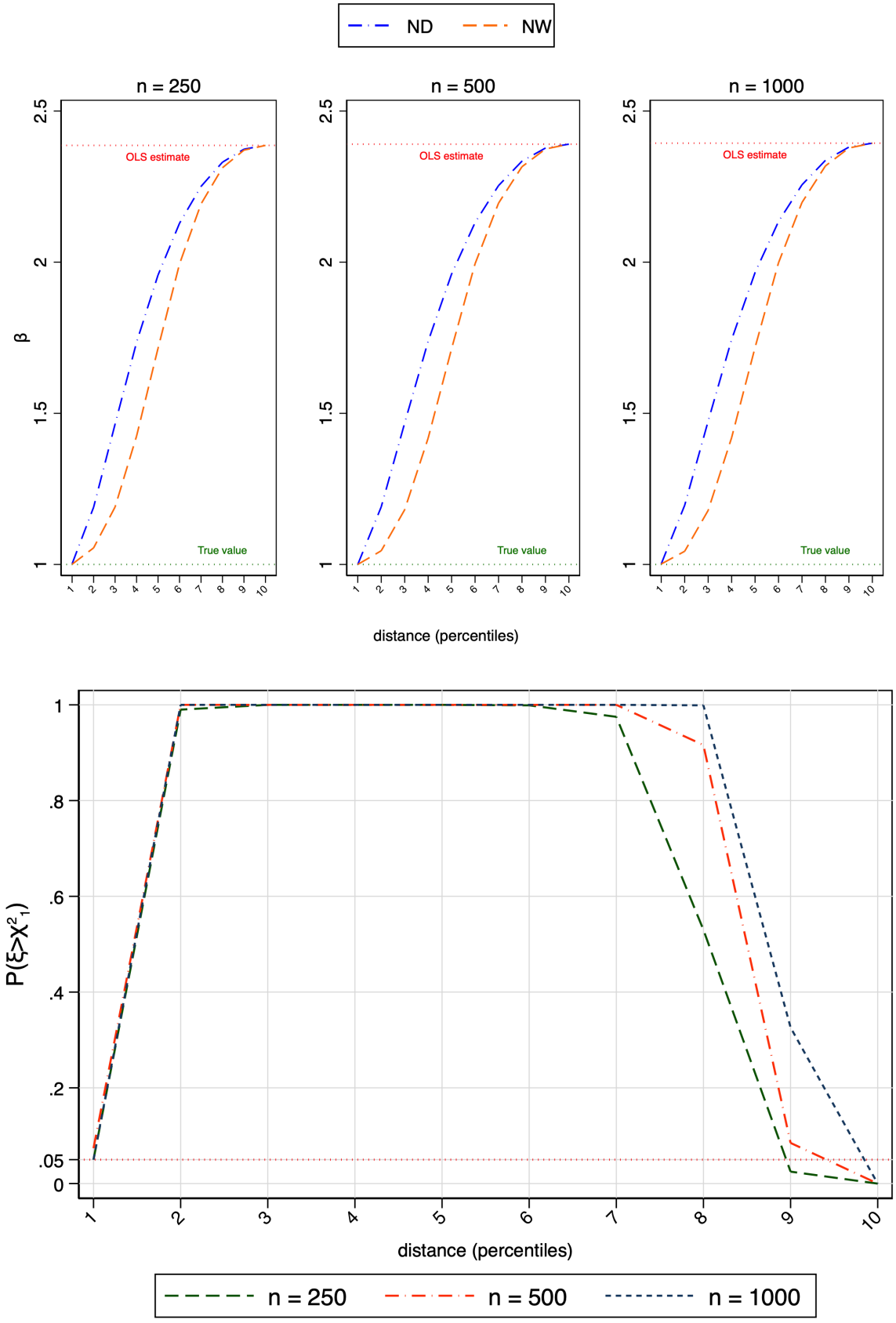
Tabord-Meehan, M. (2019). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business & Economic Statistics*, 37(4):671–680.

Figure 1: Neighborhood Selection by Threshold Distances



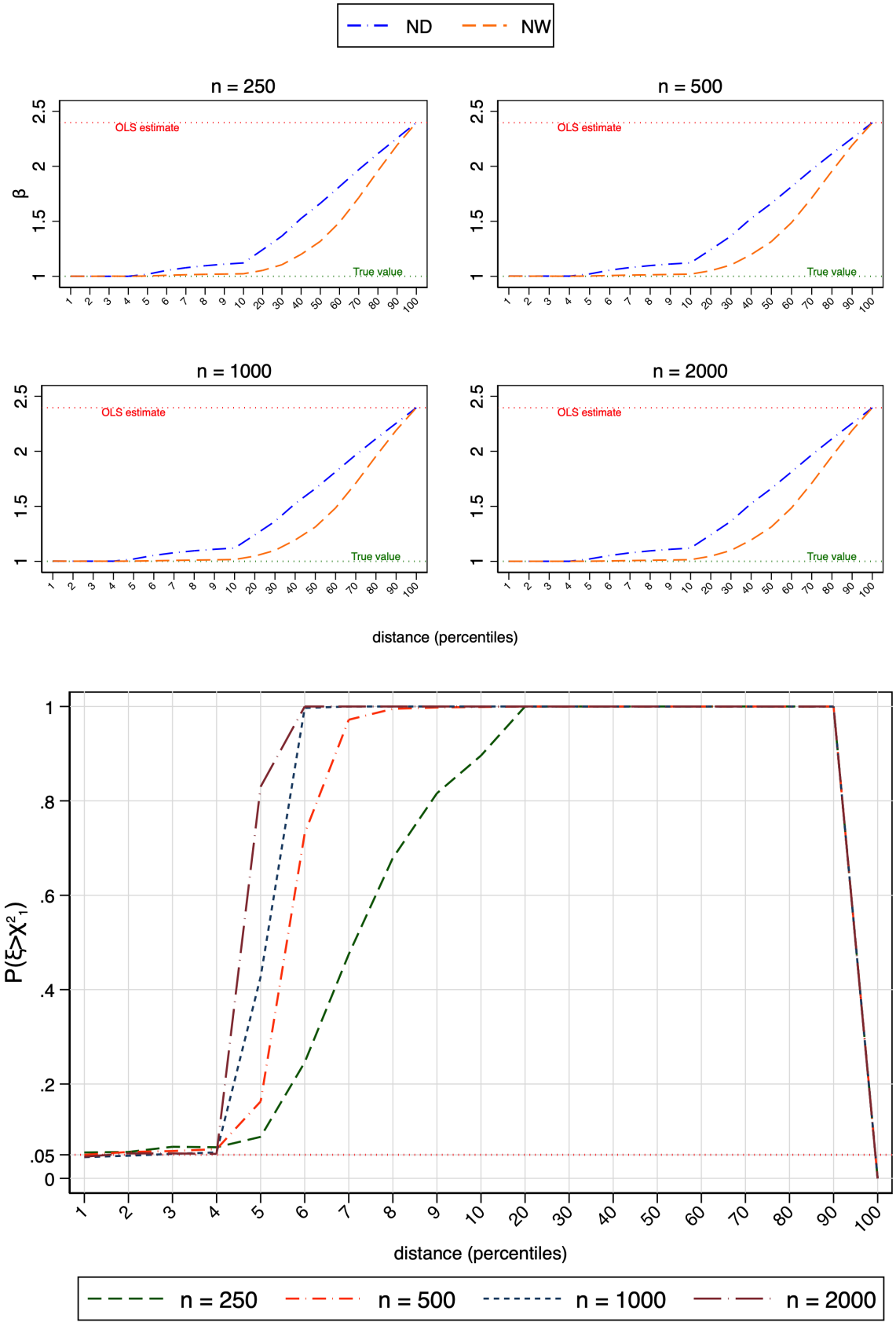
Notes. The graph reports neighborhood specifications based on two different proximity thresholds. The four neighborhoods are centers towards the units with radius d (left) and d^* (right). d^* is assumed to be the threshold that defines neighborhoods where units are more similar in terms of unobservables.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

Figure 2: Monte Carlo evidence (DGP_1): average estimated β and Hausman-like test performance (1000 replications)



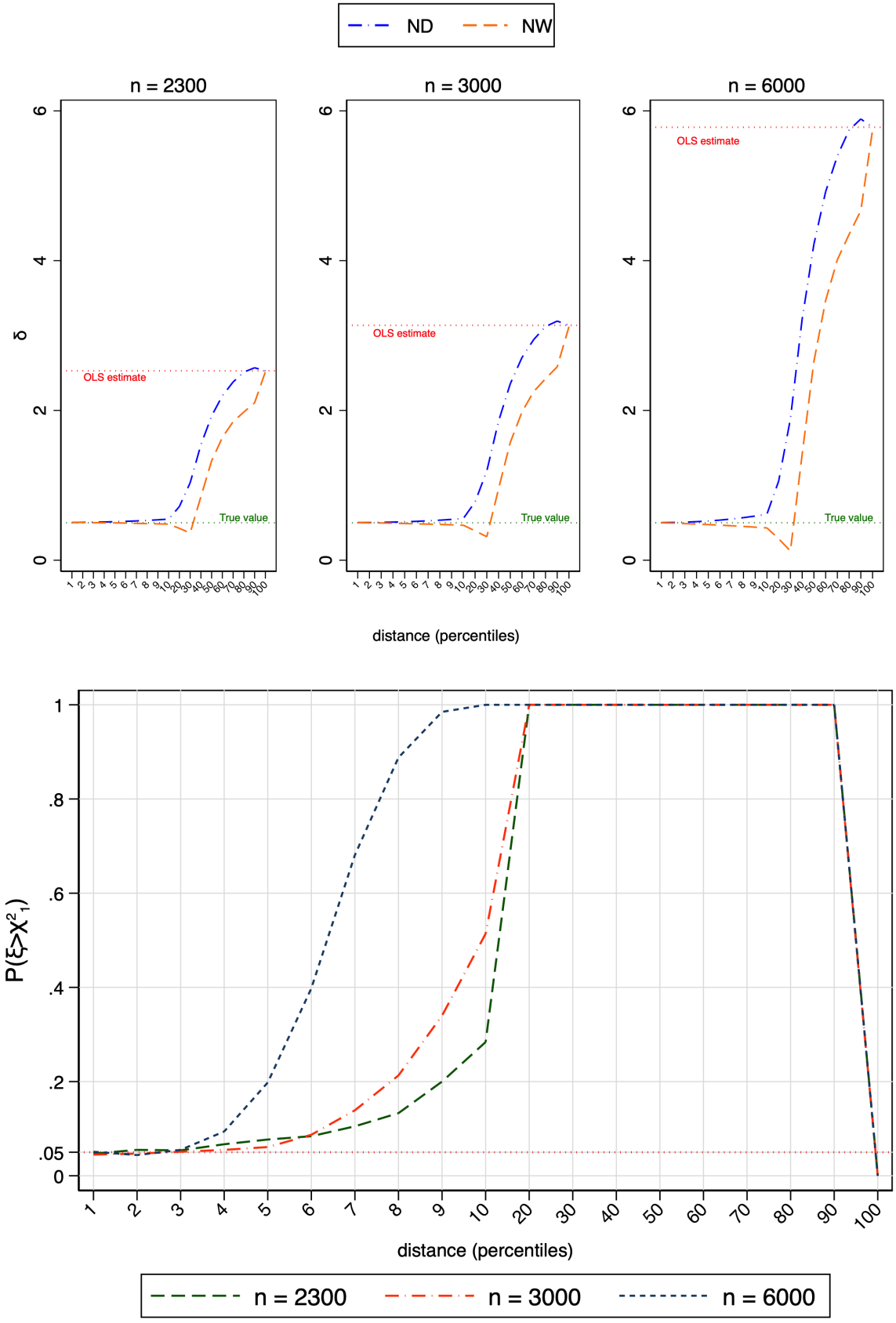
Notes. The top graph reports average (over Monte Carlo replications) estimates for β_{ND} and β_{NW} obtained at different distance thresholds in presence of unobserved neighborhood heterogeneity for $n = 250, 500, 1000$. The first (red) dotted line reports the biased β_{OLS}. The second (green) dotted line reports the true value for β. The bottom graph shows the Hausman-like test performance at different distance thresholds.

Figure 3: Monte Carlo evidence (DGP_2): average estimated β and Hausman-like test performance (1000 replications)



Notes. The top graph reports average (over Monte Carlo replications) estimates for β_{ND} and β_{NW} obtained at different distance threshold percentiles in presence of unobserved unit-specific heterogeneity for $n = 250, 500, 1000, 2000$. The first (red) dotted line reports the biased β_{OLS}. The second (green) dotted line reports the true value for β. The bottom graph shows the Hausman-like test performance at different distance threshold percentiles.

Figure 4: Monte Carlo evidence (DGP_3): average estimated δ and Hausman-like test performance (1000 replications)



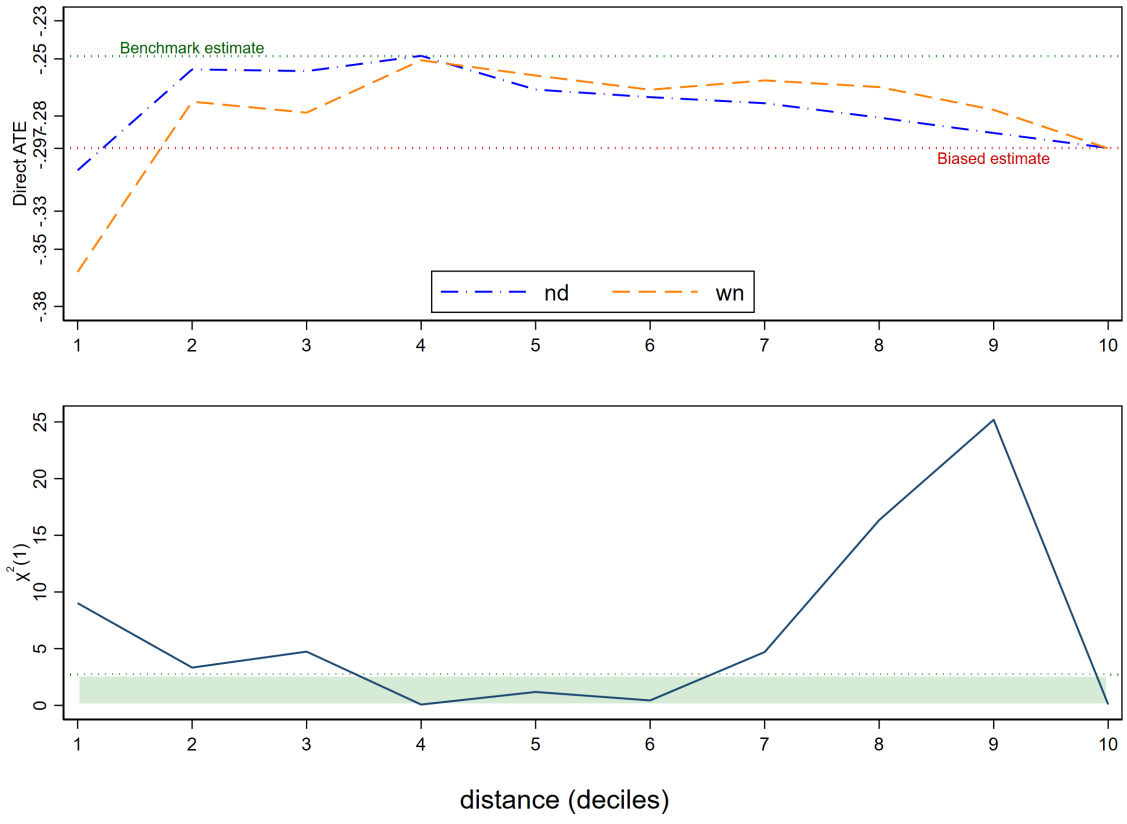
Notes. The top graph reports average (over Monte Carlo replications) estimates for δ_{ND} and δ_{NW} obtained at different distance threshold percentiles assuming unobserved spillovers for $n = 2300, 3000, 6000$. The first dotted line is the biased δ_{OLS}. The second dotted line is the true value for δ. The bottom graph shows the Hausman-like test performance at different distance threshold percentiles.

Table 1: Replication of Miguel and Kremer (2004): Deworming health direct effects and externalities across schools (Any moderate-heavy helminth infection, 1999)

Dependent Variable: Any moderate-heavy helminth infection indicator	
Indicator for Group 1 (1998 Treatment) School	-0.2486***
	(0.0592)
Group 1 pupils within 3 km (per 1000 pupils)	-0.1938**
	(0.0957)
Group 1 pupils within 3-6 km (per 1000 pupils)	-0.0693
	(0.0696)

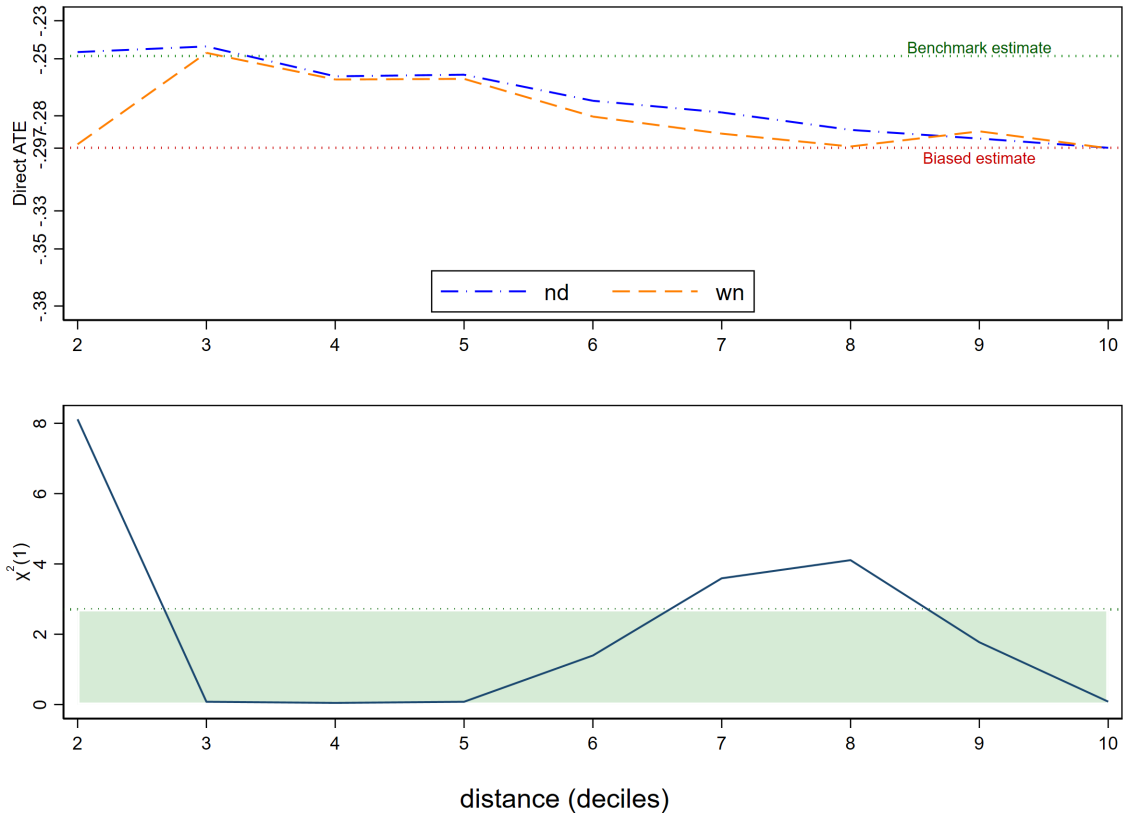
Notes: Grade 3-8 pupils. Unweighted linear probability model. It includes grade indicators, school assistance controls, district exam scores and the total number of children attending primary school within a certain distance from the school according to the model. Cluster robust standard errors at school level in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence.

Figure 5: Hausman-like test and direct ATE - Any moderate-heavy helminth infection (6 km radius)



Notes. The top graph reports estimates for β_{ND} and β_{NW} at different distance threshold deciles assuming unobserved spillovers. The first dotted line is the LPM estimates of model (26). The second dotted line is the biased β_{OLS}. The bottom graph shows the Hausman-like test statistics δ at different distance threshold deciles.

Figure 6: Hausman-like test and direct ATE - Any moderate-heavy helminth infection (3 km radius)



Notes. The top graph reports estimates for β_{ND} and β_{NW} at different distance threshold deciles assuming unobserved spillovers. The first dotted line is the LPM estimates of model (26). The second dotted line is the biased β_{OLS}. The bottom graph shows the Hausman-like test statistics δ at different distance threshold deciles.

Appendix A. Appendix

Appendix A.1. Proofs of propositions

Following Tabord-Meehan (2019), we use the CLT for dependency graphs proposed by Janson et al. (1988) to show convergence in distribution.

Theorem 1 (Janson, 1988 Theorem 2). *Suppose that, for each n, $\{X_{ni}\}$, $i = 1, \dots, n$ is a family of bounded random variables, $|X_{ni}| < c_n$, with dependency graph G_n. The maximal degree of G_n is Deg_n. Let $S_n = \sum_i X_{ni}$ and $\sigma_n^2 = var(S_n)$. If there exists an integer $l \geq 3$ such that*

$$R_n = \frac{\left(\frac{n}{Deg_n}\right)^{1/l} 2(Deg_n)c_n}{\sigma_n} \to 0 \text{ as } n \to \infty, \tag{A.1}$$

then

$$\frac{S_n - E(S_n)}{\sigma_n} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

In our context, we define the dependency graph of an estimator's key statistic with maximal degree M^u under Assumptions 3 and 4 or 7. For a definition of a dependency graph we follow Janson et al. (1988) and Tabord-Meehan (2019).

Definition 1 (Dependency graph). *A graph G is a dependency graph for a family of random variables if:*

1. *There exist a one-to-one correspondence between the random variables and the vertex of the graph.*
2. *if V_1 and V_2 are two disjoint sets of vertices in G such that no edge of G one endpoint in V_1 and the other in V_2, then the corresponding random variables are independent.*

In what follows, the proof of convergence in distribution focus on the second asymptotic approximation, i.e., denseness. The first asymptotic approximation is, then, just a particular case given that M^u is bounded. We start by deriving the asymptotic distribution of the NW estimator. This derivation is more complex compared to the ND case since it requires strengthening the assumptions to satisfy condition (6).

Proof of Proposition 2 *b*).

We have

$$\sqrt{\frac{n}{n^r}}(\hat{\boldsymbol{\beta}}_{NW} - \beta) = \left[\frac{1}{n}\sum_{i=1}^{n} \Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\boldsymbol{x}_i\right]^{-1} \frac{1}{\sqrt{nn^r}}\sum_{i=1}^{n} \Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\epsilon_i. \tag{A.2}$$

First, let us show the convergence of the following statistic $\frac{1}{n}\left[\sum_{i=1}^{n} \Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\boldsymbol{x}_i\right]^{-1}$. Without loss of generality, let us focus on the component wise convergence $\frac{1}{n}\sum_{i=1}^{n}(\Delta_{in_i}x_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n_i}\sum_{j\in B_i} x_j)^2$. We know that the $E(\frac{1}{n}\sum_{i=1}^{n}(\Delta_{in_i}x_i)^2) = E(\Delta_{in_i}x_i)^2$. For the variance, we have

$$var\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\frac{1}{n_i}\sum_{j\in B_i}x_j)^2\right)=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}cov\Big((x_i-\frac{1}{n_i}\sum_{j\in B_i}x_j)^2,(x_k-\frac{1}{n_k}\sum_{l\in B_k}x_l)^2\Big).$$

Let us study how many covariances are not zero for each term. In general, for a fixed i, each agent have at most $M^u_{NW}\leq n-1$ links. Observe that the $cov(x_i^2,x_k^2)$ is uniformly bounded by Assumption 5. There are three different leading terms in the sums of covariances: $\sum_{i=1}^n\sum_{k=1}^n cov(x_i^2,x_k^2)$, $\sum_{i=1}^n\sum_{k=1}^n cov\left(\left(\frac{1}{n_i}\sum_{j\in B_i}x_j\right)^2,x_k^2\right)$ and, $\sum_{i=1}^n\sum_{k=1}^n cov\left(\left(\frac{1}{n_i}\sum_{j\in B_i}x_j\right)^2,\left(\frac{1}{n_k}\sum_{l\in B_k}x_l\right)^2\right)$. The first sum is bounded by $n(M^u_{NW})$ that is

$$\sum_{i=1}^{n}\sum_{k=1}^{n}cov(x_i^2,x_k^2)=O(n(M^u_{NW})).$$

The equality follows because each individual has at most M^u_{NW} links. Observe that also i belongs to the neighborhood (i can be equal to k). Then, we need to multiply M^u_{NW} by the number of terms in the summation to obtain the the maximum number of covariances that are different from zero. For the second term, we have

$$\sum_{i=1}^{n}\sum_{k=1}^{n}cov\left(\left(\frac{1}{n_i}\sum_{j\in B_i}x_j\right)^2,x_k^2\right)=\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{1}{n_i^2}\sum_{l\in B_i}\sum_{j\in B_i}cov\left(x_lx_j,x_k^2\right)\leq$$

$$\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{1}{M^{u2}_{NW}}\sum_{l\in B_i}\sum_{j\in B_i}cov\left(x_lx_j,x_k^2\right)\leq\frac{1}{M^{u2}_{NW}}O(nM^{u2}_{NW})=O(n).$$

For the last relevant term, we use again the bilinearity of the covariance operator as for the second term. Hence, we have

$$\sum_{i=1}^{n}\sum_{k=1}^{n}cov\left(\left(\frac{1}{n_i}\sum_{j\in B_i}x_j\right)^2,\left(\frac{1}{n_k}\sum_{l\in B_k}x_l\right)^2\right)\leq\frac{1}{M^{u4}}(n(M^{u4}_{NW}))=O(n).$$

Therefore,

$$var\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\frac{1}{n_i}\sum_{j\in B_i}x_j)^2\right)=\frac{1}{n^2}O(nM^u_{NW})=O\left(\frac{M^u_{NW}}{n}\right)=o(1),$$

where the last equality follows given that $\frac{M^u_{NW}}{n}\to 0$ by Assumption 7. Hence,

$$\frac{1}{n}\left[\sum_{i=1}^{n}\Delta_{in_i}\boldsymbol{x}_i\Delta_{in_i}\boldsymbol{x}_i\right]^{-1}\xrightarrow{p}E[(\Delta_{in_i}x_i)^2]^{-1}\tag{A.3}$$

by continuous mapping theorem given that the variance converge to zero.

Next, we study the convergence in distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_{in_i}x_i\Delta_{in_i}\epsilon_i$. We apply Janson's CLT to the family of random variables $\left\{(x_i-\frac{1}{n_i}\sum_{j\in B_i}x_j),(\epsilon_i-\frac{1}{n_i}\sum_{j\in B_i}\epsilon_j)\right\}$. A dependency graph $G_n=(V,E)$ for this family of random variables is the graph with vertex set $V=\left\{(x_i-\frac{1}{n_i}\sum_{j\in B_i}x_j)(\epsilon_i-\frac{1}{n_i}\sum_{j\in B_i}\epsilon_j)\right\}_{i=1,\dots,n,\,j\in B_i}$ and edge set

$$E=\left\{\left\{(x_i-\frac{1}{n_i}\sum_{j\in B_i}x_j)(\epsilon_i-\frac{1}{n_i}\sum_{j\in B_i}\epsilon_j),(x_k-\frac{1}{n_k}\sum_{l\in B_k}x_l)(\epsilon_k-\frac{1}{n_k}\sum_{l\in B_k}\epsilon_l)\right\}: \tag{A.4}$$
$$\Delta_{in_i}x_i\Delta_{in_i}\epsilon_i,\Delta_{kn_k}x_k\Delta_{kn_k}\epsilon_k\in V \text{ and } |B_i\cap B_k|>c_n\right\}.$$

The maximal degree is M_{NW}^u by definition and by Assumption 5 $|\Delta_{in_i}x_i\Delta_{in_i}\epsilon_i|<c$ for all n and c constant. Now, we can check the main condition of Janson's CLT theorem. Let

$$\boldsymbol{B}_n=var(\sum_i\Delta_{in_i}x_i\Delta_{in_i}\epsilon_i).$$

In this framework the key (sufficient) condition to Janson' s CLT to apply is the following

$$R_n=\frac{\left(\frac{n}{M_{NW}^u}\right)^{1/l}M_{NW}^u\,c}{\sqrt{\boldsymbol{B}_n}}=\frac{\left(\frac{n}{M_{NW}^u}\right)^{1/l}M_{NW}^u\,c}{\sqrt{nn^r}}\cdot\left(\frac{1}{nn^r}\boldsymbol{B}_n\right)^{-1/2}.$$

The second term of the product converge to $\boldsymbol{B}^{1/2}$ by Assumption 8. The first term can be written

$$\frac{\left(\frac{n}{M_{NW}^u}\right)^{1/l}}{\sqrt{n^r}}\frac{M_{NW}^u c}{\sqrt{n}},$$

$\frac{M_{NW}^u\,c}{\sqrt{n}}\to\bar{c}<\infty$ by Assumption 7. Hence,

$$\frac{\left(\frac{n}{M_{NW}^u}\right)^{1/l}}{\sqrt{n^r}}\to 0$$

if

$$\frac{n^{\frac{1}{l}}}{n^{\frac{r}{2}}}\to 0,$$

given Assumption 7 (n is faster than M_{NW}^u). This means choosing an r and l sufficiently large such that $1/l<r/2$. So $R_n\to 0$ and the condition for the Janson's CLT are satisfied.
Hence, we have

$$\frac{\sum_i\Delta_{in_i}x_i\Delta_{in_i}\epsilon_i}{\sqrt{\boldsymbol{B}_n}}\xrightarrow{d}N(0,1),$$

that can be written as,

$$\frac{1}{\sqrt{nn^r}}\sum_i \Delta_{in_i} x_i \Delta_{in_i}\epsilon_i \left(\frac{1}{nn^r}\boldsymbol{B}_n\right)^{-1/2}.$$

Therefore, by Assumption 8 we have

$$\frac{1}{\sqrt{nn^r}}\sum_i \Delta_{in_i} x_i \Delta_{in_i}\epsilon_i \xrightarrow{d} N(0, \boldsymbol{B}_W).$$

We then apply the Slutsky's Theorem to each component and the Cramer-Wold device together with the Janson's CLT to (A.2) and conclude that

$$\sqrt{\frac{n}{n^r}}(\hat{\boldsymbol{\beta}}_{NW} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{A}_W^{-1}\boldsymbol{B}_W\boldsymbol{A}_W^{-1}).$$

■

Proof of Proposition 2 *a*).

Just a particular case of the **Proof of Proposition 2** b) with $M_{NW}^u < c$ and a different rate of convergence for the distribution of the estimator. ■

Proof of Proposition 1 *b*).

We can write the ND estimator as

$$\sqrt{\frac{N}{n}}(\hat{\boldsymbol{\beta}}_{ND} - \beta) = \left[\frac{1}{N}\sum_{i<j,j\in B_i\setminus i} \Delta_{ij}\boldsymbol{x}_i \Delta_{ij}\boldsymbol{x}_i\right]^{-1} \sqrt{\frac{N}{n}}\frac{1}{N}\sum_{i<j\in B_i\setminus i} \Delta_{ij}\boldsymbol{x}_i \Delta_{ij}\epsilon_i. \tag{A.5}$$

First, let us show the convergence of thr following statistic $\left[\sum_{i<j,j\in B_i\setminus i} \Delta_{ij}\boldsymbol{x}_i \Delta_{ij}\boldsymbol{x}_i\right]^{-1}$. Without loss of generality, let us focus on the component wise convergence $\frac{1}{N}\sum_{i<j,j\in B_i\setminus i}(\Delta_{ij}x_i)^2$. We know that $E(\frac{1}{N}\sum_{i<j,j\in B_i\setminus i}(\Delta_{ij}x_i)^2) = E(\Delta_{ij}x_i)^2$. For the variance, we have

$$var\left(\frac{1}{N}\sum_{i<j,j\in B_i\setminus i}(\Delta_{ij}x_i)^2\right) = \frac{1}{N^2}\sum_{i<j,j\in B_i\setminus i}\sum_{k<l\in B_k\setminus k} cov((\Delta_{ij}x_i)^2, (\Delta_{kl}x_k)^2).$$

Given that each agent have at most M_{ND}^u links, the inner summation of the covariances is nonzero for at most $2M_{ND}^u$ units. Furthermore, the $cov((\Delta_{ij}x_i)^2, (\Delta_{kl}x_k)^2)$ is bounded by Assumption 5. Therefore,

$$var\left(\frac{1}{N}\sum_{i<j,j\in B_i\setminus i}(\Delta_{ij}x_i)^2\right) = \frac{1}{N^2}O(NM_{ND}^u) = O\left(\frac{M_{ND}^u}{N}\right) = o(1),$$

where the last equality follow $M_{ND}^u \leq n-1$ and from the fact that $N \leq 2M_{ND}^u n$.
Hence,

$$\frac{1}{N}\left(\sum_{i<j,j\in B_i\setminus i}(\Delta_{ij}x_i)^2\right)^{-1} \xrightarrow{p} E[(\Delta_{ij}x_i)^2]^{-1}, \tag{A.6}$$

by continuous mapping theorem given that the variance converge to zero.

Next, we study the convergence in distribution of $\sqrt{\frac{N}{n}}\frac{1}{N}\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}x_i\Delta_{ij}\epsilon_i$. We apply Janson's CLT to the family of random variables $\{\Delta_{ij}x_i\Delta_{ij}\epsilon_i\}$.

A dependency graph $G_n = (V, E)$ for this family of random variables is the graph with vertex set $V = \left\{(x_i - x_j)(\epsilon_i - \epsilon_j)\right\}_{i<j,\, j\in B_i}$ and edge set

$$E = \left\{\left\{(x_i - x_j)(\epsilon_i - \epsilon_j), (x_k - x_l)(\epsilon_k - \epsilon_l)\right\} : \Delta_{ij}x_i\Delta_{ij}\epsilon_i, \Delta_{kl}x_k\Delta_{kl}\epsilon_k \in V \text{ and } B_i \cap B_k \neq \emptyset\right\}.$$

The maximal degree is $M^u_{ND} \leq n-1$ by definition and by Assumption 5 $|\Delta_{ij}x_i\Delta_{ij}\epsilon_i| < c$ for all n and c constant. Now, we can check the main condition of Janson's CLT theorem. Let

$$\boldsymbol{B}_n = var\left(\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}x_i\Delta_{ij}\epsilon_i\right).$$

The key (sufficient) condition to Janson' s CLT to apply is the following

$$R_n = \frac{\left(\frac{N}{M^u_{ND}}\right)^{1/l}(M^u_{ND})c}{\sqrt{\boldsymbol{B}_n}} = \frac{\left(\frac{N}{M^u_{ND}}\right)^{1/l}M^u_{ND}c}{\sqrt{Nn^r}}\cdot\left(\frac{1}{Nn^r}\boldsymbol{B}_n\right)^{-1/2}.$$

The second term of the product converges to $\boldsymbol{B}^{1/2}$ by Assumption 6. The first term is bounded by

$$\frac{n^{\frac{1}{l}}}{n^{\frac{r}{2}}},$$

given Assumption 4. This means choosing an r and l sufficiently large such that $1/l < r/2$. So $R_n \to 0$ and the condition for the Janson's CLT are satisfied.

Hence, we have

$$\frac{\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}x_i\Delta_{ij}\epsilon_i}{\sqrt{\boldsymbol{B}_n}} \xrightarrow{d} N(0,1),$$

that can be written as,

$$\frac{1}{\sqrt{Nn^r}}\sum_{i<j,j\in B_i\setminus i}\Delta_{ij}x_i\Delta_{ij}\epsilon_i\left(\frac{1}{Nn^r}\boldsymbol{B}_n\right)^{-1/2}.$$

Therefore, by Assumption 8 we have

$$\frac{1}{\sqrt{Nn^r}} \sum_{i<j,j\in B_i\setminus i} \Delta_{ij} x_i \Delta_{ij} \epsilon_i \xrightarrow{d} N(0, \boldsymbol{B}).$$

We then apply the Slutsky's Theorem to each component and the Cramer-Wold device together with the Janson's CLT to (A.5) and conclude that

$$\sqrt{\frac{N}{n^r}}(\hat{\boldsymbol{\beta}}_{ND} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}).$$

■

Proof of Proposition 1 *a*).

Just a particular case of the **Proof of Proposition 1** *b*) with $M^u_{ND} < c$ and a different rate of convergence for the distribution of the estimator. ■

Appendix A.2. Estimator asymptotic biases

Appendix A.2.1. Distance model

We report the distance model (22) below

$$\bar{d}_{ij} = \alpha\,(\theta_i - \theta_j) + (1-\alpha^2)^{1/2}\varepsilon_{ij} \tag{A.7}$$

where $\bar{d}_{ij}$ is the natural logarithm of the distance, ε_{ij} is iid across dyads with mean zero, unit variance and not correlated with the θs. To ease the notation let us remove the conditioning event $A = 1(\bar{d}_{ij} < d)$ from the expectation operator. Using model (A.7) and $\Delta\tau$ definition in Section 4 we have

$$\Delta\tau = E\,(\theta_i - \theta_j)^2 = E\left(\frac{1}{\alpha}(\bar{d}_{ij} - (1-\alpha^2)^{1/2}\varepsilon_{ij})\right)^2 =$$

$$= \frac{1}{\alpha^2}\left(E(\bar{d}^2_{ij}) + (1-\alpha^2)E(\varepsilon^2_{ij}) - 2(1-\alpha^2)E(\varepsilon^2_{ij})\right) = \frac{1}{\alpha^2}\left(E(d^2_{ij}) - (1-\alpha^2)E(\varepsilon^2_{ij})\right).$$

The asymptotic bias of the ND estimator can be characterized as

$$\hat{\beta}_{ND} - \beta \xrightarrow{p} \frac{\phi\frac{1}{\alpha^2}\left(E(\bar{d}^2_{ij}) - (1-\alpha^2)E(\varepsilon^2_{ij})\right)}{\phi^2\frac{1}{\alpha^2}\left(E(\bar{d}^2_{ij}) - (1-\alpha^2)E(\varepsilon^2_{ij})\right) + (1-\phi^2)}.$$

For the NW estimator, we can rewrite an averaged version of the distance model as follows

$$\frac{1}{n_i}\sum_{j\in B_i}\bar{d}_{ij} = \alpha\left(\theta_i - \frac{1}{n_i}\sum_{j\in B_i}\theta_j\right) + (1-\alpha^2)^{1/2}\frac{1}{n_i}\sum_{j\in n_i}\varepsilon_{ij}.$$

Hence, $\tilde{\tau}$ can be written as

$$\tilde{\tau} = E\left(\theta_i - \frac{1}{n_i}\sum_{j\in B_i}\theta_j\right)^2 = \frac{1}{\alpha^2}E\left(\frac{1}{n_i}\sum_{j\in n_i}\bar{d}_{ij} - (1-\alpha^2)^{1/2}\frac{1}{n_i}\sum_{j\in B_i}\varepsilon_{ij}\right)^2 =$$

$$\frac{1}{\alpha^2}\left\{E\left[\left(\frac{1}{n_i}\sum_{j\in B_i}\bar{d}_{ij}\right)^2\right] + (1-\alpha^2)E\left[\frac{1}{n_i^2}\sum_{j\in B_i}\varepsilon_{ij}^2\right] - 2(1-\alpha^2)E\left[\frac{1}{n_i^2}\sum_{j\in B_i}\varepsilon_{ij}^2\right]\right\}$$

$$\frac{1}{\alpha^2}\left\{E\left[\left(\frac{1}{n_i}\sum_{j\in B_i}\bar{d}_{ij}\right)^2\right] - (1-\alpha^2)E\left[\frac{1}{n_i^2}\sum_{j\in B_i}\varepsilon_{ij}^2\right]\right\}$$

The asymptotic bias of the NW estimator can be characterized as

$$\hat{\beta}_{NW} - \beta \xrightarrow{p} \frac{\phi\frac{1}{\alpha^2}\left\{E\left[\left(\frac{1}{n_i}\sum_{j\in B_i}\bar{d}_{ij}\right)^2\right] - (1-\alpha^2)E\left[\frac{1}{n_i^2}\sum_{j\in B_i}\varepsilon_{ij}^2\right]\right\}}{\phi^2\frac{1}{\alpha^2}\left\{E\left[\left(\frac{1}{n_i}\sum_{j\in B_i}\bar{d}_{ij}\right)^2\right] - (1-\alpha^2)E\left[\frac{1}{n_i^2}\sum_{j\in B_i}\varepsilon_{ij}^2\right]\right\} + (1-\phi^2)}.$$

Appendix A.2.2. Treatment spillovers

Let us assume that NUH is generated by a treatment spillover effect as follows

$$\theta_i = \gamma\sum_{j\in B_i}T_j + (1-\gamma^2)^{1/2}v_i, \tag{A.8}$$

where v_i is an i.i.d. rvs with zero mean and unit variance. In this case it can be shown that

$$\Delta\tau = \gamma^2[E(\sum_{l\in B_i}T_l - \sum_{k\in B_j}T_k)^2] + [E(v_i)^2 + E(v_j)^2](1-\gamma^2).$$

given that v_i and T_i are not correlated and v_i has mean zero. Thus,

$$\Delta\tau = \gamma^2[E(\sum_{l\in B_i}T_l - \sum_{k\in B_j}T_k)^2] + 2(1-\gamma^2),$$

where the second term of the summation comes from $E(v_i)^2 + E(v_j)^2 = 2$. Thus, the ND asymptotic bias depends on the second moment of the difference between number of treated neighbors for unit i and j. In a similar way, we can derive the asymptotic bias for $\tilde{\tau}$.

Appendix A.3. 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).[17] 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. Observe that the researcher can also prefer $\kappa \preceq \kappa - 1$ if $\kappa - 1 \geq \kappa$. In this case, one can start from a smaller enough d_1 and sequentially increases the threshold up to the first d so that it is not rejected by the data. One motivation for this order of priority is that when d is too small, too many units (isolates) are removed from the sample, resulting in a loss of efficiency for the estimators.

Appendix A.4. Set cardinalities for irregular lattices

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

Lemma 1. *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.*

[17] 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.

Supplementary Appendix

Additional Tables and Figures

Table S.1: Monte Carlo evidence (DGP_1). Average estimated β and Hausman-like test performace (1000 replications)

d	1	2	3	4	5	6	7	8	9	10
					n = 250					
ND	1.002	1.189	1.464	1.735	1.957	2.129	2.251	2.331	2.374	2.386
	(0.999)	(1.131)	(1.301)	(1.448)	(1.559)	(1.640)	(1.695)	(1.730)	(1.748)	(1.753)
NW	1.001	1.055	1.190	1.424	1.715	1.994	2.192	2.312	2.371	2.386
	(0.996)	(1.036)	(1.129)	(1.274)	(1.435)	(1.574)	(1.666)	(1.719)	(1.743)	(1.750)
χ^2	1.093	14.084	38.780	50.454	46.902	32.507	16.833	6.170	0.826	0.000
	[0.050]	[0.990]	[1.000]	[1.000]	[1.000]	[0.999]	[0.975]	[0.530]	[0.025]	[0.000]
					n = 500					
ND	1.000	1.190	1.468	1.738	1.961	2.131	2.253	2.334	2.378	2.390
	(1.001)	(1.136)	(1.307)	(1.454)	(1.565)	(1.645)	(1.700)	(1.735)	(1.754)	(1.759)
NW	1.000	1.045	1.181	1.418	1.714	1.995	2.195	2.316	2.374	2.390
	(1.000)	(1.034)	(1.128)	(1.276)	(1.440)	(1.580)	(1.672)	(1.726)	(1.751)	(1.758)
χ^2	1.164	32.369	83.866	105.743	96.740	66.716	34.263	11.977	1.542	0.000
	[0.074]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.916]	[0.085]	[0.000]
					n = 1000					
ND	1.002	1.196	1.474	1.744	1.966	2.135	2.256	2.337	2.381	2.394
	(0.999)	(1.135)	(1.306)	(1.453)	(1.564)	(1.642)	(1.697)	(1.732)	(1.751)	(1.757)
NW	1.002	1.044	1.179	1.419	1.718	1.998	2.197	2.318	2.378	2.394
	(0.998)	(1.029)	(1.124)	(1.274)	(1.439)	(1.578)	(1.670)	(1.723)	(1.749)	(1.756)
χ^2	1.016	71.515	176.294	216.993	196.484	135.604	70.067	24.960	3.632	0.000
	[0.050]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.999]	[0.326]	[0.000]

Notes: The table reports average (over Monte Carlo replications) estimates for β_{ND} and β_{NW} and the Hausman-like test statistics performance at different distance thresholds d (percentiles of the distance) for $n = 250, 500, 1000$. Root mean squared errors and p-values for the tests are reported in round and squared brackets, respectively. The true value for $\beta = 1$.

Table S.2: Monte Carlo evidence (DGP_2). Average estimated β and Hausman-like test performance (1000 replications)

d	1	2	3	4	5	6	7	8	9	10	50	100
						n = 250						
ND	1.000	1.001	1.000	1.000	1.019	1.054	1.078	1.096	1.110	1.121	1.664	2.397
	(0.995)	(0.996)	(0.996)	(0.996)	(1.011)	(1.036)	(1.053)	(1.066)	(1.076)	(1.084)	(1.410)	(1.756)
NW	1.001	1.000	1.002	1.000	1.003	1.009	1.014	1.018	1.021	1.023	1.318	2.397
	(0.990)	(0.993)	(0.994)	(0.995)	(0.997)	(1.002)	(1.005)	(1.008)	(1.010)	(1.012)	(1.209)	(1.753)
χ^2	1.061	1.051	1.135	1.220	1.299	2.709	4.312	5.838	7.239	8.335	57.145	0.000
	[0.055]	[0.056]	[0.067]	[0.066]	[0.088]	[0.246]	[0.475]	[0.679]	[0.816]	[0.896]	[1.000]	[0.000]
						n = 500						
ND	1.003	1.002	1.002	1.002	1.022	1.056	1.080	1.098	1.112	1.123	1.663	2.396
	(1.002)	(1.002)	(1.002)	(1.002)	(1.016)	(1.041)	(1.058)	(1.071)	(1.081)	(1.088)	(1.412)	(1.759)
NW	1.002	1.002	1.002	1.001	1.003	1.008	1.011	1.014	1.017	1.020	1.316	2.396
	(1.001)	(1.001)	(1.001)	(1.001)	(1.003)	(1.006)	(1.008)	(1.011)	(1.013)	(1.014)	(1.213)	(1.757)
χ^2	1.010	1.104	1.093	1.123	2.055	6.537	10.797	14.438	17.434	19.822	116.611	0.000
	[0.050]	[0.056]	[0.058]	[0.062]	[0.163]	[0.731]	[0.972]	[0.995]	[0.998]	[0.999]	[1.000]	[0.000]
						n = 1000						
ND	1.002	1.001	1.002	1.001	1.021	1.054	1.078	1.096	1.110	1.121	1.661	2.396
	(0.999)	(0.999)	(0.999)	(0.999)	(1.014)	(1.038)	(1.055)	(1.068)	(1.078)	(1.085)	(1.411)	(1.759)
NW	1.002	1.002	1.002	1.001	1.002	1.005	1.008	1.011	1.014	1.016	1.312	2.396
	(0.998)	(0.998)	(0.998)	(0.998)	(0.999)	(1.001)	(1.004)	(1.006)	(1.008)	(1.010)	(1.210)	(1.758)
χ^2	0.970	0.958	0.969	1.017	3.971	14.804	24.486	32.187	38.351	43.325	234.901	0.000
	[0.045]	[0.048]	[0.052]	[0.055]	[0.427]	[0.997]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.000]
						n = 2000						
ND	1.001	1.001	1.001	1.001	1.021	1.053	1.077	1.095	1.109	1.120	1.662	2.396
	(0.999)	(0.998)	(0.998)	(0.998)	(1.014)	(1.037)	(1.054)	(1.067)	(1.077)	(1.085)	(1.410)	(1.757)
NW	1.001	1.001	1.001	1.001	1.001	1.004	1.007	1.010	1.013	1.015	1.313	2.396
	(0.998)	(0.998)	(0.998)	(0.998)	(0.999)	(1.001)	(1.003)	(1.005)	(1.007)	(1.009)	(1.210)	(1.757)
χ^2	1.012	0.979	0.970	1.012	9.189	32.976	53.360	69.371	81.792	91.705	472.790	0.000
	[0.046]	[0.053]	[0.053]	[0.052]	[0.830]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.000]

Notes: The table reports average (over Monte Carlo replications) estimates for β_{ND} and β_{NW} and the Hausman-like test statistics performance at different distance thresholds d (percentiles of the distance) for $n = 250, 500, 1000, 2000$. Root mean squared errors and p-values for the tests are reported in round and squared brackets, respectively. The true value for $\beta = 1$.

Table S.3: Monte Carlo evidence (DGP_3). Average estimated β and Hausman-like test performance (1000 replications)

d	1	2	3	4	5	6	7	8	9	10	50	100
						n = 2300						
ND	0.504	0.507	0.508	0.511	0.515	0.519	0.525	0.532	0.540	0.549	1.926	2.529
	(0.500)	(0.500)	(0.501)	(0.501)	(0.502)	(0.503)	(0.504)	(0.505)	(0.507)	(0.509)	(0.816)	(1.499)
NW	0.505	0.504	0.502	0.500	0.498	0.496	0.492	0.489	0.485	0.482	1.326	2.516
	(0.499)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.501)	(0.501)	(0.661)	(1.500)
χ^2	0.952	1.004	1.040	1.089	1.122	1.200	1.393	1.700	2.154	2.869	192.959	-9.664
	[0.047]	[0.055]	[0.054]	[0.067]	[0.077]	[0.084]	[0.105]	[0.133]	[0.200]	[0.284]	[1.000]	[0.000]
						n = 3000						
ND	0.503	0.503	0.506	0.509	0.513	0.519	0.526	0.535	0.545	0.558	2.356	3.136
	(0.500)	(0.500)	(0.501)	(0.502)	(0.503)	(0.505)	(0.507)	(0.509)	(0.512)	(0.516)	(0.980)	(1.908)
NW	0.503	0.499	0.498	0.495	0.490	0.486	0.482	0.478	0.473	0.467	1.573	3.119
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.501)	(0.501)	(0.502)	(0.502)	(0.755)	(1.910)
χ^2	0.935	0.970	1.009	1.038	1.154	1.394	1.801	2.424	3.439	4.961	262.224	-10.894
	[0.045]	[0.047]	[0.051]	[0.055]	[0.061]	[0.087]	[0.139]	[0.213]	[0.340]	[0.514]	[1.000]	[0.000]
						n = 6000						
ND	0.502	0.504	0.509	0.515	0.524	0.536	0.551	0.569	0.590	0.615	4.218	5.781
	(0.500)	(0.501)	(0.504)	(0.507)	(0.512)	(0.518)	(0.526)	(0.536)	(0.547)	(0.560)	(1.758)	(3.721)
NW	0.501	0.496	0.490	0.483	0.476	0.469	0.460	0.451	0.442	0.432	2.653	5.747
	(0.500)	(0.500)	(0.500)	(0.501)	(0.501)	(0.502)	(0.503)	(0.505)	(0.507)	(0.509)	(1.236)	(3.726)
χ^2	0.959	0.963	1.085	1.500	2.297	3.808	6.325	10.113	15.484	22.468	545.615	-18.373
	[0.051]	[0.044]	[0.055]	[0.094]	[0.198]	[0.398]	[0.681]	[0.888]	[0.985]	[1.000]	[1.000]	[0.000]

Notes: The table reports average (over Monte Carlo replications) estimates for β_{ND} and β_{NW} and the Hausman-like test statistics performance at different distance thresholds d (percentiles of the distance) for $n = 2300, 3000, 6000$. Root mean squared errors and p-values for the tests are reported in round and squared brackets, respectively. The true value for $\beta = 1$.