Essays in Econometrics of Networks

by

Kensuke Sakamoto

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2026

Date of final oral examination: 04/09/2026

The dissertation is approved by the following members of the Final Oral Committee:

Jack Porter, Professor, Economics

Bruce Hansen, Professor, Economics

Harold Chiang, Assistant Professor, Economics

Yinqiu He, Assistant Professor, Statistics

To my wife, Hinako.

Acknowledgements

First of all, I would like to express my sincere gratitude to my advisor, Jack Porter, for his invaluable guidance and support throughout my Ph.D. program. Without his encouragement, insights, generosity, and patience, this dissertation would not have been possible. I am also grateful to Bruce Hansen for his helpful comments and suggestions on my research. His emphasis on the importance of connecting to economic applications has been a guiding principle for my work. I am indebted to Harold Chiang for his encouragement and feedback. Because of his support, I was able to begin my research in econometrics without prior experience and to continue developing it. I would also like to thank Yong Cai for his helpful comments, especially on the first chapter. I am also grateful to Xiaoxia Shi and Kohei Yata for their helpful comments and suggestions. I would like to thank Yinqiu He for serving as a committee member for this dissertation.

My friends and colleagues at the University of Wisconsin–Madison have been a great source of support and inspiration. I am grateful to my co-author, Yuya Shimizu, for his collaboration on the second chapter. I would like to thank my office mates, Andrew Smith and Ryan Mather, for their friendship and support not only in research but also in life. I am also grateful to the econometrics student group for their constructive feedback and discussions. I would like to thank Naoki Aizawa for his support and encouragement from the very beginning of my Ph.D. program.

Finally, I would like to express my deepest gratitude to my family. My father encouraged me to pursue a Ph.D. in economics and has supported me in every way possible. My mother is unfortunately no longer with us, but she raised me to be who I am today, and I hope she would be proud of me. My wife, Hinako, has been my greatest source of support and happiness throughout this journey. We spent the last four years together in Madison, and she has provided me with support, encouragement, and fresh perspectives that I could not have found elsewhere. I dedicate this dissertation to her.

Abstract

This dissertation studies econometric methods for network data, with a focus on estimation and inference under dependence, interference, and endogenous network formation.

The first chapter studies fixed-effect regressions on network data with dependent observations. Fixed effects are widely used to capture unobserved heterogeneity in economic applications, yet most existing methods rely on conditional independence assumptions that may fail in network settings. I consider a framework in which errors arise from both node- and edge-level shocks that are not fully absorbed by the fixed effects. I show that the least-squares estimator of the fixed effects can be inconsistent due to a persistent noise component induced by the dependence structure. To address this issue, I propose inference methods that leverage regression residuals to explicitly account for dependence, and I introduce a bias-correction procedure for estimating the sample variance of the fixed effects.

The second chapter (co-authored with Yuya Shimizu) develops a framework for linear regression analysis of network experiments. Ordinary least squares (OLS) estimators are widely used in this context to estimate spillover effects. We study the causal interpretation of, and inference for, the OLS estimator under both design-based uncertainty from random treatment assignment and sampling-based uncertainty in network links. We show that correlations among regressors capturing exposure to neighbors' treatments can induce contamination bias, preventing OLS from aggregating heterogeneous spillover effects

into a parameter with a clear causal interpretation. We derive the asymptotic distribution of the OLS estimator and propose a network-robust variance estimator.

The third chapter develops estimation and inference methods for dyadic regression models with endogenous link formation, which induces sample selection. Dyadic data often exhibit a large number of zero outcomes. I explicitly model these zero outcomes as arising from a link formation process jointly determined with the outcome process. I propose a consistent semiparametric estimator and show that it exhibits two distinct asymptotic regimes depending on the underlying dependence structure. I then develop a bias-corrected confidence interval and a variance estimator that adapt to both regimes.

Contents

List of Tables

List of Figures

Chapter 1

Network Robust Inference for Fixed-Effect Regressions

1.1 Introduction

Network data have become increasingly prevalent in applied economics, providing additional dimensions that are not available in individual-level data. One important application of network data is the identification and estimation of agent-level fixed effects to capture unobserved heterogeneity that is crucial for understanding economic phenomena. For example, Card et al. (2013) estimate the contribution of firm-level heterogeneity to wage inequality using matched employer-employee data, exploiting job transitions and the corresponding wage changes to separate firm-specific from worker-specific effects.[1] In this setting, firms are linked through workers' mobility, forming a network of firms in which firm-level fixed effects can be estimated from workers' wage changes between origin and destination firms. Such network structures naturally induce cross-sectional dependence in outcomes, as agent-level shocks are shared across linked observations. This paper examines how dependence in network data affects the estimation and inference of fixed effects and proposes a novel inference method that explicitly accounts for the dependence structure.

[1] Other examples include studies in labor: Abowd et al. (1999); Sorkin (2018); Engbom and Moser (2022); Bonhomme et al. (2023), education: Jackson (2013); Bacher-Hicks and Koedel (2023), health: Finkelstein et al. (2016, 2021), trade: Bernard et al. (2022), and industrial organization: Alviarez et al. (2025).

Despite the networked structure of the data, most state-of-the-art econometric methods for fixed-effects estimation rely on the assumption that outcomes are independent, or at most weakly dependent, conditional on fixed effects. For example, Engbom and Moser (2022) estimate changes in workers' and firms' contributions to wage determination using the methodology of Kline et al. (2020), which assumes conditional independence of wages. However, the plausibility of this assumption is application-dependent. For example, in matched employer-employee data, the wage a worker receives from firm j may be correlated with the wage received by another worker at the same firm due to a firm-level shock (e.g., a change in management style) that is not fully captured by time-invariant firm fixed effects. Such shocks can induce correlation in wage differences between firm j and other firms, violating the conventional independence assumption. The presence of network dependence can alter the behavior of fixed-effect estimators and invalidate current econometric approaches.

In this paper, we study fixed-effect regressions on network data with dependent errors. The data is associated with a graph whose nodes correspond to economic agents and whose edges represent interactions; the unit of observation is an edge, and each edge has an associated scalar outcome. We consider a cross-sectional linear model in which an edge outcome is driven by the difference between the incident nodes' fixed effects, edge-level covariates, and an edge-level error term that may be correlated across edges. Our parameters of interest are the node-level fixed effects, which capture unobserved heterogeneity at the node level. For example, in labor economics applications, the graph's nodes are firms and an edge links a pair of firms when a worker moves between them; the edge outcome is the wage change for that mover, determined in part by the difference in the two firms' fixed effects, which capture firm-level wage premiums and underlying productivity differences.

In this high-dimensional setting, the dependence across outcomes poses significant challenges for inference on fixed effects. To deal with these challenges and for analytical tractability, we assume that errors are generated by both node-level and edge-level shocks, which are not controlled for by the fixed effects, introducing conditional dependence among the errors. This model allows for dependence between two edges that share at least one node, capturing for example the correlation in movers' wage changes sharing the same origin or destination firm due to exposure to firm-level shocks. Although higher-order dependence between two edges that do not directly share a node is ruled out, this dependence structure allows for independence as a special case.

In this setting, we find that both the connectivity of the graph and the strength of the dependence structure jointly determine the behavior of the least-squares estimator of fixed effects. The underlying graph must be sufficiently well-connected to consistently estimate the fixed effects; otherwise, there is insufficient information to identify them. This issue, known as "limited mobility" (e.g., Andrews et al., 2008; Bonhomme et al., 2023), is well documented in the literature. However, when allowing for dependence, a new challenge emerges: if the graph is too densely connected, the resulting dependence can become too strong, which in turn hinders precise estimation of the fixed effects. As a result, strategies aimed at improving only connectivity by removing bottlenecks or grouping fixed effects (as in Andrews et al., 2008; Bonhomme et al., 2019) may not be effective in this context because they can increase graph density and further strengthen dependence.

Our main results are as follows. First, we derive a first-order approximation for the least-squares estimator of fixed effects, showing that the estimator can be inconsistent due to a persistent noise term arising from node-level shocks that induce dependence and fail to average out. This inconsistency poses a fundamental challenge for inference, as the usual asymptotic normal approximation does not hold and conventional adjustments to standard errors for dependence do not restore validity. To address this issue, we propose a non-standard inference method à la Conley and Taber (2011), which exploits information in the regression residuals to estimate the distribution of the persistent noise term. We show that this estimated distribution is uniformly consistent for the true one. Based on this result, we construct confidence intervals for individual fixed effects and joint hypothesis tests for groups of fixed effects, and we prove that these procedures are asymptotically valid. Simulation exercises confirm that our inference methods perform well in finite samples when connectivity and dependence are sufficiently balanced.

Second, we introduce a new bias-correction method for estimating moments of fixed effects under the dependence structure. Rather than focusing on each fixed effect individually, it is often of interest to estimate distributional features of the fixed effects, such as their variance, which captures the extent of unobserved heterogeneity. However, the plug-in estimator for these distributional features can be substantially biased due to the nonlinearity of the transformation and the presence of estimation error. Focusing specifically on the variance of fixed effects, we characterize the bias inherent in the plug-in estimator and decompose it into two components: one due to estimation error under independence, or the "limited mobility bias," and the other due to the dependence structure. We show that the latter component can dominate the former when both limited mobility and dependence are moderate. We then develop a new bias-corrected estimator that primarily addresses the network-dependence bias and prove its consistency. Simulation results demonstrate that our bias-corrected estimator outperforms the plug-in estimator and existing bias-correction methods that adjust only for limited mobility bias.

We demonstrate the usefulness of our results in the context of the Veneto Worker History (VWH) file, a comprehensive matched employer-employee dataset from the Veneto region in Italy. We construct a mobility network of firms based on workers' job transitions between 1999 and 2001 and estimate firm fixed effects from workers' wage changes between origin and destination firms in a two-period AKM framework (Abowd et al., 1999). We then implement our inference and bias-correction procedures. For inference, we construct a novel confidence interval for the fixed effect of the most central firm in the mobility network and find that this firm does not exhibit a significantly larger effect than the average firm, reversing the conclusion obtained under conventional confidence intervals that assume independence. This result aligns with recent arguments that non-wage amenities play a significant role in attracting workers (Sorkin, 2018; Mas, 2025) and highlights the importance of accounting for dependence. For variance estimation, our bias-corrected estimator of the variance of firm effects is about 30% smaller than the plug-in estimator and marginally smaller than the bias-corrected estimator of Andrews et al. (2008), which assumes independence. These findings suggest that the plug-in estimator substantially overestimates dispersion in firm effects due to network dependence and that accommodating dependence is crucial for accurate estimation of the variance.

Our paper contributes to the literature on fixed-effects estimation in network data (Abowd et al., 1999; Andrews et al., 2008; Graham, 2017; Jochmans and Weidner, 2019; Bonhomme et al., 2019; Kline et al., 2020; Bonhomme et al., 2023). This body of work generally assumes conditional independence of outcomes. The most closely related work is Jochmans and Weidner (2019), which establishes a first-order approximation for the least-squares estimator of fixed effects under at most weak dependence, and its asymptotic normality under independence. In contrast, our paper introduces a strong dependence structure and provides a first-order approximation of the estimator, along with alternative inference methods that explicitly account for dependence. Relatedly, Bonhomme et al. (2019) propose grouping fixed effects to make mobility networks denser and improve estimation, allowing for network dependence across groups. Unlike their approach, we do not require grouping of fixed effects and instead develop inference methods that can potentially test the validity of such grouping strategies.

The estimation of moments of fixed effects is also a key topic in this literature. Andrews et al. (2008) provide a bias-correction approach for the estimation of variance components under independence and homoskedasticity. Kline et al. (2020) extend this method to accommodate heteroskedasticity, maintaining the independence assumption or allowing at most weak cluster dependence. Our paper is the first to introduce a bias-correction approach for variance components estimation under the strong dependence structure, showing that the bias stemming from network dependence is non-negligible and can dominate the bias addressed by the existing methods. Although our current focus is on the sample variance of fixed effects, our approach can be extended to other distributional features such as covariance between worker and firm effects.

We also relate to the literature on two-way/dyadic cluster-robust inference (Cameron et al., 2011; Cameron and Miller, 2014; Tabord-Meehan, 2019; Verdier, 2020; Davezies et al., 2021; Menzel, 2021; Chiang et al., 2024). While our dependence structure shares features with these papers, our focus differs: we center the analysis on inference of fixed effects, which this literature typically treats as nuisance parameters. For example, Verdier (2020) studies inference for common regression coefficients in matched employer-employee/student-teacher data, allowing outcomes to be correlated if they share a firm/teacher, similarly to our dependence structure. In that setting, they establish the asymptotic normality for the estimator of common coefficients and propose a standard error that is robust to both the dependence structure and to high-dimensional fixed effects. By contrast, under our dependence structure, inference on fixed effects themselves faces a fundamental difficulty: the least-squares estimator is inconsistent and its asymptotic distribution is non-normal. We develop valid inference procedures that address these challenges.

1.2 Model

In this section, we introduce the data structure and the linear model we consider in this paper. We then introduce the dependence structure of error terms, along with a useful decomposition to understand the behavior of the estimator.

1.2.1 Setup

We consider network data based on a directed multigraph $\mathcal{G} = (V, E, s, t)$, consisting of a node set V, an edge set E, and source and target functions $s, t : E \to V$. The node set $V = \{1, ..., n\}$ contains n nodes, and the edge set $E = \{1, ..., m\}$ contains m edges. We use e to denote a typical edge in E. For each edge $e \in E$, the source function $s(e)$ specifies its origin node, while the target function $t(e)$ specifies its destination node.

Because multiple edges can exist between the same pair of nodes, we define an edge subset $E_{(i,j)} = E_{(j,i)} \subset E$ as the set of edges whose two endpoints are nodes i and j:

$$E_{(i,j)} = \{e \in E : s(e) = i, t(e) = j \text{ or } s(e) = j, t(e) = i\} \text{ for each } i, j \in V.$$

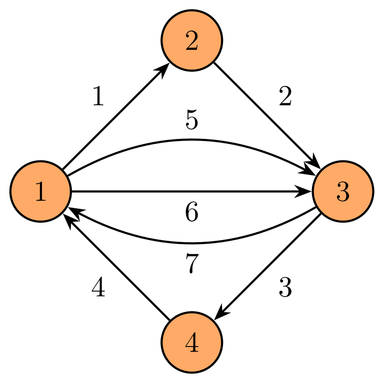
We allow $E_{(i,j)}$ to be empty when there are no edges between nodes i and j. By convention, we do not allow self-loops, i.e., $E_{(i,i)} = \emptyset$ for all $i \in V$. Additionally, for each $i \in V$, let $E_i^s = \{e \in E : s(e) = i\}$ be the set of edges that originate from node i and $E_i^t = \{e \in E : t(e) = i\}$ be the set of edges that terminate at node i. Then, $E_i = E_i^s \cup E_i^t$ is the set of edges incident to node i. We define the out-degree of node i as $d_i^s = |E_i^s|$, the number of edges flowing out of node i, and the in-degree as $d_i^t = |E_i^t|$, the number of edges flowing into node i. We refer to $d_i = d_i^s + d_i^t$ as the degree of node i, which is the number of edges incident to node i.

To illustrate the notation, consider the directed multigraph $\mathcal{G}$ shown in Figure 1.1, with nodes $V = \{1, 2, 3, 4\}$ and edges $E = \{1, ..., 7\}$. Focus on node 1: this node is the source of edges $1 \in E_{(1,2)} = \{1\}, 5, 6 \in E_{(1,3)} = \{5, 6, 7\}$, and the target of edges $4 \in E_{(1,4)} = \{4\}$ and $7 \in E_{(1,3)}$. Consequently, we have

$$s(1) = s(5) = s(6) = t(4) = t(7) = 1,$$

and so $E_1^s = \{1, 5, 6\}$ and $E_1^t = \{4, 7\}$. The out-degree of node 1 is $d_1^s = 3$, the in-degree is $d_1^t = 2$, and thus the degree is $d_1 = d_1^s + d_1^t = 5$.

Figure 1.1: A directed multigraph $\mathcal{G}$ with $n = 4$ nodes and $m = 7$ edges.



Note: Arrows indicate the direction of edges. For example, edge 1 originates from node 1 and terminates at node 2.

The following matrices associated with graph $\mathcal{G}$ are crucial for our analysis. First, the $n \times n$ (symmetric) adjacency matrix $\mathbf{A}$ is defined as

$$\mathbf{A}_{i,j} = |E_{(i,j)}|,$$

for $i, j \in V$. Note that $\mathbf{A}$ is the adjacency matrix of the undirected graph derived from the directed graph $\mathcal{G}$ by ignoring the direction of edges and counting the number of edges between each pair of nodes. The degree of node i is also defined as $d_i = \sum_{j \neq i} \mathbf{A}_{i,j}$. Let $\mathbf{D}$ be the $n \times n$ diagonal matrix with degrees d_i placed along its diagonal.

Next, define the $m \times n$ incidence matrix $\mathbf{B}$ as follows:

$$\mathbf{B}_{e,i} = \begin{cases} 1 \text{ if } t(e) = i \\ -1 \text{ if } s(e) = i \\ 0 \text{ otherwise} \end{cases}$$

for $e \in E$ and $i \in V$. The incidence matrix $\mathbf{B}$ maps a node-level vector to an edge-level vector of differences. For example, for a vector $\boldsymbol{v} \in \mathbb{R}^n$, the e-th element of $\mathbf{B}\boldsymbol{v}$ equals the difference $v_{t(e)} - v_{s(e)}$. Finally, these matrices are connected through the graph Laplacian $\mathbf{L}$, defined as

$$\mathbf{L} = \mathbf{B}'\mathbf{B} = \mathbf{D} - \mathbf{A}.$$

As we will see later, the graph Laplacian plays a crucial role in our analysis.

Remark 1.1. *We focus on unweighted directed multigraphs for simplicity, but our analysis can be extended to weighted graphs as in Jochmans and Weidner (2019). In a weighted graph, each edge $e \in E$ is associated with a positive weight $w_e > 0$. The adjacency matrix $\mathbf{A}$ is then defined as $\mathbf{A}_{i,j} = \sum_{e \in E_{(i,j)}} w_e$ for $i, j \in V$, and the incidence matrix $\mathbf{B}$ is defined as $\mathbf{B}_{e,i} = \sqrt{w_e}$ if $t(e) = i$, $\mathbf{B}_{e,i} = -\sqrt{w_e}$ if $s(e) = i$, and $\mathbf{B}_{e,i} = 0$ otherwise. Focusing on unweighted graphs simplifies the notation and exposition, and does not affect the generality of our results when applied to Example 1.1 and our empirical application where each edge corresponds to a single worker transition.*

1.2.2 Linear Model

Suppose that each edge $e \in E$ is associated with a scalar outcome y_e and a p-dimensional covariate X_e. These outcomes and covariates are stacked into the m-dimensional vector $\mathbf{y} = (y_e)_{e=1}^m$ and the $m \times p$ matrix $\mathbf{X} = (X_e)_{e=1}^m$, respectively. We consider the following linear model:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\alpha} = (\alpha_i)_{i=1}^n$ is the n-dimensional vector of node-level fixed effects, $\boldsymbol{\beta}$ is the p-dimensional vector of coefficients for the covariates, and $\boldsymbol{\epsilon} = (\epsilon_e)_{e=1}^m$ is the m-dimensional vector of edge-level errors. For now, we focus on estimation of $\boldsymbol{\alpha}$ and leave out covariates or equivalently treat $\boldsymbol{\beta}$ as known and redefine $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ as $\mathbf{y}$.[2] Then, the model can be simplified to

$$\mathbf{y} = \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{1.1}$$

which we refer to as our regression model.

Element-wise, the regression model (1.1) is written as

$$y_e = \alpha_{t(e)} - \alpha_{s(e)} + \epsilon_e,$$

for $e \in E$. Thus, each outcome is driven by the difference between the destination node fixed effect $\alpha_{t(e)}$ and the origin node effect $\alpha_{s(e)}$. The following example illustrates how such differences in fixed effects naturally arise in

[2] Appendix A.1 provides formal results on the case where $\boldsymbol{\beta}$ is unknown and needs to be estimated.

economic applications.

Example 1.1. *(Two-Period AKM model): An important example of the model* (1.1) *arises from the so-called AKM model (Abowd et al., 1999), which links worker and firm effects to wages. Leaving out covariates, in this model, worker* g *is employed by firm* $J(g,t)$ *at time* t*, and the (log) wage of worker* g *at time* t *is given by*

$$w_{g,t} = \phi_g + \alpha_{J(g,t)} + u_{g,t}, \tag{1.2}$$

where ϕ_g *is the worker fixed effect,* $\alpha_{J(g,t)}$ *is the firm fixed effect corresponding to the worker's employer at time* t*, and* $u_{g,t}$ *is the idiosyncratic error. Suppose that the economy lasts for two periods, as considered in Section 2 of Kline et al. (2020). Then, 'movers,' i.e., workers* g *with* $J(g,1) \neq J(g,2)$*, form an edge* e *by transitioning from firm* $s(e) = J(g,1)$ *to firm* $t(e) = J(g,2)$*. Then, we can write the wage difference as*

$$\underbrace{w_{g,2} - w_{g,1}}_{\equiv y_e} = \underbrace{\alpha_{J(g,2)} - \alpha_{J(g,1)}}_{=\alpha_{t(e)} - \alpha_{s(e)}} + \underbrace{u_{g,2} - u_{g,1}}_{\equiv \epsilon_e},$$

which corresponds to the model (1.1).

Theorem 1 of Jochmans and Weidner (2019) shows that when the graph $\mathcal{G}$ is connected and the fixed effects $\boldsymbol{\alpha}$ are normalized such that $\sum_{i \in V} \alpha_i d_i = 0$, the least-squares estimator for $\boldsymbol{\alpha}$ is uniquely given by:

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{B}'\mathbf{B})^{*}\mathbf{B}'\mathbf{y} = \mathbf{L}^{*}\mathbf{B}'\mathbf{y}, \tag{1.3}$$

where, for any $n \times n$ matrix $\boldsymbol{C}$, $\boldsymbol{C}^{*}$ denotes the pseudo-inverse defined as:

$$\boldsymbol{C}^{*} = \boldsymbol{D}^{-1/2}(\boldsymbol{D}^{-1/2}\boldsymbol{C}\boldsymbol{D}^{-1/2})^{+}\boldsymbol{D}^{-1/2}$$

with $^{+}$ representing the Moore-Penrose inverse. Throughout the paper, we maintain such assumptions on the graph structure and the normalization of fixed effects that justifies the use of the estimator (1.3):

Assumption 1.1. *The graph $\mathcal{G}$ is connected in the following sense: for any $i, j \in V$, there exists a path between i and j on the undirected graph derived from $\mathcal{G}$. Furthermore, the fixed effects $\boldsymbol{\alpha}$ are normalized such that $\sum_{i \in V} \alpha_i d_i = 0$.*

If the graph is not connected, the normalization condition must be applied separately to each connected component of the graph, and the subsequent analysis can then be applied to each component individually. Therefore, it will not be possible to compare fixed effects across different components. In practice, researchers typically focus on the largest connected component if it constitutes a significant portion of the entire graph (e.g., Engbom and Moser, 2022).

The specific normalization employed in Assumption 1.1 is not crucial for our analysis, and alternative normalizations can be used to obtain similar results. For example, we could normalize the fixed effects by imposing $\sum_{i \in V} \alpha_i = 0$, which leads to a least-squares estimator with a different pseudo-inverse:

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{B}'\mathbf{B})^{+}\mathbf{B}'\mathbf{y}.$$

Another common normalization is to set $\alpha_i = 0$ for a specific node $i \in V$. In this case, the least-squares estimator can be written as

$$\widehat{\boldsymbol{\alpha}}_{-i} = (\mathbf{B}'_{-i}\mathbf{B}_{-i})^{-1}\mathbf{B}'_{-i}\mathbf{y},$$

where $\widehat{\boldsymbol{\alpha}}_{-i}$ is the $(n-1)$-dimensional vector and $\mathbf{B}_{-i}$ is the $m \times (n-1)$ matrix obtained by removing the i-th column of $\mathbf{B}$. Although this normalization is often used in practice, it complicates theoretical analysis by breaking the direct link to graph-related objects such as the graph Laplacian $\mathbf{L}$. As discussed in Kline (2024), one can nonetheless relate $(\mathbf{B}'_{-i}\mathbf{B}_{-i})^{-1}$ to $\mathbf{L}^*$ using the results of Bozzo (2013). We leave the exploration of that connection to future work, focusing here on the normalization in Assumption 1.1.

Remark 1.2. *Note that randomness in our setup arises solely from the error vector $\boldsymbol{\epsilon}$; the fixed effects $\boldsymbol{\alpha}$ and the graph $\mathcal{G}$, and hence the matrices $\mathbf{A}, \mathbf{B}, \mathbf{D}, \mathbf{L}$, are treated as fixed. Accordingly, all probabilistic statements are understood as conditional on $\boldsymbol{\alpha}$ and $\mathcal{G}$. This conditional approach is standard in the literature on fixed-effect estimation with network data (e.g., Jochmans and Weidner, 2019; Kline et al., 2020; Bonhomme et al., 2023).*

1.2.3 Dependence Structure

Next, we introduce a model of dependence for the edge-level errors ϵ_e. Previous literature has typically imposed independence of the error terms across edges, or at most weak dependence, meaning that the covariance matrix of $\boldsymbol{\epsilon}$ is nearly diagonal. However, since outcomes and errors are defined at the edge level, more realistic dependence structures would allow for dependence between edges that share a node, which is ruled out by the independence or weak dependence assumptions.

Here, we focus on a first-order strong dependence structure: we allow for sources of dependence including (i) node-level shocks, which induce dependence between edges sharing a node; and (ii) edge-level shocks, which induce dependence between edges in the same edge subset. These shocks are shared across edges that either share a node or belong to the same edge subset, respectively, and can generate strong dependence among the errors at the first order.

Specifically, for each edge $e \in E$, we assume the following structure for ϵ_e:

$$\epsilon_e = f(U_{s(e)}, U_{t(e)}, V_e); \quad \mathbb{E}[\epsilon_e] = 0, \tag{1.4}$$

where f is a measurable function unknown to the researcher and $U_{s(e)}, U_{t(e)}$ and V_e are random vectors. Specifically, $U_{s(e)}$ and $U_{t(e)}$ represent node-level shocks associated with the source and target nodes of edge e, while V_e is an edge-level shock associated with edge e. Since our analysis does not depend on their dimension, in the following, we treat each U_i and V_e as one-dimensional random variables unless otherwise specified. Importantly, we do not require prior knowledge of the form of f: it may be additive, interactive, or of any other form.

We impose the following structure on (1.4) to facilitate the analysis:

Assumption 1.2. *The error term* $\boldsymbol{\epsilon} = (\epsilon_e)_{e=1}^m$ *satisfies the following conditions:*

(i) For each $e \in E$, ϵ_e *follows* (1.4), $\min_{e\in E} \mathbb{E}[\epsilon_e^2] \geq C_1 > 0$, *and* $\max_{e\in E} \mathbb{E}[\epsilon_e^4] \leq C_2 < \infty$ *for some absolute constants* $C_1, C_2 > 0$.

(ii) The node-level shocks $U_i, i \in V$ *are independently and identically distributed. The edge-level shocks* $V_e, e \in E$ *satisfy the following independence structure:*

$$V_e \perp\!\!\!\perp V_{e'} \text{ if } \{s(e), t(e)\} \neq \{s(e'), t(e')\}.$$

(iii) $(U_i)_{i=1}^n$ *is independent of* $(V_e)_{e=1}^m$

Assumption 1.2 imposes a specific dependence structure on the error terms $\boldsymbol{\epsilon}$. Part (i) ensures that the errors are non-degenerate and possess uniformly bounded fourth moments. Part (ii) requires that the node-level shocks U_i be homogeneously distributed across nodes, which is a strong restriction but simplifies the analysis. In contrast, the edge-level shocks V_e are allowed to have heterogeneous distributions and may exhibit correlation among them if two edges belong to the same edge subset $E_{(i,j)}$. For example, one might write the covariance structure of V_e as follows:

$$\text{Cov}(V_e, V_{e'}) = \begin{cases} v_e^2 \text{ if } e = e'; \\ v_{e,e'} \text{ if } e, e' \in E_{(i,j)} \text{ for some } i, j \in V; \\ 0 \text{ otherwise,} \end{cases}$$

where v_e^2 and $v_{e,e'}$ can be different across $e, e' \in E$. Finally, part (iii), the mutual independence between the node-level shocks $(U_i)_{i \in V}$ and the edge-level shocks $(V_e)_{e \in E}$, is a common assumption in the network and two-way clustering literature (e.g., Graham, 2017; Menzel, 2021; Chiang et al., 2024). This assumption is not as restrictive as it might seem because the function f in (1.4) can incorporate both U_i and V_e in a nonlinear fashion, for example, by allowing them to interact with each other.

An important implication of the dependence structure in (1.4) under Assumption 1.2 is that the error terms $\boldsymbol{\epsilon}$ are dependent across edges either through (i) shared node-level shocks U_i when two edges share one common node, or (ii) correlation induced by edge-level shocks when two edges belong to the same edge subset $E_{(i,j)}$. Consequently, we can write the covariance structure of $\boldsymbol{\epsilon}$ as

$$\text{Cov}(\epsilon_e, \epsilon_{e'}) = \begin{cases} \sigma_e^2 \text{ if } e = e'; \\ \sigma_{e,e'} \text{ if } e \neq e' \text{ and } e, e' \in E_i \text{ for some } i \in V; \\ 0 \text{ otherwise} \end{cases}, \tag{1.5}$$

for $e, e' \in E$. Note that we are allowing for heteroskedasticity in the error terms, as the variance σ_e^2 and covariance $\sigma_{e,e'}$ can differ across edges due to the heterogeneity of the edge-level shocks V_e.

Note that the dependence structure in (1.4) rules out higher-order correlations among edges that are indirectly connected through common neighbors. While this restriction simplifies the analysis, it is flexible enough to accommodate a broad range of first-order correlations, and it subsumes the traditional assumption of independence as a special case.

We can connect the dependence structure in (1.4) to the AKM model in Example 1.1:

Example 1.2. *(Two-Period AKM model, continued). Consider the setting described in Example 1.1. Recall that the error term in this model is given by* $\epsilon_e = u_{g,2} - u_{g,1}$ *for a mover* g*. This error term can absorb misspecification in the linear model* (1.2)*, such as time-varying firm effects (Engbom et al., 2023; Lachowska et al., 2023) and match-specific heterogeneity (Bonhomme et al., 2019) that depends on both destination and origin firms, as captured in search models with on-the-job search (e.g., Postel-Vinay and Robin, 2002; Bagger et al., 2014; Di Addario et al., 2023). For instance, if firm effects vary over time as* $\alpha_{J(g,t),t} = \alpha_{J(g,t)} + U_{J(g,t),t}$ *for* $t = 1, 2$*, where* $U_{J(g,t),t}$ *is a firm-level shock at time* t*, then the error term contains* $U_{J(g,2),2} - U_{J(g,1),1}$ *and can be expressed in the form of* (1.4) *by setting* $U_i = (U_{i,1}, U_{i,2})$ *for each firm* $i \in V$ *and* $V_e = u_{g,2} - u_{g,1}$*. Similarly, if there is match-specific heterogeneity that depends on interactions among the worker, the origin firm, and the destination firm, we may specify the error term as* $\epsilon_e = U_{J(g,1)} \times U_{J(g,2)} - V_e$*, where* $U_{J(g,1)}$ *and* $U_{J(g,2)}$ *are shocks to firms' willingness to pay for the position, and* V_e *is a mover-specific preference shock; this specification is a special case of* (1.4)*. The* *interaction between $U_{J(g,1)}$ and $U_{J(g,2)}$ captures the idea that when both firms' willingness to pay is high, competition for the worker intensifies, raising the wage offer.*

1.2.4 Decomposition

To isolate the impact of the dependence structure on the fixed-effect estimator $\widehat{\boldsymbol{\alpha}}$, we propose the following decomposition. For each $e \in E$, define the functions:

$$\tau_e^s(\cdot) \equiv \mathbb{E}[\epsilon_e | U_{s(e)} = \cdot], \quad \tau_e^t(\cdot) \equiv \mathbb{E}[\epsilon_e | U_{t(e)} = \cdot].$$

These functions capture, respectively, the effects of the origin and destination node-level shocks $U_{s(e)}$ and $U_{t(e)}$ on the error term ϵ_e. Note that they are not necessarily equal; the function f in (1.4) can be asymmetric in its arguments $U_{s(e)}$ and $U_{t(e)}$.

For example, suppose we have

$$\epsilon_e = U_{t(e)} U_{s(e)} + U_{t(e)} - 2U_{s(e)},$$

and assume that $\mathbb{E}[U_i] = 1$. Then we can show that

$$\tau_e^s(u) = -u + 1 \text{ and } \tau_e^t(u) = 2u - 2,$$

for any $u \in \mathcal{U}$, the support of U_i. This example illustrates that the influence of the node-level shocks on the error term can differ depending on whether the shock comes from the origin or destination node.

To reduce dimensionality, we impose the following homogeneity assumption on these functions:

Assumption 1.3. *For any $e, e' \in E$, we have*

$$\tau_e^s = \tau_{e'}^s \equiv \tau^s, \quad \tau_e^t = \tau_{e'}^t \equiv \tau^t.$$

This assumption ensures that τ_e^t and τ_e^s do not depend on the particular edge e. In practice, under Assumption 1.2, this condition is satisfied if V_e is identically distributed across edges. More generally, it is satisfied in a broad class of models where there is a certain separability between the node-level shocks U_i and the edge-level shocks V_e even when V_e is heterogeneously distributed across edges.

A straightforward case is the additively separable model:

$$\epsilon_e = U_{s(e)} + U_{t(e)} + V_e,$$

which implies that $\tau^s(u) = \tau^t(u) = u$ for any $u \in \mathcal{U}$. We can also allow for interactions between U_i and V_e, provided V_e is not transformed nonlinearly. For instance, if we assume that $\mathbb{E}[U_i] = 0, \mathbb{E}[V_e] = 1$ and consider the model

$$\epsilon_e = (U_{s(e)} + U_{t(e)}) \times V_e,$$

then it follows that $\tau^s(u) = \tau^t(u) = u$ for any $u \in \mathcal{U}$.

In contrast, Assumption 1.3 is violated when heteroskedastic V_e is nonlinearly transformed in the interaction. For example, if we have

$$\epsilon_e = (U_{s(e)} + U_{t(e)})V_e^2$$

then

$$\mathbb{E}[\epsilon_e | U_{s(e)} = u] = \mathbb{E}[\epsilon_e | U_{t(e)} = u] = u \times v_e^2,$$

which depends on the edge e through $v_e^2 = \mathbb{E}[V_e^2]$.

Notice that we can decompose the error term ϵ_e as follows:

$$\epsilon_e = \tau^s(U_{s(e)}) + \tau^t(U_{t(e)}) + \epsilon_e - \tau^s(U_{s(e)}) - \tau^t(U_{t(e)}).$$

This decomposition is useful because it expresses the error term ϵ_e as a sum of two uncorrelated components: the first component is the sum of the node-level shocks $\tau^s(U_{s(e)})$ and $\tau^t(U_{t(e)})$, while the second component is the residual error term $\epsilon_e - \tau^s(U_{s(e)}) - \tau^t(U_{t(e)})$, and these two components are uncorrelated:

$$\begin{aligned}
&\mathbb{E}[\tau^o(U_{o(e)}) \times (\epsilon_e - \tau^s(U_{s(e)}) - \tau^t(U_{t(e)}))] \\
&= \mathbb{E}[\tau^o(U_{o(e)})\epsilon_e] - \mathbb{E}[(\tau^o(U_{o(e)}))^2] \because \mathbb{E}[\tau^s(U_{s(e)})\tau^t(U_{t(e)})] = 0 \\
&= \mathbb{E}[\tau^o(U_{o(e)})\mathbb{E}[\epsilon_e | U_{o(e)}]] - \mathbb{E}[(\tau^o(U_{o(e)}))^2] \\
&= 0,
\end{aligned}$$

for $o \in \{s, t\}$.

In matrix form, we can write this decomposition as:

$$\boldsymbol{\epsilon} = \mathbf{F}^s\boldsymbol{\tau}^s + \mathbf{F}^t\boldsymbol{\tau}^t + \left(\boldsymbol{\epsilon} - \mathbf{F}^s\boldsymbol{\tau}^s - \mathbf{F}^t\boldsymbol{\tau}^t\right),$$

where $\mathbf{F}^s$ and $\mathbf{F}^t$ are the $m \times n$ matrices defined as:

$$\mathbf{F}^s_{e,i} = \begin{cases} 1 \text{ if } s(e) = i \\ 0 \text{ otherwise} \end{cases}, \quad \mathbf{F}^t_{e,i} = \begin{cases} 1 \text{ if } t(e) = i \\ 0 \text{ otherwise} \end{cases},$$

and $\boldsymbol{\tau}^s = (\tau^s(U_i))_{i \in V}$ and $\boldsymbol{\tau}^t = (\tau^t(U_i))_{i \in V}$ are the n-dimensional vectors of node-level origin and destination shocks, respectively.

Now we can decompose $\widehat{\boldsymbol{\alpha}}$ as follows:[3]

$$\widehat{\boldsymbol{\alpha}} = \mathbf{L}^*\mathbf{B}'\mathbf{y} = \boldsymbol{\alpha} + \mathbf{L}^*\mathbf{B}'(\mathbf{F}^s\boldsymbol{\tau}^s + \mathbf{F}^t\boldsymbol{\tau}^t) + \mathbf{L}^*\mathbf{B}'(\boldsymbol{\epsilon} - \mathbf{F}^s\boldsymbol{\tau}^s - \mathbf{F}^t\boldsymbol{\tau}^t). \tag{1.6}$$

The first term in (1.6) represents the true fixed effects, the second term captures the stochastic shock arising from the node-level dependence structure, and the third term is the remaining noise term net of the node-level shocks.

1.2.5 Variance Bound

Since the second term and third term in (1.6) are orthogonal to each other, i.e., uncorrelated element-wise, the covariance matrix of $\widehat{\boldsymbol{\alpha}}$ is given by:

$$\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}) = \mathbf{L}^*\mathbf{B}'\boldsymbol{\Omega}_1\mathbf{B}\mathbf{L}^* + \mathbf{L}^*\mathbf{B}'\boldsymbol{\Omega}_2\mathbf{B}\mathbf{L}^*, \tag{1.7}$$

[3]Note that $\mathbf{L}^*\mathbf{L}\boldsymbol{\alpha} = \boldsymbol{\alpha}$ as $\mathbf{L}^*\mathbf{L}$ works as a projection matrix on the null space of $\mathbf{d}$ and $\boldsymbol{\alpha}$ is in the null space of $\mathbf{d}$: $\mathbf{d}'\boldsymbol{\alpha} = 0$ under Assumption 1.1

where

$$\begin{aligned}\mathbf{\Omega}_1 &= \mathbb{E}[\tau^s(U_i)^2]\mathbf{F}^s(\mathbf{F}^s)' \\ &\quad + \mathbb{E}[\tau^s(U_i)\tau^t(U_i)](\mathbf{F}^s(\mathbf{F}^t)' + \mathbf{F}^t(\mathbf{F}^s)') + \mathbb{E}[\tau^t(U_i)^2]\mathbf{F}^t(\mathbf{F}^t)', \\ \mathbf{\Omega}_2 &= \mathrm{Cov}(\boldsymbol{\epsilon} - \mathbf{F}^s\boldsymbol{\tau}^s - \mathbf{F}^t\boldsymbol{\tau}^t).\end{aligned}$$

The matrix $\mathbf{\Omega}_1$ in the first term can be bounded as follows:

$$\mathbf{\Omega}_1 \precsim \sigma_\tau^2 \times \mathbf{F}\mathbf{F}' \precsim \sigma_\tau^2 \times \lambda_{n,F} \times \mathbf{I}_m,$$

where $\sigma_\tau^2 = \max_{o\in\{s,t\}} \mathbb{E}[\tau^o(U_i)^2]$, $\mathbf{F} = \mathbf{F}^s + \mathbf{F}^t = abs(\mathbf{B})$ is the signless incidence matrix, $\lambda_{n,F}$ is the largest eigenvalue of $\mathbf{F}\mathbf{F}'$, and $\precsim$ indicates that the right-hand side multiplied by some constant upper bounds the left-hand side in the positive semidefinite ordering. Thus, the first term in (1.7) is bounded by

$$\mathbf{L}^*\mathbf{B}'\mathbf{\Omega}_1\mathbf{B}\mathbf{L}^* \precsim \sigma_\tau^2 \times \lambda_{n,F} \times \mathbf{L}^*.$$

This $\lambda_{n,F}$ captures the strength of the dependence structure induced by the node-level shocks U_i as $\mathbf{F}\mathbf{F}'$ is the version of the adjacency matrix of the line graph derived from $\mathcal{G}$: non-diagonal entries of edge-edge matrix $\mathbf{F}\mathbf{F}'$ are non-zero if two edges share at least one common node. Thus, if many edges share nodes in common and dependence is strong, $\lambda_{n,F}$ will be large and vice versa.

The matrix $\mathbf{\Omega}_2$ in the second term is block-diagonal because of the cluster-like dependence structure induced by V_e in Assumption 1.2 (ii):

$$\mathbf{\Omega}_2 = \begin{pmatrix} \ddots & & \\ & \mathbf{\Omega}_{(i,j)} & \\ & & \ddots \end{pmatrix},$$

$$\mathbf{\Omega}_{(i,j)} = \begin{pmatrix} \sigma^2_{e_k} & \sigma_{e_k,e_l} & \cdots \\ \sigma_{e_k,e_l} & \sigma^2_{e_l} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} - (\mathbb{E}[\tau^s(U_i)^2 + \tau^t(U_i)^2])\boldsymbol{\iota}_{|E_{(i,j)}|}\boldsymbol{\iota}'_{|E_{(i,j)}|},$$

for each (i,j) such that $E_{(i,j)} \neq \emptyset$. Since the largest eigenvalue of $\mathbf{\Omega}_2$ is the maximum of the largest eigenvalues of each block $\mathbf{\Omega}_{(i,j)}$, we have

$$\mathbf{\Omega}_2 \precsim \widetilde{\sigma}^2_n \times \mathbf{I}_m,$$

where $\widetilde{\sigma}^2_n \equiv \max_{i,j} \mathbf{A}_{i,j} \max_{e \in E_{(i,j)}} (\sigma^2_e - \mathbb{E}[\tau^s(U_i)^2 + \tau^t(U_i)^2])$. Thus, the second term in (1.7) is bounded by

$$\mathbf{L}^*\mathbf{B}'\mathbf{\Omega}_2\mathbf{B}\mathbf{L}^* \precsim \widetilde{\sigma}^2_n \times \mathbf{L}^*.$$

Hence, the upper bound on the variance of the fixed-effect estimator is proportional to

$$\text{Cov}(\widehat{\boldsymbol{\alpha}}) \precsim (\sigma^2_\tau \times \lambda_{n,F} + \widetilde{\sigma}^2_n) \times \mathbf{L}^*.$$

Note that $\mathbf{L}^*$ is proportional to the covariance matrix in the absence of the dependence structure and is analyzed in Jochmans and Weidner (2019): it is inversely proportional to the connectivity of the graph. Thus, the covariance of the estimator is amplified by the strong dependence structure, as captured by $\sigma_\tau^2 \times \lambda_{n,F}$, in addition to the weak dependence structure, as captured by $\widetilde{\sigma}_n^2$.

To measure the connectivity, as in Jochmans and Weidner (2019), let $\lambda_{2,L}$ be the second smallest eigenvalue of the normalized Laplacian

$$\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}.$$

Additionally, for each $i \in V$, define

$$h_i = \left(\frac{1}{d_i} \sum_{j \neq i} \frac{\mathbf{A}_{i,j}^2}{d_j} \right)^{-1},$$

which is a harmonic mean of the degrees of the neighbors of node i. These two objects reflect the graph's connectivity: $\lambda_{2,L}$ captures the global connectivity, while h_i characterizes the local connectivity around node i.

The following proposition provides a finite-sample upper bound on each diagonal element of $\mathrm{Cov}(\widehat{\boldsymbol{\alpha}})$.

Proposition 1.1. *Under Assumptions 1.1-1.3, the variance of each fixed-effect estimator $\widehat{\alpha}_i$ is bounded as follows:*

$$\mathrm{Var}(\widehat{\alpha}_i) \leq \left(\widetilde{\sigma}_n^2 + \lambda_{n,F}\sigma_\tau^2\right) \times \left\{ \frac{1}{d_i}\left(1 + \frac{1}{\lambda_{2,L} h_i}\right) - \frac{2}{\sum_{j \in V} d_j} \right\}.$$

Proposition 1.1 highlights that the worst-case variance of the fixed-effect estimator depends critically on the interplay between graph connectivity (as captured by $\lambda_{2,L}$) and the density of the line graph (as measured by $\lambda_{n,F}$). In particular, when the node-level shocks U_i are degenerate (i.e., $\sigma_\tau^2 = 0$), this bound recovers the variance bound found in Jochmans and Weidner (2019) if we further assume that V_e is independently distributed across edges with same variance (i.e., $\widetilde{\sigma}_n^2 = \sigma^2$):

$$\operatorname{Var}(\widehat{\alpha}_i) \leq \sigma^2 \times \left\{ \frac{1}{d_i} \left(1 + \frac{1}{\lambda_{2,L} h_i} \right) - \frac{2}{\sum_{j \in V} d_j} \right\},$$

found in Theorem 2 of Jochmans and Weidner (2019).

It is well known in the spectral graph literature that (Cvetković et al., 2007)

$$2 \min_{k \in V} d_k \leq \lambda_{n,F} \leq 2 \max_{k \in V} d_k,$$

Thus, the upper bound in Proposition 1.1 is of order

$$O(\lambda_{n,F}/d_i + \lambda_{n,F}/(d_i \lambda_{2,L} h_i)),$$

which is $O(1)$ if $d_i \propto \lambda_{n,F}$ and $\lambda_{2,L} h_i$ is bounded away from zero. This implies that even when the graph is well-connected, the consistency of the fixed-effect estimator is not guaranteed in the worst-case scenario. In the following example, $\lambda_{n,F}$ is proportional to d_i and the bound in Proposition 1.1 is shown to be tight up to a constant factor.

Example 1.3. *Consider a star graph with one central node (node i) connected to $n-1$ peripheral nodes, each of which is connected only to the central node. Suppose there are no multiple edges and $s(e) = i$ for all edges $e \in E$. Suppose further that $\tau^t = \tau^s = \tau$. In this case, if V_e is i.i.d. across edges, we can* *directly compute the variance of $\widehat{\alpha}_i$ as follows:*

$$\mathrm{Var}(\widehat{\alpha}_i) = \frac{(n-1)\mathbb{E}[\epsilon_e^2]}{n^2} + \frac{(n-1)(n-2)\mathbb{E}[\tau(U_i)^2]}{n^2},$$

which is $O(1)$ and bounded away from zero as $n \to \infty$ if and only if $\mathbb{E}[\tau(U_i)^2] > 0$. Also, in this case, we have $\lambda_{n,F} = n$, $\lambda_{2,L} = 1$, $\sum_{j \in V} d_j = 2(n-1)$, and $h_i = 1$. Thus, the upper bound in Proposition 1.1 is given by

$$\mathrm{Bound}_{star} \equiv \frac{\mathbb{E}[\epsilon_e^2]}{n-1} + \frac{(n-2)\mathbb{E}[\tau(U_i)^2]}{n-1},$$

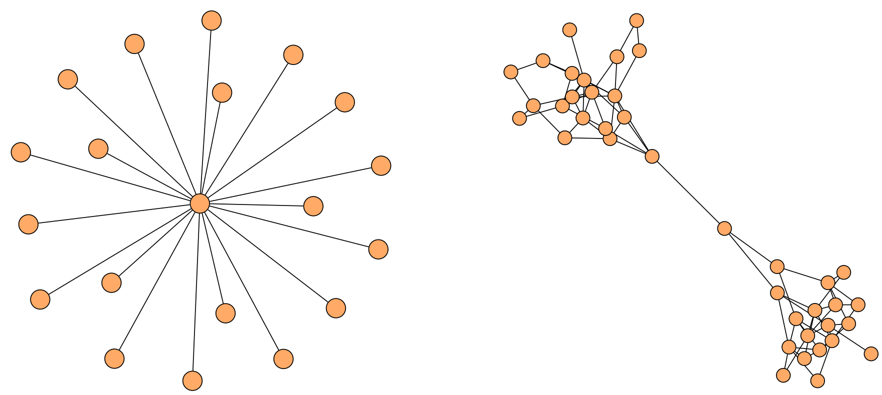
and a straightforward calculation shows that

$$\frac{\mathrm{Bound}_{star}}{4} \leq Var(\widehat{\alpha}_i) \leq \mathrm{Bound}_{star}.$$

Therefore, the variance bound in Proposition 1.1 is tight up to a constant factor in this example.

Remark 1.3. *Note that $\lambda_{n,F}$ and $\lambda_{2,L}$ can vary independently. For instance, in Figure 1.2, panel (a) illustrates a star graph with $\lambda_{n,F}/n = 1$ and $\lambda_{2,L} = 1 \gg 0$, whereas panel (b) depicts two highly connected clusters linked by a single edge, yielding $\lambda_{n,F}/n \approx 1$ and $\lambda_{2,L} \approx 0$. Thus, a high $\lambda_{2,L}$ does not imply a low $\lambda_{n,F}$, and vice versa. There can, however, be a trade-off in some cases. For example, aggregating low-degree nodes (e.g., firms with fewer than 15 movers) into a single supernode to improve connectivity, as in Andrews et al. (2008) and Alviarez et al. (2025), tends to raise $\lambda_{2,L}$, but it also raises $\lambda_{n,F}$ because the supernode connects to many other nodes, thereby strengthening the dependence* *structure. Similarly, grouping fixed effects into fewer classes (e.g., Bonhomme et al., 2019) can increase both* $\lambda_{n,F}$ *and* $\lambda_{2,L}$*. Consequently, these aggregation strategies need not reduce the variance of the fixed-effect estimator: they may simultaneously improve connectivity while amplifying dependence.*

Figure 1.2: Comparison of $\lambda_{n,F}$ and $\lambda_{2,L}$



(a) High $\lambda_{n,F}$ and high $\lambda_{2,L}$ (b) High $\lambda_{n,F}$ and low $\lambda_{2,L}$

1.3 Asymptotic Theory

The variance decomposition in (1.7) and the finite-sample variance bound in Proposition 1.1 help clarify the behavior of the fixed-effect estimator under the assumed dependence structure. Notably, the estimator can be inconsistent even when the graph is well-connected, contrasting with the consistency results under independence (see Jochmans and Weidner, 2019). However, the finite-sample theory above provides only a worst-case bound and does not sharply characterize the estimator's behavior, except for specific graph structures like Example 1.3. Therefore, in this section, we develop a first-order approximation for the fixed-effect estimator and propose a new inference method for fixed effects under an asymptotic framework, where the sequence of graphs $\mathcal{G}_1, \mathcal{G}_2, \ldots$ grows both locally around each node of interest and globally in terms of the number of nodes and edges.

1.3.1 First-Order Approximation

We begin with the following first-order approximation for the fixed-effect estimator. The idea is to decompose the estimator as follows:

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \mathbf{D}^{-1}\mathbf{B}'\boldsymbol{\epsilon} + \underbrace{\mathbf{D}^{-1}\mathbf{A}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})}_{\equiv \boldsymbol{r}}.$$

The remainder term $\boldsymbol{r}$ aggregates the deviation of the fixed-effect estimator from the true effects, and it will be shown to be negligible, depending on both the connectivity of the graph and the dependence structure as seen in the variance bound in Proposition 1.1.

Element-wise, the first term of the decomposition for node i is given by

$$\frac{1}{d_i} \sum_{e \in E_i} \mathbf{B}_{e,i} \epsilon_e.$$

Thus, each fixed-effect estimator $\widehat{\alpha}_i$ is locally driven by the average of the error terms ϵ_{e_l} over the set E_i of edges incident to node i.

We can further decompose the estimator as follows:

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \mathbf{D}^{-1}\mathbf{B}'(\mathbf{F}^s\boldsymbol{\tau}^s + \mathbf{F}^t\boldsymbol{\tau}^t) + \mathbf{D}^{-1}\mathbf{B}'(\boldsymbol{\epsilon} - \mathbf{F}^s\boldsymbol{\tau}^s - \mathbf{F}^t\boldsymbol{\tau}^t) + \boldsymbol{r}. \tag{1.8}$$

The first term on the right-hand side of (1.8) can be expressed as

$$\left(\frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i) - \frac{1}{d_i}\sum_{e\in E_i^s}\tau^t(U_{t(e)}) + \frac{1}{d_i}\sum_{e\in E_i^t}\tau^s(U_{s(e)})\right)_{i\in V}.$$

Note that the term

$$\frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i)$$

is generally persistent as d_i increases as at least one of d_i^t/d_i or d_i^s/d_i converges to a non-zero constant. On the other hand, since $\tau^s(U_j)$ and $\tau^t(U_j)$ are mean-zero random variables for any $j \in V$, the averages

$$\frac{1}{d_i}\sum_{e\in E_i^s}\tau^t(U_{t(e)}), \quad \frac{1}{d_i}\sum_{e\in E_i^t}\tau^s(U_{s(e)})$$

are of smaller order than the persistent term. Furthermore, the second and third terms in (1.8) are also of smaller order than the persistent term if the graph is well-connected. The following assumption ensures that the remaining terms are negligible and simplifies the asymptotic analysis:

Assumption 1.4. *The number of edges between each pair of nodes is absolutely bounded, i.e., there exists an absolute constant* $C > 0$ *such that* $\max_{i,j\in V}|E_{(i,j)}| \le C$ *for all* $n \in \mathbb{N}$.

Assumption 1.4 can be relaxed to allow for a growing number of edges between each pair of nodes, provided that the growth is sufficiently slow. For example, if node i is of interest and $d_i \to \infty$, we can allow for $\max_{i,j\in V}|E_{(i,j)}| \to$

∞ as long as the growth rate is slower than d_i. Also, we can force this assumption to hold by randomly removing a subset of multiple edges between each pair of nodes such that the number of remaining edges is bounded.

The following result summarizes the above discussion and provides a first-order approximation for the fixed-effect estimator $\widehat{\boldsymbol{\alpha}}$:

Theorem 1.1. *Under Assumptions 1.1-1.4, for each $i \in V$, we have*

$$\widehat{\alpha}_i - \alpha_i = \frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i) + O_p\left(\sqrt{\frac{\lambda_{n,F}}{d_i \lambda_{2,L} h_i}}\right),$$

as $d_i \to \infty$.

Theorem 1.1 shows that the fixed-effect estimator is approximated by the true fixed effect plus a mean-zero noise term, expressed as a weighted difference in the origin and destination effects, $\tau^s(U_i)$ and $\tau^t(U_i)$. Thus, the estimator is unbiased but inconsistent. The inconsistency arises because the fixed-effect estimator is, approximately, a local average of the error terms in which the node-level random effects are not fully averaged out. That is why the noise term is closely related to the proportions of edges where node i appears as an origin (d_i^s/d_i) versus as a destination (d_i^t/d_i). Since the noise does not vanish, this result is consistent with the intuition that the fixed-effect estimator is potentially inconsistent even in a well-connected graph due to the dependence structure.

Also note that Theorem 1.1 extends Theorem 4 in Jochmans and Weidner (2019), which establishes the first-order approximation of $\widehat{\alpha}_i$ under a weak dependence structure. In Jochmans and Weidner (2019), this weak dependence structure is characterized by requiring that the largest eigenvalue of $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']$ remains bounded by a positive absolute constant. In our framework, however, the dependence structure is sufficiently strong that the largest eigenvalue of $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']$ can diverge with d_i. This divergence leads directly to the inconsistency of the fixed-effect estimator in our setting.

The remainder term is negligible if the graph is well-connected, such that $\lambda_{2,L}$ does not decay too quickly and $\lambda_{2,L} h_i \to \infty$, and the graph itself grows sufficiently slowly relative to d_i, ensuring that $\lambda_{n,F}/d_i$ does not grow too rapidly as d_i increases. In that case, we have

$$\widehat{\alpha}_i - \alpha_i = \frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i) + o_p(1),$$

so that the fixed-effect estimator converges to

$$\alpha_i + \frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i).$$

A sufficient condition for this is that

$$\lambda_{n,F}/(d_i \lambda_{2,L} h_i) \to 0.$$

This condition is satisfied in the following examples:[4]

Example 1.4. *Suppose that $\mathcal{G}$ is a complete graph. Then, $d_i = n-1, h_i = n-1$ for each $i \in V$ and $\lambda_{n,F} = 2(n-1), \lambda_{2,L} = 1$. Thus, $\lambda_{n,F}/(d_i \lambda_{2,L} h_i) =$*

[4]The dependence measure $\lambda_{n,F}$ can be conservative in highly heterogeneous graphs where a few nodes have a large number of links. Improving this bound will be an important direction for future work.

$2/(n-1) \to 0$ *as* $n \to \infty$.

Example 1.5. *Suppose that* $\mathcal{G}$ *is an Erdős–Rényi random graph with edge probability* $p_n = c\log(n)/n$ *for some constant* $c > 1$. *Then,* $d_i/\log(n) \to c, h_i/\log(n) \to c$ *for each* $i \in V$ *and* $\lambda_{n,F}/\log(n) \to c, \lambda_{2,L} \to 1$ *almost surely as* $n \to \infty$. *Thus,* $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) \approx 1/\log(n) \to 0$ *almost surely as* $n \to \infty$.

A counterexample that violates the condition $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) \to 0$ is given below:

Example 1.6. *Suppose that* $\mathcal{G}$ *is a stochastic-block random graph composed of two blocks of size* $n/2$. *Let the intra-block edge probability be* $p_n = p_0\log(n)/n$ *for some constant* $p_0 > 2$, *and let the inter-block edge probability be* $q_n = q_0/n$ *for some constant* $q_0 < p_0$. *Then, for each* $i \in V$, $d_i/\log(n) \to p_0/2$ *and* $h_i/\log(n) \to p_0/2$, *and* $\lambda_{n,F}/\log(n) \to p_0/2, \lambda_{2,L} \times \log(n) = O(1)$ *almost surely as* $n \to \infty$. *Thus, we have* $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) = O(1)$ *almost surely as* $n \to \infty$.[5]

In Example 1.6, the graph is not well-connected: the connections between the two blocks are sparse, making it easier for the graph to break into two disconnected components as n increases. In contrast, within each block, the graph is densely connected, and the dependence structure remains non-negligible. Consequently, the first-order approximation in Theorem 1.1 may fail under such conditions.[6] Indeed, without the dependence structure, $\lambda_{n,F}$ would be replaced

[5]This result is based on Deng et al. (2021), who characterized the asymptotic behavior of $\lambda_{2,L}$ in the stochastic-block random graphs

[6]Note that this condition is a sufficient but not necessary condition for the negligibility

by 1, so even with the stochastic-block structure, the fixed-effect estimator would be consistent for α_i as $1/(d_i\lambda_{2,L}h_i) \to 0$ as $d_i \to \infty$.

Remark 1.4. *Attempting to improve connectivity by aggregating nodes into supernodes or grouping fixed effects into fewer classes, as discussed in Remark 1.3, may help achieve the negligibility of the remainder term in Theorem 1.1. For instance, when grouping fixed effects into K classes as in Bonhomme et al. (2019), the mobility network likely becomes closer to a complete graph, as in Example 1.4, with $\lambda_{2,L}$ bounded away from zero and $\lambda_{n,F}/d_i$ close to one, especially when K is small and mobility across classes is dense (e.g., $K = 10$). As long as these aggregation strategies are aligned with the data generating process, Theorem 1.1 can provide a useful approximation for the fixed-effect estimator in such cases after aggregation.*

1.3.2 Inference

The first-order approximation in Theorem 1.1 suggests a potential approach for performing inference on the fixed effect α_i by estimating the distribution of

$$\frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i).$$

However, the challenge is that we must either directly estimate the differences across nodes or estimate each $\tau^s(U_i)$ and $\tau^t(U_i)$ separately. This is in general not feasible without additional assumptions as α_i is unknown and only the sum $\tau^t(U_i) + \tau^s(U_i)$ can be consistently estimated.

of the remainder term. Thus, even if $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) \nrightarrow 0$, the remainder term may still be negligible depending on the specific graph structure and dependence structure.

To illustrate this point, consider the residual $\widehat{\boldsymbol{\epsilon}}$ defined as

$$\widehat{\boldsymbol{\epsilon}} \equiv \mathbf{y} - \mathbf{B}\widehat{\boldsymbol{\alpha}} = \boldsymbol{M}_{\mathbf{B}}\mathbf{y},$$

where $\boldsymbol{M}_{\mathbf{B}} = \mathbf{I}_m - \mathbf{B}\mathbf{L}^*\mathbf{B}'$ projects onto the orthogonal complement of the column space of $\mathbf{B}$. For each $i \in V$, let c_i be defined as

$$c_i = \frac{1}{d_i}\boldsymbol{f}_i'\boldsymbol{M}_{\mathbf{B}}\boldsymbol{f}_i,$$

where $\boldsymbol{f}_i$ is the i-th column of $\mathbf{F}$. Note that $c_i \in [0,1]$ because $\boldsymbol{f}_i'\boldsymbol{f}_i = d_i$ and $\boldsymbol{M}_{\mathbf{B}}$ is a projection matrix. Intuitively, c_i is a measure of balance between the inflow and outflow of edges incident to node i: if $d_i^t - d_i^s$ is close to zero, then c_i is close to one, while if $|d_i^t - d_i^s|$ is close to d_i, then c_i is close to zero. In fact, we can show that

$$c_i \approx 1 - \left(\frac{d_i^t - d_i^s}{d_i}\right)^2,$$

when the graph is well-connected with large $\lambda_{2,L}h_i$.

By locally averaging the residuals, we can obtain the following:

Proposition 1.2. *Under Assumptions 1.1-1.4, for each $i \in V$, we have*

$$\frac{1}{d_i}\sum_{e\in E_i}\widehat{\epsilon}_e = c_i \times \frac{\tau^t(U_i) + \tau^s(U_i)}{2} + O_p\left(\sqrt{\frac{1}{d_i}}\right),$$

as $d_i \to \infty$.

Proposition 1.2 shows that if node i has sufficient balance between its inflow and outflow of edges, the local average of the residuals $\widehat{\epsilon}_e$ for edges incident to node i contains information about the average $(\tau^t(U_i) + \tau^s(U_i))/2$. Specifically, if $c_i > 0$, we have

$$\widehat{\tau}_i \equiv \frac{1}{d_i c_i} \sum_{e \in E_i} \widehat{\epsilon}_e = \frac{\tau^t(U_i) + \tau^s(U_i)}{2} + O_p\left(\sqrt{\frac{1}{c_i^2 d_i}}\right). \tag{1.9}$$

Thus, as long as the balance measure c_i does not converge to zero quickly, we can consistently estimate the average $(\tau^t(U_i) + \tau^s(U_i))/2$. An extreme case would be when $c_i = 0$, which occurs when $d_i^t = 0$ or $d_i^s = 0$. In this case, the local average of the residuals is exactly zero and uninformative about the average $(\tau^t(U_i) + \tau^s(U_i))/2$.

Remark 1.5. *The convergence rate of $\widehat{\tau}_i$ in* (1.9) *is notable because it does not depend on $\lambda_{n,F}$ or $\lambda_{2,L}h_i$. In other words, the convergence rate is determined solely by the balance and the number of edges incident to node i, regardless of the overall connectivity of the graph or the strength of the dependence structure. This implies that even if the graph is not well-connected and the dependence structure is strong, so the first-order approximation in Theorem 1.1 may not hold, as in Example 1.6, we can still consistently estimate the average $(\tau^t(U_i) + \tau^s(U_i))/2$ as long as the edges incident to node i are sufficiently balanced.*

As argued above, the average $(\tau^t(U_i)+\tau^s(U_i))/2$ is not sufficient for performing inference on the fixed effect α_i because essentially we have three unknowns

$(\alpha_i, \tau^t(U_i), \tau^s(U_i))$ for two equations:

$$\widehat{\alpha}_i \approx \alpha_i + \frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i)$$
$$\widehat{\tau}_i \approx \frac{\tau^t(U_i) + \tau^s(U_i)}{2}.$$

To facilitate inference, we impose the following additional structure on the $\tau^s(\cdot)$ and $\tau^t(\cdot)$ functions:

Assumption 1.5. *For any $u \in \mathcal{U}$, we have*

$$\tau^s(u) = \tau^t(u) \equiv \tau(u).$$

Assumption 1.5 is satisfied in a broad class of models where the dependence structure is symmetric in $U_{s(e)}$ and $U_{t(e)}$, such as the additively separable model or the interactive model discussed above following Assumption 1.3. Importantly, this assumption is weaker than requiring symmetry of f in (1.4) with respect to the origin and destination shocks, i.e., $f(u, u', v) = f(u', u, v)$ for all u, u', v in their supports. The following example illustrates that Assumption 1.5 can hold even when f is not symmetric in $U_{s(e)}$ and $U_{t(e)}$: If

$$\epsilon_e = f(U_{s(e)}, U_{t(e)}, V_e) = U_{s(e)} + U_{t(e)} + U_{s(e)} \times \left(U_{t(e)}^2 - \frac{1}{3}\right)$$

with $U_i \sim \text{Unif}[-1, 1]$, then

$$\tau^s(u) = \tau^t(u) = u$$

for all $u \in \mathcal{U}$, even though $(U_{s(e)}, U_{t(e)}) = (1,0)$ and $(0,1)$ yield $\epsilon_e = 2/3$ and 1, respectively, which breaks the symmetry of f. Thus, we can allow for asymmetric responses of the error term to the origin and destination shocks, while still satisfying Assumption 1.5.

When Assumption 1.5 is violated, ϵ_e is necessarily asymmetric in $U_{s(e)}$ and $U_{t(e)}$. For example, consider the model where

$$\epsilon_e = U_{s(e)}U_{t(e)} + U_{t(e)} - 2U_{s(e)},$$

and suppose $\mathbb{E}[U_i] = 1$. In this case, one can verify that

$$\tau^s(u) = -u + 1, \quad \tau^t(u) = 2u - 2,$$

so that $\tau^s(u) \neq \tau^t(u)$ for any $u \in \mathcal{U}$ except for $u = 1$.

The following example connects Assumption 1.5 to the AKM model discussed earlier:

Example 1.7. *(Two-Period AKM model, revisited). Consider the two-period AKM model described in Examples 1.1 and 1.2. In the time-varying firm effect model with $\alpha_{j,t} = \alpha_j + U_{j,t}$ for firm j and period $t = 1, 2$, the error term can be specified as*

$$\epsilon_e = U_{t(e),2} - U_{s(e),1} + V_e, \quad \mathbb{E}[U_{j,t}] = \mathbb{E}[V_e] = 0.$$

To reduce the dimension of the shocks, suppose $U_{j,2} = \rho U_{j,1}$ for some constant $\rho \in \mathbb{R}$. Then, $\tau^s(u) = -u$ and $\tau^t(u) = \rho u$. Thus, Assumption 1.5 holds if *and only if $\rho = -1$, i.e., the random component of the firm effect is perfectly negatively correlated over time. In contrast, in the model with match-specific heterogeneity, if we specify the error term as*

$$\epsilon_e = U_{s(e)} U_{t(e)} - V_e, \quad \mathbb{E}[U_j] = \mathbb{E}[V_e] = 1,$$

then we have $\tau^s(u) = \tau^t(u) = u - 1$, so Assumption 1.5 holds.

Remark 1.6. *Instead of imposing Assumption 1.5, we may consider alternative normalizations for the functions $\tau^s(\cdot)$ and $\tau^t(\cdot)$, such as setting $\tau^t(\cdot) = 0$ or $\tau^s(\cdot) = 0$. In these cases, the nonzero function can be denoted by τ, and inference can proceed as described below. More generally, if there is a linear relationship between $\tau^t(\cdot)$ and $\tau^s(\cdot)$, we may set $\tau(\cdot) = \tau^t(\cdot)$ and $\tau^s(\cdot) = c\tau(\cdot)$, or $\tau(\cdot) = \tau^s(\cdot)$ and $\tau^t(\cdot) = c\tau(\cdot)$ for some constant c. The time-varying firm effect model in Example 1.7 is an instance of this, with $\tau^t(\cdot) = -\rho\tau^s(\cdot)$. See Appendix A.3 for inference under such settings.*

Under Assumption 1.5, Proposition 1.2 implies that we can consistently estimate $\tau(U_i)$ as follows:

$$\widehat{\tau}_i \approx \tau(U_i),$$

if $c_i^2 d_i \to \infty$ as $d_i \to \infty$. Moreover, under Assumption 1.5, Theorem 1.1 implies that

$$\widehat{\alpha}_i - \alpha_i \approx \frac{d_i^t - d_i^s}{d_i} \tau(U_i),$$

provided that $\lambda_{n,F}/(d_i \lambda_{2,L} h_i) \to 0$ as $d_i \to \infty$. Thus, we can leverage the consistency of $\widehat{\tau}_i$ to estimate the distribution of $\tau(U_i)$ and to perform inference on fixed effects.

Remark 1.7. *Since $\widehat{\tau}_i$ is a consistent estimator for $\tau(U_i)$, one might consider correcting the fixed-effect estimator as follows:*

$$\widehat{\alpha}_i - \frac{d_i^t - d_i^s}{d_i}\widehat{\tau}_i \approx \alpha_i.$$

For point estimation, this correction would yield a consistent estimator for α_i if the remainder term in Theorem 1.1 is negligible and i has sufficient balance. Relatedly, Appendix A.2 discusses an alternative estimator for α_i that directly incorporates this correction. For inference, however, we do not pursue this approach for two reasons. First, this correction is not feasible for every node $i \in V$ when $c_i = 0$ or close to zero, as $\tau(U_i)$ cannot be consistently estimated in these cases. This issue is particularly relevant when nodes of interest, or a non-negligible number of nodes, have c_i close to zero, as observed in the empirical application in Section 1.6 and in Figure 1.3. Second, the correction does not guarantee asymptotic normality, since the remainder term $\boldsymbol{r}$ in (1.8) can be dominant with order $O_p(\sqrt{\lambda_{n,F}/(d_i\lambda_{2,L}h_i)})$, and the central limit theorem does not directly apply to this term. Moreover, estimating the standard error of the corrected estimator is not straightforward without further assumptions on the distribution of $(V_e)_{e\in E}$. These considerations motivate the development of a new inference method based on the empirical distribution of $\widehat{\tau}_i$ for $i \in \mathcal{C}_n$, as defined below.

For this purpose, we impose the following assumption on the distribution of $\tau(U_i)$:

Assumption 1.6. *$\tau(U_i)$ has a continuous cumulative distribution function F_τ.*

Assumption 1.6 ensures that the distribution of $\tau(U_i)$ is well-behaved. This excludes degenerate cases and guarantees that the dependence structure remains relevant as the graph size n increases. Since $\tau(U_i)$ is mean-zero, degeneracy would imply $\tau(U_i) = 0$ almost surely for all $i \in V$, in which case the standard inference results from Jochmans and Weidner (2019) would apply.

We define the set of nodes useful for estimating $\boldsymbol{\tau}$ as $\mathcal{C}_n \equiv \{i \in V : c_i > c\}$, where $c > 0$ is an absolute constant.[7] Figure 1.3 illustrates the determination of $\mathcal{C}_n$ using data from the empirical application in Section 1.6. Let $\widehat{F}_{n,\tau}$ be the empirical distribution function of $\widehat{\tau}_i$ for $i \in \mathcal{C}_n$, defined as

$$\widehat{F}_{n,\tau}(t) \equiv \frac{1}{|\mathcal{C}_n|} \sum_{i \in \mathcal{C}_n} \mathbb{I}\left(\widehat{\tau}_i \leq t\right),$$

where $\mathbb{I}(\cdot)$ is the indicator function. Also, define the following object to globally control the estimation errors for $|\mathcal{C}_n|$ nodes:

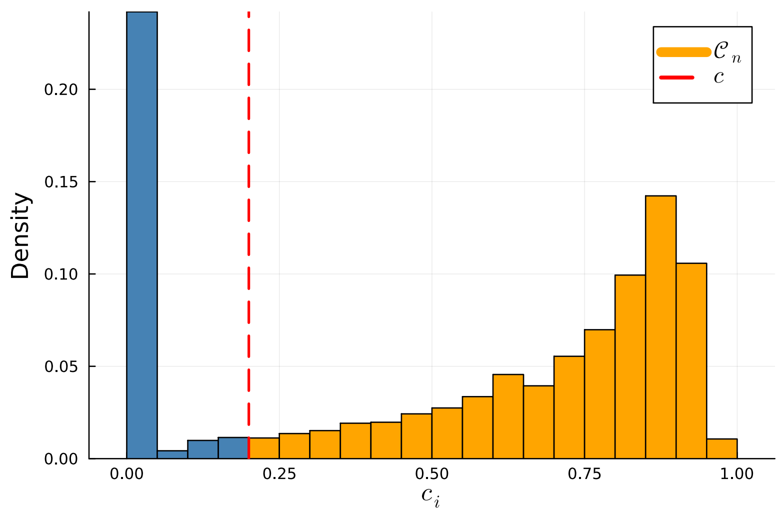
$$\eta_n \equiv \frac{1}{|\mathcal{C}_n|} \sum_{i \in \mathcal{C}_n} \frac{1}{d_i},$$

which tends to zero as $n \to \infty$ provided that nodes in $\mathcal{C}_n$ have growing d_i on average.

Then, we have the following result:

[7] We fix c as an absolute constant for simplicity. At the cost of complicating the proofs, we can allow c to depend on n and converge to zero sufficiently slowly as $n \to \infty$.

Figure 1.3: Determination of $\mathcal{C}_n$



Notes: This figure presents the distribution of the balance measure c_i using data from the empirical application in Section 1.6. The orange bins correspond to nodes in the set $\mathcal{C}_n$ with $c_i > c$ and the red dash line indicates the threshold $c = 0.2$.

Theorem 1.2. *Under Assumptions 1.1-1.6, if $|\mathcal{C}_n| \to \infty$ and $\eta_n \to 0$ as $n \to \infty$, we have*

$$\sup_{t \in \mathbb{R}} |\widehat{F}_{n,\tau}(t) - F_\tau(t)| \to_p 0,$$

as $n \to \infty$.

The additional conditions $|\mathcal{C}_n| \to \infty$ and $\eta_n \to 0$ ensure that there are enough informative nodes to reliably estimate the distribution of $\tau(U_i)$. These conditions are satisfied in many graphs with growing degrees, where nodes have sufficient inflow and outflow of edges. For example, in Examples 1.4-1.6, for each edge, if source and target nodes are assigned independently at random with probability $q \in (0, 1)$ and $1 - q$, respectively, then the conditions are likely satisfied when q is an absolute constant.

Conditioning on $\mathcal{C}_n$ does not introduce selection bias because $\mathcal{C}_n$ is determined solely by the graph structure $\mathcal{G}_n$, which is treated as fixed and independent of the error terms $(\epsilon_e)_{e\in E}$ in our analysis. However, if the graph structure is endogenous to $(\epsilon_e)_{e\in E}$, conditioning on $\mathcal{C}_n$ could distort inference about the distribution of $\tau(U_i)$. Addressing such endogeneity is beyond the scope of this paper and is left for future research.

Theorem 1.2 is useful for conducting inference on fixed effects in a manner similar to that of Conley and Taber (2011), who developed a non-standard inference under small treatment group asymptotics. Noting that

$$\left(\frac{d_i^t - d_i^s}{d_i}\right)^{-1}(\widehat{\alpha}_i - \alpha_i) \approx \tau(U_i),$$

for $i \in V$ such that $|d_i^t - d_i^s| > d_i c$ for some absolute constant $c > 0$, we can construct a confidence interval for α_i by inverting the test based on $\widehat{F}_{n,\tau}$. Specifically, for a given confidence level $1-\alpha$, let $\widehat{c}_{\alpha/2}$ and $\widehat{c}_{1-\alpha/2}$ be the $(\alpha/2)$-th and $(1-\alpha/2)$-th quantiles of $\widehat{F}_{n,\tau}$, respectively. Then, the $1-\alpha$ confidence interval for α_i is given by

$$CI_{i,1-\alpha} \equiv \begin{cases} \left[\widehat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\widehat{c}_{1-\alpha/2}, \widehat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\widehat{c}_{\alpha/2}\right] & \text{if } d_i^t - d_i^s > 0; \\ \left[\widehat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\widehat{c}_{\alpha/2}, \widehat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\widehat{c}_{1-\alpha/2}\right] & \text{if } d_i^t - d_i^s < 0. \end{cases} \tag{1.10}$$

Another useful application of Theorem 1.2 is joint hypothesis testing for fixed effects $\boldsymbol{\alpha}$. For example, consider testing the null hypothesis that a subset

V_0 of nodes share the same fixed effect:

$$H_0 : \alpha_i = \alpha_j \text{ for all } i, j \in V_0 \quad \text{versus} \quad H_1 : \alpha_i \neq \alpha_j \text{ for some } i, j \in V_0.$$

We can test this hypothesis using the following test statistic:

$$T = \widehat{\boldsymbol{\alpha}}'_{V_0} \boldsymbol{M}_{n_0} \widehat{\boldsymbol{\alpha}}_{V_0} \tag{1.11}$$

where $\boldsymbol{\alpha}_{V_0} = (\alpha_i)_{i \in V_0}$, $n_0 = |V_0|$, and $\boldsymbol{M}_{n_0}$ is the demeaning matrix of size n_0. Under the null hypothesis,

$$\boldsymbol{\alpha}'_{V_0} \boldsymbol{M}_{n_0} \boldsymbol{\alpha}_{V_0} = 0,$$

and the distribution of T can be approximated by simulating $\widehat{\boldsymbol{\alpha}}_{V_0}$ as $((d_i^t - d_i^s)/d_i \times \tau_i^{(m)})_{i \in V_0}$ for M repetitions, where each $\tau_i^{(m)}$ is drawn from the empirical distribution $\widehat{F}_{n,\tau}$. Let $\widehat{c}^T_{1-\alpha}$ denote the $(1-\alpha)$ quantile of the simulated distribution of T under the null. The test rejects H_0 if $T > \widehat{c}^T_{1-\alpha}$.

More generally, we can test linear hypotheses of the form $H_0 : \boldsymbol{R}\boldsymbol{\alpha} = 0$ for some known $q \times n$ matrix $\boldsymbol{R}$ with rank q using a similar approach and constructing the test statistic as $g_n(\widehat{\boldsymbol{\alpha}}_{V_0})$ for some known function $g_n(\cdot)$ such that the null distribution can be simulated from $\widehat{F}_{n,\tau}$.

The following result establishes the asymptotic validity of the confidence interval in (1.10) and the test based on the statistic T in (1.11):[8]

[8] Our result here only establishes pointwise asymptotic validity. Establishing uniform asymptotic validity over a class of DGPs and graph sequences is an important direction for future work.

Proposition 1.3. *Suppose that the conditions of Theorem 1.2 hold. Suppose also that F_τ is strictly increasing around quantiles of interest. For each $i \in V$ such that $(d_i^t - d_i^s)/d_i = O(1)$ and $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) \to 0$ as $n \to \infty$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\alpha_i \notin CI_{i,1-\alpha}\right) = \alpha.$$

Also, if for each $i \in V_0$, $(d_i^t - d_i^s)/d_i \to c_i^{st}$, $\min_{i\in V_0} |c_i^{st}| > 0$, $\max_{i\in V_0} \lambda_{n,F}/(d_i\lambda_{2,L}h_i) \to 0$ as $n \to \infty$, and n_0 is fixed, then

$$\lim_{n,M\to\infty} \mathbb{P}(T > \widehat{c}^{T}_{1-\alpha}) = \alpha,$$

under the null hypothesis that $\alpha_i = \alpha_j$ for all $i, j \in V_0$.

Remark 1.8. *The proposed inference methods rule out the case where nodes of interest have almost perfect balance between their inflow and outflow of edges, i.e., $d_i^t \approx d_i^s$. In such cases, we have*

$$\widehat{\alpha}_i - \alpha_i \approx 0$$

and little information on $\tau(U_i)$ remains in $\widehat{\alpha}_i$. The confidence interval in (1.10) *and the distribution of the test statistic T will be degenerate, leading to unreliable inference. If all nodes of interest have such balance, we can at least acknowledge that the estimates for these nodes are consistent without correcting as done in Remark 1.7.*

1.4 Variance Components Estimation

In the previous sections, we have focused on the estimation and inference of fixed effects $\boldsymbol{\alpha}$. However, in empirical applications, researchers are often interested in distributional properties of the fixed effects. For example, Card et al. (2013) estimate the sample variance of firm fixed effects in the AKM model to assess the contribution of workplace heterogeneity to rising wage inequality. In this section, we focus on estimating the sample variance of the fixed effects $\boldsymbol{\alpha}$ and address the bias in the sample variance estimator that arises from the dependence structure. See Appendix A.4 for estimation of the covariance between two sets of fixed effects.

1.4.1 Estimation

Our parameter of interest is the sample variance of $\boldsymbol{\alpha}$, given by

$$
\begin{aligned}
V_{\boldsymbol{\alpha}} &\equiv \frac{1}{n}\sum_{i\in V}(\alpha_i - \overline{\boldsymbol{\alpha}})^2 \\
&= \frac{\boldsymbol{\alpha}'\boldsymbol{M}_n\boldsymbol{\alpha}}{n},
\end{aligned}
$$

where $\boldsymbol{M}_n = \mathbf{I}_n - 1/n \times \boldsymbol{\iota}_n\boldsymbol{\iota}_n'$ is the demeaning matrix, $\boldsymbol{\iota}_n$ is an n-dimensional vector of ones, and $\overline{\boldsymbol{\alpha}} = \boldsymbol{\iota}_n'\boldsymbol{\alpha}/n$ is the average of $\boldsymbol{\alpha}$. We can estimate $V_{\boldsymbol{\alpha}}$ by plugging in $\widehat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$:

$$
\widehat{V}_{\boldsymbol{\alpha}} = \frac{\widehat{\boldsymbol{\alpha}}'\boldsymbol{M}_n\widehat{\boldsymbol{\alpha}}}{n}.
$$

It is well known that $\widehat{V}_{\boldsymbol{\alpha}}$ is biased upward due to the estimation error in $\widehat{\boldsymbol{\alpha}}$ even in cases where dependence is ruled out (Andrews et al., 2008). To see this, from (1.3), we have

$$\begin{aligned}\mathbb{E}[\widehat{V}_{\boldsymbol{\alpha}}] &= V_{\boldsymbol{\alpha}} + \frac{\mathbb{E}[\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^{*}\boldsymbol{M}_{n}\mathbf{L}^{*}\mathbf{B}'\boldsymbol{\epsilon}]}{n} \\ &= V_{\boldsymbol{\alpha}} + \frac{\mathrm{tr}(\mathbf{B}\mathbf{L}^{*}\boldsymbol{M}_{n}\mathbf{L}^{*}\mathbf{B}'\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'])}{n}, \end{aligned} \tag{1.12}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Thus, the bias in $\widehat{V}_{\boldsymbol{\alpha}}$ is given by the second term in (1.12).

The bias term in (1.12) is nonzero in general and known as "limited mobility bias" in the literature. As an illustration, suppose that $\boldsymbol{\epsilon}$ is independent and identically distributed. Then, the bias term simplifies to

$$\mathbb{E}[\epsilon_e^2] \times \frac{\mathrm{tr}(\boldsymbol{M}_n\mathbf{L}^{*})}{n},$$

which is inversely proportional to the connectivity of the graph.

1.4.2 Bias Correction

To correct the bias in $\widehat{V}_{\boldsymbol{\alpha}}$, several bias-correction methods have been proposed in the literature. When $\boldsymbol{\epsilon}$ is independent and identically distributed, the bias can be estimated by the following formula (Andrews et al., 2008):

$$\widehat{\sigma}^2 \times \frac{\mathrm{tr}(\boldsymbol{M}_n\mathbf{L}^{*})}{n},$$

where $\widehat{\sigma}^2$ is an estimator consistent for $\mathbb{E}[\epsilon_e^2]$. Kline et al. (2020) extends this bias-correction method to accommodate the case where $\boldsymbol{\epsilon}$ is independent but not identically distributed.

In our setting, however, $\boldsymbol{\epsilon}$ is not independent, so the above methods are not directly applicable. To address this issue, note that we can decompose the covariance matrix of $\boldsymbol{\epsilon}$ as

$$\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2,$$

as in (1.7). Thus, under Assumption 1.5, the bias term in (1.12) can be rewritten as:[9]

$$\mathbb{E}[\tau(U_i)^2] \times \frac{\mathrm{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} + \frac{\mathrm{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\boldsymbol{\Omega}_2)}{n}. \tag{1.13}$$

Here, the first term in (1.13) is the new bias term due to the node-level dependence structure, while the second term corresponds to the limited mobility bias in the literature. Note that when there are few multiple edges between any given pair of nodes, the second term is proportional to $\mathrm{tr}(\boldsymbol{M}_n\mathbf{L}^*)/n$.

Heuristically, we can assess relative severity of the new bias term compared to the second term by comparing

$$\frac{\mathrm{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} \text{ and } \frac{\mathrm{tr}(\boldsymbol{M}_n\mathbf{L}^*)}{n}.$$

These quantities can be computed from the graph structure. In a well-connected graph, the latter will be small in comparison to the former, so the bias arising from the dependence structure will dominate. Conversely, in graphs that are

[9] Since the distributions of $\tau^t(U_i)$ and $\tau^s(U_i)$ are not separately identified as shown above, in this section we focus on the symmetric case and maintain Assumption 1.5. Extensions to asymmetric cases are possible following the discussion in Appendix A.3.

not well-connected, both terms will be significant.

Hypothetically, we can estimate the bias term in (1.13) by estimating $\mathbb{E}[\tau(U_i)^2]$ and $\mathbf{\Omega}_2$ separately. In particular, we can estimate $\mathbb{E}[\tau(U_i)^2]$ via the following estimator:

$$\widehat{\sigma}_\tau^2 = \frac{1}{|\mathcal{C}_n|} \sum_{i \in \mathcal{C}_n} \left(\frac{1}{d_i c_i} \sum_{e \in E_i} \widehat{\epsilon}_e \right)^2,$$

whose consistency is anticipated by Proposition 1.2. However, estimating $\mathbf{\Omega}_2$, without imposing additional restrictions on the dependence structure or heteroskedasticity is challenging as $\widehat{\boldsymbol{\alpha}}$ is not, in general, a consistent estimator of $\boldsymbol{\alpha}$, and the correlation structure within each block of $\mathbf{\Omega}_2$ is too flexible to estimate.

Instead of estimating $\mathbf{\Omega}_2$ directly, we propose the following bias-corrected estimator for $V_{\boldsymbol{\alpha}}$ by correcting only the first term in (1.13):

$$\widehat{V}_{\boldsymbol{\alpha}}^{bc} = \widehat{V}_{\boldsymbol{\alpha}} - \widehat{\sigma}_\tau^2 \times \frac{\mathrm{tr}(\mathbf{B}\mathbf{L}^* \boldsymbol{M}_n \mathbf{L}^* \mathbf{B}' \mathbf{F}\mathbf{F}')}{n}.$$

We need the following additional regularity conditions to establish the consistency of $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$:

Assumption 1.7. *The following conditions hold:*

(i) The fixed effects $\boldsymbol{\alpha}$ are uniformly bounded, i.e., $sup_n \sup_{i \in V} |\alpha_i| < C$ for some absolute constant $C > 0$.

(ii) The following holds as $n \to \infty$:

$$\frac{\lambda_{n,F}}{n} \sum_{i \in V} \frac{1}{d_i} = O(1); \quad \frac{\lambda_{n,F}}{n} \sum_{i \in V} \frac{1}{d_i \, \lambda_{2,L} \, h_i} = o(1).$$

Part (i) of Assumption 1.7 excludes the case in which some nodes have unbounded fixed effects, a relatively mild requirement. Part (ii) consists of technical conditions that ensure the bias correction leaves a negligible remainder. These conditions hold, for example, in well-connected, homogeneous graphs such as the complete graph of Example 1.4 and the Erdős–Rényi random graph of Example 1.5, but fail if the graph is poorly connected or if edges are concentrated among only a few node pairs such as the stochastic block model of Example 1.6.

Then, we have the following consistency result for the bias-corrected estimator:

Theorem 1.3. *Under Assumptions 1.1-1.5 and 1.7, if $|\mathcal{C}_n| \to \infty$ and $\eta_n \to 0$ as $n \to \infty$, we have*

$$\widehat{V}_{\boldsymbol{\alpha}}^{bc} - V_{\boldsymbol{\alpha}} \to_p 0,$$

as $n \to \infty$.

Note that although Theorem 1.3 establishes the consistency of the bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$, in finite sample, it is typically still biased upward as we have not corrected the second term in (1.13). Nonetheless, our finding suggests that ignoring the bias arising from the dependence structure can be more severe than ignoring the second term when the graph is well-connected and homogeneous.

To account for the finite-sample bias, we propose the following rule-of-thumb modification to the bias-corrected estimator:

$$\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod} = \widehat{V}_{\boldsymbol{\alpha}} - \widehat{\sigma}_{\tau}^2 \times \frac{\mathrm{tr}(\mathbf{B}\mathbf{L}^* \boldsymbol{M}_n \mathbf{L}^* \mathbf{B}' \mathbf{F}\mathbf{F}')}{n} - \max\{0, \widehat{\sigma}^2 - 2\widehat{\sigma}_{\tau}^2\} \times \frac{\mathrm{tr}(\boldsymbol{M}_n \mathbf{L}^*)}{n},$$

where

$$\widehat{\sigma}^2 = \frac{1}{m} \sum_{e \in E} \widehat{\epsilon}_e^2.$$

The second term in $\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$ serves as the bias-correction term for the second term in (1.13) when the errors are assumed to be homoskedastic and the within-block correlation is neglected. If the block correlation induces positive bias, then this modification will be conservative.

Also, note that $\widehat{\sigma}^2$ is not consistent for $\mathbb{E}[\epsilon_e^2]$ in general. However, if the errors are homoskedastic, it can be approximated by

$$\mathbb{E}[\widehat{\sigma}^2] \approx \mathbb{E}[\tau^2(U_i)] \times \left[\frac{4}{\sum_{i \in V} d_i^t} \sum_{i \in V} \frac{d_i^t d_i^s}{d_i} - 2 \right] + \mathbb{E}[\epsilon_e^2] \leq \mathbb{E}[\epsilon_e^2]$$

so that $\widehat{\sigma}^2$ is downward biased relative to $\mathbb{E}[\epsilon_e^2]$, leading to a conservative modification of the bias-corrected estimator.

Remark 1.9. *The rule-of-thumb modification may result in over-correction, especially when the errors are heteroskedastic, as observed in the simulation results in the next section. This is because the approximation of* $\widehat{\sigma}^2$ *above assumes homoskedasticity. Using a more robust estimator for* $\widehat{\sigma}^2$*, such as the* *one proposed by Kline et al. (2020), could potentially improve the performance of the modified estimator. We leave a detailed investigation of this approach to future work.*

We can compare our bias-corrected estimators with those previously proposed in the literature. The key difference is that our estimator corrects for bias by targeting the first term in (1.13), which arises from the node-level dependence structure, and it does so without requiring the independence assumption on the error term $\boldsymbol{\epsilon}$. By contrast, existing bias-corrected estimators, such as those of Andrews et al. (2008) and Kline et al. (2020), correct only for the second term in (1.13) and are consistent if $\boldsymbol{\epsilon}$ is independent.[10]

For example, Andrews et al. (2008)'s bias-corrected estimator is given by

$$\widehat{V}_{\boldsymbol{\alpha}}^{a} = \widehat{V}_{\boldsymbol{\alpha}} - \widehat{\sigma}^2 \times \frac{\operatorname{tr}(\boldsymbol{M}_n \mathbf{L}^*)}{n}.$$

If the graph is not well-connected and the contribution of the node-level shocks (i.e., $\mathbb{E}[\tau^2(U_i)]$) is small relative to $\mathbb{E}[\epsilon_e^2]$, this correction may work well as the first term in (1.13) may be negligible.

However, if the node-level dependence structure is strong with significant $\mathbb{E}[\tau(U_i)^2]$ relative to $\mathbb{E}[\epsilon_e^2]$, then the bias stemming from this dependence structure becomes dominant, as shown in Theorem 1.3. In this scenario, the Andrews et al. (2008)-type bias correction, which ignores this component, will be inconsistent. Similarly, the bias-correction method proposed by Kline et al. (2020), which also hinges on the independence of $\boldsymbol{\epsilon}$ or weak dependence,

[10]While Kline et al. (2020) permits weak dependence in their model, their framework does not accommodate the strong dependence structure we consider here.

suffers from the same limitation and will generally be inconsistent when the dependence structure is strong.

Remark 1.10. *Relatedly, strategies to alleviate limited mobility bias by aggregating nodes into clusters (e.g., Andrews et al., 2008; Bonhomme et al., 2019), as discussed in Remark 1.3, are likely to make the bias from the dependence structure more pronounced, thereby making our bias-correction method more essential. This is because such strategies improve the connectivity of the graph, which reduces the limited mobility bias but can increase the dependence measure* $\lambda_{n,F}$ *and thus amplify the bias from the dependence structure. Moreover, graphs resulting from such aggregation are more likely to satisfy the regularity conditions in Assumption 1.7, as in the case of complete graphs, thereby increasing the applicability of our bias-correction method.*

1.5 Simulation

In this section, we conduct a simulation study to illustrate our inference procedure and the finite-sample properties of the bias-correction method.

1.5.1 Design

We first generate an undirected graph $\mathcal{G} = (V, E)$ with $|V| = n$ nodes from the stochastic block model discussed in Example 1.6. Recall that the stochastic block model is a random graph model where nodes are partitioned into K blocks, and edges are formed between nodes in the same block with probability p_n and between nodes in different blocks with probability q_n. We set $p_n = 10\log(n)/n$ and $q_n = 2/(\log(n)n)$ for each n. We then extract the largest connected component from the generated graph.[11] We vary the number of blocks $K = 1, 2$ and the number of nodes $n = 500, 1000, 2500,$ and 5000 to compare the performance of our inference procedure under different graph structures.

Given the connected graph $\mathcal{G} = (V, E)$, we generate and fix the true fixed effects $\boldsymbol{\alpha}$ as follows:

$$v_i \sim \text{Uniform}[-1, 1] \text{ for } i \in V;$$
$$\boldsymbol{\alpha} = \boldsymbol{v} - (\boldsymbol{v}'\mathbf{d}/\mathbf{d}'\mathbf{d}) \times \mathbf{d}$$

which ensures that $\mathbf{d}'\boldsymbol{\alpha} = 0$.

We generate the error terms $\boldsymbol{\epsilon}$ according to

$$\epsilon_e = U_{s(e)} + U_{t(e)} + V_e \text{ for } e \in E,$$

where $U_i \sim N(0, 1)$, and

$$V_e \sim N(0, 1 + |\alpha_{s(e)}| + |\alpha_{t(e)}|)$$

independently. Note that this structure satisfies Assumptions 1.3 and 1.5 with $\tau(U_i) = U_i$ for each $i \in V$ and allows for heteroskedasticity in V_e.

Using $\boldsymbol{\alpha}$, we randomly assign the inflow and outflow of each edge $e \in E_{(i,j)}$

[11] In our simulation setting, randomly generated graphs are typically connected, and even if they are not, the largest connected component covers a large proportion of the nodes. Thus, we do not distinguish the original $\mathcal{G}$ from the largest connected component, and we keep denoting it as $\mathcal{G}$ for simplicity.

as follows:

$$t(e) = i \text{ with probability } \frac{|\alpha_i|}{|\alpha_i| + |\alpha_j|},$$

independently for each edge $e \in E_{(i,j)}$. We can interpret this assignment as a process where a node with large effect is more likely to attract the inflow of an edge. Then, with (V, E, s, t), we can construct the incidence matrix $\mathbf{B}$ and generate the outcome vector according to (1.1). For each iteration, we compute the least-squares estimator $\widehat{\boldsymbol{\alpha}}$ and the empirical distribution of $\tau(U_i)$ using Theorem 1.2, by setting $\mathcal{C}_n = \{i \in V : c_i > 0.2\}$.

Table 1.1 reports the degree distribution of the generated graphs. For each n, the generated graphs exhibit a balanced degree distribution, avoiding extremes of sparsity and density, regardless of the underlying generating process.

Table 1.2 reports the global measures of the generated graphs. The connectivity measure $\lambda_{2,L}$ is well bounded away from zero for $K = 1$, while it converges to zero for $K = 2$, reflecting that the latter graph is much easier to partition into separate components. In both models, the dependence measure $\lambda_{n,F}$ increases slowly as n increases. Moreover, more than 90% of the nodes are in the set $\mathcal{C}_n$, suggesting there are enough informative nodes to estimate the distribution of $\tau(U_i)$. The convergence measure η_n approaches zero for both models, indicating that Theorem 1.2 provides a good approximation of the distribution of $\tau(U_i)$ for large n. The other convergence measure, $H_n \equiv \lambda_{n,F}\lambda_{n,L}^{-1} n^{-1} \sum_{i \in V} d_i^{-1} h_i^{-1}$, converges to zero for $K = 1$ but not for $K = 2$, reflecting that the stochastic block model with $K = 2$ is not well-connected while the dependence structure remains non-negligible.

Table 1.1: Degree Distributions of the Generated Graphs

	min	**Q1**	**Q3**	**max**	**mean**
			Panel A: $K = 1$		
$n = 500$	87	110	123	148	116.944
$n = 1000$	100	126	141	174	133.868
$n = 2500$	120	146	162	199	154.169
$n = 5000$	124	160	177	217	168.815
			Panel B: $K = 2$		
$n = 500$	41	54	62	76	58.208
$n = 1000$	44	62	72	92	67.362
$n = 2500$	43	71	83	108	77.154
$n = 5000$	56	78	91	124	84.68

Panel A reports the degree distributions when the number of blocks is $K = 1$ (Erdős–Rényi model), and Panel B reports the degree distributions when the number of blocks is $K = 2$. The first column reports the number of nodes n, and the second to the fifth columns report the minimum, 25th percentile, 75th percentile, maximum, and mean of the degree.

Table 1.2: Global Measures of the Generated Graphs

	$\lambda_{2,L}$	$\lambda_{n,F}$	$\lvert\mathcal{C}_n\rvert$	η_n	H_n
			Panel A: $K = 1$		
$n = 500$	0.842	236.914	496	0.009	0.021
$n = 1000$	0.842	271.302	993	0.008	0.018
$n = 2500$	0.846	311.924	2486	0.007	0.016
$n = 5000$	0.849	341.375	4974	0.006	0.014
			Panel B: $K = 2$		
$n = 500$	0.013	119.039	495	0.017	2.852
$n = 1000$	0.008	138.749	989	0.015	4.088
$n = 2500$	0.007	158.181	2478	0.013	3.987
$n = 5000$	0.005	173.399	4938	0.012	4.648

Note: Panel A reports the global measures when the number of blocks is $K = 1$ (Erdős–Rényi model), and Panel B reports the global measures when the number of blocks is $K = 2$. The first column reports the number of nodes n, the second column reports the connectivity measure $\lambda_{2,L}$, the third column reports the dependence measure $\lambda_{n,F}$, and the fourth column reports the number of nodes in $\mathcal{C}_n = \{i \in V : c_i > 0.2\}$. The last two columns report the convergence measures η_n and H_n.

1.5.2 Results: Inference

In this exercise, we construct 95% confidence intervals for α_i based on (1.10) and evaluate the coverage probability of these confidence intervals. We also construct 95% confidence intervals for α_i based on Theorem 5 in Jochmans and Weidner (2019), which shows the asymptotic normality of $\widehat{\alpha}_i$ when $\boldsymbol{\epsilon}$ are independent. Specifically, the standard error of $\widehat{\alpha}_i$ in the independent case is given by

$$\frac{\sqrt{\sum_{e \in E_i} \widehat{\epsilon}_e^2}}{d_i}.$$

This procedure is repeated 2000 times to evaluate the coverage probability. See Appendix A.5 for additional simulation results, including results for joint hypothesis tests discussed above.

Table 1.3 summarizes the Monte Carlo simulation results. The confidence intervals based on (1.10) achieve coverage probabilities close to the nominal 95% level for both $K = 1$ and $K = 2$. While this is expected for $K = 1$, it is notable for $K = 2$, where the graph is not well-connected and $H_1 = \lambda_{n,F}/(d_1 \lambda_{2,L} h_1)$ does not converge to zero, so the first-order approximation underlying (1.10) is not theoretically guaranteed. The strong performance in this case suggests that the approximation remains accurate in finite samples, even when the graph is not well-connected.

In contrast, confidence intervals based on the asymptotic normality of $\widehat{\alpha}_i$ under independence of $\boldsymbol{\epsilon}$ (based on Jochmans and Weidner, 2019) show substantial under-coverage in both $K = 1$ and $K = 2$, highlighting the importance of accounting for the dependence structure in the error terms. This under-coverage is more pronounced in a single block ($K = 1$) than in two blocks ($K = 2$), likely because the relative contribution of the independent component of the error term is smaller in this case than in $K = 2$.

Table 1.3: Coverage Probability of the Confidence Intervals

	α_1	d_1	$(d_1^t - d_1^s)/d_1$	H_1	**95%**	**Normal 95%**
			Panel A: $K = 1$			
$n = 500$	-0.596	119	-0.697	0.020	0.946	0.440
$n = 1000$	0.314	143	-0.986	0.016	0.954	0.278
$n = 2500$	0.281	161	-0.652	0.014	0.948	0.370
$n = 5000$	0.539	159	-0.635	0.016	0.944	0.403
			Panel B: $K = 2$			
$n = 500$	-0.594	49	-0.878	3.886	0.948	0.516
$n = 1000$	0.312	67	-1.000	3.966	0.954	0.365
$n = 2500$	0.278	72	-0.583	4.407	0.944	0.565
$n = 5000$	0.542	93	-0.613	3.721	0.931	0.508

Note: Panel A reports the results for $K = 1$ (Erdős–Rényi model), and Panel B reports the results for $K = 2$. The first column reports the number of nodes n, the second column reports the true value of α_1, the third and fourth columns report node 1's degree and the coefficient, respectively. The fifth column reports the convergence measure $H_1 = \lambda_{n,F}/(d_1\lambda_{2,L}h_1)$. The sixth column reports the coverage probability of the confidence intervals based on (1.10), and the seventh column reports the coverage probability of the confidence intervals based on Jochmans and Weidner (2019)'s asymptotic normality.

1.5.3 Results: Variance Components

In this exercise, we evaluate the performance of our proposed bias-correction method for estimating the sample variance of the fixed effects $\boldsymbol{\alpha}$. We compare the true variance $V_{\boldsymbol{\alpha}}$ with the plug-in estimator $\widehat{V}_{\boldsymbol{\alpha}}$, the bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$, the rule-of-thumb modified bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$, and the Andrews et al. (2008)-type bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{a}$.[12] We simulate each of

[12]We do not consider the bias-correction method proposed by Kline et al. (2020) here, as it requires leave-one-out connected graphs, which can be different from the graphs generated

these variance estimators 2000 times and compute the mean of the estimated variance.

Table 1.4 summarizes the results and Figures 1.4 and 1.5 present histograms of the simulated variance estimators for $n = 2500$. The plug-in estimator $\widehat{V}_{\boldsymbol{\alpha}}$ exhibits substantial upward bias. The bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$ is much closer to the true value $V_{\boldsymbol{\alpha}}$, demonstrating the effectiveness of our bias correction. The rule-of-thumb estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$ performs well for $K = 1$, but tends to overcorrect and display downward bias for $K = 2$, indicating that more adaptive or refined bias-correction methods may be needed for certain graph structures. The Andrews et al. (2008)-type bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{a}$ also reduces bias, but its remaining bias is larger than that of $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$ and $\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$ due to unaccounted dependence.

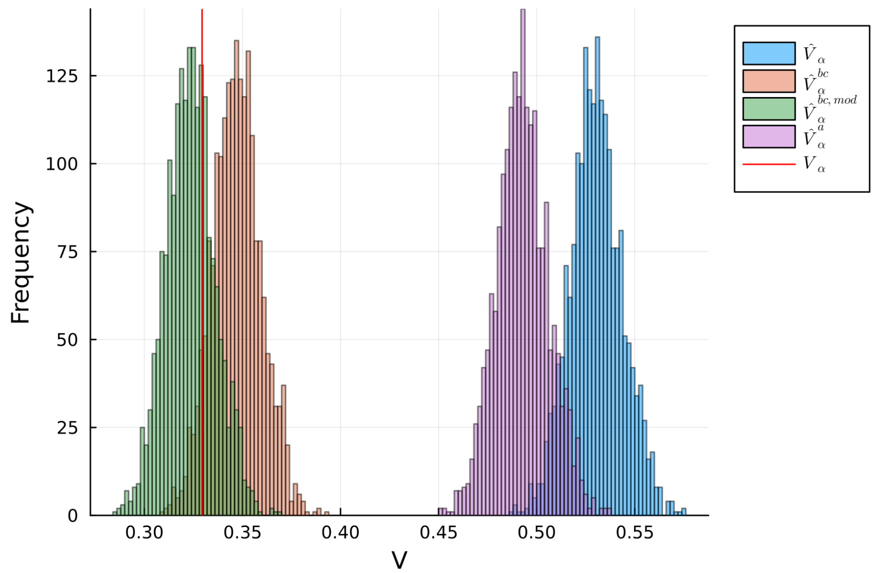
Table 1.4: Variance Estimation

	$V_{\boldsymbol{\alpha}}$	$\widehat{V}_{\boldsymbol{\alpha}}$	$\widehat{V}_{\boldsymbol{\alpha}}^{bc}$	$\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$	$\widehat{V}_{\boldsymbol{\alpha}}^{a}$
			Panel A: $K = 1$		
$n = 500$	0.334	0.544	0.359	0.328	0.494
$n = 1000$	0.340	0.549	0.361	0.333	0.505
$n = 2500$	0.329	0.530	0.347	0.323	0.492
$n = 5000$	0.336	0.528	0.353	0.33	0.494
			Panel B: $K = 2$		
$n = 500$	0.334	0.609	0.388	0.32	0.492
$n = 1000$	0.340	0.602	0.388	0.329	0.504
$n = 2500$	0.329	0.574	0.356	0.309	0.495
$n = 5000$	0.336	0.568	0.362	0.319	0.496

Note: Panel A reports the results for $K = 1$ (Erdős–Rényi model), and Panel B reports the results for $K = 2$. The first column reports the number of nodes n, the second column reports the true variance of $\boldsymbol{\alpha}$, the third to fifth columns report the mean of $\widehat{V}_{\boldsymbol{\alpha}}, \widehat{V}_{\boldsymbol{\alpha}}^{bc}, \widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$, and $\widehat{V}_{\boldsymbol{\alpha}}^{a}$, respectively.

in our simulation design. We leave a complete comparison including their method as in Bonhomme et al. (2023) to future work.

Figure 1.4: Histogram of the simulated $\widehat{V}_{\boldsymbol{\alpha}}$: Erdős–Rényi Model

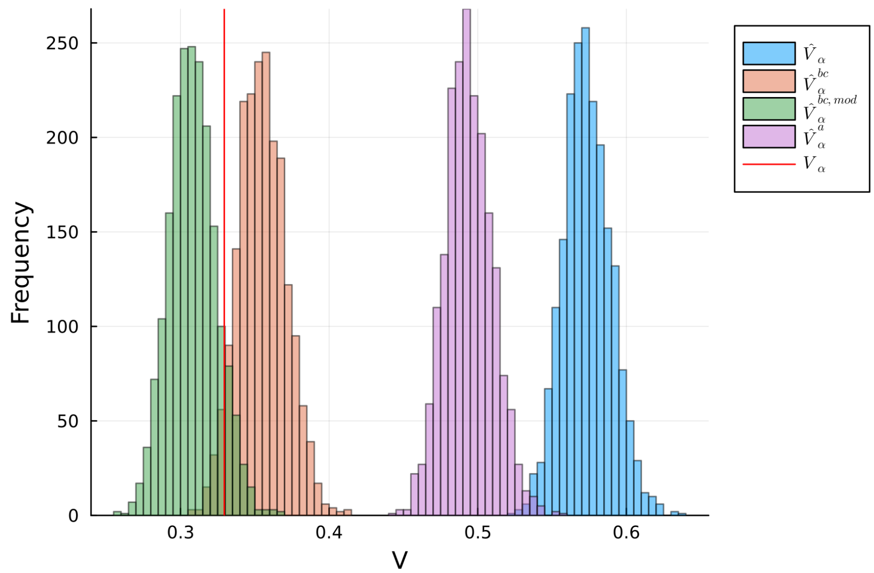


Note: the red line represents the true variance of $\boldsymbol{\alpha}$. The blue, orange, green, purple bins represent the simulated $\widehat{V}_{\boldsymbol{\alpha}}, \widehat{V}_{\boldsymbol{\alpha}}^{bc}, \widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}, \widehat{V}_{\boldsymbol{\alpha}}^{a}$, respectively. $n = 2500$.

1.6 Empirical Application

In this section, we apply our methods to an Italian matched employer-employee dataset. First, we construct a confidence interval for the fixed effect of a "central" firm in the mobility network. Next, we estimate the variance of the firm fixed effects and report the corresponding bias-corrected variance using our proposed bias-correction method.

Figure 1.5: Histogram of the simulated $\widehat{V}_{\boldsymbol{\alpha}}$: Stochastic-Block Model



Note: the red line represents the true variance of $\boldsymbol{\alpha}$. The blue, orange, green, purple bins represent the simulated $\widehat{V}_{\boldsymbol{\alpha}}, \widehat{V}_{\boldsymbol{\alpha}}^{bc}, \widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}, \widehat{V}_{\boldsymbol{\alpha}}^{a}$, respectively. $n = 2500$.

1.6.1 Data

The data are drawn from the Veneto Worker History (VWH) file.[13] The VWH file contains annual job spells for all workers employed in the Italian region of Veneto. For each spell, annual wages and the number of days worked per year are reported for the period 1975-2001. We follow the same sample selection strategy as in Kline et al. (2020). Specifically, we focus on periods 1999 and 2001, on workers whose ages are between 18 and 64, and on "dominant jobs" (i.e., spells) where the worker earned the most in a given year. We also exclude workers who work in the public sector and whose wages and days worked are

[13]This dataset was developed by the Economics Department in Università Ca' Foscari Venezia under the supervision of Giuseppe Tattara.

outliers.[14]

From this sample, we construct mobility network data where nodes represent firms and edges represent movers between origin and destination firms. The construction is similar to the procedure we discussed in Example 1.1; our handling of the data corresponds to the two-period AKM model. We focus on movers who change firms from i to j between 1999 and 2001. We exclude firms with fewer than 15 movers, similarly to the procedure in Bonhomme et al. (2023). The mobility network is then defined by the set of firms and the movers, and we extract the largest connected component from this network.[15] The resulting graph represents our $\mathcal{G}$, and the wage differences constitute the outcome vector $\mathbf{y}$ in our model.

The resulting graph consists of $|V| = 4394$ firms and $|E| = 51{,}878$ movers. Table 1.5 presents the degree distribution of the firm-firm network, which is sparse with a mean degree of 18. Table 1.6 provides global measures of the network. Notably, the connectivity measure $\lambda_{2,L}$ is very small and close to zero, indicating that the graph is not well-connected and can be easily separated into components. The dependence measure $\lambda_{n,F}$ exceeds the maximum degree, and since $\lambda_{2,L}$ is small while $\lambda_{n,F}$ is large, the global measure H_n is also large. This suggests that, on average, the first-order approximation in Theorem 1.1 may not be valid, and the finite-sample bias in $\widehat{V}^{bc}_{\boldsymbol{\alpha}}$ is likely to be non-negligible.

[14]Specifically, "outliers" are defined as workers who (i) report a daily wage less than 5 euros or have zero days worked, (ii) report a log daily wage change one year to the next greater than 1 in absolute value, (iii) have more than 10 jobs in any year or have missing gender.

[15]When extracting the largest connected component, we also discard some movers who move between firms that are not in the largest connected component. Thus, the final mobility network contains firms with movers less than 15 as indicated in Table 1.5

Table 1.5: Degree Distribution for the Firm-Firm Network

min	Q1	Q3	max	mean
1.0	6.0	17.0	1092.0	18.748

Note: The first and fourth columns report the minimum and maximum degree of the firm-firm network. The second and third columns report the 25th and 75th percentiles of the degree distribution. The fifth column reports the average degree of the firm-firm network.

Table 1.6: Global Measures for the Firm-Firm Network

$\lambda_{2,L}$	$\lambda_{n,F}$	$\lvert C_n \rvert$	η_n	H_n
0.006	1779.6	4611.0	0.121	6248.04

Note: The first column reports the second smallest eigenvalue of the normalized Laplacian matrix. The second column reports the largest eigenvalue of the signless Laplacian matrix. The third column reports the number of nodes with $c_i > 0.2$. The fourth and fifth columns report the global measures of convergence.

1.6.2 Results: Inference for Central Firms

First, we identify the most central firm in the firm-firm mobility network by computing the PageRank centrality for each firm. PageRank centrality measures the importance of a node based on its connections to other highly central nodes.[16] In the context of the mobility network, a firm with high PageRank centrality is one that receives many workers from other firms that themselves attract many movers. We are interested in whether the fixed effect of this central firm is significantly different from others, as this may indicate that firm fixed effects play an important role in shaping the mobility network. As a related contribution, Sorkin (2018) provides a structural interpretation of PageRank centrality when applied to a mobility network.

Applying the PageRank algorithm to our firm-firm mobility network, we find that the most central firm is firm 931 in our dataset. Table 1.8 summarizes this

[16] See Newman (2018) for the precise definition and discussion of PageRank centrality.

Table 1.7: Distribution of the Firm Fixed Effects

min	**Q1**	**Q3**	**max**	**mean**
-1.41	-1.102	0.096	0.868	-0.017

Note: The first column reports the minimum of the firm fixed effects, the second and third columns report the 25th and 75th percentiles of the firm fixed effects, respectively. The fourth column reports the maximum of the firm fixed effects, and the last column reports the mean of the firm fixed effects.

firm's key characteristics. Its estimated fixed effect $\widehat{\alpha}_{931}$ is in approximately the top 34% of the distribution of firm fixed effects. As a central firm, it has a larger number of inflows d_{931}^t than outflows d_{931}^s, which is consistent with the idea that central firms attract more workers. The ratio $H_{931} = \lambda_{n,F}/(d_{931}\lambda_{2,L}h_{931})$ is about 2.90, which is much smaller than the average value $H_n \approx 6248$ and comparable to what we observed in the simulation study when the graph is not well-connected (see Table 1.3). This suggests that our inference procedure based on the first-order approximation in Theorem 1.1 may be valid for this central firm, and we can construct a confidence interval for α_{931}.

Table 1.8: Central Firm Information

$\widehat{\alpha}_{931}$	d_{931}	d_{931}^t	d_{931}^s	H_{931}
0.058	455	309	146	2.90

Note: The first column reports the estimated fixed effect of the central firm, the second to fourth columns report the firm's degree, the number of inflows, and the number of outflows, respectively. The last column reports the convergence measure $H_{931} = \lambda_{n,F}/(d_{931}\lambda_{2,L}h_{931})$.

The 95% confidence interval for α_{931} is:

$$CI_{931,0.95} = \left[\underbrace{\widehat{\alpha}_{931}}_{=0.058} - \underbrace{\left(\frac{d^t_{931} - d^s_{931}}{d_{931}} \right)}_{=0.358} \underbrace{\widehat{c}_{0.975}}_{=0.234}, \quad \widehat{\alpha}_{931} - \left(\frac{d^t_{931} - d^s_{931}}{d_{931}} \right) \underbrace{\widehat{c}_{0.025}}_{=-0.201} \right]$$

$$= [-0.026, \quad 0.130].$$

Note that the mean of the firm fixed effects is -0.017 (see Table 1.7), which lies within this confidence interval. This suggests that the central firm's fixed effect is not significantly larger than the average firm in the network. Since a higher firm fixed effect is typically associated with higher wages, this finding implies that the central firm does not pay significantly higher wages than the average firm, despite its central position reflecting its attractiveness to workers. This is consistent with recent literature suggesting that non-wage amenities—such as location, working environment, and firm culture—may play a more significant role in shaping worker mobility than wages alone (see Mas, 2025 for a recent discussion).

For comparison, we can also construct confidence intervals based on the asymptotic normality of $\widehat{\alpha}_i$ under the assumption of independent errors, as in the simulation section. The 95% confidence interval for $\widehat{\alpha}_{931}$ using this conventional method is

$$CI^{\text{Normal}}_{931,0.95} = [0.031, \quad 0.085].$$

This interval does not include the mean of the firm fixed effects, suggesting that the central firm's fixed effect is significantly different from the average firm.

However, given the under-coverage observed for this method in the simulation study, it likely overstates the significance of the central firm's fixed effect compared to our proposed approach. This highlights the empirical relevance of dependence in the error terms and the importance of accounting for it in inference.

This exercise demonstrates the practical applicability of our inferential method for fixed effects, which is rarely implemented in empirical work. Another potential application is hypothesis testing to assess whether the fixed effects of firms within a particular group (e.g., firms with similar observable characteristics) are statistically indistinguishable. Such a test could help determine whether grouping fixed effects, as suggested by Bonhomme et al. (2019), is appropriate in empirical analyses.

1.6.3 Results: Variance Estimation

We estimate the variance of the firm fixed effects $\widehat{V}_{\boldsymbol{\alpha}}$ and several bias-corrected estimators: $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$, $\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$, and $\widehat{V}_{\boldsymbol{\alpha}}^{a}$. Table 1.9 presents the estimated variance of the firm fixed effects $\boldsymbol{\alpha}$ and the bias-corrected estimators.

Our results show that the bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$ reduces the plug-in estimator $\widehat{V}_{\boldsymbol{\alpha}}$ by approximately 28%, which is a larger reduction than that achieved by the Andrews et al. (2008)-type bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{a}$ (about 23%). The modified bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$ yields an even lower estimate. These findings indicate that the dependence structure in the data is substantial and that our proposed bias-correction methods more effectively account for this dependence, resulting in more accurate estimation of the variance of the firm fixed effects.

Table 1.9: Variance Component Estimation on the VWH Data

$\widehat{V}_{\boldsymbol{\alpha}}$	$\widehat{V}_{\boldsymbol{\alpha}}^{bc}$	$\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$	$\widehat{V}_{\boldsymbol{\alpha}}^{a}$
0.039	0.028	0.025	0.030

Note: This table reports the estimated variances of the firm fixed effects $\boldsymbol{\alpha}$ via the plug-in estimator $\widehat{V}_{\boldsymbol{\alpha}}$, the bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc}$, the modified bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{bc,mod}$, and the Andrews et al. (2008)-type bias-corrected estimator $\widehat{V}_{\boldsymbol{\alpha}}^{a}$.

If we report the two components of the bias in (1.13) separately, we obtain

$$\underbrace{\widehat{\sigma}_{\tau}^{2}}_{=0.015} \times \underbrace{\frac{\operatorname{tr}(\mathbf{B}\mathbf{L}^{*}\boldsymbol{M}_{n}\mathbf{L}^{*}\mathbf{B}'\mathbf{F}\mathbf{F}')}{n}}_{=0.695}, \quad \underbrace{(\widehat{\sigma}^{2} - 2\widehat{\sigma}_{\tau}^{2})}_{=0.022} \times \underbrace{\frac{\operatorname{tr}(\boldsymbol{M}_{n}\mathbf{L}^{*})}{n}}_{=0.204}.$$

Note that $\widehat{\sigma}_{\tau}^{2}$, which is an estimator for $\mathbb{E}[\tau(U_i)^2]$, is about 0.015, which is roughly one-quarter of $\widehat{\sigma}^{2}$, the biased estimator for $\mathbb{E}[\epsilon_e^2]$. This disparity suggests that in our data the dependence structure plays a non-negligible role in generating the error terms, and this effect is transmitted to the bias in the variance estimator. Moreover, although there are uncertainties in estimating $\mathbb{E}[\tau(U_i)^2]$ and $\mathbb{E}[\epsilon_e^2]$, a heuristic comparison of the two quantities characterizing the graph structure

$$\frac{\operatorname{tr}(\mathbf{B}\mathbf{L}^{*}\boldsymbol{M}_{n}\mathbf{L}^{*}\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} \text{ and } \frac{\operatorname{tr}(\boldsymbol{M}_{n}\mathbf{L}^{*})}{n}$$

indicates that the first term is more than three times larger than the second term. Thus, even if $\widehat{\sigma}_{\tau}^{2}$ were estimated to be smaller than the current estimate, the bias originating from the dependence structure would remain significant.

This exercise demonstrates that the bias in variance estimation due to dependence, which has been largely overlooked in the literature, can be substantial in empirical applications. It also provides evidence that the dependence structure in the data is non-negligible, highlighting the practical usefulness of our bias-correction method. See Appendix A.4 for additional results on covariance estimation between worker and firm fixed effects.

1.7 Conclusion

This paper proposes a new inference method for fixed effects in network regressions with dependent errors. We derive a first-order approximation to the fixed-effect estimator and show that, under node- and edge-level dependence, the estimator can be inconsistent. Building on this approximation, we develop a non-standard inference procedure that exploits the empirical distribution of appropriately constructed residual averages to obtain valid confidence intervals and hypothesis tests for fixed effects.

We also study estimation of distributional moments of the fixed effects. We characterize the bias of the plug-in variance estimator under the dependence structure and propose a bias correction that is consistent in the presence of node-level dependence.

Monte Carlo simulations and an application to Italian matched employer-employee data illustrate the finite-sample performance of the methods. The results show that network dependence can substantially distort inference for fixed effects and is an important source of bias in variance estimation.

Future research could explore formal extensions to models with multiple types of fixed effects such as worker and firm effects in two-way models, and develop methods for estimating covariance between different fixed-effect types, where preliminary results are provided in Appendix A.4. Another avenue is to devise inference procedures for variance components that go beyond the bias-correction proposed here. Finally, it would be useful to extend our methods to high-dimensional, team-based settings in which nodes may belong to multiple teams (e.g., Bonhomme, 2021), thereby accounting for dependence within and across teams.

Chapter 2

Design-Based and Network Sampling-Based Uncertainties in Network Experiments

2.1 Introduction

Network experiments, or randomized controlled trials (RCTs) on networks, have become increasingly common in applied economics (e.g., Cai et al., 2015a; Dizon et al., 2020; Carter et al., 2021a; Fernando, 2021; Beaman et al., 2021a). A central objective of these experiments is to estimate the "spillover effect" of policy interventions as they propagate through networks. For example, Cai et al. (2015a) estimate spillover effects from randomly assigned information sessions on rice farmers' decisions to purchase a weather insurance product in Chinese villages. In this paper, we develop a comprehensive theoretical framework for ordinary least squares (OLS) estimators in network experiments, explicitly accounting for both design-based uncertainty, arising from randomness in treatment assignment, and sampling-based uncertainty, arising from randomness in sampling units and network links. Our theory is motivated by two key gaps between empirical practice in applied work and existing econometric theory.

The first gap lies in the choice of estimator. In applications, researchers predominantly use OLS estimators to estimate spillover effects, employing exposure mappings that summarize treatment status and network structure. In our survey of 29 papers analyzing network experiments, published in the "top

5" economics journals and two leading field journals, all of the studies report using the OLS estimator, while only two papers use propensity score-based estimators.[1] This pattern stands in contrast to the theoretical literature on inference in network experiments (e.g., Aronow and Samii, 2017; Leung, 2022a; Gao and Ding, 2023), which provides inference results for inverse probability weighting (IPW) estimators that directly estimate average spillover effects.

The other gap is due to ignoring a source of randomness. In many applied cases, researchers need to collect network information through surveys. This collection process can introduce an extra layer of uncertainty beyond design-based uncertainty. Moreover, the collected network may only partially capture the true network governing the propagation mechanism. By contrast, the theoretical literature on causal inference in network experiments typically abstracts away from sampling-based uncertainty, assuming that the data correspond to the entire population and that the observed network is complete.

To address these gaps, we make three contributions. First, we develop a novel framework that jointly incorporates design-based randomness in treatment assignment and sampling-based randomness in network links. Our framework considers a finite population of n units, from which units are randomly sampled and treatments are assigned. We explicitly model the network sampling process, focusing on two common sampling methods: (i) induced subgraph sampling,

[1] Specifically, we considered papers published from April 2010 through April 2025 in the following journals: American Economic Review, Econometrica, Quarterly Journal of Economics, Journal of Political Economy, Review of Economic Studies, American Economic Journal: Applied Economics, and Journal of Development Economics. We searched for articles that listed "networks" and either "field experiments" or "randomized trial" as keywords on the Web of Science platform. This search resulted in 52 papers, of which 29 conducted network experiments and are mentioned in the text. These papers are referenced in Appendix B.5.

where each sampled unit reports friends within the sample, and (ii) star sampling, where each sampled unit reports friends from the entire population. In this setup, unlike in non-network experiments, sampling-based uncertainty arises from two sources: (i) which units are sampled, and (ii) which links are observed. We consider potential outcomes that depend on the entire treatment vector, thus violating the Stable Unit Treatment Value Assumption (SUTVA). To address the resulting dimensionality problem, we assume that the potential outcomes are linear in an exposure mapping, a set of user-specified sufficient statistics summarizing treatment status and network structure. For example, a common exposure mapping includes the fraction of one's friends who are treated. Importantly, we do not assume that the user-specified exposure mapping is correctly specified; it may differ from the true exposure mapping in both functional form and dimension. This flexibility also allows us to incorporate censored network links in a unified way.

As our second contribution, we investigate whether the estimands associated with the OLS estimator can be interpreted as causal spillover effects. We distinguish between two causal targets: a population-level estimand and a sample-level estimand. The population-level estimand is defined as the weighted average of the treatment effect vector across the entire population, including those who are not sampled, with complete network information. On the other hand, the sample-level estimand is defined as the sample average of the treatment effect vector across the sampled units, with the sampled network information. We show that both types of estimands can be contaminated: each element of the multi-dimensional estimands may reflect causal effects from other elements of the exposure mapping vector. With heterogeneous treatment effects, correlations among elements in the exposure mapping vector (e.g., the proportion of treated friends and the proportion of friends' treated friends) blur the distinction between the true causal effects in one element and those in another. Although the population-level causal estimand can be free from contamination if the exposure mapping is defined such that there is no correlation among its elements, the sample-level causal estimand can still be subject to contamination, and thus lacks causal interpretability due to network sampling. Missing links can create undesirable correlations between the observed and true exposure mapping across different elements. As a result, the two estimands can remain distinct even in large samples unless the exposure mapping is correctly specified and the network links for the neighborhood are completely sampled.

In our third contribution, we derive asymptotic theory for the OLS estimator and find conditions under which the OLS estimator approximates the estimands. We show that the OLS estimator is consistent for the sample-level causal estimand, conditionally or unconditionally on the sampling uncertainty. However, because the sample-level causal estimand generally lacks causal interpretability, results from OLS estimation should be interpreted with caution. If the exposure mapping is correctly specified and there is no potential correlation between the true and observed exposure mappings, the sample-level causal estimand is consistent for the population-level causal estimand; thus, we can guarantee a clear interpretation of the OLS estimator. We further derive the estimator's asymptotic distribution and provide a conservative network heteroskedasticity and autocorrelation consistent (HAC) variance estimator.

This paper contributes to the literature on design-based inference in network experiments (Aronow and Samii, 2017; Leung, 2022a; Gao and Ding, 2023; Hoshino and Yanagi, 2024). Previous works have primarily focused on design-based uncertainty, where treatment assignment is the only source of randomness and complete network information is assumed to be available without sampling uncertainty. Additionally, these works have mainly considered IPW estimators, which allow for direct estimation of causal spillover effects, while the OLS estimator has received less attention. To focus on IPW estimators, these works typically assume that the exposure mapping takes discrete values, such as an indicator of whether a unit has at least one treated friend.[2] In contrast, this paper considers both design-based and sampling-based uncertainties with an explicit network collection process, and focuses on the OLS estimator with exposure mappings as regressors, which is widely used in empirical applications and allows for continuous exposure mappings.

This paper also relates to the literature on simultaneous design-based and sampling-based inference (see Abadie et al., 2020; Xu and Wooldridge, 2022; Abadie et al., 2023; Viviano, 2024). Our framework extends the approach of Abadie et al. (2020) to the network setting by allowing for both design-based and sampling-based uncertainties in network experiments, and by focusing on both population-level and sample-level estimands. We differ from Abadie et al. (2020) in several important respects. First, we explicitly model network sampling, where the observed network may be only partially observed. Second, we study the OLS estimator with exposure mappings as regressors, which induces dependence among outcomes and between regressors and sampling

[2] Gao and Ding (2023) discuss the potential application of IPW-based estimators to continuous exposure mappings.

indicators, features not present in their analysis. Third, we provide an element-wise causal interpretation of the estimands and the OLS estimator, which is not addressed in their work. Relatedly, Viviano (2024) also considers both design-based and sampling-based uncertainties, including uncertainty arising from network sampling. However, while his approach assumes that all relevant network information for computing the true exposure mapping is observed, our framework allows for the possibility that some relevant network information is unobserved due to sampling uncertainty. Additionally, while Viviano (2024) focuses on a sample-level estimand that maximizes a welfare measure, our study is concerned with inference for both population-level and sample-level causal estimands, emphasizing the potential divergence between the two.

This paper is also related to the literature studying the impact of network data collection on parameters of interest (Chandrasekhar and Lewis, 2016; Griffith, 2022; Lewbel et al., 2023; Hsieh et al., 2024). While these papers share a similar motivation in that the network sampling process can affect the estimation of spillover effects, they primarily focus on the potential bias of estimators with respect to homogeneous parameters due to network sampling. In contrast, this paper focuses on the causal interpretability of the OLS estimator with heterogeneous spillover effects. This distinction is important because attenuation bias, as highlighted for example in Chandrasekhar and Lewis (2016), does not necessarily hinder learning about spillover effects if the estimator preserves the sign of the underlying effects. However, we show that the OLS estimator with exposure mappings may not preserve the sign of the true spillover effects due to contamination bias, potentially leading to misleading conclusions.

More broadly, this paper contributes to the literature on the causal interpretability of estimators in linear regressions with heterogeneous treatment effects (Angrist, 1998; Goldsmith-Pinkham et al., 2022; Borusyak and Hull, 2024). In particular, Goldsmith-Pinkham et al. (2022) show that the OLS estimator with multi-dimensional treatment indicators can be contaminated in the presence of heterogeneous treatment effects, which aligns with our findings in Corollary 2.1. There are two important differences. First, we consider a finite population model, whereas Goldsmith-Pinkham et al. (2022) focus on an infinite population model, making it nontrivial to extend their results to our setting. Second, we allow for general exposure mappings as regressors, while Goldsmith-Pinkham et al. (2022) restrict attention to mutually exclusive multi-dimensional treatment indicators. In our context, contamination bias arises from overlaps in the treatment status across elements of the exposure mapping, whereas such overlaps are not possible in the non-network setup of Goldsmith-Pinkham et al. (2022).

The remainder of this paper is organized as follows. Section 2.2 introduces the framework for network sampling, the model, and assumptions. Section 2.3 presents the main results, including a causal interpretation and asymptotic theory. Section 2.4 proposes a network heteroskedasticity and autocorrelation consistent (HAC) estimator for the standard errors. Section 2.5 provides a simulation study to illustrate the finite sample properties of the proposed estimator. Section 2.6 applies the proposed method to a real-world dataset. Section 2.7 concludes the paper and provides a flowchart (Figure 2.3) outlining recommended steps for conducting inference in network experiments using the OLS estimator. Appendix B.1 discusses how to estimate the nuisance parameters consistently, Appendix B.2 contains technical lemmas, Appendix

B.3 contains proofs, Appendix B.4 presents additional simulation results, and Appendix B.5 lists the papers included in the survey of network experiment research presented in the Introduction.

2.2 Model

In this section, we first outline our framework for modeling network experiments. We then introduce the estimands of interest, which are defined both for the entire population and for the sampled group, as well as the OLS estimator used to estimate these estimands.

2.2.1 Population

As in Abadie et al. (2020), we consider a sequence of finite populations. There are finitely many units ($n < \infty$) in the population, denoted by $\mathcal{N}_n = \{1, ..., n\}$. These units are connected through the network represented by an adjacency matrix $\mathbf{A}_n = [A_{n,i,j}]_{i,j \in \mathcal{N}_n} \in \{0,1\}^{n \times n}$. We define $A_{n,i,j} = 1$ if there is a network link between units i and j, and $A_{n,i,j} = 0$ otherwise. We assume that the network is undirected ($A_{n,i,j} = A_{n,j,i}$) and has no self-loops ($A_{n,i,i} = 0$). Each unit i is characterized by a vector of covariates $Z_{n,i} \in \mathcal{Z}_n \subset \mathbb{R}^{dz}$, potential outcomes $Y^*_{n,i}(\cdot) \in \mathcal{Y}_n \subset \mathbb{R}$ that depend on the entire vector of binary treatments $\mathbf{D}_n = [D_{n,i}]_{i \in \mathcal{N}_n} \in \{0,1\}^n$. We consider the setup where the researcher assigns treatments only to the sampled units, but spillovers to non-sampled units are allowed. The covariates $Z_{n,i}$ include both network information (e.g., number of i's neighbors, degree: $\text{deg}_{n,i} = \sum_{j \neq i} A_{n,i,j}$) and individual information (e.g., i's age). Also, the potential outcomes may violate the Stable Unit Treatment

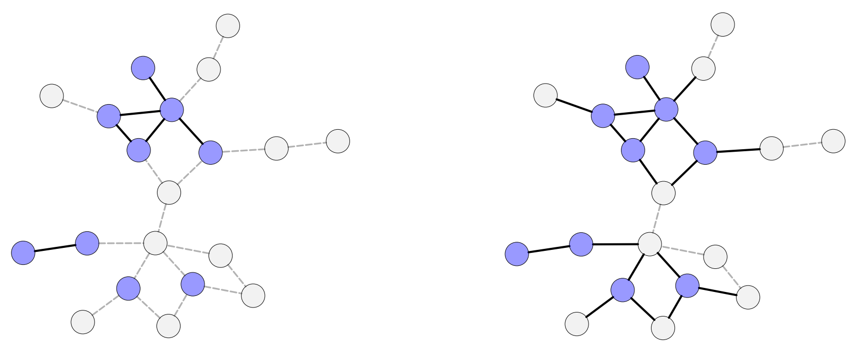
Value Assumption (SUTVA) by allowing for others' treatment status as inputs.

2.2.2 Sampling

From a finite population of n units, we draw a sample of $N = \sum_{i=1}^{n} R_{n,i}$ units (hence $n \geq N$), where $R_{n,i} \in \{0,1\}$ is the sampling indicator for the i-th unit: $R_{n,i} = 1$ if i is in the sample and otherwise $R_{n,i} = 0$. Given the sampling indicator vector $\boldsymbol{R}_n$, partial elements of the true network $\mathbf{A}_n$ are sampled. We denote the sampled network, given the sampling indicator vector $\boldsymbol{R}_n$, as $\widetilde{\mathbf{A}}_n(\boldsymbol{R}_n)$. When the dependence on $\boldsymbol{R}_n$ is clear from context, we simply write $\widetilde{\mathbf{A}}_n$. The sampled adjacency matrix $\widetilde{\mathbf{A}}_n = [\widetilde{A}_{n,i,j}]_{i,j \in \mathcal{N}_n} \in \{0,1\}^{n \times n}$ has (i,j)-element $\widetilde{A}_{n,i,j}$, which equals one if there is a true network link between units i and j and the link is sampled, and zero otherwise.

In this paper, we focus on two canonical network sampling methods: (i) *induced subgraph sampling*, and (ii) *star sampling*. In the induced subgraph sampling case, we sample $\widetilde{\mathbf{A}}_n = \boldsymbol{R}_n \boldsymbol{R}_n' \odot \mathbf{A}_n$ where $\odot$ is the element-wise product and the (i,j)-element of $\widetilde{\mathbf{A}}_n$, $\widetilde{A}_{n,i,j} = R_{n,i} R_{n,j} A_{n,i,j}$ represents a network link between the units i and j, which is sampled if both units are sampled. In the star sampling case, we sample $\widetilde{\mathbf{A}}_n = (\mathbf{1}_n \mathbf{1}_n' - (\mathbf{1}_n - \boldsymbol{R}_n)(\mathbf{1}_n - \boldsymbol{R}_n)') \odot \mathbf{A}_n$, where $\widetilde{A}_{n,i,j} = \max\{R_{n,i}, R_{n,j}\} A_{n,i,j}$ represents a network link between the units i and j, which is sampled if at least one of the two units is sampled. Sampled networks under induced subgraph and star sampling are illustrated in Figure 2.1. In the figure, the sampled units are in blue, and the non-sampled units are in light gray. The sampled links are shown as solid black lines, and the non-sampled links as dashed gray lines. In practice, if the researcher asks the sampled units to list their friends *from the list of sampled units*, the induced subgraph sampling network is sampled (e.g., Conley and Udry, 2010; Dizon et al., 2020; Carter et al., 2021a). If the researcher asks the sampled units to list their friends *from the population*, the star sampling network is sampled (e.g., Banerjee et al., 2013; Cai et al., 2015a; Beaman et al., 2021a). See Section 5.3 of Kolaczyk and Csárdi (2014) for further examples of network sampling.

Figure 2.1: Comparison of induced subgraph sampling (left) and star sampling (right).



(a) Induced subgraph sampling (b) Star sampling

Note: Blue nodes indicate sampled units, while light gray nodes denote non-sampled units. Solid black links are observable to the researcher; dashed gray links are unobserved.

We denote the observed covariates by $\widetilde{Z}_{n,i}$, which may differ from $Z_{n,i}$ due to network sampling. For example, if $Z_{n,i}$ includes i's degree, then $\widetilde{Z}_{n,i}$ contains i's degree computed from the sampled network $\widetilde{\mathbf{A}}_n$: $\widetilde{\deg}_{n,i} = \sum_{j \neq i} \widetilde{A}_{n,i,j}$. Note that we allow both $Z_{n,i}$ and $\widetilde{Z}_{n,i}$ to depend on $\boldsymbol{R}_n$.

Throughout the paper, we maintain the following assumption regarding the sampling process and the assignment mechanism.

Assumption 2.1. *(i) Random sampling:*

$$R_{n,i} \sim \textit{Bernoulli}(\rho_n) \ \textit{i.i.d.},$$

where $\rho_n \in (0,1]$ is a sequence of sampling probabilities such that $\rho_n \to \rho \in (0,1]$.

(ii) Network sampling: Given a fixed entire network sequence $\mathbf{A}_n$, the (i,j)-element of sampled network $\widetilde{\mathbf{A}}_n$ is generated by the induced subgraph sampling $\widetilde{A}_{n,i,j} = R_{n,i}R_{n,j}A_{n,i,j}$ or the star sampling $\widetilde{A}_{n,i,j} = \max\{R_{n,i}, R_{n,j}\} \times A_{n,i,j}$.

(iii) Treatment assignment mechanism: Let $\boldsymbol{R}_{n,-i}$ denote the vector $\boldsymbol{R}_n$ excluding the i-th element, $R_{n,i}$. The assignment mechanism $D_{n,i}$ is independent of $\boldsymbol{R}_{n,-i}$ and drawn independently (but not necessarily identically) from a known distribution. The distribution of $D_{n,i}$ is degenerate at 0 if and only if $R_{n,i} = 0$.

Assumption 2.1 (iii) implies $D_{n,i} = 0$ if $R_{n,i} = 0$, which means we treat only the sampled units. The simplest example is

$$D_{n,i} \sim \text{Bernoulli}(R_{n,i}p_{n,i}) \text{ independently.} \tag{2.1}$$

While we use the Bernoulli assignment in (2.1) for all illustrations in the paper, our theoretical results accommodate more general assignment mechanisms, as specified in Assumption 2.1 (iii). Since Assumption 2.1 (iii) does not require the identical draws, $p_{n,i}$ could depend on $\mathbf{A}_n$, $Z_{n,i}$ or other observed characteristics of unit i. We can equivalently write Assumption 2.1 (iii) as $D_{n,i} = R_{n,i}D^*_{n,i}$, where $D^*_{n,i}$ is defined as the latent treatment indicator generated by $D^*_{n,i} \sim \text{Bernoulli}(p_{n,i})$ (or more general distribution satisfying the assumption) independently. Note also that the treatment assignment mechanism is known to the researcher, which is satisfied in a randomized controlled trial and commonly assumed in the design-based inference literature.

Remark 2.1. *Assumption 2.1 (i) rules out cluster sampling and multi-wave network sampling, because in such designs* $R_{n,i}$ *may depend on* $R_{n,j}$ *for some* $j \neq i$ *through the cluster or the network, respectively. Assumption 2.1 (ii) prohibits censoring of* $\widetilde{\mathbf{A}}_n$*; when censoring occurs, it can be treated as misspecification of the exposure mapping (see Example 2.4). Assumption 2.1 (iii) excludes complex assignment schemes, such as matched-pair or blocked randomization.*

2.2.3 Potential Outcome

As discussed above, each unit's potential outcome $Y^*_{n,i}(\cdot)$ is a function of the full treatment vector $\mathbf{D}_n$. By Assumption 2.1 (iii), we can write $\mathbf{D}_n = \boldsymbol{R}_n \odot \mathbf{D}^*_n$, where $\mathbf{D}^*_n = [D^*_{n,i}]_{i \in \mathcal{N}_n}$. Following the literature (e.g., Aronow and Samii, 2017), we assume that there is an exposure mapping $T_{n,i} \in \mathcal{T}_n \subset \mathbb{R}^{d_T}$ that essentially determines i's potential outcome by summarizing the network structure and the treatment status vector. See Section 2.2.4 below for a detailed definition and discussion on the exposure mapping. We consider a linear potential outcome model, so that for each $t \in \mathcal{T}_n$, $Y^*_{n,i}(t)$ is defined as follows.

Assumption 2.2. *For all* $t \in \mathcal{T}_n$*,*

$$Y^*_{n,i}(t) = t'\theta_{n,i} + \nu_{n,i},$$

where $\theta_{n,i}$ *and* $\nu_{n,i}$ *are non-stochastic.*

Note that $\theta_{n,i}$ is a vector of heterogeneous treatment effects. Each element $\theta_{n,i,(k)}$ represents the marginal effect of the k-th component of the exposure mapping $T_{n,i}$ on the potential outcome $Y^*_{n,i}(t)$. For example, if the k-th component of $T_{n,i}$ is the share of treated friends, then $\theta_{n,i,(k)}$ captures the causal spillover effect from i's treated friends on i's outcome. Since $\theta_{n,i}$ is non-stochastic, it may depend on the population network $\mathbf{A}_n$, allowing for heterogeneity based on network structure and unit i's position.

Although a linear model may seem restrictive, when $|\mathcal{T}_n|$ is finite (e.g., $\mathcal{T}_n = \{0,1\}^2$), this assumption is without loss of generality as discussed in Abadie et al. (2020). The realized outcome is $Y_{n,i} = Y^*_{n,i}(T_{n,i})$. Thus, the outcome depends on $\mathbf{D}_n$ only via the exposure mapping $T_{n,i}$.

2.2.4 Exposure Mapping

Let the true exposure mapping be $T_{n,i} = g(i, \mathbf{D}_n, \mathbf{A}_n) \in \mathcal{T}_n \subset \mathbb{R}^{d_T}$, where $g : \mathcal{N}_n \times \{0,1\}^n \times \{0,1\}^{n \times n} \to \mathcal{T}_n$ is a function that generates the true exposure mapping for each unit. Specifically, for unit i, it takes (i) i's index, (ii) the treatment vector $\mathbf{D}_n$, and (iii) the true network $\mathbf{A}_n$ as inputs, and returns a lower-dimensional vector of summary statistics for the outcome. For example, applied researchers use the presence of i's treated friends and the share of i's treated friends as the exposure mapping.

This paper allows the researcher to misspecify the functional form of g. For example, if the researcher uses the presence of treated friends as the exposure mapping, while the true potential outcome is linear in the share of treated friends, then the exposure mapping is misspecified. We denote this misspecified exposure mapping function by $\widetilde{g}_n : \mathcal{N}_n \times \{0,1\}^n \times \{0,1\}^{n \times n} \to \widetilde{\mathcal{T}}_n$, where

$\widetilde{\mathcal{T}}_n \in \mathbb{R}^{d_{\widetilde{T}}}$. Note that the dimensions d_T and $d_{\widetilde{T}}$ may differ. The functional form $\widetilde{g}_n$ could depend on the sample size n (as in Example 2.3), but for notational simplicity, we omit the subscript n. We assume that dimensions d_T and $d_{\widetilde{T}}$ are constants independent of n.

If $\widetilde{g} = g$, then the observed exposure mapping $\widetilde{T}_{n,i}$ can be written as $\widetilde{T}_{n,i} = g(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n)$. That is, the only difference between the true exposure mapping and the observed exposure mapping is the network input, between $\mathbf{A}_n$ and $\widetilde{\mathbf{A}}_n$. More generally, if the researcher misspecifies g as $\widetilde{g}$, then the observed exposure mapping is $\widetilde{T}_{n,i} = \widetilde{g}(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n)$. In this case, the dimensions d_T and $d_{\widetilde{T}}$ may differ.

Below, we provide some examples of exposure mappings.

Example 2.1. *Suppose that the true exposure mapping is i's own treatment indicator:*

$$T_{n,i} = g(i, \mathbf{D}_n, \mathbf{A}_n) = D_{n,i} = R_{n,i} D^*_{n,i}.$$

Note that the exposure mapping does not depend on the network information, and as long as the researcher correctly specifies the exposure mapping $g = \widetilde{g}$, *we have* $T_{n,i} = \widetilde{T}_{n,i}$ *for all* $i \in \mathcal{N}_n$.

Example 2.2. *Suppose that the true exposure mapping is an indicator of the* *existence of at least one treated friend:*

$$\begin{aligned} T_{n,i} &= g(i, \mathbf{D}_n, \mathbf{A}_n) \\ &= \mathbb{1}\left\{ \sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j} > 0 \right\}, \end{aligned}$$

and the researcher correctly specifies the exposure mapping as $\widetilde{T}_{n,i} = g(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n)$*. Thus, for the induced subgraph sampling case (*$\widetilde{A}_{n,i,j} = R_{n,i} R_{n,j} A_{n,i,j}$*),*

$$\begin{aligned} \widetilde{T}_{n,i} &= \mathbb{1}\left\{ \sum_{j \neq i} R_{n,i} R_{n,j} A_{n,i,j} R_{n,j} D^*_{n,j} > 0 \right\} \\ &= \mathbb{1}\left\{ R_{n,i} \sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j} > 0 \right\}. \end{aligned}$$

Thus, when $R_{n,i} = 1$*, we have* $T_{n,i} = \widetilde{T}_{n,i}$*. For the star sampling case (*$\widetilde{A}_{n,i,j} = \max\{R_{n,i}, R_{n,j}\} A_{n,i,j}$*),*

$$\begin{aligned} \widetilde{T}_{n,i} &= \mathbb{1}\left\{ \sum_{j \neq i} \max\{R_{n,i}, R_{n,j}\} A_{n,i,j} R_{n,j} D^*_{n,j} > 0 \right\} \\ &= \mathbb{1}\left\{ \sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j} > 0 \right\}, \end{aligned}$$

and we have $T_{n,i} = \widetilde{T}_{n,i}$ *for all* $i \in \mathcal{N}_n$*.*

Although the two preceding examples correctly specify the exposure mapping, the subsequent example fails to do so.

Example 2.3. *Suppose that the true exposure mapping is a vector of a direct* *treatment, a spillover treatment through a fraction of treated peers, and their interaction term:*

$$\begin{aligned} T_{n,i} &= g(i, \mathbf{D}_n, \mathbf{A}_n) \\ &= \left(R_{n,i} D^*_{n,i}, \frac{\sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j}}{\sum_{j \neq i} A_{n,i,j}}, R_{n,i} D^*_{n,i} \times \frac{\sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j}}{\sum_{j \neq i} A_{n,i,j}} \right). \end{aligned}$$

By convention, we usually define $0/0 = 0$ *in the case of isolated units. Suppose that the researcher misspecifies* $\widetilde{g}$ *as*

$$\widetilde{T}_{n,i} = \widetilde{g}(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n) = \left(R_{n,i} D^*_{n,i}, \mathbb{1} \left\{ \sum_{j \neq i} \widetilde{A}_{n,i,j} R_{n,j} D^*_{n,j} > 0 \right\} \right).$$

In this specification, it is evident that $g \neq \widetilde{g}$ *because* $d_T > d_{\widetilde{T}}$*. The misspecified* $\widetilde{g}$ *accounts only for the direct effect and the spillover effect represented by an indicator of the presence of at least one treated friend. Consequently, not only do the dimensions differ, but the structures of the variables capturing spillover effects are also distinct.*

2.2.5 Censored Network

We can also treat censoring on a sampled network as arising from a misspecified exposure mapping as $\widetilde{g}$ can specify which links in a sampled network $\widetilde{\mathbf{A}}$ to be used to compute the exposure mapping. This is empirically relevant as in practice, some studies impose a cap on the number of links each sampled unit can report, leading to a discrepancy between the sampled and censored networks. For example, in Cai et al. (2015a), each sampled unit was asked to report up to five closest friends, which potentially introduces censoring in the observed network. See also Griffith (2022) for further examples and a detailed discussion of censoring in network data collection. The following example illustrates how censoring can be framed as a misspecified exposure mapping:

Example 2.4. *Let g be the same as in Example 2.2. Suppose that the researcher misspecifies $\widetilde{g}$ due to the censoring as*

$$\begin{aligned}\widetilde{T}_{n,i} &= \widetilde{g}(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n) \\ &= g(i, \mathbf{D}_n, \boldsymbol{C}_n(\widetilde{\mathbf{A}}_n) \odot \widetilde{\mathbf{A}}_n) = \mathbb{1}\left\{\sum_{j \neq i} C_{n,i,j}(\widetilde{\mathbf{A}}_n)\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j} > 0\right\},\end{aligned}$$

where $\boldsymbol{C}_n(\widetilde{\mathbf{A}}_n)$ is the censoring indicator matrix whose (i,j)-element is $C_{n,i,j}(\widetilde{\mathbf{A}}_n) \in \{0,1\}$, a binary variable that indicates whether unit j is censored from i's perspective. The censoring indicator can be a random variable, as we allow it to be an unknown function of the sampled network $\widetilde{\mathbf{A}}_n$. For example, $C_{n,i,j} = 1$ when unit i (or j) with $R_{n,i} = 1$ (or $R_{n,j} = 1$) is asked to list their five closest friends and j (or i) is one of them.[3] *In this example, $g \neq \widetilde{g}$ in general and misspecification occurs due to the censoring.*

We distinguish between the sampled network $\widetilde{\mathbf{A}}_n$ and the censored network $\boldsymbol{C}_n(\widetilde{\mathbf{A}}_n) \odot \widetilde{\mathbf{A}}_n$, and the discrepancy is framed as the misspecification of the exposure mapping. This framework is useful for separating the sampling effect from the censoring. In the extreme case with $\rho_n = 1$, we sample the entire network $\widetilde{\mathbf{A}}_n = \mathbf{A}_n$, but the censoring still matters as we observe $\boldsymbol{C}_n(\mathbf{A}_n) \odot \mathbf{A}_n$. For convenience, we will omit the notational dependence of $\boldsymbol{C}_n$ on $\widetilde{\mathbf{A}}_n$.

[3] We can define $C_{n,i,i}(\widetilde{\mathbf{A}}_n)$ arbitrarily because $A_{n,i,i} = 0$.

The dependence of $\boldsymbol{C}_n$ on $\widetilde{\mathbf{A}}_n$ is motivated as follows. In practice, the censored induced subgraph sampling network is observed if the researcher asks the sampled unit to list a fixed number of closest friends *from the sampled friends*. Thus, it usually depends on $[\widetilde{A}_{n,i,j}]_{j\in\mathcal{N}_n}$. The censored star sampling network is observed if the researcher asks i with $R_{n,i} = 1$ to list a fixed number of closest friends *from their friends in population* $[A_{n,i,j}]_{j\in\mathcal{N}_n}$. Since $\widetilde{A}_{n,i,j} = A_{n,i,j}$ holds for $R_{n,i} = 1$ for the star sampling network, the censoring depends on $[\widetilde{A}_{n,i,j}]_{j\in\mathcal{N}_n}$. We also allow the arbitrary dependence of $\boldsymbol{C}_n$ on other deterministic variables, such as individuals' preferences regarding their friends, which is a benefit of the design-based framework.

2.2.6 Estimands and Estimator

To facilitate the introduction of our estimands and OLS estimator, we first transform the exposure mappings. Recall that the exposure mappings $T_{n,i}$ and $\widetilde{T}_{n,i}$ are random vectors that depend on $\boldsymbol{R}_n$ and $\mathbf{D}_n$, and the covariates $Z_{n,i}$ and $\widetilde{Z}_{n,i}$ are random vectors that depend only on $\boldsymbol{R}_n$. Define

$$X_{n,i} = T_{n,i} - \Lambda_n Z_{n,i}, \quad \text{and} \quad \widetilde{X}_{n,i} = \widetilde{T}_{n,i} - \widetilde{\Lambda}_n \widetilde{Z}_{n,i},$$

where

$$\Lambda_n = \left(\sum_{i=1}^{n} \mathbb{E}[T_{n,i} Z'_{n,i}]\right)\left(\sum_{i=1}^{n} \mathbb{E}[Z_{n,i} Z'_{n,i}]\right)^{-1},$$

and

$$\widetilde{\Lambda}_n = \left(\sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n]\widetilde{Z}'_{n,i}\right)\left(\sum_{i=1}^{n} R_{n,i}\widetilde{Z}_{n,i}\widetilde{Z}'_{n,i}\right)^{-1}.$$

That is, $X_{n,i}$ is the population residual of the regression of $T_{n,i}$ on $Z_{n,i}$, and $\widetilde{X}_{n,i}$ is the residual of the regression of $\widetilde{T}_{n,i}$ on $\widetilde{Z}_{n,i}$ using sampled units. Since we know the treatment assignment distribution with known $p_{n,i}$ and observe $\boldsymbol{R}_n$, we can calculate $\mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n]$ analytically.

Table 2.1 summarizes the conditional expectation of widely used exposure mappings when the assignment probability is homogeneous: $D^*_{n,i} \sim$ Bernoulli(p_n) i.i.d. The table focuses on the case where the exposure mapping is scalar. The researcher applies it element-wise for multi-dimensional cases. For the second neighborhood, the expectation can be calculated similarly. See also Example 2.10 below for the modification on multi-dimensional cases with the second neighborhood.

Table 2.1: Conditional Expectation of Exposure Mappings Frequently Used in Applied Research

Exposure Mapping	$\widetilde{T}_{n,i} = g(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n)$	$\mathbb{E}\left[\widetilde{T}_{n,i} \mid \boldsymbol{R}_n\right]$
Individual Treatment	$R_{n,i}D^*_{n,i}$	$R_{n,i}p_n$
Treated Friends Share	$\frac{\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j}}{\sum_{j\neq i}\widetilde{A}_{n,i,j}}$	$p_n \times \frac{\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}}{\sum_{j\neq i}\widetilde{A}_{n,i,j}}$
Treated Friends Number	$\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j}$	$p_n \times \sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}$
Treated Friends Existence	$\mathbb{1}\left\{\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j} > 0\right\}$	$1 - q_n^{\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}}$

Note: Assume that $R_{n,i} \sim$ Bernoulli(ρ_n) i.i.d. and $D^*_{n,i} \sim$ Bernoulli(p_n) i.i.d. In the last row, $q_n = 1 - p_n$ is the probability of not being treated. By convention, we usually set $\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j}/\sum_{j\neq i}\widetilde{A}_{n,i,j} = 0$ if $\sum_{j\neq i}\widetilde{A}_{n,i,j} = 0$.

To summarize relevant moments of the data, define the population matrix

Ω_n and the sample matrices $\widetilde{Q}_n$ and $\widetilde{\Omega}_n$:

$$\Omega_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\begin{pmatrix} Y_{n,i} \\ X_{n,i} \\ Z_{n,i} \end{pmatrix}\begin{pmatrix} Y_{n,i} \\ X_{n,i} \\ Z_{n,i} \end{pmatrix}'\right] \equiv \begin{pmatrix} \Omega_n^{YY} & \Omega_n^{YX} & \Omega_n^{YZ} \\ \Omega_n^{XY} & \Omega_n^{XX} & \Omega_n^{XZ} \\ \Omega_n^{ZY} & \Omega_n^{ZX} & \Omega_n^{ZZ} \end{pmatrix},$$

$$\widetilde{Q}_n = \frac{1}{N}\sum_{i=1}^{n} R_{n,i}\begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}\begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}' \equiv \begin{pmatrix} \widetilde{Q}_n^{YY} & \widetilde{Q}_n^{YX} & \widetilde{Q}_n^{YZ} \\ \widetilde{Q}_n^{XY} & \widetilde{Q}_n^{XX} & \widetilde{Q}_n^{XZ} \\ \widetilde{Q}_n^{ZY} & \widetilde{Q}_n^{ZX} & \widetilde{Q}_n^{ZZ} \end{pmatrix},$$

and

$$\widetilde{\Omega}_n = \frac{1}{N}\sum_{i=1}^{n} R_{n,i}\mathbb{E}\left[\begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}\begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}' \mid \boldsymbol{R}_n\right] \equiv \begin{pmatrix} \widetilde{\Omega}_n^{YY} & \widetilde{\Omega}_n^{YX} & \widetilde{\Omega}_n^{YZ} \\ \widetilde{\Omega}_n^{XY} & \widetilde{\Omega}_n^{XX} & \widetilde{\Omega}_n^{XZ} \\ \widetilde{\Omega}_n^{ZY} & \widetilde{\Omega}_n^{ZX} & \widetilde{\Omega}_n^{ZZ} \end{pmatrix}.$$

Note that the expectation for Ω_n is taken over $\mathbf{D}_n$ and $\boldsymbol{R}_n$ while the conditional expectation for $\widetilde{\Omega}_n$ is taken over $\mathbf{D}_n$ conditional on $\boldsymbol{R}_n$.

Our estimands of interest are

$$\begin{pmatrix} \theta_n^{\text{causal}} \\ \gamma_n^{\text{causal}} \end{pmatrix} = \begin{pmatrix} \Omega_n^{XX} & \Omega_n^{XZ} \\ \Omega_n^{ZX} & \Omega_n^{ZZ} \end{pmatrix}^{-1}\begin{pmatrix} \Omega_n^{XY} \\ \Omega_n^{ZY} \end{pmatrix}, \tag{2.2}$$

and

$$\begin{pmatrix} \theta_n^{\text{causal,sample}} \\ \gamma_n^{\text{causal,sample}} \end{pmatrix} = \begin{pmatrix} \widetilde{\Omega}_n^{XX} & \widetilde{\Omega}_n^{XZ} \\ \widetilde{\Omega}_n^{ZX} & \widetilde{\Omega}_n^{ZZ} \end{pmatrix}^{-1}\begin{pmatrix} \widetilde{\Omega}_n^{XY} \\ \widetilde{\Omega}_n^{ZY} \end{pmatrix}. \tag{2.3}$$

These are causal estimands in the sense specified by Abadie et al. (2020).

$(\theta_n^{\text{causal}}, \gamma_n^{\text{causal}})'$ concerns the population-level causal effects of intervention while $(\theta_n^{\text{causal,sample}}, \gamma_n^{\text{causal,sample}})$ concerns the sample-level causal effects when the sampling is governed by $\boldsymbol{R}_n$. $(\theta_n^{\text{causal}}, \gamma_n^{\text{causal}})'$ is a solution for the population moment condition:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\begin{pmatrix} X_{n,i} \\ Z_{n,i} \end{pmatrix}\left(Y_{n,i} - X_{n,i}'\theta_n^{\text{causal}} - Z_{n,i}'\gamma_n^{\text{causal}}\right)\right] = 0, \tag{2.4}$$

and $(\theta_n^{\text{causal,sample}}, \gamma_n^{\text{causal,sample}})$ is a solution for the sample moment condition:

$$\frac{1}{N}\sum_{i=1}^{n}R_{n,i}\mathbb{E}\left[\begin{pmatrix} \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}\left(Y_{n,i} - \widetilde{X}_{n,i}'\theta_n^{\text{causal,sample}} - \widetilde{Z}_{n,i}'\gamma_n^{\text{causal,sample}}\right) \mid \boldsymbol{R}_n\right] = 0. \tag{2.5}$$

We study (i) whether the sample-level estimand can be estimated consistently (internal validity), and, if so, (ii) how closely it approximates the population-level estimand (external validity). We will also discuss whether each element of these estimands admits a causal interpretation, namely, whether each OLS coefficient represents a convex combination of the corresponding heterogeneous treatment effects, which is not discussed in Abadie et al. (2020).

For the sample-level causal estimand, we consider the ordinary least squares estimator:

$$\begin{pmatrix} \widehat{\theta}_n \\ \widehat{\gamma}_n \end{pmatrix} = \begin{pmatrix} \widetilde{Q}_n^{XX} & \widetilde{Q}_n^{XZ} \\ \widetilde{Q}_n^{ZX} & \widetilde{Q}_n^{ZZ} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{Q}_n^{XY} \\ \widetilde{Q}_n^{ZY} \end{pmatrix}. \tag{2.6}$$

Equivalently, the moment condition is

$$\frac{1}{n}\sum_{i=1}^{n} R_{n,i} \begin{pmatrix} \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix} \left(Y_{n,i} - \widetilde{X}'_{n,i}\theta - \widetilde{Z}'_{n,i}\gamma\right) = 0. \tag{2.7}$$

An alternative approach is to use the inverse probability weighting (IPW) estimator (e.g., Leung, 2022a; Gao and Ding, 2023). A usual condition for the IPW estimator to work in a network experimental setting is the individual-level overlapping condition; in our notation, we need to have $\mathbb{P}[\widetilde{T}_{n,i} = t|\boldsymbol{R}_n] \in (\eta, 1-\eta)$ almost surely for all $i \in \mathcal{N}_n$ and $t \in \mathcal{T}_n$ for some $\eta \in (0, 1/2)$. This overlapping condition is difficult to maintain in the network sampling framework. For example, consider a population of two connected units. Suppose the first unit is sampled, while the second is not. The exposure mapping is defined as the number of treated neighbors. In this case, $\mathbb{P}[\widetilde{T}_{n,1} = 1|\boldsymbol{R}_n] = 0$, thereby violating the overlapping condition. Also, it is notable that the IPW estimator typically targets a quantity that differs from our estimands, which are defined through moment conditions in (2.4) and (2.5).

Throughout this section, we have defined the network sampling framework, the exposure mapping, and the potential outcome model. We have also defined the population- and sample-level estimands, which are the solutions to the population and sample moment conditions, respectively. The next section provides our main theoretical results within this framework.

2.3 Main Results

In this section, we present the main results of this paper. We first discuss the population- and sample-level estimands' causal interpretation, then derive the asymptotic properties of the OLS estimator for both. Since we have assumed that the sequence of sampling probabilities ρ_n is bounded away from 0 (Assumption 2.1-(i)), it follows that $N > 0$ a.s. for large enough n (Lemma B.4 in Appendix B.2). Thus, there is no additional concern for the degeneracy of the estimands and the OLS estimator in a large population, relative to the standard design-based setting with $\rho_n = 1$.[4]

2.3.1 Interpretability of the Causal Estimands

We impose the following regularity conditions for the causal estimands to be well-defined. These conditions require boundedness of the outcome, exposure mappings, and covariates, as well as full rank of the exposure mappings and covariates.

Assumption 2.3.

(i) (Uniform Boundedness): The sequence of potential outcomes $Y^*_{n,i}(\cdot)$ is uniformly bounded, i.e., there exists some constant $\overline{Y} > 0$ such that $|Y^*_{n,i}(t)| \leq \overline{Y} < \infty$ for all n, $i \in \mathcal{N}_n$, and $t \in \mathcal{T}$.

(ii) The sequences of exposure mappings $T_{n,i}$ and $\widetilde{T}_{n,i}$ satisfy the following.

[4] This is why we have a stronger statement than Abadie et al. (2020) who allow $\rho_n \to 0$ as $n \to \infty$ and use "with probability approaching 1" instead of "almost surely" in their results. Note that Lemma B.4 allows $\rho_n \to 0$ as long as $\rho_n n \to \infty$.

(a) *(Uniform Boundedness): There exists some constant* $\overline{T}$ *such that* $\|T_{n,i}\|, \|\widetilde{T}_{n,i}\| \leq \overline{T} < \infty$ *almost surely for all* $n, i \in \mathcal{N}_n$.

(b) *(Variation):* $(1/n) \times \sum_{i \in \mathcal{N}_n} \mathrm{Var}(T_{n,i})$ *is invertible and* $(1/N) \times \sum_{i \in \mathcal{N}_n} R_{n,i} \mathrm{Var}(\widetilde{T}_{n,i} \mid \boldsymbol{R}_n)$ *is almost surely invertible for large enough* n.

(iii) *The sequences of covariates* $Z_{n,i}$ *and* $\widetilde{Z}_{n,i}$ *satisfy the following.*

(a) *(Uniform Boundedness): There exists some constant* $\overline{Z}$ *such that* $\|Z_{n,i}\|, \|\widetilde{Z}_{n,i}\| \leq \overline{Z} < \infty$ *almost surely for all* $n, i \in \mathcal{N}_n$.

(b) *(Full Rank):* $(1/n) \times \sum_{i=1}^{n} Z_{n,i} Z_{n,i}'$ *is almost surely full-rank for large enough* n, *and* $(1/N) \times \sum_{i=1}^{n} R_{n,i} \widetilde{Z}_{n,i} \widetilde{Z}_{n,i}'$ *is almost surely invertible for large enough* n.

Assumption 2.3 (iii) implies that the sequences of residualized exposure mappings $X_{n,i}$ and $\widetilde{X}_{n,i}$ satisfy the following.

(a) (Uniform Boundedness): There exists some constant $\overline{X}$ such that

$\|X_{n,i}\|, \|\widetilde{X}_{n,i}\| \leq \overline{X} < \infty$ almost surely for all $n, i \in \mathcal{N}_n$.

(b) (Full Rank): $(1/n) \times \sum_{i \in \mathcal{N}_n} \mathbb{E}[X_{n,i} X_{n,i}']$ is invertible and

$(1/N) \times \sum_{i=1}^{n} R_{n,i} \mathbb{E}[\widetilde{X}_{n,i} \widetilde{X}_{n,i}' | \boldsymbol{R}_n]$ is almost surely invertible for large enough n.

The uniform boundedness of the potential outcomes in Assumption 2.3 (i) is a standard assumption in the literature (e.g., Leung, 2022a; Gao and Ding, 2023). Assumption 2.3 (ii-a) rules out some network statistics in a large, dense network (e.g., a diverging degree). Assumption 2.3 (ii-b) requires that the exposure mappings are not degenerate across the units. For example, in Example 2.2, Assumption 2.3 (ii-b) is violated if the network is empty, $A_{n,i,j} = 0$ for all $i, j \in \mathcal{N}_n$, as $\mathbb{1}\{\sum_{j \neq i} R_{n,j} A_{n,i,j} D^*_{n,j} > 0\} = 0$ for all $i \in \mathcal{N}_n$. Assumption 2.3 (iii-b) does not exclude the constant term in $Z_{n,i}$ and $\widetilde{Z}_{n,i}$. Assumption 2.3 (ii-b) and (iii-b) are not as restrictive as they seem since we have $N > 0$ a.s. for large enough n.

We impose an additional condition on the exposure mapping:

Assumption 2.4. *There exists a sequence of matrices L_n such that*

$$\mathbb{E}[T_{n,i} | \boldsymbol{R}_n] = L_n Z_{n,i} \quad a.s.$$

for large enough n. Similarly, there exists a sequence of matrices $\widetilde{L}_n$ measurable with respect to $\sigma(\boldsymbol{R}_n)$ such that

$$\mathbb{E}[\widetilde{T}_{n,i} | \boldsymbol{R}_n] = \widetilde{L}_n \widetilde{Z}_{n,i} \quad a.s.$$

for large enough n.

This assumption is fairly weak, as it is automatically satisfied if $\mathbb{E}[T_{n,i} | \boldsymbol{R}_n]$ and $\mathbb{E}[\widetilde{T}_{n,i} | \boldsymbol{R}_n]$ are included in $Z_{n,i}$ and $\widetilde{Z}_{n,i}$, respectively. Typically, in a field experiment, the experimenter knows the assignment mechanism, so $\mathbb{E}[\widetilde{T}_{n,i} | \boldsymbol{R}_n]$ can be computed either analytically or numerically and included as covariates. As the following example shows, in some cases, it is sufficient to include some network statistics in the covariates to satisfy this assumption.

Example 2.5. *Consider a variant of the exposure mapping in Miguel and Kremer (2004) that counts the number of treated friends:*

$$T_{n,i} = g(i, \mathbf{D}_n, \mathbf{A}_n) = \sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j}.$$

If there is no censoring, then $\widetilde{g} = g$. *The conditional expectations of exposure mappings are derived as* $\mathbb{E}[T_{n,i}|\boldsymbol{R}_n] = \sum_{j \neq i} A_{n,i,j} R_{n,j} p_{n,j}$, *and* $\mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n] = \sum_{j \neq i} \widetilde{A}_{n,i,j} R_{n,j} p_{n,j} = \sum_{j \neq i} A_{n,i,j} R_{n,j} p_{n,j}$ *for* $R_{n,i} = 1$. *Thus, Assumption 2.4 holds if the weighted degree* $\sum_{j \neq i} A_{n,i,j} R_{n,j} p_{n,j}$ *is included in* $Z_{n,i}$ *and* $\widetilde{Z}_{n,i}$.

We obtain the following transformations of the estimands in terms of the individual causal effects $\theta_{n,i}$ in the linear potential outcome model in Assumption 2.2:

Theorem 2.1. *Under Assumptions 2.1 to 2.4, for large enough n,*

$$\theta_n^{\text{causal}} = \left(\sum_{i=1}^n \mathbb{E}[X_{n,i} X'_{n,i}] \right)^{-1} \sum_{i=1}^n \mathbb{E}[X_{n,i} X'_{n,i}] \theta_{n,i},$$

and

$$\theta_n^{\text{causal,sample}} = \left(\sum_{i=1}^n R_{n,i} \mathbb{E}[\widetilde{X}_{n,i} \widetilde{X}'_{n,i} | \boldsymbol{R}_n] \right)^{-1} \sum_{i=1}^n R_{n,i} \mathbb{E}[\widetilde{X}_{n,i} X'_{n,i} | \boldsymbol{R}_n] \theta_{n,i} \quad a.s.$$

Theorem 2.1 shows that θ_n^{causal} is expressed as a weighted sum of causal effects $\theta_{n,i}$ induced by the exposure mapping. On the other hand, $\theta_n^{\text{causal,sample}}$ is not necessarily a weighted sum of $\theta_{n,i}$ because of the difference in $X_{n,i}$ and $\widetilde{X}_{n,i}$ in the numerator. Moreover, the dimension of $\theta_n^{\text{causal,sample}}$ is $d_{\widetilde{T}}$, which can be different from d_T, the dimension of $\theta_{n,i}$.

In the absence of Assumption 2.4, it is known that the formula in Theorem 2.1 does not hold due to the omitted variable bias (OVB). Assumption 2.4 and Theorem 2.1 suggest a takeaway for practitioners: *under the linear propensity scores, the researcher can select necessary controls easily to avoid the OVB.*

The linear propensity score assumption Assumption 2.4 is a weak assumption in design-based causal inference. This assumption also appears in Abadie et al. (2020) and Borusyak and Hull (2023a). In the latter, the OVB is removed by using the recentered instruments. Theoretically, including the controls and using the recentered instruments are equivalent, but including the controls is more frequently used in practice. While Borusyak and Hull (2023a) focuses on homogeneous treatment effects, this paper allows for heterogeneous treatment effects.

Note that in general, the k-th elements of θ_n^{causal} and $\theta_n^{\text{causal,sample}}$ do not directly correspond to the causal effect of changes in the k-th element of the exposure mapping on the outcomes. For example, if the exposure mapping is two-dimensional, we could have the first element of θ_n^{causal} to be negative while the first element of $\theta_{n,i}$ is positive for all $i \in \mathcal{N}_n$ if the second element of it is significantly negative.

2.3.2 Causal Interpretation

To provide a causal interpretation for each element $\theta_{n,(k)}^{\text{causal}}$ and $\theta_{n,(k)}^{\text{causal,sample}}$, we develop an element-wise version of Theorem 2.1. To this end, we let $T_{n,i,(k)}$ denote the k-th element of $T_{n,i}$. Similarly, we write $\widetilde{T}_{n,i,(k)}, X_{n,i,(k)}, \widetilde{X}_{n,i,(k)}$. For each k, let $U_{n,i,(k)}$ be the residual when projecting $X_{n,i,(k)}$ onto the $X_{n,i,(-k)} = (X_{n,i,(l)})_{l \neq k}$:

$$\begin{aligned} U_{n,i,(k)} &= X_{n,i,(k)} \\ &- \left(\sum_{i=1}^{n} \mathbb{E}[X_{n,i,(k)} X'_{n,i,(-k)}] \right) \\ &\times \left(\sum_{i=1}^{n} \mathbb{E}[X_{n,i,(-k)} X'_{n,i,(-k)}] \right)^{-1} X_{n,i,(-k)}. \end{aligned}$$

Similarly, define

$$\begin{aligned} \widetilde{U}_{n,i,(k)} &= \widetilde{X}_{n,i,(k)} \\ &- \left(\sum_{i=1}^{n} R_{n,i} \mathbb{E}[\widetilde{X}_{n,i,(k)} \widetilde{X}'_{n,i,(-k)} | \boldsymbol{R}_n] \right) \\ &\times \left(\sum_{i=1}^{n} R_{n,i} \mathbb{E}[\widetilde{X}_{n,i,(-k)} \widetilde{X}'_{n,i,(-k)} | \boldsymbol{R}_n] \right)^{-1} \widetilde{X}_{n,i,(-k)}. \end{aligned}$$

Then, we have the following decompositions:

Corollary 2.1. *Under Assumptions 2.1 to 2.4, for large enough n,*

$$\theta_{n,(k)}^{\text{causal}} = \frac{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} X_{n,i,(k)}] \theta_{n,i,(k)}}{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)}^2]} + \frac{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} X'_{n,i,(-k)}] \theta_{n,i,(-k)}}{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)}^2]} \tag{2.8}$$

for each $k = 1, ..., d_T$, and

$$\theta_{n,(k)}^{\text{causal,sample}} = \frac{\sum_{i=1}^{n} R_{n,i} \mathbb{E}[\widetilde{U}_{n,i,(k)} X'_{n,i} | \boldsymbol{R}_n] \theta_{n,i}}{\sum_{i=1}^{n} R_{n,i} \mathbb{E}[\widetilde{U}_{n,i,(k)}^2 | \boldsymbol{R}_n]} \quad a.s. \tag{2.9}$$

for each $k = 1, ..., d_{\widetilde{T}}$. *Under an additional assumption* $d_{\widetilde{T}} = d_T$, *we can simplify it into*

$$\theta_{n,(k)}^{\text{causal,sample}} = \frac{\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n]\theta_{n,i,(k)}}{\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}^2|\boldsymbol{R}_n]} + \frac{\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}X'_{n,i,(-k)}|\boldsymbol{R}_n]\theta_{n,i,(-k)}}{\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}^2|\boldsymbol{R}_n]} \quad a.s. \tag{2.10}$$

for each $k = 1, ..., d_T$.

Corollary 2.1 shows that $\theta_{n,(k)}^{\text{causal}}$ and $\theta_{n,(k)}^{\text{causal,sample}}$ can be influenced by effects from other elements $\theta_{n,i,(l)}$ with $l \neq k$. However, the residualization does not eliminate contamination bias, because the definition of $U_{n,i,(k)}$ and $\widetilde{U}_{n,i,(k)}$ only implies

$$\sum_{i=1}^n \mathbb{E}[U_{n,i,(k)}X'_{n,i,(-k)}] = 0 \text{ and } \sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}X'_{n,i,(-k)}|\boldsymbol{R}_n] = 0,$$

respectively. Moreover, $\mathbb{E}[U_{n,i,(k)}X_{n,i,(k)}]$ and $\mathbb{E}[\widetilde{U}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n]$ are not guaranteed to be non-negative.

Example 2.6. *Suppose the true exposure mapping is the number of treated friends,* $T_{n,i} = \sum_{j\neq i} A_{n,i,j}R_{n,j}D^*_{n,j}$, *and that* $T_{n,i}$ *takes three possible values* 1, 2, *or* 3. *Suppose the researcher misspecifies the exposure mapping as dummy variables:* $\widetilde{T}_{n,i} = (\widetilde{T}_{n,i,(1)}, \widetilde{T}_{n,i,(2)}, \widetilde{T}_{n,i,(3)})$, *where* $\widetilde{T}_{n,i,(k)} = \mathbb{1}\{\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j} = k\}$. *Thus,* $d_{\widetilde{T}} = 3 > d_T = 1$. *For simplicity, consider star sampling, which provides all network links in the first neighborhood. Then,* $\widetilde{T}_{n,i,(k)} = \mathbb{1}\{T_{n,i} = k\}$. *Equation (2.9) in Corollary 2.1 implies that the coefficient for* $\widetilde{T}_{n,i,(k)}$ *has no* *contamination term because* $X_{n,i}$ *is a scalar in this example. However, each element of* $\theta_n^{\text{causal,sample}}$ *captures a different weighted sum of* $\theta_{n,i}$*; thus, the interpretation is unclear.*

Remark 2.2. ***(i)*** *Assuming* $d_{\widetilde{T}} = d_T$ *requires the researcher to correctly specify the dimension of the exposure mapping* ($d_T = d_{\widetilde{T}}$)*. However, this assumption allows the researcher to misspecify the shape of* $\widetilde{g} \neq g$ *or mismeasure the network.*

(ii) *Our result for* θ_n^{causal} *is a design-based analogue of Proposition 1 in Goldsmith-Pinkham et al. (2022). The main differences are that their analysis is model-based and focuses on mutually exclusive treatment indicators (e.g., K-arms).*[5,6] *In contrast, we allow more flexible treatments, including network spillovers. Our decomposition for* $\theta_n^{\text{causal,sample}}$ *additionally accommodates both misspecification of the exposure mapping and mismeasurement of the network.*

(iii) *If the distribution of* $T_{n,i}$ *does not depend on* i*, a result in Corollary 2.1 can be strengthened to*

$$\theta_{n,(k)}^{\text{causal}} = \frac{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} X_{n,i,(k)}] \theta_{n,i,(k)}}{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)}^2]}$$

for any k*. That is, we do not have a contamination bias. However, the weight can be negative. Moreover, the homogeneous requirement of the treatment*

[5] Mutually exclusive treatments guarantee that each treatment's own effect receives a non-negative weight.

[6] Goldsmith-Pinkham et al. (2022) propose three approaches to eliminate contamination bias, but all require modeling the conditional expectation of heterogeneous treatment effects based on observed covariates. In a design-based setting with deterministic treatment effects $\theta_{n,i}$, such modeling is not appropriate. Even if the modeling assumption is justified, their methods may be unreliable for network experiments due to weak overlap in propensity scores, which is often violated for common exposure mappings.

variable $T_{n,i}$ is usually violated in design-based network experiments since the exposure mapping depends on the network information for each i and the population network $\mathbf{A}_n$ is treated as non-random.

(iv) *The weight for θ_n^{causal} is clearly non-negative if the dimension of the treatment variable $T_{n,i}$ is one ($d_T = 1$) because no contamination occurs when $d_T = 1$. This result is consistent with Borusyak and Hull (2024), but our result in Corollary 2.1 is more general ($d_T > 1$).*

2.3.3 When Can We Avoid the Contamination Bias?

The following statement provides sufficient conditions to avoid contamination bias. Define the conditional covariance for random variables W_1 and W_2 given $\boldsymbol{R}_n$ as $\text{Cov}(W_1, W_2|\boldsymbol{R}_n) = \mathbb{E}[(W_1 - \mathbb{E}[W_1|\boldsymbol{R}_n])(W_2 - \mathbb{E}[W_2|\boldsymbol{R}_n])|\boldsymbol{R}_n]$.

Corollary 2.2. *Assume that Assumptions 2.1 to 2.4 and $d_{\widetilde{T}} = d_T$ hold. Suppose that $\mathbb{E}[\text{Cov}(T_{n,i,(k)}, T_{n,i,(l)}|\boldsymbol{R}_n)] = 0$ for all $i \in \mathcal{N}_n$ and for any $l \neq k$. Then, for large enough n, there is no contamination bias for $\theta_{n,(k)}^{\text{causal}}$, i.e.,*

$$\theta_{n,(k)}^{\text{causal}} = \frac{\sum_{i=1}^n \mathbb{E}[X_{n,i,(k)}^2]\theta_{n,i,(k)}}{\sum_{i=1}^n \mathbb{E}[X_{n,i,(k)}^2]}$$

for each $k = 1, ..., d_T$. Suppose that $\text{Cov}(\widetilde{T}_{n,i,(k)}, T_{n,i,(l)}|\boldsymbol{R}_n) = 0$ for all $i \in \mathcal{N}_n$ with $R_{n,i} = 1$ and for any $l \neq k$. Then, for large enough n, there is no contamination bias for $\theta_{n,(k)}^{\text{causal,sample}}$, i.e.,

$$\theta_{n,(k)}^{\text{causal,sample}} = \frac{\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{X}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n]\theta_{n,i,(k)}}{\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{X}_{n,i,(k)}^2|\boldsymbol{R}_n]} \quad a.s.$$

for each $k = 1, ..., d_T$.

The weights of $\theta^{\text{causal}}_{n,(k)}$ for $\theta_{n,i,(k)}$ are always non-negative. If we further assume that $\text{Cov}(\widetilde{T}_{n,i,(k)}, T_{n,i,(k)}|\boldsymbol{R}_n) \geq 0$ a.s. for all $i \in \mathcal{N}_n$ with $R_{n,i} = 1$ and for all $k = 1, ..., d_T$, then the weights of $\theta^{\text{causal,sample}}_{n,(k)}$ for $\theta_{n,i,(k)}$ are non-negative, i.e.,

$$\frac{R_{n,i}\mathbb{E}[\widetilde{X}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n]}{\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{X}^2_{n,i,(k)}|\boldsymbol{R}_n]} \geq 0 \quad a.s.$$

for all $i \in \mathcal{N}_n$ and each $k = 1, ..., d_T$.

The zero conditional covariance assumption is satisfied if elements of $T_{n,i}$ and $\widetilde{T}_{n,i}$ are mutually independent. The positive conditional covariance assumption is satisfied under the censored network (see Example 2.7 below).

Remark 2.3. *Under homogeneous treatment effects $\theta_{n,i} = \theta_n$, we have $\theta^{\text{causal}}_n = \theta_n$, but*

$$\begin{aligned}&\theta^{\text{causal,sample}}_n \\ &= \theta_n - \left(\sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{X}_{n,i}\widetilde{X}'_{n,i}|\boldsymbol{R}_n]\right)^{-1} \sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{X}_{n,i}(X_{n,i} - \widetilde{X}_{n,i})'|\boldsymbol{R}_n]\theta_n.\end{aligned}$$

Thus, θ^{causal}_n does not have contamination bias for homogeneous treatment effects, but $\theta^{\text{causal,sample}}_n$ does. Under homogeneous treatment effects and $X_{n,i} = \widetilde{X}_{n,i}$, we have $\theta^{\text{causal}}_n = \theta^{\text{causal,sample}}_n = \theta_n$.

Example 2.7. *Consider the exposure mapping in Example 2.4. The misspecified exposure mapping is $\widetilde{T}_{n,i} = \widetilde{g}(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n) = \mathbb{1}\left\{\sum_{j \neq i} C_{n,i,j}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j} > 0\right\}$. Assume that $D^*_{n,i} \sim$ Bernoulli(p_n) for $i = 1, \ldots, n$ independently. By adapting*

Corollary 2.2, $\theta_n^{\text{causal,sample}}$ is a convex combination of $\theta_{n,i}$. Indeed, we can calculate

$$\theta_n^{\text{causal,sample}} = \frac{\sum_{i=1}^{n} R_{n,i}\left(1-(1-p_n)^{\sum_{j\neq i} C_{n,i,j}\widetilde{A}_{n,i,j}R_{n,j}}\right)\theta_{n,i,(1)}}{\sum_{i=1}^{n} R_{n,i}\left(1-(1-p_n)^{\sum_{j\neq i} C_{n,i,j}\widetilde{A}_{n,i,j}R_{n,j}}\right)},$$

and the weights are non-negative. In general, if both mappings $\widetilde{T}_{n,i,(k)}$ and $T_{n,i,(k)}$ are weakly increasing (or both weakly decreasing) in $\{D^*_{n,i}\}_{i\in\mathcal{N}_n}$, then the weights are non-negative. Thus, censoring does not cause negative weight problems when g is weakly monotone on $\{D^*_{n,i}\}_{i\in\mathcal{N}_n}$ for the first neighborhood exposure mapping.

2.3.4 More Examples

Example 2.8. *Consider a general form of exposure mapping. For some function $q:\mathbb{R}^2\to\mathbb{R}$, let $T_{n,i}=(R_{n,i}D^*_{n,i}, q(\sum_{j\neq i}A_{n,i,j}R_{n,j}D^*_{n,j}, \sum_{j\neq i}A_{n,i,j}R_{n,j}))$ and $\widetilde{T}_{n,i}=(R_{n,i}D^*_{n,i}, q(\sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j}, \sum_{j\neq i}\widetilde{A}_{n,i,j}R_{n,j}))$. For example, the share of treated friends is covered by the following q:*

$$q\left(\sum_{j\neq i}A_{n,i,j}R_{n,j}D^*_{n,j}, \sum_{j\neq i}A_{n,i,j}R_{n,j}\right)=\frac{\sum_{j\neq i}A_{n,i,j}R_{n,j}D^*_{n,j}}{\sum_{j\neq i}A_{n,i,j}R_{n,j}}.$$

It also covers the indicator function as in Example 2.2. Since $D^*_{n,i}\perp\!\!\!\perp D^*_{n,j}$, this satisfies the no-correlation conditions. If q is non-decreasing with respect to the first argument, then $\widetilde{T}_{n,i}$ and $T_{n,i}$ are positively correlated, giving $\theta_n^{\text{causal,sample}}$ a clear causal interpretation. This type of exposure mapping is used in Cai et al. (2015a) and Carter et al. (2021a). As we illustrated above in the special case, the censoring $\widetilde{T}_{n,i}=(R_{n,i}D^*_{n,i}, q(\sum_{j\neq i}C_{n,i,j}\widetilde{A}_{n,i,j}R_{n,j}D^*_{n,j}, \sum_{j\neq i}C_{n,i,j}\widetilde{A}_{n,i,j}R_{n,j}))$

does not cause negative weight problems since the exposure mapping g is weakly monotone on $\{D^*_{n,i}\}_{i\in\mathcal{N}_n}$.

Example 2.9. *Let $T_{n,i} = (R_{n,i}D^*_{n,i}G_{n,i}, R_{n,i}D^*_{n,i}(1-G_{n,i}), (1-R_{n,i}D^*_{n,i})G_{n,i})$, where $G_{n,i} = \mathbb{1}\{\sum_{j\neq i} A_{n,i,j}R_{n,j}D^*_{n,j} > 0\}$. The exposure mapping categorizes each unit i into one of three mutually exclusive exposure types, based on their own treatment status and the presence of treated friends.*[7] *The elements are mutually exclusive but dependent, so the no-correlation conditions are violated, and we have a contamination bias. This exposure mapping is used in Aronow and Samii (2017). For the exposure mapping with dependence among its elements, we recommend using the inverse propensity score weighting (IPW) estimators to avoid contamination bias.*

Remark 2.4. *(Comparison with IPW estimators) The causal estimand for the IPW estimators is the average treatment effect (ATE), $(1/n)\sum_{i=1}^n Y^*_{n,i}(t)$ for each t. In other words, the IPW estimator and the regression estimator are for different causal estimands. While the IPW estimator works well for cases like Example 2.9, it is not suitable for cases like Example 2.10 because the overlapping condition of the propensity score is easily violated. For example, suppose that $T_{n,i}$ is the treated friends share $(\sum_{j\neq i} A_{n,i,j}R_{n,j}D^*_{n,j})/(\sum_{j\neq i} A_{n,i,j})$, and there are two units having three and two friends in the population network, respectively. The former can take $T_{n,i} = 1/3$ with positive probability, but the latter never takes the value. Thus, the overlapping condition fails to hold. Moreover, the overlapping condition can be violated in the sampled network even if it is satisfied*

[7] The slope of the OLS estimator captures the effect associated with the group of units that are untreated and have no treated friends.

in the population network, since the sampled network is a sub-network of the population one.

The choice between the IPW estimator and the regression should be decided by the exposure mapping formula that the researcher wants to use. We recommend using the IPW estimators to avoid contamination bias when the overlapping condition is satisfied. On the other hand, if there is any doubt about the overlapping condition or the exposure mapping takes (nearly) continuous values, we suggest using the regression model since it does not require the overlapping condition. We leave a more detailed comparison between the IPW estimator and the OLS estimator for future research.

Example 2.10. *Consider an exposure mapping*

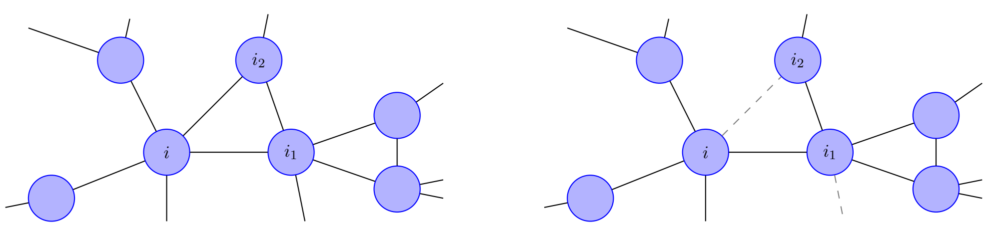
$$T_{n,i} = \left(R_{n,i}D^*_{n,i}, \frac{\sum_{j\neq i} A_{n,i,j}R_{n,j}D^*_{n,j}}{\sum_{j\neq i} A_{n,i,j}}, \frac{\sum_{j\neq i}\sum_{k\neq i,j} A_{n,i,j}A_{n,j,k}R_{n,k}D^*_{n,k}}{\sum_{j\neq i}\sum_{k\neq i,j} A_{n,i,j}A_{n,j,k}} \right),$$

where the first element indicates whether unit i is directly treated or not, the second element captures the treated friends share among i's first neighbors, and the third element captures the treated friends share among i's second neighbors. There are overlaps in $\mathbf{D}_n$ in the second and third elements if there are triangles in the network, so no-correlation conditions are generally violated. Figure 2.2a shows an example of a network with triangles. The second element of $T_{n,i}$ is the average of the neighbors' treatment status including D_{n,i_1} and D_{n,i_2}. The third element is the average of the first neighbors' treatment status, including D_{n,i_1} and D_{n,i_2}, again. Thus, the second and third elements are correlated. This setting is employed in Cai et al. (2015a). An easy way to avoid contamination *bias is to modify the exposure mapping g to eliminate the double counting. For example, we can use*

$$T_{n,i} = \Bigg(R_{n,i} D^*_{n,i}, \frac{\sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j}}{\sum_{j \neq i} A_{n,i,j}}, \frac{\sum_{j \neq i} \sum_{k \neq i,j} A_{n,i,j} A_{n,j,k} (1 - A_{n,i,k}) R_{n,k} D^*_{n,k}}{\sum_{j \neq i} \sum_{k \neq i,j} A_{n,i,j} A_{n,j,k} (1 - A_{n,i,k})} \Bigg), \tag{2.11}$$

instead. Although we miss some of the second-order links, we still manage to avoid the double counting and hence contamination bias.

Figure 2.2: Networks with triangle links



(a) Without censored links (b) With censored (dashed) links

Example 2.11. *Consider the setup in Example 2.10 but with censoring caused by naming up to four friends. As illustrated in Figure 2.2b, suppose that the sampled network link between i_1 and i_2 is not observed due to the censoring. Then, i_2 is misclassified as a second neighborhood friend in the observed network while i_2 is a first neighborhood friend in the population network. Thus, if we consider the true exposure mapping $T_{n,i}$ as in (2.11), and a misspecified exposure* *mapping for the sampled network*

$$\widetilde{T}_{n,i} = \Bigg(R_{n,i} D^*_{n,i}, \frac{\sum_{j \neq i} C_{n,i,j} \widetilde{A}_{n,i,j} R_{n,j} D^*_{n,j}}{\sum_{j \neq i} C_{n,i,j} \widetilde{A}_{n,i,j}}, \\ \frac{\sum_{j \neq i} \sum_{k \neq i,j} C_{n,i,j} \widetilde{A}_{n,i,j} C_{n,j,k} \widetilde{A}_{n,j,k} (1 - C_{n,i,k} \widetilde{A}_{n,i,k}) R_{n,k} D^*_{n,k}}{\sum_{j \neq i} \sum_{k \neq i,j} C_{n,i,j} \widetilde{A}_{n,i,j} C_{n,j,k} \widetilde{A}_{n,j,k} (1 - C_{n,i,k} \widetilde{A}_{n,i,k})} \Bigg),$$

then, there is a correlation between $T_{n,i,(2)}$ and $\widetilde{T}_{n,i,(3)}$. An easy way to avoid contamination bias is to modify the exposure mapping $\widetilde{g}$ so that $\widetilde{T}_{n,i,(3)}$ equals zero for individuals subject to censoring. For example, if the censoring happens by asking up to four friends, we can eliminate the individuals with four observed links from consideration:

$$\widetilde{T}_{n,i,(3)} \times \mathbb{1} \left\{ \sum_{j \neq i} C_{n,i,j} \widetilde{A}_{n,i,j} < 4 \right\}.$$

Note that the censoring for i does not matter for the first neighborhood element $\widetilde{T}_{n,i,(2)}$ by the same logic as Example 2.8. Moreover, the censoring for i_1 does not matter for the second neighborhood element $\widetilde{T}_{n,i,(3)}$ of i because it does not introduce any misclassification.

2.3.5 Asymptotic Theory

We mostly follow the notation of Kojevnikov et al. (2021a). Let $\mathcal{N}_n = \{1, ..., n\}$ be the set of population units and $d_n(i,j)$ be the shortest distance between $i, j \in \mathcal{N}_n$ on $\mathbf{A}_n$ (set $d_n(i,i) = 0$; set $d_n(i,j) = \infty$ if there are no paths between i and j). Define $\mathcal{L}_v = \{\mathcal{L}_{v,a} : a \in \mathbb{N}\}$, where $\mathcal{L}_{v,a} = \{f : \mathbb{R}^{v \times a} \to \mathbb{R} : \|f\|_\infty < \infty, \text{Lip}(f) < \infty\}$, $\|\cdot\|_\infty$ is the sup-norm, and $\text{Lip}(f)$ is the Lipschitz constant of f. Let $\mathcal{P}_n(a,b;s) = \{(A,B) : A,B \subset \mathcal{N}_n, |A| = a, |B| = b, d_n(A,B) \geq s\}$, where $d_n(A,B) = \min_{i\in A}\min_{j\in B} d_n(i,j)$. For each $A \subset \mathcal{N}_n$ and triangular array $(U_{n,i})$, let us write $U_{n,A} = (U_{n,i})_{i\in A}$.

Definition 2.1. *A triangular array $\{U_{n,i}\}, n \geq 1, U_{n,i} \in \mathbb{R}^v$, is called conditionally ψ-dependent given $\boldsymbol{R}_n$, if for each $n \in \mathbb{N}$, there exists a $\sigma(\boldsymbol{R}_n)$-measurable sequence $\xi_n = \{\xi_{n,s}\}_{s\geq 0}, \xi_{n,0} = 1$, and a collection of nonrandom functions $(\psi_{a,b})_{a,b\in\mathbb{N}}, \psi_{a,b} : \mathcal{L}_{v,a} \times \mathcal{L}_{v,b} \to [0,\infty)$ such that for all $(A,B) \in \mathcal{P}_n(a,b;s)$ with $s > 0$ and all $f \in \mathcal{L}_{v,a}$ and $g \in \mathcal{L}_{v,b}$,*

$$|\operatorname{Cov}(f(U_{n,A}), g(U_{n,B}))| \leq \psi_{a,b}(f,g)\xi_{n,s} \quad a.s.$$

Define

$$\mathcal{N}_n(i;s) = \{j \in \mathcal{N}_n : d_n(i,j) \leq s\},$$

which is the set of i's neighborhood within s-distance. First, we assume that the network dependence of the exposure mappings is local.

Assumption 2.5. *There exists some $K \in \mathbb{N}$ such that for any $i \in \mathcal{N}_n, n \in \mathbb{N}$ and $\mathbf{d}_n, \mathbf{d}'_n \in \{0,1\}^n$ such that $\mathbf{d}_{n,\mathcal{N}_n(i,K)} = \mathbf{d}'_{n,\mathcal{N}_n(i,K)}$,*

$$g(i,\mathbf{d}_n,\mathbf{A}_n) = g(i,\mathbf{d}'_n,\mathbf{A}_n), \quad \text{and} \quad \widetilde{g}(i,\mathbf{d}_n,\widetilde{\mathbf{A}}_n) = \widetilde{g}(i,\mathbf{d}'_n,\widetilde{\mathbf{A}}_n) \quad a.s.$$

Let $\widetilde{d}_n(i,j)$ be the shortest distance between $i,j \in \mathcal{N}_n$ on $\widetilde{\mathbf{A}}_n$. Assumptions 2.1 and 2.5 imply that $T_{n,i} \perp\!\!\!\perp T_{n,j}$ if $d_n(i,j) > 2K$. They also imply that $\widetilde{T}_{n,i} \perp\!\!\!\perp \widetilde{T}_{n,j}$ if $d_n(i,j) > 2K$ because $\widetilde{d}_n(i,j) \geq d_n(i,j)$ almost surely and because i and j do not share $R_{n,k}$ and $D^*_{n,k}$ for any $k \neq i,j$ in their K-neighborhoods.

Under the correctly specified exposure mapping, $g = \widetilde{g}$, the condition $g(i, \mathbf{d}_n, \mathbf{A}_n) = g(i, \mathbf{d}'_n, \mathbf{A}_n)$ automatically implies $\widetilde{g}(i, \mathbf{d}_n, \widetilde{\mathbf{A}}_n) = \widetilde{g}(i, \mathbf{d}'_n, \widetilde{\mathbf{A}}_n)$ a.s. because the distance on a sampled network is always weakly longer than that on the population network: $\widetilde{d}_n(i,j) \geq d_n(i,j)$. For the same reason, the distance on a censored network is always weakly longer than that on sampled or population networks.

Define $\mathcal{N}_n^{\partial}(i;s) = \{j \in \mathcal{N}_n : d_n(i,j) = s\}$, which is the set of i's neighborhood with exact s-distance, and its p-th sample moment $\delta_n^{\partial}(s;p) = n^{-1}\sum_{i \in \mathcal{N}_n} |\mathcal{N}_n^{\partial}(i;s)|^p$. The next assumption requires that the sum of these p-th sample moments within $2K$-distance is bounded.

Assumption 2.6. *The sequence of networks* $(\mathbf{A}_n)$ *satisfies*

$$\sum_{1 \leq s \leq 2K} \delta_n^{\partial}(s;1) = O(1).$$

By a simple calculation and $\rho > 0$, we can show that Assumption 2.6 is equivalent to $(n\rho_n)^{-1}\sum_{i=1}^{n}\sum_{j \in \mathcal{N}_n(i;2K)} 1 = O(1)$. Also note that Assumption 2.6 is weaker than the bounded network degree since this assumption only requires the boundedness on average.

Then, we show that our estimator is consistent for the sample-level causal estimand:

Theorem 2.2. *Under Assumptions 2.1 to 2.6,*

$$\widehat{\theta}_n - \theta_n^{\text{causal,sample}} \xrightarrow{p^R} 0 \quad \textit{and} \quad \widehat{\theta}_n - \theta_n^{\text{causal,sample}} \xrightarrow{p} 0,$$

where $\xrightarrow{p^R}$ *denotes convergence in probability conditional on* $\boldsymbol{R}_n$*, that is, for any* $\varepsilon > 0$*,*

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta_n^{\text{causal,sample}}\| \leq \varepsilon \mid \boldsymbol{R}_n\right) \xrightarrow{a.s.} 1$$

as $n \to \infty$*.*

Theorem 2.2 establishes the internal validity of our network experiment. However, in general, $\widehat{\theta}_n - \theta_n^{\text{causal}} \not\xrightarrow{p} 0$ because $\theta_n^{\text{causal}} - \theta_n^{\text{causal,sample}} \not\xrightarrow{p} 0$ due to misspecification of the exposure mapping. Moreover, as shown in Corollary 2.1, $\theta_n^{\text{causal,sample}}$ does not have a clear causal interpretation. Consequently, Theorem 2.2 does not guarantee the external validity of our network experiment.

Ideally, our network experiment would satisfy $\widehat{\theta}_n - \theta_n^{\text{causal}} \xrightarrow{p} 0$ so that each element of $\widehat{\theta}_n$ can be interpreted as a causal spillover effect. We show that this consistency is achieved when there is no misspecification and no mismeasurement ($\widetilde{T}_{n,i} = T_{n,i}$ for each $i \in \mathcal{N}_n$) and the observed covariates coincide with those in the population ($\widetilde{Z}_{n,i} = Z_{n,i}$ for each $i \in \mathcal{N}_n$). We are essentially assuming that each $\widetilde{T}_{n,i}$ is computed by $g(i, \mathbf{D}_n, \mathbf{A}_n) = T_{n,i}$ where we replace $\widetilde{g}$ with g and $\widetilde{\mathbf{A}}_n$ with $\mathbf{A}_n$. Under the linear propensity scores, we can show that $X_{n,i} = \widetilde{X}_{n,i}$ a.s. (Lemma B.7).

Assumption 2.7.

1. *We have the following equalities almost surely for* $R_{n,i} = 1$*:* $\widetilde{T}_{n,i} = T_{n,i}$

and $\widetilde{Z}_{n,i} = Z_{n,i}$ *for all* $i \in \mathcal{N}_n$ *and* $n \in \mathbb{N}$.

2. *Each element of* $T_{n,i}$ *and* $Z_{n,i}$ *either does not depend on* $R_{n,i}$*, or depends on it only through a multiplicative form.*

3. *At most one element of* $T_{n,i}$ *depends on* i*'s own treatment* $R_{n,i}D^*_{n,i}$ *and the element does not depend on* $R_{n,j}$ *and* $D_{n,j}$ *for any* $j \neq i$.

Assumption 2.7 (i) holds when there is no misspecification and no mismeasurement for sampled units, i.e., $g = \widetilde{g}$ and $\widetilde{\mathbf{A}}_n = \mathbf{A}_n$ locally. For example, under star sampling with an exposure mapping restricted to the first neighborhood, all relevant links are correctly observed for sampled units. However, Assumption 2.7 (i) may not hold for exposure mappings with higher-order dependence, since more global measurement of the network is then required up to the relevant order. Nonetheless, in such cases, researchers can apply our results to $\theta_n^{\text{causal,sample}}$ instead of θ_n^{causal} and interpret it as a convex combination of heterogeneous treatment effects. Assumption 2.7 (ii) means some components of the covariate vector are independent of the sampling indicator, while others incorporate $R_{n,i}$ in a multiplicative way—for example, $Z_{n,i,(k)} = R_{n,i}p_{n,i}$. We can always pick covariates $Z_{n,i}$ having Assumption 2.7 (ii) since $R_{n,i}$ enters only multiplicatively for $T_{n,i}$ by Assumption 2.1 if we include the direct effect without any transformation. Thus, we can choose covariates $Z_{n,i}$ satisfying Assumption 2.7 and Assumption 2.4 simultaneously. Assumption 2.7 (iii) is satisfied if we do not include the cross term of the direct effect $R_{n,i}D^*_{n,i}$ and a spillover effect. Excluding the cross term is also used to guarantee no contamination (Corollary 2.2).[8]

Under Assumption 2.7, we can show the consistency of the OLS estimator $\widehat{\theta}_n$ for the population-level causal estimand θ_n^{causal}:

Theorem 2.3. *Under Assumptions 2.1 to 2.7,*

$$\widehat{\theta}_n - \theta_n^{\text{causal}} \xrightarrow{p} 0.$$

It is worth noting that Theorem 2.3 does not hold if $\widetilde{Z}_{n,i} \neq Z_{n,i}$, since we cannot ensure $\widetilde{X}_{n,i} \sim X_{n,i}$ asymptotically. Instead, under no misspecification, Corollary 2.1 implies

$$\theta_{n,(k)}^{\text{causal,sample}} = \frac{\sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}^{2}|\boldsymbol{R}_n]\theta_{n,i}}{\sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}^{2}|\boldsymbol{R}_n]}$$

for each $k = 1, ..., d_{\widetilde{T}}$. Thus, although the consistency for θ_n^{causal} may fail in this setting, the absence of misspecification alone recovers the causal interpretability of $\theta_n^{\text{causal,sample}}$, and by extension, that of $\widehat{\theta}_n$.

Next, we consider the asymptotic distribution of $\widehat{\theta}_n$. Now, we introduce additional dependence measures of the network. Define

$$\Delta_n(s, m; k) = \frac{1}{n} \sum_{i \in \mathcal{N}_n} \max_{j \in \mathcal{N}_n^{\partial}(i;s)} |\mathcal{N}_n(i; m) \setminus \mathcal{N}_n(j; s-1)|^k,$$

and

$$c_n(s, m; k) = \inf_{\alpha > 1} [\Delta_n(s, m; k\alpha)]^{1/\alpha} \left[\delta_n^{\partial}\left(s; \frac{\alpha}{\alpha - 1}\right)\right]^{1-1/\alpha}.$$

[8]We can allow the violation of Assumption 2.7 (iii) if we modify $\widehat{\theta}_n$ in the same manner as $\widetilde{\gamma}_n$ in Appendix B.1.

$c_n(s, m; k)$ measures the density of the network and is used as a sufficient condition for the CLT.

Define

$$\widetilde{\varepsilon}_{n,i} = Y_{n,i} - \widetilde{X}'_{n,i}\theta_n^{\text{causal,sample}} - \widetilde{Z}'_{n,i}\gamma_n^{\text{causal,sample}},$$
$$\varepsilon_{n,i} = Y_{n,i} - X'_{n,i}\theta_n^{\text{causal}} - Z'_{n,i}\gamma_n^{\text{causal}},$$

and

$$\widetilde{\Sigma}_n = \operatorname{Var}\left(\sum_{i=1}^{n} R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} \mid \boldsymbol{R}_n\right), \qquad \Sigma_n = \operatorname{Var}\left(\sum_{i=1}^{n} R_{n,i}X_{n,i}\varepsilon_{n,i}\right).$$

We impose the following assumption, which requires a weak dependence structure in the network and rules out overly dense networks.

Assumption 2.8. *There exists a positive sequence $m_n \to \infty$ such that for $p = 1, 2$,*

$$n\widetilde{\Sigma}_n^{-(1+p/2)}\sum_{s=0}^{2K} c_n(s, m_n; p) \xrightarrow{a.s.} 0, \qquad n\Sigma_n^{-(1+p/2)}\sum_{s=0}^{2K} c_n(s, m_n; p) \to 0.$$

Then, we show that $\widehat{\theta}_n$ is asymptotically normal relative to $\theta_n^{\text{causal,sample}}$:

Theorem 2.4. *Under Assumptions 2.1 to 2.6 and 2.8,*

$$\widetilde{\Sigma}_n^{-1/2}\widetilde{Q}_n^{XX}(\widehat{\theta}_n - \theta_n^{\text{causal,sample}}) \xrightarrow{d^R} \mathrm{N}(0, I_{d_{\widetilde{T}}}), \quad \textit{and}$$
$$\widetilde{\Sigma}_n^{-1/2}\widetilde{Q}_n^{XX}(\widehat{\theta}_n - \theta_n^{\text{causal,sample}}) \xrightarrow{d} \mathrm{N}(0, I_{d_{\widetilde{T}}}),$$

where $\overset{d^R}{\to}$ *denotes convergence in distribution conditional on* $\boldsymbol{R}_n$*, that is,*

$$\left|\mathbb{P}\left(\widetilde{\Sigma}_n^{-1/2}\widetilde{Q}_n^{XX}(\widehat{\theta}_n - \theta_n^{\text{causal,sample}}) \leq t \mid \boldsymbol{R}_n\right) - F(t)\right| \xrightarrow{a.s.} 0$$

as $n \to \infty$ *for any* $t \in \mathbb{R}^{d_{\widetilde{T}}}$ *letting* $F(t)$ *be the distribution function of* $\mathrm{N}(0, I_{d_{\widetilde{T}}})$.

We also show that the absence of misspecification and access to the variables in the population yield asymptotic normality of $\widehat{\theta}_n$ relative to θ_n^{causal}:

Theorem 2.5. *Under Assumptions 2.1 to 2.8, we have*

$$\Sigma_n^{-1/2}\widetilde{Q}_n^{XX}(\widehat{\theta}_n - \theta_n^{\text{causal}}) \xrightarrow{d} \mathrm{N}(0, I_{d_T}).$$

Remark 2.5. *When we have a homogeneous effect* $\theta_{n,i} = \theta_n$*, we have* $\theta_n^{\text{causal,sample}} = \theta_n^{\text{causal}}$ *a.s. for large enough* n *under* $X_{n,i} = \widetilde{X}_{n,i}$*. Hence, we can use the same asymptotic distribution among them.*

2.4 Variance Estimation

In this section, we provide a conservative network heteroskedasticity- and autocorrelation-consistent (HAC) variance estimator for $\widehat{\theta}_n$. Note that even when treatments and samples are randomly assigned and drawn, dependence can persist within a $2K$-neighborhood because exposure mappings $T_{n,i}$ may share elements of $\mathbf{D}_n$ and $\boldsymbol{R}_n$. As a result, the variance estimator must account for this local dependence structure. However, for any pair i, j with $d_n(i,j) > 2K$, the exposure mappings $T_{n,i}$ and $T_{n,j}$ are independent. When the exposure mapping is correctly specified ($\widetilde{g} = g$), the researcher can directly choose a finite K based on the functional form. If there is potential misspecification in $\widetilde{g}$, K should be selected conservatively, reflecting the maximum range over which the exposure mapping may induce dependence.

Define

$$\widetilde{\mathcal{N}}_n(i;s) = \{j \in \mathcal{N}_n : \widetilde{d}_n(i,j) \leq s\},$$

which is the set of i's neighborhood within s-distance on a sampled network $\widetilde{\mathbf{A}}_n$. Note that $\widetilde{\mathcal{N}}_n(i;s)$ is a random set because $\widetilde{d}_n(i,j)$ is a random variable depending on $\boldsymbol{R}_n$. On the other hand, $d_n(i,j)$ and $\mathcal{N}_n(i;s)$ are non-random. Recall that we also have $\widetilde{d}_n(i,j) \geq d_n(i,j)$ a.s., thus, $\widetilde{\mathcal{N}}_n(i;s) \subseteq \mathcal{N}_n(i;s)$ a.s.

Let

$$\widehat{\varepsilon}_{n,i} = Y_{n,i} - \widetilde{X}'_{n,i}\widehat{\theta}_n - \widetilde{Z}'_{n,i}\widetilde{\gamma}_n,$$

$\Psi_{n,i} = X_{n,i}\varepsilon_{n,i}$, $\widetilde{\Psi}_{n,i} = \widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i}$, and $\widehat{\Psi}_{n,i} = \widetilde{X}_{n,i}\widehat{\varepsilon}_{n,i}$, where we define $\widetilde{\gamma}_n$ later in Theorems 2.6 and 2.7. By orthogonality conditions, $\sum_{i=1}^n \mathbb{E}\left[\Psi_{n,i}\right] = 0$, $\sum_{i=1}^n R_{n,i}\mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right] = 0$, and $\sum_{i=1}^n R_{n,i}\widehat{\Psi}_{n,i} = 0$.

Then, the variances of interest can be written as

$$\begin{aligned}
\frac{1}{n\rho_n}\widetilde{\Sigma}_n &= \operatorname{Var}\left(\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^n R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} \mid \boldsymbol{R}_n\right) \\
&= \frac{1}{n\rho_n}\sum_{i=1}^n \sum_{j \in \widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\mathbb{E}\Bigg[\left(\widetilde{\Psi}_{n,i} - \mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\right) \\
&\times \left(\widetilde{\Psi}_{n,j} - \mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]\right)' \mid \boldsymbol{R}_n\Bigg],
\end{aligned}$$

and

$$\frac{1}{n\rho_n}\Sigma_n = \mathrm{Var}\left(\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^{n} R_{n,i}X_{n,i}\varepsilon_{n,i}\right)$$
$$= \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)} \mathbb{E}\Big[\left(R_{n,i}\Psi_{n,i} - \rho_n\mathbb{E}\left[\Psi_{n,i}\right]\right)$$
$$\times\left(R_{n,j}\Psi_{n,j} - \rho_n\mathbb{E}\left[\Psi_{n,j}\right]\right)'\Big].$$

We consider the following feasible estimator:

$$\frac{1}{N}\widehat{\Sigma}_n = \frac{1}{N}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\widehat{\Psi}_{n,i}\widehat{\Psi}'_{n,j}.$$

To show the consistency of the variance estimator, we assume an additional sparsity condition. The assumption requires a few more notations. Let $\delta_n(s;p)$ be the p-th sample moment of the set of i's neighborhood within s-distance: $\delta_n(s;p) = n^{-1}\sum_{i=1}^{n}|\mathcal{N}_n(i;s)|^p$. We also define $\mathcal{J}_n(s,m)$ as the set of quadruples (i,j,i',j') such that i' and j' are m-neighbors of i and j, respectively, and the distance between i and j is exactly s:

$$\mathcal{J}_n(s,m) = \left\{(i,j,i',j')\in\mathcal{N}_n^4 : i'\in\mathcal{N}_n(i,m), j'\in\mathcal{N}_n(j,m), d_n(i,j)=s\right\}.$$

Assumption 2.9. *(i)* $\delta_n(2K;2) = o(n)$. *(ii)* $\sum_{s=0}^{2K}|\mathcal{J}_n(s,2K)| = o(n^2)$.

Assumption 2.9 is a version of Assumptions 7c and 7d of Leung (2022a). This assumption is satisfied if network links are not too dense.

Theorem 2.6. *Let* $\widetilde{\gamma}_n = \widehat{\gamma}_n$. *Under Assumptions 2.1 to 2.6, 2.8 and 2.9, we* *have*

$$\frac{1}{N}\widehat{\Sigma}_n = \frac{1}{n\rho_n}\widetilde{\Sigma}_n + \widetilde{B}_n + o_{p^R}(1),$$

where $U_n = o_{p^R}(1)$ *means* $U_n \xrightarrow{p^R} 0$*, and*

$$\widetilde{B}_n = \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]'.$$

Let $\widetilde{\gamma}_n = \gamma_n^{\text{causal}} + o_p(1)$*. If, in addition, we assume Assumption 2.7 and* $\widetilde{d}_n(i,j) = d_n(i,j)$ *a.s. for all* $(i,j) \in \mathcal{N}_n^2$ *with* $R_{n,i} = 1$ *and* $R_{n,j} = 1$ *and for all* $n \in \mathbb{N}$*, then,*

$$\frac{1}{N}\widehat{\Sigma}_n = \frac{1}{n\rho_n}\Sigma_n + \widehat{B}_n + o_p(1),$$

where

$$\widehat{B}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)} \rho_n\mathbb{E}\left[\Psi_{n,i}\right]\mathbb{E}\left[\Psi_{n,j}\right]'.$$

An estimator satisfying $\widetilde{\gamma}_n = \gamma_n^{\text{causal}} + o_p(1)$ is given in Appendix B.1. In general, $\widehat{\gamma}_n \neq \gamma_n^{\text{causal}} + o_p(1)$, and we need a modification on $\widehat{\gamma}_n$. The condition $\widetilde{d}_n(i,j) = d_n(i,j)$ a.s. for units with $R_{n,i} = 1$ and $R_{n,j} = 1$ is satisfied, for example, when the network is sampled using star sampling and the exposure mapping is restricted to the first neighborhood.

Theorem 2.6 implies that we can only estimate the variance up to the ones with bias terms $\widetilde{B}_n$ and $\widehat{B}_n$ since there is no hope to estimate each heterogeneous expectation consistently. This bias is inevitable in heterogeneous treatment effect settings (Abadie et al., 2020; Leung, 2020; Gao and Ding, 2023). Combining this convergence and the asymptotic normality, we can estimate the variance of $\widehat{\theta}_n$ by

$$\left(\widetilde{Q}_n^{XX}\right)^{-1}\left(\frac{1}{N}\widehat{\Sigma}_n\right)\left(\widetilde{Q}_n^{XX}\right)^{-1}. \tag{2.12}$$

The above variance estimator has a problem because we cannot guarantee conservativeness. Indeed, bias matrices $\widehat{B}_n$ and $\widetilde{B}_n$ are not necessarily positive semi-definite.[9] Conservative guarantee modification is possible. We can write $(1/N)\widehat{\Sigma}_n = (1/N)\widehat{R\Psi}'_n\widetilde{K}_n\widehat{R\Psi}_n$, where

$$\begin{aligned}\widehat{R\Psi}_n &= \left(R_{n,1}\widetilde{X}_{n,1}\widehat{\varepsilon}_{n,1}, \cdots, R_{n,n}\widetilde{X}_{n,n}\widehat{\varepsilon}_{n,n}\right)',\\ \widetilde{K}_n &= [\mathbb{1}\{\widetilde{d}_n(i,j) \leq 2K\}]_{i,j}.\end{aligned}$$

Eigendecomposition gives $\widetilde{K}_n = \mathcal{Q}_n\Xi_n\mathcal{Q}_n'$. By replacing $\widetilde{K}_n$ by $\widetilde{K}_n^+ = \mathcal{Q}_n\max\{0, \Xi_n\}\mathcal{Q}_n'$ (max is taken element-wise), the variance matrix estimator

$$\frac{1}{N}\widehat{\Sigma}_n^+ = \frac{1}{N}\widehat{R\Psi}'_n\widetilde{K}_n^+\widehat{R\Psi}_n = \frac{1}{N}\sum_{i=1}^n\sum_{j=1}^n R_{n,i}R_{n,j}\widehat{\Psi}_{n,i}\widehat{\Psi}'_{n,j}\widetilde{K}^+_{n,i,j}.$$

becomes positive semi-definite. We also have $\widetilde{K}_n^- = \mathcal{Q}_n|\min\{0,\Xi_n\}|\mathcal{Q}_n' = \widetilde{K}_n^+ - \widetilde{K}_n$. This modification is provided by Gao and Ding (2023). The

[9] Alternatively, we can implement the randomized inference as Borusyak and Hull (2023a). For multidimensional $\widehat{\theta}_n$, the randomized inference do not guarantee conservativeness, too.

modified variance estimator is given by

$$\left(\widetilde{Q}_n^{XX}\right)^{-1}\left(\frac{1}{N}\widehat{\Sigma}_n^{+}\right)\left(\widetilde{Q}_n^{XX}\right)^{-1}. \tag{2.13}$$

Define $K_n = [\mathbb{1}\{d_n(i,j) \leq 2K\}]_{i,j}$, and define K_n^+ and K_n^- in a similar manner to $\widetilde{K}_n^+$ and $\widetilde{K}_n^-$. Define

$$\widetilde{\delta}_n^-(2K;p) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}|\widetilde{K}_{n,i,j}^-|\right)^p, \qquad \delta_n^-(2K;p) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}|K_{n,i,j}^-|\right)^p,$$

and

$$|\widetilde{\mathcal{J}}_n^-(s,2K)| = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{1}\{d_n(i,j)=s\}\left(\sum_{i'=1}^{n}|\widetilde{K}_{n,i,i'}^-|\right)\left(\sum_{j'=1}^{n}|\widetilde{K}_{n,j,j'}^-|\right),$$
$$|\mathcal{J}_n^-(s,2K)| = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{1}\{d_n(i,j)=s\}\left(\sum_{i'=1}^{n}|K_{n,i,i'}^-|\right)\left(\sum_{j'=1}^{n}|K_{n,j,j'}^-|\right).$$

Assumption 2.10. *(i)* $\widetilde{\delta}_n^-(2K;1) = O_{a.s.}(1)$ *and* $\delta_n^-(2K;1) = O(1)$. *(ii)* $\widetilde{\delta}_n^-(2K;2) = O_{a.s.}(n)$ *and* $\delta_n^-(2K;2) = O(n)$. *(iii)* $\sum_{s=0}^{2K}|\widetilde{\mathcal{J}}_n^-(s,2K)| = O_{a.s.}(n^2)$ *and* $\sum_{s=0}^{2K}|\mathcal{J}_n^-(s,2K)| = O(n^2)$.

Assumption 2.10 is a version of Assumptions 7b-7d of Gao and Ding (2023). The assumption is a modified version of Assumption 2.9 for the eigenvalue modification.

Theorem 2.7. *Let* $\widetilde{\gamma}_n = \widehat{\gamma}_n$. *Under Assumptions 2.1 to 2.6, 2.8 and 2.10, we* *have*

$$\frac{1}{N}\widehat{\Sigma}_n^+ = \frac{1}{n\rho_n}\widetilde{\Sigma}_n + \widetilde{B}_n^+ + o_p^R(1),$$

where

$$\begin{aligned}\widetilde{B}_n^+ =& \frac{1}{n\rho_n}\sum_{i=1}^n\sum_{j=1}^n R_{n,i}R_{n,j}\mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]'\widetilde{K}_{n,i,j}^+ \\ &+ \frac{1}{n\rho_n}\sum_{i=1}^n\sum_{j=1}^n R_{n,i}R_{n,j}\mathbb{E}\Bigg[\left(\widetilde{\Psi}_{n,i} - \mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\right) \\ &\times \left(\widetilde{\Psi}_{n,j} - \mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]\right)' \mid \boldsymbol{R}_n\Bigg]\widetilde{K}_{n,i,j}^-\end{aligned}$$

Let $\widetilde{\gamma}_n = \gamma_n^{\text{causal}} + o_p(1)$*. If, in addition, we assume Assumption 2.7 and* $\widetilde{d}_n(i,j) = d_n(i,j)$ *a.s. for all* $(i,j) \in \mathcal{N}_n^2$ *with* $R_{n,i} = 1$ *and* $R_{n,j} = 1$ *and for all* $n \in \mathbb{N}$*, then,*

$$\frac{1}{N}\widehat{\Sigma}_n^+ = \frac{1}{n\rho_n}\Sigma_n + \widehat{B}_n^+ + o_p(1),$$

where

$$\begin{aligned}\widehat{B}_n^+ =& \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \rho_n\mathbb{E}\left[\Psi_{n,i}\right]\mathbb{E}\left[\Psi_{n,j}\right]' K_{n,i,j}^+ \\ &+ \frac{1}{n\rho_n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left[\left(R_{n,i}\Psi_{n,i} - \rho_n\mathbb{E}\left[\Psi_{n,i}\right]\right)\left(R_{n,j}\Psi_{n,j} - \rho_n\mathbb{E}\left[\Psi_{n,j}\right]\right)'\right] K_{n,i,j}^-.\end{aligned}$$

2.5 Simulation

In this section, we conduct a simulation exercise to illustrate the potential severity of contamination bias. We focus on a case where $T_{n,i} \neq \widetilde{T}_{n,i}$ and contamination bias can arise. See Appendix B.4 for results when $T_{n,i} = \widetilde{T}_{n,i}$, where no contamination bias occurs and our inference procedure is valid under correct model specification.

In the following exercise, we use network link information from Banerjee et al. (2013) to simulate variables based on a real-world network structure, rather than on an artificially generated population network. That study conducted a network survey among randomly selected respondents across 75 villages in rural southern India, where respondents were asked to name 5 to 8 contacts across 12 interaction dimensions (e.g., house visits, borrowing goods). We focus on the borrowing network among individuals, specifically whether a person borrows rice or kerosene from others.[10] To illustrate the applicability of our framework to a single large network without relying on many clusters, we focus on the largest village and use its borrowing network as the population network $\mathbf{A}_n$. Basic network statistics for this village are presented in Table 2.2:

Table 2.2: Network Information

Nodes	**Edges**	**Mean Degree**	**Mean 2nd Order Degree**
1770	5556	6.28	11.44

Notes: **Nodes** reports the number of individuals in the village; **Edges** reports the number of links based on borrowing relationships; **Mean Degree** reports the mean degree; **Mean 2nd Order Degree** reports the mean count of friends-of-friends not directly connected to node i.

In this exercise, we consider a scenario in which the true and observed

[10] Banerjee et al. (2013) also collected household-level network data; we use individual-level network data, which is sparser than the household-level networks.

exposure mappings differ. The main objective is to quantify the severity of contamination bias. Specifically, we focus on a case where there is no contamination bias at the population level, but bias can arise due to the choice of $\widetilde{g}$. We specify the exposure mapping as in Example 2.10:

$$\begin{aligned} T_{n,i} = \Bigg(& R_{n,i} D^*_{n,i}, \frac{\sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j}}{\sum_{j \neq i} A_{n,i,j}}, \\ & \frac{\sum_{j \neq i} \sum_{k \neq i,j} A_{n,i,j} A_{n,j,k} (1 - A_{n,i,k}) R_{n,k} D^*_{n,k}}{\sum_{j \neq i} \sum_{k \neq i,j} A_{n,i,j} A_{n,j,k} (1 - A_{n,i,k})} \Bigg) \\ & =: (D_{n,i}, \mathrm{net}_{n,i}, \mathrm{weak}_{n,i}), \end{aligned}$$

and $\widetilde{T}_{n,i}$ is the same as $T_{n,i}$ except that its second and third elements are replaced by

$$\begin{aligned} \widetilde{\mathrm{net}}_{n,i} &= \frac{\sum_{j \neq i} A_{n,i,j} R_{n,j} D^*_{n,j}}{\sum_{j \neq i} R_{n,j} A_{n,i,j}}; \\ \widetilde{\mathrm{weak}}_{n,i} &= \frac{\sum_{j \neq i} \sum_{k \neq i,j} R_{n,j} A_{n,i,j} A_{n,j,k} (1 - A_{n,i,k}) R_{n,k} D^*_{n,k}}{\sum_{j \neq i} \sum_{k \neq i,j} R_{n,j} A_{n,i,j} R_{n,k} A_{n,j,k} (1 - A_{n,i,k})}. \end{aligned}$$

For comparison, we also consider $\widetilde{T}^{\mathrm{overlap}}_{n,i}$, which is the same as $\widetilde{T}_{n,i}$ except that each $1 - A_{n,i,k}$ in $\widetilde{\mathrm{weak}}_{n,i}$ is replaced by 1. As discussed in Example 2.10, due to overlaps in the second and third elements, the sample-level causal estimand based on $\widetilde{T}^{\mathrm{overlap}}_{n,i}$ will be contaminated. In contrast, the estimands based on $T_{n,i}$ and $\widetilde{T}_{n,i}$ are not, as they are free of such overlaps and correlations.

We implement the following simulation design. First, we set individual-specific parameters as $\theta_{n,i,(1)} \sim \mathrm{Exponential}(1/3)$ i.i.d., $\theta_{n,i,(2)} = M_{n,i}$, $\theta_{n,i,(3)} = 0$, and $\nu_{n,i} \sim N(0, 2)$ i.i.d., where $M_{n,i}$ is a clustering coefficient given by $M_{n,i} = (100/n) \times \sum_{k \neq i} \left(\sum_{j \neq i,k} A_{n,i,j} A_{n,j,k} \right)^2$. Specifically, we draw these $\theta_{n,i}$

and $\nu_{n,i}$ once and treat them as fixed for each Monte Carlo iteration to simulate design-based and sampling-based uncertainties. We choose $\theta_{n,i,(2)} = M_{n,i}$ to mechanically maximize contamination bias, as $M_{n,i}$ correlates with the contamination weights appearing in Corollary 2.1. The average spillover effect from $\text{net}_{n,i}$ (i.e., the average of $M_{n,i}$) is about 1/2. We also set $\theta_{n,i,(3)} = 0$ for all i, so any deviation from 0 can be interpreted as contamination bias. Given the fixed population adjacency matrix $\mathbf{A}_n$ from Banerjee et al. (2013), we can calculate the population-based causal estimand θ_n^{causal}.

Next, for each iteration, we draw $D_{n,i}^* \sim \text{Bernoulli}(0.5)$ i.i.d., and $R_{n,i} \sim \text{Bernoulli}(\rho_n)$ i.i.d. for varying sampling probabilities $\rho_n \in \{0.1, 0.5, 1.0\}$ to see the impact of sampling uncertainty on inference. For each realization of $\boldsymbol{R}_n$, we compute $\theta_n^{\text{causal,sample}}$. Subsequently, using each realization of $\boldsymbol{R}_n$ and $\mathbf{D}_n$, we estimate $\widehat{\theta}_n$ from the regression $Y_{n,i} \sim \widetilde{X}_{n,i} + \widetilde{Z}_{n,i}$, where $\widetilde{Z}_{n,i} = (R_{n,i}p_n, p_n\mathbb{1}\{\sum_{j\neq i} R_{n,j}A_{n,i,j} > 0\}, p_n\mathbb{1}\{\sum_{j\neq i}\sum_{k\neq i,j} R_{n,j}A_{n,i,j}R_{n,k}A_{n,j,k}(1 - A_{n,i,k}) > 0\})$, restricted to units with $R_{n,i} = 1$. Finally, we compute the standard errors based on (2.13) with $\widetilde{\gamma}_n = \widehat{\gamma}_n$ for $\theta^{\text{causal,sample}}$ and with $\widetilde{\gamma}_n$ from Appendix B.1 for θ^{causal}, as well as the conventional Eicker-Huber-White (EHW) standard errors, which are computed from the following variance estimator:

$$\left(\widetilde{Q}_n^{XX}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{n} R_{n,i}\widetilde{X}_{n,i}\widetilde{X}_{n,i}'\widehat{\varepsilon}_{n,i}^2\right)\left(\widetilde{Q}_n^{XX}\right)^{-1}.$$

When computing the standard errors based on (2.13), we use the observed network $\widetilde{\mathbf{A}}_n = [R_{n,i} \times R_{n,j} \times A_{n,i,j}]_{i,j}$, which is the sampled network with induced subgraph links. We repeat this process 2,000 times. The overlapping case is implemented in the same manner, except that we use $\widetilde{T}_{n,i}^{\text{overlap}}$ instead of $\widetilde{T}_{n,i}$, and the third element of $\widetilde{Z}_{n,i}$ is replaced by $p_n \mathbb{1}\{\sum_{j\neq i}\sum_{k\neq i,j} R_{n,j}A_{n,i,j}R_{n,k}A_{n,j,k} > 0\}$.

Simulation results for $\rho_n \in \{0.1, 0.5, 1.0\}$ are summarized in Tables 2.3 to 2.5. In Panel A, we use $\widetilde{T}_{n,i}$ whose $\widetilde{\text{weak}}_{n,i}$ does not have an overlap in $D^*_{n,j}$ for any j with $\widetilde{\text{net}}_{n,i}$. In Panel B, we use $\widetilde{T}^{\text{overlap}}_{n,i}$ whose $\widetilde{\text{weak}}^{\text{overlap}}_{n,i}$ does share some $D^*_{n,j}$ with $\widetilde{\text{net}}_{n,i}$. Also note that, in both panels, the true exposure mapping is fixed to $T_{n,i}$ defined above. Hence, the population-level causal estimands θ^{causal}_n are the same regardless of which $\widetilde{T}_{n,i}$ or $\widetilde{T}^{\text{overlap}}_{n,i}$ is used.

From Panel A, we can observe that the sample-level estimand and estimator largely deviate from the population-level estimand for $\text{net}_{n,i}$. This deviation is driven not by contamination, but by the difference between $\text{net}_{n,i}$ and $\widetilde{\text{net}}_{n,i}$:

$$\text{net}_{n,i} = \frac{\sum_{j\neq i} A_{n,i,j}R_{n,j}D^*_{n,j}}{\sum_{j\neq i} A_{n,i,j}} \neq \frac{\sum_{j\neq i} A_{n,i,j}R_{n,j}D^*_{n,j}}{\sum_{j\neq i} R_{n,j}A_{n,i,j}} = \widetilde{\text{net}}_{n,i}.$$

When ρ_n is small, the denominator of $\widetilde{\text{weak}}_{n,i}$ tends to be smaller than that of $\text{net}_{n,i}$, which results in a downward bias.

Because of the bias, the coverage probabilities against θ^{causal}_n are close to 0 with both EHW standard errors and those based on (2.13), especially when ρ_n is small. However, as ρ_n increases, the bias and coverage probabilities tend to improve with our proposed standard errors (2.13) because the difference between $T_{n,i}$ and $\widetilde{T}_{n,i}$ becomes smaller and the standard errors are designed to be conservative. In contrast, the EHW standard errors fail to capture the dependence structure and thus severely under-cover the causal estimands as ρ_n increases.

From Panel B (with overlap case), we can observe a similar pattern as in

Panel A when ρ_n is small. However, a crucial difference arises when $\rho_n = 1.0$. We can observe that $\theta_{n,(3)}^{\text{causal,sample}}$ and $\widehat{\theta}_{n,(3)}$ are largely biased downward compared with $\theta_{n,(3)}^{\text{causal}}$, with a magnitude similar to that of $\theta_{n,(2)}^{\text{causal}}$. Since the true $\theta_{n,i,(3)} = 0$ for all i, this bias is mainly driven by contamination, as suggested by Corollary 2.1. The contamination bias is also reflected in the average absolute deviation of the estimator and the coverage probabilities against $\theta_{n,(3)}^{\text{causal}}$ for both EHW standard errors and those based on (2.13), resulting in under-coverage.

In summary, the simulation results in Tables 2.3 to 2.5 show that the deviation of $\widetilde{T}_{n,i}$ from $T_{n,i}$ can lead to severe bias and under-coverage for the population causal estimands. The results also highlight the potential severity of contamination bias when there is a small overlap in elements of $\widetilde{T}_{n,i}$, whose size can be comparable to the true spillover effects. This emphasizes the importance of choice of $\widetilde{g}$ in practice and calls for caution when interpreting the results based on the linear regression framework. In the next section, we discuss whether the contamination bias is present in the real data application.

2.6 Empirical Illustration

In an influential study, Cai et al. (2015a) conducted a large-scale network experiment in which they randomly assigned information sessions on weather insurance products to rice farmers in rural villages in China. Out of 185 randomly selected villages, all rice farmers were invited to participate, and approximately 90% agreed to attend. The researchers administered both a household survey (to gather farmer characteristics) and a network survey (to collect friendship links). In the network survey, household heads were asked

Table 2.3: Simulation Results: $T_{n,i} \neq \widetilde{T}_{n,i}$, $\rho = 0.1$

	Panel A: No Overlaps			**Panel B: With Overlaps**		
	D	net	weak	D	net	weak
θ^{causal}	0.348	0.567	0.000	0.348	0.567	0.000
$\theta^{\text{causal,sample}}$	0.347	0.153	0.000	0.347	0.149	0.032
$\widehat{\theta}$	0.347	0.139	0.009	0.347	0.135	0.037
SE EHW	0.163	0.251	0.475	0.163	0.269	0.447
SE mod, θ^{causal}	0.165	0.263	0.549	0.165	0.279	0.492
SE mod, $\theta^{\text{causal,sample}}$	0.163	0.248	0.398	0.163	0.265	0.416
$\lvert\widehat{\theta} - \theta^{\text{causal}}\rvert$	0.182	0.470	0.696	0.181	0.476	0.590
$\lvert\widehat{\theta} - \theta^{\text{causal,sample}}\rvert$	0.180	0.289	0.696	0.179	0.295	0.589
Cov. EHW, θ^{causal}	0.844	0.560	0.703	0.845	0.584	0.756
Cov. EHW, $\theta^{\text{causal,sample}}$	0.846	0.819	0.703	0.840	0.845	0.752
Cov. mod, θ^{causal}	0.847	0.577	0.768	0.848	0.597	0.800
Cov. mod, $\theta^{\text{causal,sample}}$	0.844	0.813	0.618	0.840	0.837	0.714

Note: This table reports simulation results for $\rho = 0.1$. Panel A reports results when $\widetilde{T}_{n,i}$ is used, while Panel B reports results when $\widetilde{T}_{n,i}^{\text{overlap}}$ is used. The first three rows report the averages of the population-level causal estimand θ^{causal}, the sample-level causal estimand $\theta^{\text{causal,sample}}$, and the OLS estimator $\widehat{\theta}$. The next three rows report the average standard errors: the Eicker–Huber–White (EHW) standard errors and the proposed standard errors based on (2.13), evaluated at θ^{causal} and $\theta^{\text{causal,sample}}$. The following two rows report the average absolute deviations of the estimator from the two causal estimands. The final four rows report the coverage probabilities of the 95% confidence intervals constructed using the EHW standard errors and the proposed standard errors for both causal estimands.

Table 2.4: Simulation Results: $T_{n,i} \neq \widetilde{T}_{n,i}$, $\rho = 0.5$

	Panel A: No Overlaps			**Panel B: With Overlaps**		
	D	net	weak	D	net	weak
θ^{causal}	0.348	0.567	0.000	0.348	0.567	0.000
$\theta^{\text{causal,sample}}$	0.348	0.282	0.000	0.348	0.279	0.008
$\widehat{\theta}$	0.346	0.280	-0.004	0.346	0.280	0.000
SE EHW	0.087	0.113	0.128	0.087	0.155	0.176
SE mod, θ^{causal}	0.102	0.135	0.175	0.102	0.186	0.220
SE mod, $\theta^{\text{causal,sample}}$	0.100	0.133	0.153	0.100	0.180	0.204
$\lvert\widehat{\theta} - \theta^{\text{causal}}\rvert$	0.080	0.292	0.159	0.080	0.297	0.204
$\lvert\widehat{\theta} - \theta^{\text{causal,sample}}\rvert$	0.080	0.124	0.159	0.080	0.147	0.204
Cov. EHW, θ^{causal}	0.908	0.335	0.797	0.908	0.528	0.828
Cov. EHW, $\theta^{\text{causal,sample}}$	0.909	0.836	0.797	0.910	0.896	0.828
Cov. mod, θ^{causal}	0.942	0.443	0.922	0.938	0.653	0.918
Cov. mod, $\theta^{\text{causal,sample}}$	0.939	0.898	0.870	0.936	0.933	0.886

Note: This table reports simulation results for $\rho = 0.5$. See Table 2.3 for details.

Table 2.5: Simulation Results: $T_{n,i} \neq \widetilde{T}_{n,i}$, $\rho = 1.0$

	Panel A: No Overlaps			**Panel B: With Overlaps**		
	D	net	weak	D	net	weak
θ^{causal}	0.348	0.567	0.000	0.348	0.567	0.000
$\theta^{\text{causal,sample}}$	0.348	0.567	0.000	0.348	0.773	-0.356
$\widehat{\theta}$	0.347	0.565	-0.010	0.347	0.783	-0.374
SE EHW	0.068	0.110	0.111	0.068	0.249	0.264
SE mod, θ^{causal}	0.108	0.191	0.233	0.105	0.419	0.452
SE mod, $\theta^{\text{causal,sample}}$	0.104	0.198	0.173	0.104	0.394	0.395
$\lvert\widehat{\theta} - \theta^{\text{causal}}\rvert$	0.058	0.141	0.153	0.058	0.279	0.439
$\lvert\widehat{\theta} - \theta^{\text{causal,sample}}\rvert$	0.058	0.141	0.153	0.058	0.211	0.295
Cov. EHW, θ^{causal}	0.938	0.775	0.740	0.936	0.854	0.650
Cov. EHW, $\theta^{\text{causal,sample}}$	0.938	0.775	0.740	0.936	0.928	0.834
Cov. mod, θ^{causal}	0.997	0.968	0.964	0.996	0.987	0.902
Cov. mod, $\theta^{\text{causal,sample}}$	0.995	0.971	0.915	0.995	0.995	0.960

Note: This table reports simulation results for $\rho = 1.0$. See Table 2.3 for details.

to list their five closest friends with whom they discussed rice production and financial matters, which provides a star sampling network. They were allowed to list friends outside of their village.[11]

The information sessions were conducted in two rounds (first and second) and with varying intensity (simple or intensive). Farmers were randomly assigned to one of four possible sessions. The main outcome here, $Y_{n,i}$, is a test score measuring understanding of the insurance product, taking 10 values between 0 and 1 (**test**). The treatment variable, $D_{n,i}$, indicates whether a farmer was assigned to an intensive session (**intensive**). To measure the spillover/diffusion effects of the information sessions on farmers' knowledge, the researchers focused on a subsample of farmers who were not invited in the first round and defined (i) the fraction of a farmer's friends who attended an

[11] Cai et al. (2015a) conducted a pilot network survey in two villages without limiting the number of friends, but found that most farmers listed five or fewer friends. We take this analysis at face value and assume that there is no concern about censoring the number of friends.

intensive session in the first round (**net**) and (ii) the fraction of those friends' friends who attended an intensive session in the first round (**weak**).

As discussed in Example 2.10 and the simulation section, including first-order overlaps between **net** and **weak** can significantly affect inference through induced contamination bias.[12] Here, we empirically examine whether such overlaps make a significant difference by comparing results when these overlaps are included or excluded in **net** and **weak**. Specifically, we run the following regression for the overlap and no-overlap specifications:[13]

$$\text{test} \sim \text{intensive} + \text{net} + \text{weak} + \text{controls}.$$

For estimation, unlike in the simulation exercise above, we use all the available villages in the sample, as done in Cai et al. (2015a). We control for household characteristics, village fixed effects, and network information (degree dummy) to satisfy Assumption 2.4. Standard errors are calculated via our proposed method (2.13), with $K = 2$.

Table 2.6 reports the OLS estimator $\widehat{\theta}_n$ and its standard errors, both with and without overlaps in the exposure mappings. When overlaps are included, the coefficient for **net** remains largely unchanged, but the estimate for **weak** becomes substantially more negative. Specifically, the coefficient on **weak** is statistically significant at the 95% confidence level under the overlap

[12] We found that Cai et al. (2015a) included such overlaps in their version of **weak**; see the data/do/rawnet.do file in their replication folder: `https://www.openicpsr.org/openicpsr/project/113593/version/V1/view;jsessionid=743ABAC8AEBB3E612D4250D02BE40429`.

[13] Note that Cai et al. (2015a) specified the exposure mapping as either (intensive, net) or (weak), running regressions separately. Here, we consider a hypothetical scenario where both **net** and **weak** are included in the regression simultaneously, rather than replicating their original results.

Table 2.6: Regression Results for Cai et al. (2015a)'s data

	With Overlaps	No Overlaps
intensive	0.0752	0.0734
	(0.0159)	(0.0164)
net	0.3110	0.2879
	(0.0527)	(0.0500)
weak	-0.1511	-0.0741
	(0.0453)	(0.0383)

Notes: The number of villages is 47, and the total sample size is 1247. The first and second columns report estimates with and without overlaps in first-order links between **net** and **weak**. All regressions include household characteristics, village fixed effects, and network information as controls. Standard errors, computed using our proposed method (2.13) with $\widetilde{\gamma}_n = \widehat{\gamma}_n$, are reported in parentheses.

specification, and its magnitude nearly doubles compared to the no-overlap specification—becoming comparable in size (but opposite in sign) to that of **net**. This highlights the risk of overstating the effect of weak connections due to contamination bias, even when the true effect may be small or absent.

This pattern in the empirical results is consistent with the simulation findings in Table 2.5, where overlaps in the exposure mapping lead to substantial contamination bias in the estimates of **weak**, while the estimates of **net** remain largely unaffected. Overall, this exercise highlights that correlations among elements of the exposure mapping can potentially lead to misleading assessments of causal spillover effects.

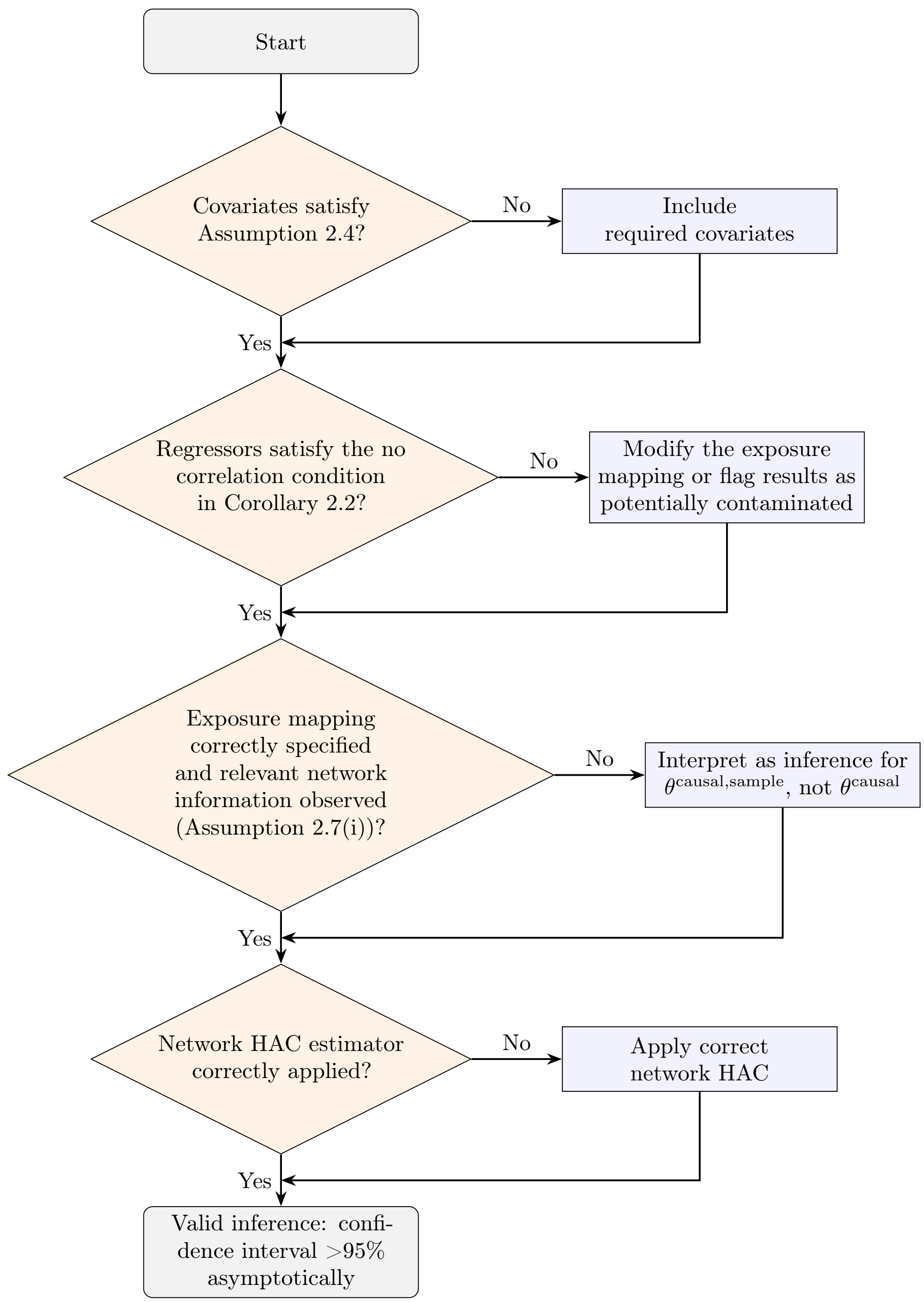
2.7 Conclusion

In this paper, we study a linear regression framework for estimating causal spillover effects in network experiments. We show that, due to contamination bias, the OLS estimator for spillover effects does not bear a causal interpretation unless the exposure mapping is free of correlation among its elements. We also develop a novel asymptotic theory for inference on causal spillover effects, allowing for explicit sampling of units and networks, as well as network dependence.

Based on our theoretical analysis and simulation/empirical exercises, we recommend that researchers follow the flowchart in Figure 2.3 when estimating causal spillover effects in network experiments using linear regression. A crucial step is to ensure that the exposure mapping is free of correlations among its elements to avoid contamination bias and to ensure a causal interpretation of the OLS estimator. If the exposure mapping implied by plausible economic theories is not free of correlations but is sufficiently discrete (e.g., binary) to satisfy the overlap condition, we suggest avoiding the OLS estimator and instead using alternative methods, such as inverse probability weighting (e.g., Aronow and Samii, 2017; Leung, 2022a; Gao and Ding, 2023), to directly estimate the causal treatment effects.

While this paper establishes a comprehensive framework for network experiments on sampled networks, several avenues for future research emerge. First, relaxing the sampling assumptions to accommodate cluster and multi-wave designs, as well as allowing more complex assignment mechanisms, would broaden applicability. The present analysis permits assignment conditional on observed covariates but excludes matched-pair and blocked randomization. Second, a systematic comparison between regression-based estimators and inverse-probability-weighting approaches for spillover effects in network experiments is important, but lies beyond the scope of this paper.

Figure 2.3: Flowchart for Valid Inference with Linear Regression



Chapter 3

Dyadic Regression with Sample Selection

3.1 Introduction

Dyadic data describe pairwise outcomes, such as trade volume between countries. Numerous applications have analyzed such data using the regression model, referred to as dyadic regression. Examples include gravity equations in trade, migration, and urban economics (Helpman et al., 2008; Moretti and Wilson, 2017; Monte et al., 2018), and risk-sharing networks in development economics (Fafchamps and Gubert, 2007). One of the prominent features of dyadic data is the non-negligible number of zeros in the outcomes of interest, [1] possibly due to economic mechanisms such as prohibitive fixed costs. This paper deals with panel dyadic data, where zeros are prevalent both across cross-sections and over time.

How should we treat zeros in dyadic regression? In applications, zeros are often discarded due to the log-linear specification (Moretti and Wilson, 2017). The Poisson pseudo-maximum-likelihood (PPML) estimator is also frequently used to avoid discarding zeros and address issues related to log-linearization (Silva and Tenreyro, 2006). These approaches implicitly assume that zeros occur exogenously. Since a zero in a pairwise outcome results from

[1] Helpman et al. (2008) document that there was no trade among roughly 50% of country pairs from 1970 to 1997. In 2017, there was no migration among about 60% of country pairs (the author calculated using the data available from the World Bank (`https://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data`).

no link between two units, we can associate zeros with the underlying network formation mechanism that determines which pairs appear in a sample. If the network is formed endogenously as a result of an interaction between two agents, the empirical practices mentioned above can be subject to sample selection bias, as in Heckman (1979).

This paper has two primary objectives. First, we aim to jointly model network formation and the outcome generation on such networks. This joint modeling allows identification of the effects of changes in pair-level or individual-level characteristics, separating them from the effects caused by changes in networks. In contrast, the dyadic regression literature has primarily focused on regression with fixed or exogenous networks. Second, we develop a robust inference method that accounts for the dyadic dependence structure. Pairwise outcomes are likely to be dependent on each other through common shocks to individuals. This dyadic dependence can be especially important in the presence of zeros and the network formation because a few individuals can have significantly more links than others, [2] which strengthens the influence of shocks to those individuals on the dyadic dependence. At the same time, it is known that with dyadic data, we can have different asymptotic regimes depending on the nature of those individual-level shocks (Menzel, 2021). To be practitioner-friendly, our inference method needs to consider the dyadic dependence and ensure adaptivity to different resulting asymptotic regimes.

Our setup will be a linear panel dyadic regression model, featuring the

[2] For example, in Moretti and Wilson (2017)'s migration flow data, star scientists' migration from or to California constituted approximately 14% of the links in the sample on average. This percentage is much higher than the expected 2% when considering all potential links in the sample.

network formation process as a sample selection mechanism that generates both zeros and unobservable outcomes. To capture the dyadic dependence structure, we incorporate two types of unobservable individual heterogeneity into the model: time-invariant fixed effects and time-varying random effects, which is a new modeling strategy in the literature. We extend Kyriazidou (1997)'s identification argument, originally designed for individualistic data, to dyadic data, and correspondingly propose a semiparametric, kernel-based estimator that assigns weights to pairs whose selection index remains stable over time. A significant challenge we face when analyzing our estimator is the need to address the dependence structure caused by node-level shocks, which is absent in individualistic data models analyzed in Kyriazidou (1997). To control for this type of dependence, we utilize the U-statistic-like structure of our estimator, which gives us a mutually uncorrelated decomposition into the node-level Hájek projection part and the dyad-level projection error part.

We show that our estimator is asymptotically normal with two different convergence rates depending on the nature of errors. If the Hájek projection is non-degenerate (i.e., each summand has positive variance), our estimator achieves $\sqrt{n}$-asymptotic normality, where n is the number of nodes. In this case, we not only have zero asymptotic bias but also share the same convergence rates as the usual fixed effect estimator and PPML estimator when its leading term is also non-degenerate. The latter point implies that there is no loss in effective sample sizes with our estimator for using a kernel-based local method compared with the usual non-weighted estimator. If the Hájek projection is degenerate, our estimator achieves $\sqrt{Nh_n}$-asymptotic normality, where $N \approx n^2$ is the number of dyads and h_n is a bandwidth. While the usual fixed effect estimator and the PPML estimator can be non-Gaussian in the limit (Menzel, 2021), our estimator is guaranteed to be asymptotically normal regardless of degeneracy. This result is analogous to Hall (1984)'s central limit theorem for degenerate U-statistics, allowing common statistics of interest, such as confidence intervals, to be constructed in a standard manner. In the degenerate case, our estimator exhibits asymptotic bias, which motivates us to introduce a bias correction.

We propose a variance estimator and bias-corrected confidence intervals that adapt to the degeneracy. Our variance estimator is similar to the one proposed by Graham et al. (2019) for nonparametric dyadic density estimation. We show that our estimator is consistent for the asymptotic variances in both non-degenerate and degenerate cases, after being rescaled by $\sqrt{n}$ or $\sqrt{Nh_n}$, respectively. For the bias correction, we use a consistent estimator for the asymptotic bias in the degenerate case. We show that the correction term is negligible in the non-degenerate case after being rescaled by $\sqrt{n}$. Combining both bias-corrected estimator and variance estimator, we can construct bias-corrected confidence intervals for our estimator. These intervals have asymptotically correct sizes regardless of the (non-)degeneracy of the leading term in our estimator.

We conduct a simple simulation exercise to demonstrate the performance of our estimators compared to the usual fixed effect estimator and PPML estimator, as we vary the fraction of selected dyads from 10% to 90%. Our proposed estimator exhibits better finite sample properties than the other two estimators. Our bias-corrected confidence intervals also outperform the alternatives in coverage probabilities, regardless of degeneracy. This result underscores the importance of bias correction in finite samples, even though the asymptotic bias is zero in the non-degenerate case, which is a new finding in the literature.

We apply our estimator to the regression specification proposed by Moretti and Wilson (2017), which estimates the effects of state tax differences on the internal migration flows within the U.S. Comparing our proposed estimator with Moretti and Wilson (2017)'s, we find that their conclusion, which suggests that state tax differences have a significant impact on internal migration, may not be robust in the presence of a dyadic dependence structure and sample selection biases.

This paper is closely related to the growing literature on dyadic regression (Cameron and Miller, 2014; Tabord-Meehan, 2019; Bonhomme, 2020; Zeleneev, 2020; Graham, 2020; Graham et al., 2021; Sassi, 2023). With the exception of Bonhomme (2020) and Zeleneev (2020), most of these papers do not address non-random sample selection, but instead focus on the consequences of dyadic dependence. Bonhomme (2020) primarily studies cases where selection is conditionally random with random effects, and briefly discusses conditionally non-random selection without providing a theoretical analysis. Zeleneev (2020) investigates identification and estimation in cross-sectional dyadic regression models with more flexible combinations of node-level fixed effects, including fixed selection effects as a special case. However, this flexibility comes at the cost of more complex inference, which is not covered in their paper. In contrast, our paper focuses on models with an additional time dimension, enabling us to develop asymptotic distribution theory and a practical inference method that adapts to degeneracy.

This paper also contributes to the literature on econometric analysis of models with endogenous network formation. Examples include Johnsson and Moon (2021), Auerbach (2022), and Jochmans (2023). While these papers study social interaction/peer effects type models where outcomes of interest are individualistic, our paper studies the direct consequence of network formation on dyadic outcomes.

3.2 Model

3.2.1 Setup

There are n nodes in the data (e.g., states, countries), indexed by $i = 1, ..., n$. Let $\{(X_{it}, Z_{it})_{t=1,...,T}\}_{i=1}^n$ be a node-level observation, where $X_{it} \in \mathbb{R}^{q_x}$ and $Z_{it} \in \mathbb{R}^{q_z}$. For each dyad ij and time t, $Y_{ijt} \in \mathbb{R}$ is a main outcome, and we observe a binary variable $d_{ijt} \in \{0, 1\}$, which indicates that Y_{ijt} is observable only if $d_{ijt} = 1$.[3] We can interpret the adjacency matrix $D_t \equiv [d_{ijt}]_{i,j=1,...,n}$ as a network that summarizes the existence of interactions between nodes. In this paper, we restrict our attention to a model with $T = 2$ and an undirected graph where $Y_{ijt} = Y_{jit}$, $d_{ijt} = d_{jit}$ for all i, j, t. We also rule out self-loops by convention: $Y_{iit} = d_{iit} = 0$ for all i, t. An extension to $T > 2$ and a directed graph is discussed in Section 3.4.1.

[3] Since we focus on a linear model, we can interchange unobservability with zero. Alternatively, we can interpret Y_{ijt} as the logarithm of $\widetilde{Y}_{ijt} \geq 0$.

The data is generated according to the following model:

$$W_{ijt} = w(X_{it}, X_{jt}), R_{ijt} = r(Z_{it}, Z_{jt}), \tag{3.1}$$

$$Y^*_{ijt} = W'_{ijt}\beta + \psi(A_i, A_j) + \epsilon_{ijt}, \tag{3.2}$$

$$d_{ijt} = \mathbf{1}\{R'_{ijt}\gamma + \varphi(B_i, B_j) - \eta_{ijt} \geq 0\}, \tag{3.3}$$

$$Y_{ijt} = \begin{cases} Y^*_{ijt} \text{ if } d_{ijt} = 1 \\ \text{unobserved if } d_{ijt} = 0 \end{cases}. \tag{3.4}$$

The regressors $W_{ijt} \in \mathbb{R}^{q_w}$ and $R_{ijt} \in \mathbb{R}^{q_r}$ are constructed from some user-specified symmetric functions $w : \mathbb{R}^{q_x} \times \mathbb{R}^{q_x} \to \mathbb{R}^{q_w}$ and $r : \mathbb{R}^{q_z} \times \mathbb{R}^{q_z} \to \mathbb{R}^{q_r}$ such that $w(x, y) = w(y, x)$ and $r(x', y') = r(y', x')$ for any $x, y \in \mathbb{R}^{q_x}$ and $x', y' \in \mathbb{R}^{q_z}$. For example, we can specify w to be a pairwise summation $w(x, y) = x + y$. The symmetry in these functions is needed as our graphs are undirected; we can relax this requirement with directed graphs, as discussed in Section 3.4.1. The node-level fixed effects $A_i, B_i \in \mathbb{R}$ are unobservable, and we allow them to correlate with the regressors, as in the usual fixed effect model. The functions $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are unknown symmetric functions that capture the interaction between two nodes through their fixed effects.

We specify the structure of errors $\epsilon_{ijt}, \eta_{ijt}$ as follows: For $1 \leq i < j \leq n$,

$$(\epsilon_{ij1}, \epsilon_{ij2}, \eta_{ij1}, \eta_{ij2}) = \tau(U_{i1}, U_{i2}, U_{j1}, U_{j2}, U_{ij1}, U_{ij2}), \tag{3.5}$$

where $U_i \equiv (U_{i1}, U_{i2})$ and $U_{ij} \equiv (U_{ij1}, U_{ij2})$ are node-level and dyad-level random vectors, respectively, and τ is an unknown multivariate function.[4]

Let $\xi_i \equiv (X_{i1}, X_{i2}, Z_{i1}, Z_{i2}, A_i, B_i)$ be a vector that contains observed and unobserved information in the two periods with respect to node i. We impose the following distributional assumption:

Assumption 3.1.

(1) ξ_i*,* $i = 1, ..., n$ *are independently and identically distributed.*

(2) $(\epsilon_{ijt}, \eta_{ijt})_{t=1,2}, 1 \leq i < j \leq n$ *are generated according to (3.5).*

(3) Conditionally on $\{\xi_i\}_{i=1}^n$*,* U_i*,* $i = 1, ..., n$ *are independent,* U_{ij}*,* $1 \leq i < j \leq n$ *are independent, and both of them are mutually independent.*

(4) For $i < j$*,* (U_i, U_j, U_{ij}) *conditional on* $\{\xi_i\}_{i=1}^n$ *has the same distribution as* (U_i, U_j, U_{ij}) *conditional on* ξ_i, ξ_j*.*

(5) For $i < j < k$*, if* $\xi_i = \xi_j = \xi_k$*,* (U_i, U_j, U_{ij}) *and* (U_i, U_k, U_{ik}) *has the same distribution conditionally on* ξ_i, ξ_j, ξ_k*.*

Part (1) imposes homogeneity on the node-level data-generating process. Parts (2) and (3) are new to the literature on dyadic regression with fixed effects. While the previous literature assumes conditional independence of dyadic-level errors (Graham, 2017; Zeleneev, 2020; Candelaria, 2020), our error structure (3.5) allows for the conditional dependence between errors with a common node (e.g., ϵ_{ij1} and ϵ_{ik1}) through U_i, but also includes conditional independence as a special case where node-level random vectors U_{it}, U_{jt} are

[4] Here, we need not specify the dimensions of those vectors and the function since the following results do not depend on them as long as those dimensions are fixed.

degenerate given $\{\xi_i\}_{i=1}^n$. Part (4) is the standard assumption in the literature and excludes "externalities," where dyad ij can be affected by nodes other than i or j. Part (5) ensures the conditional exchangeability of $(\epsilon_{ijt}, \eta_{ijt})_{t=1,2}$ across dyads.

3.2.2 Identification

Let $\Delta A = A_1 - A_2$ be a time difference of a vector A_t between two periods. Also, let $d_{ij} = d_{ij1} \times d_{ij2}$ be an indicator for the dyad ij being observed in both periods.

The following two assumptions are crucial for the identification of β.

Assumption 3.2. *$(\epsilon_{ij1}, \epsilon_{ij2}, \eta_{ij1}, \eta_{ij2})$ and $(\epsilon_{ij2}, \epsilon_{ij1}, \eta_{ij2}, \eta_{ij1})$ are identically distributed conditionally on ξ_i, ξ_j.*

Assumption 3.3. *$\mathbb{E}[d_{ij}\Delta W_{ij}\Delta W_{ij}' | \Delta R_{ij}'\gamma = 0]$ is non-singular.*

Assumption 3.2 excludes cases where, for example, the conditional variance of ϵ_{ijt} depends only on period t's information: $\text{Var}(\epsilon_{ijt}|\xi_i, \xi_j) = \sigma^2 \times W_{ijt}'\beta$. However, it allows time invariant heteroskedasticity such as $\text{Var}(\epsilon_{ijt}|\xi_i, \xi_j) = \sigma^2(W_{ij1} + W_{ij2})'\beta \times A_i \times A_j$. From (3.5), this assumption is implied by the conditional exchangeability of U_{it} and U_{ijt} with respect to time and the symmetry of function τ in the sense that $\tau(u_1, u_2, v_1, v_2, w_1, w_2) = \tau(u_2, u_1, v_2, v_1, w_2, w_1)$ for any $u_1, u_2, v_1, v_2, w_1, w_2$. Assumption 3.3 excludes cases where W_{ijt} is exactly the same as R_{ijt} and implies that some variables in R_{ijt} must be excluded from W_{ijt}. Since

$$\begin{aligned}&\mathbb{E}[d_{ij}\Delta W_{ij}\Delta W_{ij}'|\Delta R_{ij}'\gamma = 0]\\&= Pr(d_{ij} = 1|\Delta R_{ij}'\gamma = 0) \times \mathbb{E}[\Delta W_{ij}\Delta W_{ij}'|d_{ij} = 1, \Delta R_{ij}'\gamma = 0],\end{aligned}$$

this assumption also implies that the networks D_1, D_2 are locally dense across time in the sense that $Pr(d_{ij} = 1|\Delta R_{ij}'\gamma = 0) > 0$.

Our identification argument is summarized in the following two steps, similarly to Kyriazidou (1997). First, take the time-difference on observed outcomes (dyads with $d_{ij1} = d_{ij2} = 1$) to eliminate the fixed effects:

$$\Delta Y_{ij} = \Delta W_{ij}'\beta + \epsilon_{ij1} - \epsilon_{ij2}.$$

If we take expectation of both sides conditionally on $d_{ij1} = d_{ij2} = 1$ and ξ_i, ξ_j,

$$\mathbb{E}[\Delta Y_{ij}|d_{ij} = 1, \xi_i, \xi_j] = \Delta W_{ij}'\beta + \underbrace{\mathbb{E}[\epsilon_{ij1} - \epsilon_{ij2}|d_{ij} = 1, \xi_i, \xi_j]}_{\text{Sample selection effect}}.$$

Note that, in general, the sample selection effect is not 0.

Second, we seek to find conditions to eliminate the selection effect. Assumption 3.2 is equivalent to

$$F(\epsilon_{ij1}, \epsilon_{ij2}, \eta_{ij1}, \eta_{ij2}|\xi_i, \xi_j) = F(\epsilon_{ij2}, \epsilon_{ij1}, \eta_{ij2}, \eta_{ij1}|\xi_i, \xi_j),$$

where F is the conditional distribution of the errors given ξ_i, ξ_j. Then, for dyad

ij with $\Delta R'_{ij}\gamma = R'_{ij1}\gamma - R'_{ij2}\gamma = 0$,

$$\begin{aligned}
&\mathbb{E}[\epsilon_{ij1}|d_{ij} = 1, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] \\
&= \mathbb{E}[\epsilon_{ij1}|R'_{ij1}\gamma + \varphi(B_i, B_j) \geq \eta_{ij1}, R'_{ij2}\gamma + \varphi(B_i, B_j) \geq \eta_{ij2}, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] \\
&= \mathbb{E}[\epsilon_{ij2}|R'_{ij2}\gamma + \varphi(B_i, B_j) \geq \eta_{ij2}, R'_{ij1}\gamma + \varphi(B_i, B_j) \geq \eta_{ij1}, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] \\
&= \mathbb{E}[\epsilon_{ij2}|d_{ij} = 1, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0].
\end{aligned}$$

Hence, the conditional expectation of ΔY_{ij} given $d_{ij} = 1$, ξ_i, ξ_j, and $\Delta R'_{ij}\gamma = 0$ is

$$\mathbb{E}[\Delta Y_{ij}|d_{ij} = 1, \xi_i, \xi_j, \Delta R'_{ij}\gamma = 0] = \Delta W'_{ij}\beta.$$

Multiplying the both sides by ΔW_{ij} and aggregating ξ_i, ξ_j, we get

$$\begin{aligned}
&\mathbb{E}[\Delta W_{ij}\Delta Y_{ij}|d_{ij} = 1, \Delta R'_{ij}\gamma = 0] \\
&= \mathbb{E}[\Delta W_{ij}\Delta W'_{ij}|d_{ij} = 1, \Delta R'_{ij}\gamma = 0]\beta.
\end{aligned}$$

Then, under Assumption 3.3, β is uniquely written as

$$\beta = \mathbb{E}[d_{ij}\Delta W_{ij}\Delta W'_{ij}|\Delta R'_{ij}\gamma = 0]^{-1}\mathbb{E}[d_{ij}\Delta W_{ij}\Delta Y_{ij}|\Delta R'_{ij}\gamma = 0]. \tag{3.6}$$

3.2.3 Estimation

Estimation is done in two steps. In the first step, we estimate γ with a consistent estimator $\widehat{\gamma}_n$, and in the second step we estimate β with $\widehat{\beta}_n$, a sample analogue of the identified β with γ replaced by $\widehat{\gamma}_n$.

In the following, we focus on the second step. The sample-analogue of (3.6) is given by

$$\widehat{\beta}_n = \left[\sum_{i<j} d_{ij} \Delta W_{ij} \Delta W_{ij}' K_{h_n}(\Delta R_{ij}' \widehat{\gamma}_n) \right]^{-1} \times \left[\sum_{i<j} d_{ij} \Delta W_{ij} \Delta Y_{ij} K_{h_n}(\Delta R_{ij}' \widehat{\gamma}_n) \right],$$

where $\sum_{i<j} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n}$, $K_{h_n}(v) = h_n^{-1} K(v/h_n)$ is a kernel, and h_n is a bandwidth. The weight function is used to smooth the condition $\Delta R_{ij}' \gamma = 0$ and puts larger weight on observations with small $\Delta R_{ij}' \widehat{\gamma}_n$.

To evaluate $\widehat{\beta}_n$ in terms of β, rewrite the time-differenced model as

$$\Delta Y_{ij} = \Delta W_{ij}' \beta + \lambda_{ij} + \nu_{ij},$$

where

$$\lambda_{ij} \equiv \mathbb{E}[\epsilon_{ij1} - \epsilon_{ij2} | d_{ij} = 1, \xi_i, \xi_j]$$
$$\nu_{ij} \equiv \epsilon_{ij1} - \epsilon_{ij2} - \lambda_{ij}.$$

Note that $\mathbb{E}[\nu_{ij}|d_{ij}=1,\xi_i,\xi_j]=0$ by construction. Define

$$\widehat{S}_{WW} \equiv \frac{1}{N}\sum_{i<j} d_{ij}\Delta W_{ij}\Delta W_{ij}' K_{h_n}(\Delta R_{ij}'\widehat{\gamma}_n),$$
$$\widehat{S}_{W\lambda} \equiv \frac{1}{N}\sum_{i<j} d_{ij}\Delta W_{ij}\lambda_{ij} K_{h_n}(\Delta R_{ij}'\widehat{\gamma}_n),$$
$$\widehat{S}_{W\nu} \equiv \frac{1}{N}\sum_{i<j} d_{ij}\Delta W_{ij}\nu_{ij} K_{h_n}(\Delta R_{ij}'\widehat{\gamma}_n).$$

Substituting ΔY_{ij} into $\widehat{\beta}_n$ yields

$$\widehat{\beta}_n = \beta + \widehat{S}_{WW}^{-1}\widehat{S}_{W\lambda} + \widehat{S}_{WW}^{-1}\widehat{S}_{W\nu}.$$

The terms $\widehat{S}_{WW}^{-1}\widehat{S}_{W\lambda}$ and $\widehat{S}_{WW}^{-1}\widehat{S}_{W\nu}$ can be understood as the selection bias term and the stochastic error term of the estimator, respectively.

3.3 Asymptotic Analysis

3.3.1 Regularity Conditions

For ease of notation, we write the following conditions in terms of dyads 12 and 13, which entails no loss of generality under the undirected graph and Assumption 3.1.

Let $f_{R\gamma,2}$ be the joint density of $\Delta R_{12}'\gamma$ and $\Delta R_{13}'\gamma$ when it exists and $f_{R\gamma,2|\xi_2,U_2,\xi_3,U_3}$ be the conditional density given ξ_2, U_2, ξ_3, U_3. Let $f_{R\gamma}$ be the marginal density and $f_{R\gamma|\xi_1,U_1}$ be the conditional density given ξ_1, U_1.

Assumption 3.4. *The joint distribution of $\Delta R_{12}'\gamma$ and $\Delta R_{13}'\gamma$ is absolutely continuous, and for some $\kappa_0 > 0$, the following hold in the neighborhoods*

$(-\kappa_0, \kappa_0)^2$ *or* $(-\kappa_0, \kappa_0)$ *around* $(0, 0)$ *or* (0)*, respectively:*

(1) The density $f_{R\gamma,2}(\cdot, \cdot)$ *is* $k \geq 2$ *times continuously differentiable, and the derivatives* $\frac{\partial^2}{\partial x^p \partial y^q} f_{R\gamma,2}(\cdot, \cdot)$ *are bounded for* $p + q \leq k, p, q \geq 0$ *and bounded away from* 0*.*

(2) The conditional density $f_{R\gamma,2|\xi_2,U_2,\xi_3,U_3}(\cdot, \cdot)$ *given* ξ_2, U_2, ξ_3, U_3 *is continuous and bounded almost surely.*

(3) The marginal density $f_{R\gamma}(\cdot)$ *is bounded away from* 0*.*

(4) The conditional marginal density $f_{R\gamma|\xi_1,U_1}(\cdot)$ *given* ξ_1, U_1 *is continuous and bounded almost surely.*

Part (1) is a smoothness assumption on the density as in the nonparametric regression literature. Part (3) ensures that we observe $\Delta R_{12}'\gamma$ around 0, which is crucial for identification. Parts (2) and (4) essentially requires well-behaved $r(\cdot, \cdot)$ in (3.1).

Define $(w_1, w_2) \mapsto \Lambda(w_1, w_2, \xi_1, \xi_2)$ as

$$\Lambda(w_1, w_2, \xi_1, \xi_2) \equiv \mathbb{E}[\epsilon_{12t} | \eta_{12t} \leq w_1, \eta_{12s} \leq w_2, \xi_1, \xi_2]$$

with $t, s = 1, 2, t \neq s$. This Λ is the sample selection effect caused by the correlation between errors $\epsilon_{12t}, \epsilon_{12s}$ and η_{12t}, η_{12s}. Note that the function Λ does not depend on time t or s because of Assumption 3.2.

Assumption 3.5. *The function* $(w_1, w_2) \mapsto \Lambda(w_1, w_2, \xi_1, \xi_2)$ *is differentiable in the neighborhoods* $(-\kappa_0, \kappa_0)^2$ *around* $(0, 0)$ *for some* $\kappa_0 > 0$*.*

This assumption is essential for controlling the sample selection effect and characterizing the asymptotic bias in some cases. An implication of this assumption is that for some $\Lambda_{12} \equiv \widetilde{\Lambda}(w_1, w_2, \xi_1, \xi_2)$,

$$\Lambda(w_1, w_2, \xi_1, \xi_2) - \Lambda(w_2, w_1, \xi_1, \xi_2) = \Lambda_{12} \times (w_1 - w_2),$$

by the multivariate mean-value theorem. If we focus on the degenerate case discussed below, since the asymptotic bias is 0 in that case, we can relax the differentiability to Lipschitz-like continuity on Λ: $|\Lambda(w_1, w_2, \xi_1, \xi_2) - \Lambda(w_2, w_1, \xi_1, \xi_2)| \leq |\Lambda_{12}| \times |w_1 - w_2|$.

Let $\|\cdot\|$ denote a Euclidian norm of vectors.

Assumption 3.6. *For some $\kappa_0 > 0$, the following hold in the neighborhoods $(-\kappa_0, \kappa_0)^2$ or $(-\kappa_0, \kappa_0)$ around $(0, 0)$ or (0), respectively.*

(1) The following moments are bounded almost surely:

$$\mathbb{E}[\|\Delta W_{12}\|^8 | \Delta R'_{12}\gamma = \cdot, \xi_1, U_1], \mathbb{E}[\|\Delta R_{12}\|^6 | \Delta R'_{12}\gamma = \cdot, \xi_1, U_1],$$
$$\mathbb{E}[\nu_{12}^8 | \Delta R'_{12}\gamma = \cdot, \xi_1, U_1], \mathbb{E}[\Lambda_{12}^6 | \Delta R'_{12}\gamma = \cdot, \xi_1, U_1].$$

(2) The following moments are continuous and bounded, and the first two are positive definite:

$$\mathbb{E}[d_{12}\Delta W_{12}\Delta W'_{12} | \Delta R'_{12}\gamma = \cdot],$$
$$\mathbb{E}[d_{12}\Delta W_{12}\Delta W'_{12}\nu_{12}^2 | \Delta R'_{12}\gamma = \cdot],$$
$$\mathbb{E}[d_{12}d_{13}\Delta W_{12}\Delta W'_{13}\nu_{12}\nu_{13} | \Delta R'_{12}\gamma = \cdot, \Delta R'_{13}\gamma = \cdot].$$

(3) $g(\cdot) \equiv \mathbb{E}[d_{12}\Delta W_{12}\Lambda_{12}|\Delta R_{12}'\gamma = \cdot]f_{R\gamma}(\cdot)$ *is k-times continuously differentiable with bounded derivatives.*

(4) $g_{\xi_1,U_1}(\cdot) \equiv \mathbb{E}[d_{12}\Delta W_{12}\nu_{12}|\Delta R_{12}'\gamma = \cdot, \xi_1, U_1]f_{R\gamma|\xi_1,U_1}(\cdot)$ *is k-times continuously differentiable with bounded derivatives almost surely.*

Part (1) assumes the existence of conditional moments for the relevant variables. The conditioning on ξ_1 and U_1 is needed for controlling the dyadic dependence structure. Part (2) is crucial for obtaining the convergence results used below, and the positive definiteness is needed for ensuring the non-degeneracy of our estimator in the limit. Part (3) is used for characterizing the asymptotic bias provided below. Part (4) is essential for the negligibility of the approximation error of our variance estimator.

Assumption 3.7. *The following moments exist:*

$$\mathbb{E}[\|\Delta W_{12}\|^8], \mathbb{E}[\|\Delta R_{12}\|^8], \mathbb{E}[\Lambda_{12}^6], \mathbb{E}[\nu_{12}^6]$$

Additionally to Assumption 3.6, which restricts the moments locally around $(0, 0)$ or (0), we use the existence of these unconditional moments when bounding error terms coming from the usage of $\widehat{\gamma}_n$.

Assumption 3.8. *A kernel function* $K(\cdot)$ *satisfies the following:*

(1) For some $\kappa > 0$*,* K *is* 0 *outside of* $[-\kappa, \kappa]$*, bounded in* $[-\kappa, \kappa]$*, and three*

times continuously differentiable with bounded derivatives in $(-\kappa, \kappa)$.

(2) $\int K(s)ds = 1$.

(3) $\int s^i K(s)ds = 0$ *for* $i = 1, ..., k$.

For example, a fourth-order biweight kernel $K(x) = 106/64(1-3u^2)(1-x^2)^2\mathbf{1}\{|x|<1\}$ satisfies this assumption with $\kappa = 1$ and $k = 3$.

Assumption 3.9. *The sequence of bandwidths* $\{h_n\}$ *satisfies* $h_n \to 0$ *and* $nh_n \to \infty$ *as* $n \to \infty$.

This assumption is standard in the nonparametric regression literature. We impose further conditions on $\{h_n\}$ in each statement below.

Assumption 3.10. *The first-step estimator* $\widehat{\gamma}_n$ *satisfies* $\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma) = o_p(1)$.

This assumption requires the first-step estimator to be consistent and converge faster than our estimator. For example, if $\eta_{ijt} \sim Logistic(0,1)$ independently across ij and t, we can show that Chamberlain (1980)'s conditional logit estimator satisfies $\widehat{\gamma}_n - \gamma = O_p(1/\sqrt{N})$ so that $\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma) = O_p(\sqrt{h_n}) = o_p(1)$. In Section 3.3.6, we discuss the availability of alternative estimators for γ. We leave the case where $\widehat{\gamma}_n$ converges slower than required in this assumption for future research.

3.3.2 Asymptotic Normality

Define the following components that will appear in the asymptotic bias and variance expression:

$$\begin{aligned}
\Sigma_{WW} &\equiv f_{R\gamma}(0)\mathbb{E}[d_{12}\Delta W_{12}\Delta W_{12}'|\Delta R_{12}'\gamma = 0] \\
\Sigma_{W\lambda} &\equiv \frac{1}{k!}\frac{\partial^k g(0)}{\partial w^k}\int s^{k+1}K(s)ds, \\
\Sigma_{W\nu,1} &\equiv 4f_{R\gamma,2}(0,0)\mathbb{E}[d_{12}d_{13}\Delta W_{12}\Delta W_{13}'\nu_{12}\nu_{13}|\Delta R_{12}'\gamma = \Delta R_{13}'\gamma = 0], \\
\Sigma_{W\nu,2} &\equiv f_{R\gamma}(0)\mathbb{E}[d_{12}\Delta W_{12}\Delta W_{12}'\nu_{12}^2|\Delta R_{12}'\gamma = 0]\int K^2(s)ds.
\end{aligned}$$

We have the following result:

Theorem 3.1. *Suppose that Assumptions 3.1-3.10 hold. Fix an arbitrary non-zero vector* $c \in \mathbb{R}^{q_w}$ *and some constant* $h \in [0, \infty)$. *Let* $c_W = \Sigma_{WW}^{-1}c$. *Then, as* $n \to \infty$, *we have the following three cases:*

(1) If $Nh_n^{2k+3} \to h$ *and* $c_W'\Sigma_{W\nu,1}c_W > 0$:

$$\sqrt{n}c'(\widehat{\beta}_n - \beta) \to_d \mathcal{N}(0, c_W'\Sigma_{W\nu,1}c_W).$$

(2) If $Nh_n^{2k+3} \to h$ *and* $c_W'\Sigma_{W\nu,1}c_W = 0$:

$$\sqrt{Nh_n}c'(\widehat{\beta}_n - \beta) \to_d \mathcal{N}(\sqrt{h}c_W'\Sigma_{W\lambda}, c_W'\Sigma_{W\nu,2}c_W).$$

(3) If $Nh_n^{2k+3} \to \infty$ and $nh_n^{2k} \to \infty$:

$$h_n^{-(k+1)}(\widehat{\beta}_n - \beta) \to_p \Sigma_{WW}^{-1}\Sigma_{W\lambda}.$$

Parts (1) and (2) of Theorem 3.1 show that our estimator is asymptotically normal, with different convergence rates depending on $\Sigma_{W\nu,1}$. Part (1) differs from Kyriazidou (1997) in that the convergence rate is parametric and based on the number of nodes n, rather than the number of dyads N. When $c_W'\Sigma_{W\nu,1}c_W > 0$ in part (1), the covariance between summands sharing a common node (e.g., dyads ij and ik) does not vanish asymptotically, reducing the effective sample size to n. The leading term is an average of conditional means given ξ_i and U_i, which averages out and eliminates h_n from the convergence rate. This $\sqrt{n}$-asymptotic normality matches results in the dyadic nonparametric density estimation literature (Graham et al., 2019). When $\Sigma_{W\nu,1} = 0$, as in part (2), our result aligns with Kyriazidou (1997), with nonparametric convergence rates based on the number of dyads. This corresponds to the degenerate case for dyadic dependence, as discussed in the literature (Graham et al., 2019; Cattaneo et al., 2024). Part (3) of Theorem 3.1 shows that, with suitable normalization, our estimator converges to the asymptotic bias term regardless of degeneracy. This property is used to construct the bias-corrected estimator in the following section.

We can compare our estimator with the usual fixed effect estimator:

$$\widehat{\beta}_{FE} = \left[\sum_{i<j} d_{ij}\Delta W_{ij}\Delta W'_{ij}\right]^{-1}\left[\sum_{i<j} d_{ij}\Delta W_{ij}\Delta Y_{ij}\right], \tag{3.7}$$

which is biased because of the selection effect λ_{ij}. First, in the case of non-degeneracy, our estimator and the re-centered (infeasible) fixed effect estimator share the same convergence rates of $\sqrt{n}$ (Davezies et al., 2021). This implies that there is no reduction in the effective sample size for using our kernel-based local estimator, which amends the need for fairly large samples as discussed in Kyriazidou (1997). Second, in the case of degeneracy, the fixed effect estimator applied to our model can exhibit a non-Gaussian distribution in the limit (Menzel, 2021), while our estimator is asymptotically normal regardless of the degeneracy. This guaranteed asymptotic normality is analogous to Hall (1984)'s central limit theorem for degenerate U-statistics, and thus the common statistics of interest, such as confidence intervals, can be constructed in a standard manner.

If we interpret the structural equation (3.2) as the log-linearized version of the canonical gravity model (Silva and Tenreyro, 2006; Head and Mayer, 2014) with additive fixed effects,

$$\widetilde{Y}_{ijt} = exp(W'_{ijt}\beta + A_i + A_j) \times \underbrace{\omega_{ijt}}_{=d_{ijt}exp(\epsilon_{ijt})},$$

the Poisson pseudo-maximum-likelihood estimator (PPML) for β can be compared with our estimator. The PPML estimator $\widehat{\beta}_{PPML}$ with two-way fixed effects is defined as the solution to:

$$\sum_{i<j}\sum_{t=1}^{2}\left[\widetilde{Y}_{ijt} - exp(W'_{ijt}b + a_i + a_j)\right] W_{ijt} = 0, \tag{3.8}$$

where $a_1, ..., a_n$ satisfy

$$\sum_{j=i+1}^{n}\sum_{t=1}^{2}\left[\widetilde{Y}_{ijt} - exp(W'_{ijt}\widehat{\beta}_{PPML} + a_i + a_j)\right] = 0, \quad i = 1, ..., n-1.$$

We can make a similar comparison as in the fixed effect estimator based on the results by Davezies et al. (2021) and Menzel (2021): $\widehat{\beta}_{PPML}$ will be biased because of the misspecified errors, and the re-centered $\widehat{\beta}_{PPML}$ is asymptotically normal at the rate of $\sqrt{n}$ in the non-degenerate case and can be non-Gaussian in the degenerate case.

3.3.3 Variance Estimation

Since our estimator exhibits different asymptotic distributions depending on $\Sigma_{W\nu,1}$, it is desirable to have a variance estimator that adapts to the degeneracy.

First, we estimate $\Sigma_{W\nu,1}$. Define

$$\widehat{S}_{ij} \equiv 2d_{ij}K_{h_n}(\Delta R'_{ij}\widehat{\gamma}_n)\Delta W_{ij}\Delta\widehat{\epsilon}_{ij},$$

where $\Delta\widehat{\epsilon}_{ij}$ is a residual $\Delta Y_{ij} - \Delta W'_{ij}\widehat{\beta}_n$. Then, we propose an estimator for $\Sigma_{W\nu,1}$ as

$$\widehat{\Sigma}_{W\nu,1} = \binom{n}{3}^{-1} \Sigma_{i<j<k}\frac{1}{3}(\widehat{S}_{ij}\widehat{S}'_{ik} + \widehat{S}_{ij}\widehat{S}'_{jk} + \widehat{S}_{ik}\widehat{S}'_{jk}).$$

Next, we estimate $\Sigma_{W\nu,2}$ by

$$\widehat{\Sigma}_{W\nu,2} = \frac{h_n}{N} \sum_{i<j} d_{ij} K_{h_n} (\Delta R_{ij}' \widehat{\gamma}_n)^2 \Delta W_{ij} \Delta W_{ij}' \Delta \widehat{\epsilon}_{ij}^2.$$

The following result shows consistency of these estimators and their usefulness in adaptive variance estimation.

Proposition 3.1. *Suppose that Assumptions 3.1-3.10 hold. Let* $h_n = h \times N^{-1/(2k+3)}$ *for some* $h \in (0, \infty)$. *We have*

$$\widehat{\Sigma}_{W\nu,1} \to_p \Sigma_{W\nu,1},$$
$$\widehat{\Sigma}_{W\nu,2} \to_p \Sigma_{W\nu,2},$$

as $n \to \infty$. *If* $c_W \Sigma_{W\nu,1} c_W = 0$ *with* $c_W = \Sigma_{WW}^{-1} c$ *for some* $c \in \mathbb{R}^{q_w}$, *we have*

$$n h_n c' \widehat{S}_{WW}^{-1} \widehat{\Sigma}_{W\nu,1} \widehat{S}_{WW}^{-1} c \to_p 0,$$

as $n \to \infty$.

We now propose our variance estimator as follows:

$$\widehat{\Sigma} \equiv \widehat{S}_{WW}^{-1} \left[\frac{n-2}{n(n-1)} \widehat{\Sigma}_{W\nu,1} + \frac{1}{N h_n} \widehat{\Sigma}_{W\nu,2} \right] \widehat{S}_{WW}^{-1}.$$

We can see that this estimator is adaptive to the degeneracy: When $\Sigma_{W\nu,1}$ is positive definite, since $n/(Nh_n) = o(1)$,

$$nc'\widehat{\Sigma}c = c'\widehat{S}_{WW}^{-1}\left[\frac{n-2}{n-1}\widehat{\Sigma}_{W\nu,1} + \frac{n}{Nh_n}\widehat{\Sigma}_{W\nu,2}\right]\widehat{S}_{WW}^{-1}c \to_p c_W'\Sigma_{W\nu,1}c_W,$$

as $n \to \infty$ by Proposition 3.1 and Lemma C.1 in Appendix C.1. When $c_W'\Sigma_{W\nu,1}c_W = 0$, since $nh_nc'\widehat{S}_{WW}^{-1}\widehat{\Sigma}_{W\nu,1}\widehat{S}_{WW}^{-1}c = o_p(1)$ by Proposition 3.1,

$$Nh_nc'\widehat{\Sigma}c = c'\widehat{S}_{WW}^{-1}\left[\frac{(n-2)h_n}{2}\widehat{\Sigma}_{W\nu,1} + \widehat{\Sigma}_{W\nu,2}\right]\widehat{S}_{WW}^{-1}c \to_p c_W'\Sigma_{W\nu,2}c_W,$$

as $n \to \infty$.

Our variance estimator is adapted from the one provided in Graham et al. (2019) for a dyadic nonparametric density estimator. They show that this type of estimator can be adaptive to the "knife edge" case, where nh_n is bounded from above and below asymptotically so that $Nh_n \approx n$. Here, we additionally show that the estimator is adaptive to the degeneracy by showing that the term involving $\widehat{\Sigma}_{W\nu,1}$ decays fast enough to be negligible when the convergence rate is $\sqrt{Nh_n}$.

3.3.4 Bandwidth Selection

From the asymptotic distributional approximation result in Theorem 3.1, we can write down the mean squared error of our estimator (without negligible parts)

$$MSE(c'\widehat{\beta}_n) = h_n^{2(k+1)}(c_W'\Sigma_{W\lambda})^2 + \frac{1}{n}c_W'\Sigma_{W\nu,1}c_W + \frac{1}{Nh_n}c_W'\Sigma_{W\nu,2}c_W.$$

The optimal solution for minimizing this mean squared error with respect to h_n is given by

$$\begin{aligned} h_n^* &= \left(\frac{c_W' \Sigma_{W\nu,2} c_W}{2(k+1)N(c_W' \Sigma_{W\lambda})^2} \right)^{\frac{1}{2k+3}} \\ &= h^* N^{-\frac{1}{2k+3}}. \end{aligned}$$

We can estimate h^* by the plug-in method. By Proposition 3.1, we have a consistent estimator for the variance part. For the bias part, we use a pilot bandwidth given by

$$h_{n,\delta} = hN^{-\delta/(2k+3)},$$

for some $\delta \in (0, \frac{2k+3}{4k+4})$ and $h > 0$. Let $\widehat{\beta}_{n,\delta}$ be our estimator calculated with $h_{n,\delta}$. We can check that this bandwidth satisfies $Nh_{n,\delta}^{2k+3} \to \infty$ and $nh_{n,\delta}^{2k+2} \to \infty$. Thus, by Theorem 3.1,

$$h_{n,\delta}^{-(k+1)} (\widehat{\beta}_{n,\delta} - \beta) \to_p \Sigma_{WW}^{-1} \Sigma_{W\lambda},$$

as $n \to \infty$. By replacing β by $\widehat{\beta}_n$, calculated with $h_n = hN^{-\frac{1}{2k+3}}$, we have the following result:

Proposition 3.2. *Suppose that Assumptions 3.1-3.10 hold. Let $\widehat{\beta}_n$ and $\widehat{\beta}_{n,\delta}$ be the proposed estimators with bandwidths $h_n = hN^{-1/(2k+3)}$ and $h_{n,\delta} = hN^{-\delta/(2k+3)}$, respectively, for some $h > 0$ and $\delta \in (0, \frac{2k+3}{4k+4})$. Then,*

$$h_{n,\delta}^{-(k+1)} (\widehat{\beta}_{n,\delta} - \widehat{\beta}_n) \to_p \Sigma_{WW}^{-1} \Sigma_{W\lambda},$$

as $n \to \infty$.

Thus,

$$\widehat{h}^* = \left(\frac{c'\widehat{S}_{WW}^{-1}\widehat{\Sigma}_{W\nu,2}\widehat{S}_{WW}^{-1}c}{2(k+1)\{h_{n,\delta}^{-(k+1)}c'(\widehat{\beta}_{n,\delta} - \widehat{\beta}_n)\}^2} \right)^{\frac{1}{2k+3}}$$

is a consistent estimator for h^* by Propositions 3.1 and 3.2.

3.3.5 Bias Correction

Notice that our estimator has the asymptotic bias of $\sqrt{h}\Sigma_{WW}^{-1}\Sigma_{W\lambda}$ in the case of degeneracy, $\Sigma_{W\nu,1} = 0$ from Theorem 3.1. If the bias is non-negligible, it distorts the coverage probability of the confidence interval. Correcting the bias part is desirable as it is generally unknown whether the degeneracy occurs. Fortunately, given the similar asymptotic distributional result as Kyriazidou (1997) in the degenerate case, we can use her bias correction strategy as follows.

Note that $h_{n,\delta}^{-(k+1)}(\widehat{\beta}_{n,\delta} - \beta)$ directly estimates the asymptotic bias from Theorem 3.1. We can construct a bias-corrected estimator $\widehat{\beta}_{n,bc}(\beta)$ by subtracting this bias estimator from the original estimator with suitable normalization: Let $r_{n,\delta} = N^{(1-\delta)/(2k+3)}$. The bias-corrected estimator is given by

$$\widehat{\beta}_{n,bc}(\beta) = \widehat{\beta}_n - r_{n,\delta}^{-(k+1)}(\widehat{\beta}_{n,\delta} - \beta).$$

We can check that this estimator is asymptotically unbiased regardless of the degeneracy: When $c_W'\Sigma_{W\nu,1}c_W > 0$,

$$\begin{aligned}
&\sqrt{n}c'(\widehat{\beta}_{n,bc}(\beta) - \beta) \\
&= \sqrt{n}c'(\widehat{\beta}_n - \beta) - \sqrt{n}r_{n,\delta}^{-(k+1)}c'(\widehat{\beta}_{n,\delta} - \beta) \\
&= \underbrace{\sqrt{n}c'(\widehat{\beta}_n - \beta)}_{\to_d \mathcal{N}(0, c_W'\Sigma_{W\nu,1}c_W)} - \underbrace{\sqrt{n}h^{k+1}N^{-(k+1)/(2k+3)}}_{\to 0}\underbrace{h_{n,\delta}^{-(k+1)}c'(\widehat{\beta}_{n,\delta} - \beta)}_{\to_p c_W'\Sigma_{W\lambda}} \\
&\to_d \mathcal{N}(0, c_W'\Sigma_{W\nu,1}c_W),
\end{aligned}$$

as $n \to \infty$. When $c_W'\Sigma_{W\nu,1}c_W = 0$,

$$\begin{aligned}
&\sqrt{Nh_n}c'(\widehat{\beta}_{n,bc}(\beta) - \beta) \\
&= \sqrt{Nh_n}c'(\widehat{\beta}_n - \beta) - \sqrt{Nh_n}r_{n,\delta}^{-(k+1)}c'(\widehat{\beta}_{n,\delta} - \beta) \\
&= \underbrace{\sqrt{Nh_n}c'(\widehat{\beta}_n - \beta)}_{\to_d \mathcal{N}(\sqrt{h^{2k+3}}c_W'\Sigma_{W\lambda}, c_W'\Sigma_{W\nu,2}c_W)} - \underbrace{\sqrt{h^{2k+3}}h_{n,\delta}^{-(k+1)}c'(\widehat{\beta}_{n,\delta} - \beta)}_{\to_p \sqrt{h^{2k+3}}c_W\Sigma_{W\lambda}} \\
&\to_d \mathcal{N}(0, c_W'\Sigma_{W\nu,2}c_W),
\end{aligned}$$

as $n \to \infty$. Thus, given the adaptivity of $\widehat{\Sigma}$ to the degeneracy, we have

$$(c'\widehat{\Sigma}c)^{-1/2}c'(\widehat{\beta}_{n,bc}(\beta) - \beta) \to_d \mathcal{N}(0, 1),$$

as $n \to \infty$ for an arbitrary non-zero vector $c \in \mathbb{R}^{q_w}$.

Then, we can construct the bias-corrected confidence interval as follows: Letting $\Phi_{1-\alpha/2}^{-1}$ be $1 - \alpha/2$ quantile of the standard normal distribution, we have

$$
\begin{aligned}
&-\Phi^{-1}_{1-\alpha/2} \le (c'\widehat{\Sigma}c)^{-1/2}c'(\widehat{\beta}_{n,bc}(\beta)-\beta) \le \Phi^{-1}_{1-\alpha/2} \\
&\Longleftrightarrow CI_{L,\alpha,c} \le c'\beta \le CI_{U,\alpha,c},
\end{aligned}
$$

where

$$
\begin{aligned}
CI_{L,\alpha,c} &\equiv (1-h_{n,1-\delta}^{-(k+1)})^{-1}\left[c'\widehat{\beta}_n - h_{n,1-\delta}^{-(k+1)}c'\widehat{\beta}_{n,\delta} - (c'\widehat{\Sigma}c)^{-1/2}\Phi^{-1}_{1-\alpha/2}\right], \\
CI_{U,\alpha,c} &\equiv (1-h_{n,1-\delta}^{-(k+1)})^{-1}\left[c'\widehat{\beta}_n - h_{n,1-\delta}^{-(k+1)}c'\widehat{\beta}_{n,\delta} + (c'\widehat{\Sigma}c)^{-1/2}\Phi^{-1}_{1-\alpha/2}\right].
\end{aligned}
$$

The full inference procedure is summarized as follows:

(1) Compute the first step estimator $\widehat{\gamma}_n$.

(2) Choose $k \ge 2, \delta \in (0, \frac{2k+3}{4k+4})$, and $h > 0$ to compute $\widehat{\beta}_n$ and $\widehat{\beta}_{n,\delta}$ with bandwidths $h_n = hN^{-1/(2k+3)}$ and $h_{n,\delta} = hN^{-\delta/(2k+3)}$, respectively.

(3) Compute $\widehat{\Sigma}$ and $h_{n,\delta}^{-(k+1)}(\widehat{\beta}_{n,\delta} - \widehat{\beta}_n)$ to estimate the asymptotic variance and bias and obtain $\widehat{h}^*$.

(4) Update $\widehat{\beta}_n$ and $\widehat{\beta}_{n,\delta}$ with bandwidths $h_n = \widehat{h}^* N^{-1/(2k+3)}$ and $h_{n,\delta} = \widehat{h}^* N^{-\delta/(2k+3)}$, respectively.

(5) Construct the confidence interval by computing $CI_{L,\alpha,c}$ and $CI_{U,\alpha,c}$ from $\widehat{\beta}_n, \widehat{\beta}_{n,\delta}$, and $c'\widehat{\Sigma}c$.

3.3.6 First-step Estimator

Remember that we want to estimate γ from the selection equation or network formation process (3.3):

$$d_{ijt} = \mathbf{1}\{R'_{ijt}\gamma + \varphi(B_i, B_j) - \eta_{ijt} \geq 0\}.$$

This DGP can be interpreted as a panel discrete choice model as well as a network formation model. Estimators for discrete choice models such as Chamberlain (1980), Manski (1987), or Horowitz (1992) can be candidates for estimating γ. Also, estimators for network formation models such as Graham (2017) or Candelaria (2020) can be applicable under additional conditions.

Whether those estimators can be used as our first-step estimator $\widehat{\gamma}_n$ boils down to their convergence rates: Recall that Assumption 3.10 requires that $\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma) = o_p(1)$, which implies that the first-step estimator needs to converge faster than $\widehat{\beta}_n$. We can conjecture that, without additional conditions on η_{ijt}, the convergence rates of those estimators are $\sqrt{n}$ in worst cases due to the conditional dependence across dyads. Obviously, $\sqrt{n}$-rate is incompatible with Assumption 3.10. In the following, we discuss what kind of additional conditions are needed to ensure Assumption 3.10.

We may assume additive separability for η_{ijt}: $\eta_{ijt} = V_{it} + V_{jt} + V_{ijt}$, where conditionally on $\{\xi_i\}_{i=1}^n$, $(V_{i1}, V_{i2}), i = 1, ..., n$ is independent, $(V_{ij1}, V_{ij2}), 1 \leq i < j \leq n$ is independent, and both are mutually independent. This assumption is weaker than assuming $(\eta_{ij1}, \eta_{ij2}), 1 \leq i < j \leq n$ is conditionally independent given $\{\xi_i\}_{i=1}^n$, where V_{it} is treated as degenerate. With additional conditions, we can directly apply Graham (2017)'s joint maximum likelihood estimator or Candelaria (2020)'s semiparametric estimator, both of which leverage the cross-sectional variation in d_{ijt} and R_{ijt}. We can show that in our setting (especially Assumptions 3.1 and 3.6), the limiting networks are dense, which implies that both Graham (2017) and Candelaria (2020)'s estimators satisfy $\sqrt{N}(\widehat{\gamma}_n - \gamma) = O_p(1)$ and Assumption 3.10.

Alternatively, we may assume that $(\eta_{ij1}, \eta_{ij2}), 1 \leq i < j \leq n$ is conditionally independent given $\{\xi_i\}_{i=1}^n$ and $\varphi(B_i, B_j) = B_i + B_j$. Graham (2017) and Candelaria (2020)'s estimators still satisfy Assumption 3.10, but we can also show that Chamberlain (1980)'s conditional logit estimator and Horowitz (1992)'s smoothed maximum score estimator can satisfy Assumption 3.10. Under the conditional independence assumption, the latter two estimators can be written in an asymptotically locally linear form where the corresponding influence function is indexed by ij with 0 covariances. Thus, the convergence rates are based on N and Assumption 3.10 can be satisfied depending on the tuning parameters.

3.4 Extension

3.4.1 Directed Graph with Multiple Periods

In the above analysis, we restricted our attention to an undirected graph; the variables are all symmetric with respect to nodes (e.g., $Y_{ijt} = Y_{jit}$). Also, there were only two time periods, $t = 1, 2$. The extension to a directed graph case with $t = 1, ..., T$ $(T \geq 2)$ is straightforward; Letting $\Delta_{st} A \equiv A_s - A_t$ denote the time difference between s and t, we propose the following estimator:

$$\widehat{\beta}_n = \left[\sum_{s<t}\sum_{i=1}^{n}\sum_{j\neq i} d_{ijs}d_{ijt}\Delta_{st}W_{ij}\Delta_{st}W_{ij}'K_{h_n}(\Delta_{st}R_{ij}'\widehat{\gamma}_n)\right]^{-1} \times \left[\sum_{s<t}\sum_{i=1}^{n}\sum_{j\neq i} d_{ijs}d_{ijt}\Delta_{st}W_{ij}\Delta_{st}Y_{ij}K_{h_n}(\Delta_{st}R_{ij}'\widehat{\gamma}_n)\right].$$

The above results are not directly applicable to this estimator because there are dyads that are observed between periods s and t in only one direction (e.g., $d_{ijs} = d_{ijt} = 1$ but $d_{jis}d_{jit} = 0$). We leave the full theoretical analysis of this estimator for future work. However, with some loss of efficiency, we can discard those one-directional dyads and use only the two-directional dyads (e.g., $d_{ijs}d_{jit}d_{ijt}d_{jits} = 1$) in the estimation. Then, all the results and their proofs are valid with some modification because we can always rewrite the double sum $\sum_{i=1}^{n}\sum_{j\neq i} A_{ij}$ as $\sum_{i<j}(A_{ij} + A_{ji})$ for any variables $\{A_{ij}\}$ as if the graph were undirected. We will use this version of the estimator in our empirical application.

3.4.2 Pairwise Fixed Effects

In the model (3.2) and (3.3), all the fixed effects are node-wise. Since we are interested in coefficients on time-varying dyadic variables, it is possible to include pairwise fixed effects A_{ij} and B_{ij} in each equation, additionally to A_i, A_j and B_i, B_j. Clearly, with pairwise fixed effects, the identification and estimator will be the same as with node-wise fixed effects since we are leveraging the time variation. Thus, a similar asymptotic analysis will also hold as long as $(A_{ij}, B_{ij}), 1 \leq i < j \leq n$ are independently distributed conditionally on $\{\xi_i\}_{i=1}^n$.

Alternatively, we can also do away with the additive separability by incorporating node-wise fixed effects into pairwise ones:

$$A_{ij} = \widetilde{\tau}\left(\widetilde{A}_i, \widetilde{A}_j, \widetilde{A}_{ij}\right),$$

where $\widetilde{\tau}$ is some unknown function, $\widetilde{A}_i$ is a node-wise fixed effect, and $\widetilde{A}_{ij}$ is a pairwise fixed effect. We can impose a similar structure for B_{ij}. Again, the asymptotic analysis will hold as long as $(\widetilde{A}_{ij}, \widetilde{B}_{ij}), 1 \leq i < j \leq n$ are conditionally independent. With a more general dependence structure, we could show a similar asymptotic result using Kojevnikov et al. (2021b)'s central limit theorem for ψ-dependent data.

3.4.3 Sparsity

Above, we argue that our model and assumptions imply that the limiting networks D_1 and D_2 are locally dense around $\Delta R'_{ij}\gamma \approx 0$. Thus, we limit our attention to cases where the number of dyads in the sample must be proportional to N. Our modeling is appropriate in some applications, such as trade or migration, where the number of dyads is rather dense. However, ours can be inappropriate for some applications where the networks are sparse such as matched employer-employee, bank-firm data (e.g., Abowd et al. (1999), Jiménez et al. (2014)).

We can accommodate sparse networks by the following modification; let us modify Assumption 3.1 so that $\xi_i, i = 1, ..., n$ are drawn from some distribution that is allowed to depend on n. For example, as argued in Graham

(2017), we can consider a distribution where the fixed effects are such that $\liminf_{1\leq i\leq n} B_i = -\infty$. Then, we can discuss identification and estimation with fixed n, and the moments of interest are all dependent on n. Especially, we can consider the sequence of networks such that $Pr(d_{12} = 1|\Delta R_{12}'\gamma = 0) \to 0$ and $r_n Pr(d_{12} = 1|\Delta R_{12}'\gamma = 0) = \Omega(1)$ for some $r_n \to \infty$ to incorporate sparsity. We do not pursue sparsity in this paper and leave it for future projects.

3.5 Simulation

To see the performance of the estimator, we conduct some simulation exercises. Consider the following data-generating process:

$$
\begin{aligned}
&W_{ijt} = X_{it} + X_{jt}, R_{it} = (W_{ijt}, Z_{it} + Z_{jt})', \\
&A_i = \frac{X_{i1} + X_{i2}}{2}, B_i = \frac{Z_{i1} + Z_{i2}}{2}, \\
&d_{ijt} = \mathbf{1}\{R_{ijt}'(1,1)' + \theta \times (B_i + B_j) - \eta_{ijt} \geq 0\}, \\
&Y_{ijt} = d_{ijt}(W_{ijt} + A_i + A_j + \epsilon_{ijt})
\end{aligned}
$$

where

$$
\begin{aligned}
&X_{it}, Z_{it} \sim \mathcal{N}(2,1), i.i.d. \text{ across } i, t, \\
&\eta_{ijt} \sim Logistic(0,1), i.i.d. \text{ across } ij, t, \\
&\epsilon_{ijt} = U_{it} + U_{jt} + \eta_{ijt}, \text{ where } U_{it} \sim \mathcal{N}(0,\sigma), i.i.d. \text{ across } i, t.
\end{aligned}
$$

Note that $\beta = 1$ and $\gamma = (1,1)'$. We have $\theta \in \{-0.3, -2.0, -3.0\}$ inside of d_{ijt} to control for the fraction of zeros in the simulated data set:

$$Pr(d_{121} \times d_{122} = 0) \approx \begin{cases} 20\% \text{ if } \theta = -0.3 \\ 75\% \text{ if } \theta = -2.0 \\ 90\% \text{ if } \theta = -3.0 \end{cases}.$$

We also change $\sigma \in \{0.0, 1.0\}$ for U_{it} so that $\sigma = 0.0$ ($\sigma = 1.0$) corresponds to the degenerate (non-degenerate) case.

As described above, we can interpret this data-generating process as a log-linearized version of the canonical gravity model (Head and Mayer, 2014); by writing $\widetilde{Y}_{ijt}$ as an observable outcome, we redefine the main equation as

$$\widetilde{Y}_{ijt} = exp(W_{ijt} + A_i + A_j) \times \underbrace{\eta_{ijt}}_{= d_{ijt} exp(\epsilon_{ijt})}.$$

We can take a log and recover the original model for a unit with $d_{ijt} = 1$. This modeling allows a mass at $\widetilde{Y}_{ijt} = 0$, one important feature of dyadic data.

We conduct experiments for $n \in \{50, 100, 150, 200\}$, $\theta \in \{-0.3, -2.0, -3.0\}$, and $\sigma \in \{0.0, 1.0\}$, and iterate 2000 times for each one. We calculate $\widehat{\gamma}_n$ by Chamberlain (1980)'s conditional logit estimator:

$$\widehat{\gamma}_n = \underset{g \in \mathcal{G}}{argmax} \sum_{i<j: d_{ij1} + d_{ij2} = 1} M_{ij}(g)$$

where $\mathcal{G}$ is a compact subset of $\mathbb{R}^{q_r}$ and

$$M_{ij}(g) = \mathbf{1}\{d_{ij1} = 1\} \ln \left(\frac{exp(\Delta R'_{ij} g)}{1 + exp(\Delta R'_{ij} g)} \right) + \mathbf{1}\{d_{ij2} = 1\} \ln \left(\frac{1}{1 + exp(\Delta R'_{ij} g)} \right).$$

For $\widehat{\beta}_n$, we use a fourth-order biweight kernel for $K(\cdot)$, given by $K(x) = 106/64(1-3u^2)(1-x^2)^2\mathbf{1}\{|x|<1\}$. This choice implies that we assume that the smoothness of the model is given by $k = 3$. We set $\delta = 0.4$ and $h = 3.0$ and calculate each estimator and confidence interval according to the inference procedure discussed above.

For comparison, we calculate the fixed effect estimator $\widehat{\beta}_{FE}$ given by (3.7). The standard error is calculated by $\widehat{\Sigma}$, with $K_{h_n}(\cdot)$ replaced by 1. We also calculate the Poisson pseudo-maximum-likelihood (PPML) estimator $\widehat{\beta}_{PPML}$ given by (3.8). We compute $\widehat{\beta}_{PPML}$ and its standard error by the *penppml* package in R (Ferreras Garrucho and Zylkin, 2023). The standard error is clustered at the node level, which is close to $\widehat{\Sigma}_{WW}^{-2}\widehat{\Sigma}_{W\nu,1}$ in our setting (Graham, 2020).

The result is summarized in the following TABLE 1 and 2. In TABLE 1, we evaluate the three estimators by mean and median biases (MeanBias), root mean square error (RMSE) for $\sigma = 0, 1$. In TABLE 2, we compute 95% coverage probabilities (Coverage) of four different confidence intervals: CI_{conv} (conventional CI from $\widehat{\widehat{\beta}}_n$ and $\widehat{\Sigma}$), CI_{bc} (bias-corrected CI given by $CI_{L,0.05}$ and $CI_{U,0.05}$), CI_{FE} (conventional CI from $\widehat{\beta}_{FE}$ and $\widehat{\Sigma}$ with a flat kernel.), and CI_{PPML} (conventional CI from $\widehat{\beta}_{PPML}$ and its node-level clustered standard error).

From TABLE 3.1a and 3.1b, we can see that our estimator performs better than the fixed effect estimator and the PPML estimator in terms of bias, which shows that the weights given by the first step estimator work well in eliminating the bias. Our estimator also outperforms the competitors regarding RMSE, which implies that the loss in precision is not severe. Our estimator also performs well even when there is a large fraction of zeros in Y ($Pr(D_{ij1} \times D_{ij2}) \approx 90\%$ when $\theta = -3.0$). There is little difference between $\sigma = 0$ and $\sigma = 1$ other than added variances in the estimators.

From TABLE 3.2a and 3.2b, we can see that CI_{bc} is close to 95% regardless of the degeneracy ($\Sigma_{W\nu,1} = 0$ or > 0) while the others are off from the targeted nominal coverage. This result confirms the effectiveness of the bias correction strategy as well as the adaptivity of our variance estimator, as claimed in Section 3.3.3. Also, it is notable to see that the bias correction is important for obtaining correct coverage probabilities even though the asymptotic bias is 0 in the case of $\sigma = 1.0$ so that $\Sigma_{W\nu,1} > 0$ (Theorem 3.1) and CI_{conv} would return an asymptotically correct coverage.

3.6 Empirical Example

3.6.1 Background

As a leading application of our model, consider Moretti and Wilson (2017). They study how state-level tax differences affect migration by top scientists in the U.S. Specifically, they estimate the following model implied by their

Table 3.1: Finite sample properties of $\widehat{\beta}_n$, $\widehat{\beta}_{FE}$, and $\widehat{\beta}_{PPML}$

(a) $\sigma = 1.0$

		MeanBias			RMSE		
θ	n	$\widehat{\beta}_n$	$\widehat{\beta}_{FE}$	$\widehat{\beta}_{PPML}$	$\widehat{\beta}_n$	$\widehat{\beta}_{FE}$	$\widehat{\beta}_{PPML}$
-0.3	50	0.045	0.133	0.185	0.122	0.160	0.430
-2.0	50	0.141	0.352	0.467	0.210	0.377	0.617
-3.0	50	0.162	0.369	0.582	0.273	0.415	0.752
-0.3	100	0.038	0.136	0.195	0.087	0.148	0.376
-2.0	100	0.099	0.349	0.438	0.142	0.359	0.536
-3.0	100	0.117	0.359	0.542	0.184	0.378	0.657
-0.3	150	0.028	0.135	0.193	0.070	0.143	0.327
-2.0	150	0.075	0.346	0.427	0.112	0.353	0.496
-3.0	150	0.095	0.356	0.527	0.145	0.367	0.607
-0.3	200	0.024	0.134	0.193	0.060	0.140	0.305
-2.0	200	0.061	0.344	0.417	0.091	0.348	0.471
-3.0	200	0.076	0.352	0.510	0.118	0.360	0.572

(b) $\sigma = 0.0$

		MeanBias			RMSE		
θ	n	$\widehat{\beta}_n$	$\widehat{\beta}_{FE}$	$\widehat{\beta}_{PPML}$	$\widehat{\beta}_n$	$\widehat{\beta}_{FE}$	$\widehat{\beta}_{PPML}$
-0.3	50	0.048	0.134	0.194	0.082	0.142	0.399
-2.0	50	0.140	0.352	0.468	0.176	0.365	0.586
-3.0	50	0.161	0.369	0.581	0.229	0.397	0.714
-0.3	100	0.037	0.135	0.193	0.053	0.138	0.332
-2.0	100	0.093	0.348	0.438	0.110	0.352	0.508
-3.0	100	0.113	0.359	0.546	0.145	0.368	0.630
-0.3	150	0.028	0.135	0.191	0.039	0.136	0.301
-2.0	150	0.071	0.345	0.427	0.082	0.348	0.477
-3.0	150	0.089	0.354	0.529	0.108	0.359	0.592
-0.3	200	0.024	0.135	0.191	0.031	0.136	0.278
-2.0	200	0.058	0.345	0.415	0.067	0.347	0.451
-3.0	200	0.074	0.355	0.508	0.089	0.358	0.553

Note: MeanBias reports the mean of the difference between the corresponding estimator and the true value $\beta = 1$. RMSE reports the root mean square error. $\widehat{\beta}_n$ is our proposed estimator, $\widehat{\beta}_{FE}$ is the fixed effect estimator, and $\widehat{\beta}_{PPML}$ is the Poisson pseudo-maximum-likelihood estimator.

Table 3.2: 95% coverage probabilities of CI_{conv}, CI_{bc}, CI_{FE}, and CI_{PPML}

(a) $\sigma = 1.0$

θ	n	Coverage CI_{conv}	CI_{bc}	CI_{FE}	CI_{PPML}
-0.3	50	0.790	0.961	0.498	0.537
-2.0	50	0.646	0.963	0.150	0.236
-3.0	50	0.640	0.901	0.311	0.211
-0.3	100	0.785	0.978	0.233	0.498
-2.0	100	0.668	0.970	0.011	0.173
-3.0	100	0.674	0.953	0.072	0.143
-0.3	150	0.790	0.971	0.103	0.472
-2.0	150	0.689	0.949	0.001	0.117
-3.0	150	0.688	0.944	0.016	0.09
-0.3	200	0.817	0.964	0.040	0.426
-2.0	200	0.730	0.947	0.000	0.08
-3.0	200	0.720	0.946	0.004	0.08

(b) $\sigma = 0.0$

θ	n	Coverage CI_{conv}	CI_{bc}	CI_{FE}	CI_{PPML}
-0.3	50	0.698	0.918	0.141	0.547
-2.0	50	0.535	0.935	0.026	0.204
-3.0	50	0.592	0.869	0.168	0.182
-0.3	100	0.655	0.960	0.001	0.515
-2.0	100	0.482	0.958	0.000	0.12
-3.0	100	0.571	0.944	0.004	0.106
-0.3	150	0.673	0.977	0.000	0.45
-2.0	150	0.471	0.945	0.000	0.073
-3.0	150	0.532	0.949	0.001	0.065
-0.3	200	0.660	0.970	0.000	0.407
-2.0	200	0.444	0.939	0.000	0.052
-3.0	200	0.520	0.933	0.000	0.047

Note: Coverage reports the coverage probabilities of the corresponding confidence intervals. CI_{conv} is the proposed confidence interval without bias correction, CI_{bc} is the bias-corrected confidence interval given by $CI_{L,0.05}$ and $CI_{U,0.05}$. CI_{FE} is the conventional confidence interval from $\widehat{\beta}_{FE}$ and $\widehat{\Sigma}$ with a flat kernel. CI_{PPML} is the conventional confidence interval from $\widehat{\beta}_{PPML}$ and its node-level clustered standard error.

economic theory:

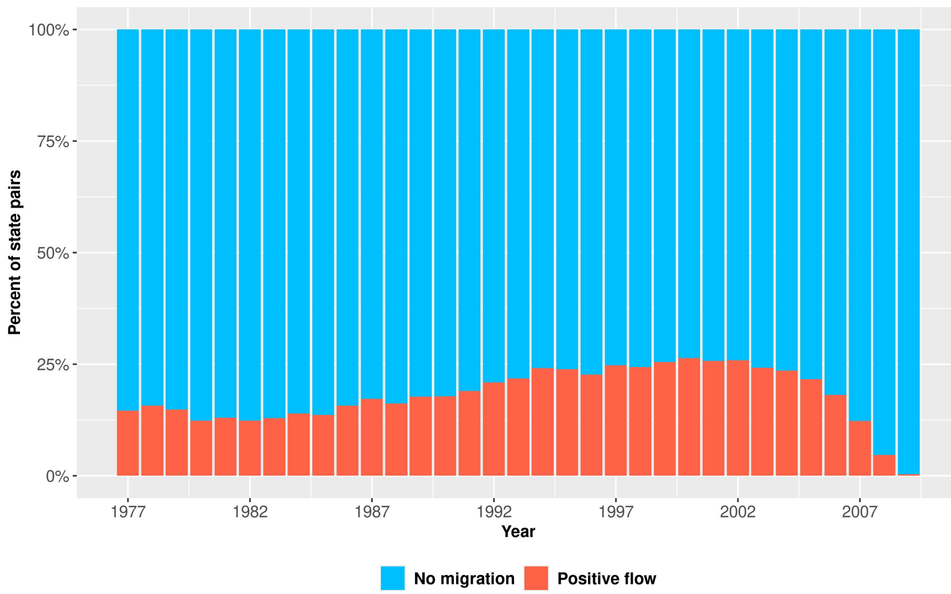
$$log(P_{ijt}/P_{iit}) = \eta\left[log(1-\tau_{jt}) - log(1-\tau_{it})\right] \\ + \eta'\left[log(1-\tau'_{jt}) - log(1-\tau'_{it})\right] + \gamma_j + \gamma_i + u_{ijt},$$

where P_{ijt} is the number of scientists migrating to state j from state i at year t, τ_{it} and τ'_{it} are personal and corporate taxes imposed in state i at year t, γ_i is a state fixed effect, and u_{ijt} is an error term.

Note that if there is no migration from i to j at year t, $P_{ijt} = 0$ and $log(P_{ijt}/P_{iit})$ is undefined. In Moretti and Wilson (2017)'s dataset, more than 70% of state-pairs exhibit no migration flow: When running a regression, they are concerned with a potential sample selection bias stemming from these undefined outcomes. They argue that if the main regressors are not systematically associated with the probability of positive migration flows, the selection bias should be minimal. Running OLS on the linear probability model, they find little correlation between the main regressors and no flow. Re-estimating their model with our method provides a check on the validity of their argument and the appropriateness of using the linear probability model.

When applying our model to their context, we must consider what R_{ijt} should be. Since Moretti and Wilson (2017)'s underlying theory is based on scientists' and firms' discrete choice, one consistent way to generate zero migration flows between some states is to consider endogenous choice sets as in Dubé et al. (2021). Formally, we can write the choice set of representative scientists in state i as $C_{it} = \{j \in \{1, ..., 51\} : d_{ijt} = 1\}$. Here, $\{d_{ijt}\}$ represents the job-market network; if $d_{ijt} = 1$, it is possible to move from i to j, and

Figure 3.1: Fraction of positive migration flows in Moretti and Wilson (2017)'s dataset.



Note: The migration flow is positive in a given year if there is at least one scientist moving from state i to j (scientists are "stars"; they are at or above 95% quantile in number of patents over the past ten years)

vice versa. We can attribute the determinants of the network to the utilities and profits of scientists and firms, as well as the matching costs between the two parties. Such costs are not present in the structural equation if those costs are not compensated through wages; The structural equation consists of the determinants of log wage differences between two states. Thus in the selection equation (3.3), in addition to W_{ijt}, we can include variables in R_{ijt} that capture non-monetary matching costs between two states i and j, which does not violate Assumption 3.3 as R_{ijt} satisfies the exclusion restriction.

3.6.2 Implementation

For W_{ijt}, as in Moretti and Wilson (2017), we include the state-to-state differences in (i) an individual income average income tax rate (ATR) faced by a hypothetical taxpayer at 99% quantile of the national income distribution, (ii) the corporate tax rate (CIT), (iii) the investment tax credit (ITC), and (iv) the R&D tax credit (R&D credit). This is the same set of regressors as Moretti and Wilson (2017)'s baseline regression. For R_{ijt}, we use W_{ijt} plus state-to-state absolute difference in the logarithm of population (POP) and a dummy variable that indicates whether i and j share their governors' political parties (GOV). The additional variables in R_{ijt} arguably measure non-monetary costs of connecting firms and workers in two states.

We implement the first step estimation as follows. We use the conditional logit estimator extended to a directed graph with multi-periods case, which is given by

$$\widehat{\gamma}_n = \underset{g \in \mathcal{G}}{argmax} \sum_{s<t} \sum_{i,j} M_{ij,st}(g),$$

where $\mathcal{G}$ is a compact subset of $\mathbb{R}^{q_r}$ and

$$\begin{aligned} & M_{ij,st}(g) \\ & = \mathbf{1}\{d_{ijs} + d_{ijt} = 1\} \times \left[\mathbf{1}\{d_{ijs} = 1\} ln(e_{ij,st}) + \mathbf{1}\{d_{ijt} = 1\} ln(1 - e_{ij,st})\right], \\ & e_{ij,st} = \frac{exp(\Delta_{st} R'_{ij} g)}{1 + exp(\Delta_{st} R'_{ij} g)}. \end{aligned}$$

TABLE 3.3 reports the first step estimation result. We can see that the coefficients on the newly added variables GOV and POP deviate from zero, which implies that a part of the identification assumptions (Assumption 3.3) is satisfied. Also, for these two variables, the estimated coefficients imply that the job market network exhibits homophily; similar states are more likely to be connected.

Table 3.3: First Step Estimation Result

Variable	$\widehat{\gamma}_n$
GOV	0.162
POP	-13.144
ATR	8.561
CIT	5.850
ITC	5.350
R&D credit	-0.230

Note: The first column reports the variable names included in R_{ijt}. The second column reports the estimated coefficients corresponding to each variable.

For the second step estimator, we use our $\widehat{\beta}_n$ defined above and extend it to the directed graph with multiple period cases, as discussed in Section 3.4.1. We use a biweight kernel for $K(\cdot)$ (so $k = 2$), choose $h = 3.0$ as an initial constant for pilot bandwidths, and use $\delta = 0.4$ for calculating $h_{n,\delta}$. We extend and use $\widehat{\Sigma}$ to calculate the standard error while taking into account the correlation across time (case 1). We also calculate the bias-corrected 95%-confidence interval by computing $CI_{L,0.05}$ and $CI_{U,0.05}$ as defined above. Also, we list $\widehat{\beta}_{MW}$, an estimate from Moretti and Wilson (2017) (page 1883, TABLE 2A, specification (3)) and calculate the conventional 95%-confidence interval based on their standard errors. Note that $\widehat{\beta}_{MW}$ is the fixed estimator, where its standard error is calculated by clustering across time and the origin, destination, and origin-destination pairs.

We summarize the result in TABLE 3.4. We can see that $\widehat{\beta}_n$ returns similar values as $\widehat{\beta}_{WM}$, which claims the robust positive effect of income and corporate-related tax differences on migration. Thus, Moretti and Wilson (2017)'s estimates are not likely to be qualitatively affected by the sample selection effects. However, while Moretti and Wilson (2017)'s estimates are statistically significant at 5% level, our confidence intervals show that all of the estimates are no longer statistically significant at that level except for ITC. Our insignificance result is driven by both the increase in standard errors [5] and the asymptotic bias correction. Thus, our exercise shows that some of the results in Moretti and Wilson (2017) are sensitive to the dyadic dependence but could be robust to the sample selection issue.

Table 3.4: Comparison of our estimator and Moretti and Wilson (2017)

(a) This paper

Variable	Estimator	CI
ATR	1.63	$[-2.66, 5.87]$
	(1.89)	
CIT	1.67	$[-2.95, 6.28]$
	(2.04)	
ITC	1.98	$[0.65, 3.29]$
	(0.58)	
R&D credit	0.43	$[-1.35, 2.12]$
	(0.78)	

(b) Moretti and Wilson (2017)

Estimator	CI
1.93	$[0.92, 2.93]$
(0.51)	
1.84	$[0.69, 2.99]$
(0.59)	
1.79	$[0.99, 2.60]$
(0.41)	
0.37	$[0.01, 0.72]$
(0.18)	

Note: The left table is our result, and the right table is Moretti and Wilson (2017)'s result (see page 1883, TABLE 2A, specification (3)). The first column reports the variable names included in W_{ijt}. Estimator reports the estimated coefficient and its standard error in parentheses. CI reports the 95% confidence interval, which is bias-corrected in our result.

[5] Our standard error is from $\widehat{\Sigma}$, which takes fully into account the dependence among pairs that share origin and destination, such as California→Wisconsin and New York→California. Moretti and Wilson (2017)'s standard error calculation ignores such dependence structure.

3.7 Conclusion

This paper studies identification and inference of a panel dyadic data sample selection model. We show that Kyriazidou (1997)'s identification strategy can be extended to our dyadic data setting, and we prove asymptotic normality of the proposed estimator.

Our estimator has some appealing properties. The distributional result implies that our estimator has the same convergence rates as the usual estimators used in practice in the non-degenerate case, and there is no loss of effective sample size for using our nonparametric type estimator. Also, our estimator is guaranteed to be asymptotically normal, while others can be non-Gaussian in the limit.

We also provide consistent estimators for asymptotic bias and variance that adapts to the degeneracy. Specifically, the bias-corrected confidence interval has an asymptotically correct size. Our simple simulation exercise confirms the validity of these estimators and highlights the importance of bias correction in both degenerate and non-degenerate cases.

Some important limitations of our analysis remain. First, we assume the availability of the first-step estimator whose convergence rate is fast enough to ignore its impact on the second-step estimator. Full treatment of the first-step estimation error is left for future work. Second, we focus on the time-invariant, node-specific fixed effects. Extending our analysis to more general forms of unobserved heterogeneity, such as time-varying and pair-specific effects, is another important avenue for future research so that our method can be applied to a broader class of models.

Appendix A

Chapter 1

A.1 Covariates

Our model with covariates is given by:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{X}$ is a matrix of covariates. Let $\boldsymbol{M}_{\mathbf{B}} \equiv \mathbf{I}_m - \mathbf{B}(\mathbf{B}'\mathbf{B})^{*}\mathbf{B}'$ and $\boldsymbol{M}_{\mathbf{X}} \equiv \mathbf{I}_m - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ be the projection matrices onto the nullspace of $\mathbf{B}$ and $\mathbf{X}$, respectively. Then, the fixed-effect estimators are given by:

$$\check{\boldsymbol{\alpha}} = (\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbf{B})^{*}\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbf{y}, \quad \widehat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{y},$$

where we assume that $rank(\mathbf{X}) = p$ and $rank((\mathbf{X}, \mathbf{B})) = p + n - 1$. Our goal is to replicate the results in Theorem 1.1 for $\check{\alpha}_i$.

Algebraically, we have

$$\check{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \mathbf{D}^{-1}\mathbf{B}'\boldsymbol{\epsilon} + \boldsymbol{r} + \widetilde{\boldsymbol{r}},$$

where

$$\boldsymbol{r} = \mathbf{D}^{-1}\mathbf{A}(\check{\boldsymbol{\alpha}} - \boldsymbol{\alpha}), \quad \widetilde{\boldsymbol{r}} = -\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

We already worked out the first term $\mathbf{D}^{-1}\mathbf{B}'\boldsymbol{\epsilon}$ in the proof of Theorem 1.1.

Thus, we need to show that both $\boldsymbol{r}$ and $\widetilde{\boldsymbol{r}}$ are negligible.

First, we show that $\widetilde{\boldsymbol{r}}$ is negligible. Note that $\mathbb{E}[\widetilde{\boldsymbol{r}}] = 0$. Observe that

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\boldsymbol{\epsilon}$$

with $\mathbb{E}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}] = 0$. Then,

$$\begin{aligned}
\mathbb{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] &= (\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']\boldsymbol{M}_{\mathbf{B}}\mathbf{X}(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1} \\
&\leq C(\mathbf{X}\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{F}\mathbf{F}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X}'(\mathbf{X}\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1} \\
&\leq \lambda_{n,F} C(\mathbf{X}\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}
\end{aligned}$$

for some absolute constant $C > 0$ under Assumption 1.1: $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] \leq C\mathbf{F}\mathbf{F}'$. Let

$$\widetilde{\rho} = \|(\mathbf{X}'\mathbf{X})(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\|_2,$$

where $\|\cdot\|_2$ is the spectral norm. By the definition of $\widetilde{\rho}$, we have $(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1} \leq \widetilde{\rho}(\mathbf{X}'\mathbf{X})^{-1}$. Thus, we have

$$\begin{aligned}
\mathbb{E}[\widetilde{\boldsymbol{r}}_i^2] &= e_i'\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}\mathbb{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})']\mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i \\
&\leq \lambda_{n,F} C e_i'\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i \\
&\leq \lambda_{n,F}\widetilde{\rho} C e_i'\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i \\
&= \frac{\lambda_{n,F}\widetilde{\rho}C}{m}\overline{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i,
\end{aligned}$$

where $\overline{\boldsymbol{x}}_i = \mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i$ and $\boldsymbol{\Omega} = \mathbf{X}'\mathbf{X}/m$. Since $\lambda_{n,F} = O(\max_{i\in V} d_i)$, as long as $\max_{i\in V} d_i/m = o(1)$ and $\widetilde{\rho} \times \overline{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i = O(1)$, we have $\widetilde{\boldsymbol{r}}_i = o_p(1)$.

Second, we show that $\boldsymbol{r}$ is negligible. Note that $\mathbb{E}[\boldsymbol{r}] = 0$. Let

$$\rho = \|(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X}\|_2.$$

We have

$$\begin{aligned}
\mathbb{E}[\boldsymbol{r}_i^2] &= e_i'\mathbf{D}^{-1}\mathbf{A}\mathbb{E}[(\check{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\check{\boldsymbol{\alpha}} - \boldsymbol{\alpha})']\mathbf{A}'\mathbf{D}^{-1}e_i \\
&= e_i'\mathbf{D}^{-1}\mathbf{A}(\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']\boldsymbol{M}_{\mathbf{X}}\mathbf{B}(\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbf{B})^*\mathbf{A}'\mathbf{D}^{-1}e_i \\
&\leq \lambda_{n,F}C \times e_i'\mathbf{D}^{-1}\mathbf{A}(\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbf{B})^*\mathbf{A}'\mathbf{D}^{-1}e_i \\
&\leq \lambda_{n,F}C \times e_i'\mathbf{D}^{-1}\mathbf{A}(\mathbf{B}'\mathbf{B})^*\mathbf{A}'\mathbf{D}^{-1}e_i \\
&+ \lambda_{n,F}C \times \rho^{-1}\mathbf{D}^{-1}\mathbf{A}(\mathbf{B}'\mathbf{B})^*\mathbf{B}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}(\mathbf{B}'\mathbf{B})^*\mathbf{A}'\mathbf{D}^{-1}e_i
\end{aligned}$$

where the last line follows from (S.14) in Jochmans and Weidner (2019). Since $(\mathbf{B}'\mathbf{B})^* \leq \lambda_{2,L}^{-1}\mathbf{D}^{-1}$ and $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \leq I$, we have

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{r}_i^2] \\
&\leq \lambda_{n,F}C \times e_i'\mathbf{D}^{-1}\mathbf{A}(\mathbf{B}'\mathbf{B})^*\mathbf{A}'\mathbf{D}^{-1}e_i + \lambda_{n,F}C \times \rho^{-1}\mathbf{D}^{-1}\mathbf{A}(\mathbf{B}'\mathbf{B})^*\mathbf{A}'\mathbf{D}^{-1}e_i \\
&\leq \frac{\lambda_{n,F}C(1+\rho)}{\rho\lambda_{2,L}} \times e_i'\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}'\mathbf{D}^{-1}e_i \\
&= \frac{\lambda_{n,F}C(1+\rho)}{\rho\lambda_{2,L}h_id_i}.
\end{aligned}$$

If $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) = o(1)$, $\boldsymbol{r}_i = o_p(1)$ as long as $\rho^{-1} = O(1)$.

In summary, as long as

- $\max_{i\in V} d_i = o(m)$,
- $\widetilde{\rho} \times \overline{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i = O(1)$

- $\rho^{-1} = O(1)$,

then Theorem 1.1 holds for $\check{\boldsymbol{\alpha}}$ as well with the same approximation error rates.

A.2 Alternative Estimator

We can consider an alternative estimator for $\boldsymbol{\alpha}$ based on the following transformation.

A.2.1 Transformation

Let $V^{st} = \{i \in V : d_i^t > 0 \text{ and } d_i^s > 0\}$ be the set of nodes with both positive in- and out-degrees. Let $V^t = \{i \in V : d_i^t > 0\}$ and $V^s = \{i \in V : d_i^s > 0\}$ be the sets of nodes with positive in-degree and out-degree, respectively. Redefine $\boldsymbol{\tau}^t$ and $\boldsymbol{\tau}^s$ as $\boldsymbol{\tau}^t = (\mathbb{I}\{i \in V^t\} \cdot \tau^t(U_i))_{i\in V}$ and $\boldsymbol{\tau}^s = (\mathbb{I}\{i \in V^s\} \cdot \tau^s(U_i))_{i\in V}$. Define the following vectors:

$$\widetilde{\boldsymbol{\tau}}^t = (\tau^t(U_i))_{i\in V^{st}}, \quad \widetilde{\boldsymbol{\tau}}^s = (\tau^s(U_i))_{i\in V^{st}},$$
$$\check{\boldsymbol{\tau}}^t = (\mathbb{I}\{i \in V^t \setminus V^s\} \cdot \tau^t(U_i))_{i\in V}, \quad \check{\boldsymbol{\tau}}^s = (\mathbb{I}\{i \in V^s \setminus V^t\} \cdot \tau^s(U_i))_{i\in V}.$$

Let $\widetilde{\mathbf{F}}$ be the $m \times |V^{st}|$ matrix defined by: for each $e \in E$ and $i \in V^{st}$,

$$\widetilde{\mathbf{F}}_{e,i} = \begin{cases} 1 & \text{if } s(e) = i \text{ or } t(e) = i, \\ 0 & \text{otherwise} \end{cases}$$

Finally, let $\widetilde{\boldsymbol{\epsilon}} = \boldsymbol{\epsilon} - \mathbf{F}^t\boldsymbol{\tau}^t - \mathbf{F}^s\boldsymbol{\tau}^s$.

Note that the model can be rewritten as

$$\begin{aligned}
\mathbf{y} &= \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \\
&= \mathbf{B}\boldsymbol{\alpha} + \mathbf{F}^t\boldsymbol{\tau}^t + \mathbf{F}^s\boldsymbol{\tau}^s + \widetilde{\boldsymbol{\epsilon}} \\
&= \mathbf{B}\left(\boldsymbol{\alpha} + \frac{\boldsymbol{\tau}^t}{2} - \frac{\boldsymbol{\tau}^s}{2}\right) + \mathbf{F}\frac{(\boldsymbol{\tau}^t + \boldsymbol{\tau}^s)}{2} + \widetilde{\boldsymbol{\epsilon}},
\end{aligned}$$

where we used the facts that $\mathbf{F}^t = (\mathbf{F} + \mathbf{B})/2$ and $\mathbf{F}^s = (\mathbf{F} - \mathbf{B})/2$. For each $e \in E$,

$$\begin{aligned}
&\left(\mathbf{F}\frac{(\boldsymbol{\tau}^t + \boldsymbol{\tau}^s)}{2}\right)_e \\
&= \begin{cases}
\frac{\tau^t_{t(e)} + \tau^s_{t(e)} + \tau^t_{s(e)} + \tau^s_{s(e)}}{2} & \text{if } s(e), t(e) \in V^{st} \\
\frac{\tau^t_{t(e)} + \tau^s_{t(e)} + \tau^s_{s(e)}}{2} & \text{if } t(e) \in V^{st},\ s(e) \notin V^{st},\ s(e) \in V^s \\
\frac{\tau^t_{t(e)} + \tau^t_{s(e)} + \tau^s_{s(e)}}{2} & \text{if } s(e) \in V^{st},\ t(e) \notin V^{st},\ t(e) \in V^t \\
\frac{\tau^t_{t(e)} + \tau^s_{s(e)}}{2} & \text{if } t(e), s(e) \notin V^{st},\ t(e) \in V^t,\ s(e) \in V^s
\end{cases}
\end{aligned}$$

This can be decomposed as

$$\left(\mathbf{F}\frac{(\boldsymbol{\tau}^t + \boldsymbol{\tau}^s)}{2}\right)_e = \left(\widetilde{\mathbf{F}}\frac{(\widetilde{\boldsymbol{\tau}}^t + \widetilde{\boldsymbol{\tau}}^s)}{2}\right)_e + \left(\mathbf{B}\frac{(\check{\boldsymbol{\tau}}^t - \check{\boldsymbol{\tau}}^s)}{2}\right)_e,$$

where

$$\left(\widetilde{\mathbf{F}}\frac{(\widetilde{\boldsymbol{\tau}}^t+\widetilde{\boldsymbol{\tau}}^s)}{2}\right)_e = \begin{cases} \frac{\tau^t_{t(e)}+\tau^s_{t(e)}+\tau^t_{s(e)}+\tau^s_{s(e)}}{2} & \text{if } s(e), t(e) \in V^{st} \\ \frac{\tau^t_{t(e)}+\tau^s_{t(e)}}{2} & \text{if } t(e) \in V^{st},\ s(e) \notin V^{st} \\ \frac{\tau^t_{s(e)}+\tau^s_{s(e)}}{2} & \text{if } s(e) \in V^{st},\ t(e) \notin V^{st} \\ 0 & \text{if } t(e), s(e) \notin V^{st} \end{cases}$$

$$\left(\mathbf{B}\frac{(\check{\boldsymbol{\tau}}^t-\check{\boldsymbol{\tau}}^s)}{2}\right)_e = \begin{cases} \frac{\tau^t_{t(e)}}{2} & \text{if } t(e) \in V^t \setminus V^s,\ s(e) \in V^{st} \\ \frac{\tau^s_{s(e)}}{2} & \text{if } s(e) \in V^s \setminus V^t,\ t(e) \in V^{st} \\ \frac{\tau^t_{t(e)}+\tau^s_{s(e)}}{2} & \text{if } t(e), s(e) \notin V^{st},\ t(e) \in V^t,\ s(e) \in V^s \\ 0 & \text{if } t(e), s(e) \in V^{st} \end{cases}$$

Therefore,

$$\frac{\mathbf{F}}{2}(\boldsymbol{\tau}^t+\boldsymbol{\tau}^s) = \widetilde{\mathbf{F}}\frac{(\widetilde{\boldsymbol{\tau}}^t+\widetilde{\boldsymbol{\tau}}^s)}{2} + \mathbf{B}\frac{(\check{\boldsymbol{\tau}}^t-\check{\boldsymbol{\tau}}^s)}{2}.$$

Substituting this into the model, we obtain

$$\mathbf{y} = \mathbf{B}\left(\boldsymbol{\alpha} + \frac{\boldsymbol{\tau}^t+\check{\boldsymbol{\tau}}^t}{2} - \frac{\boldsymbol{\tau}^s+\check{\boldsymbol{\tau}}^s}{2}\right) + \widetilde{\mathbf{F}}\frac{(\widetilde{\boldsymbol{\tau}}^t+\widetilde{\boldsymbol{\tau}}^s)}{2} + \widetilde{\boldsymbol{\epsilon}}. \tag{A.1}$$

This is the transformed model where the effects explained by the incidence matrix $\mathbf{B}$ are separated from those not explained by it.

A.2.2 Alternative Estimator

Let $\boldsymbol{M}_{\widetilde{\mathbf{F}}} = \mathbf{I}_m - \widetilde{\mathbf{F}}(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}})^{+}\widetilde{\mathbf{F}}'$, which is the projection matrix onto the nullspace of $\widetilde{\mathbf{F}}$. Premultiplying both sides of equation (A.1) by $\boldsymbol{M}_{\widetilde{\mathbf{F}}}$, we have

$$\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{y} = \boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B}\left(\boldsymbol{\alpha} + \frac{\boldsymbol{\tau}^t + \check{\boldsymbol{\tau}}^t}{2} - \frac{\boldsymbol{\tau}^s + \check{\boldsymbol{\tau}}^s}{2}\right) + \boldsymbol{M}_{\widetilde{\mathbf{F}}}\widetilde{\boldsymbol{\epsilon}}.$$

We can consider the following alternative estimator for $\boldsymbol{\alpha}$ as if $\boldsymbol{\alpha} + \frac{\boldsymbol{\tau}^t + \check{\boldsymbol{\tau}}^t}{2} - \frac{\boldsymbol{\tau}^s + \check{\boldsymbol{\tau}}^s}{2}$ were the true fixed effects:

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{alt} &= (\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})^{*}\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{y} \\ &= \boldsymbol{\alpha} + (\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})^{*}\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B}\left(\frac{\boldsymbol{\tau}^t + \check{\boldsymbol{\tau}}^t}{2} - \frac{\boldsymbol{\tau}^s + \check{\boldsymbol{\tau}}^s}{2}\right) + (\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})^{*}\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\widetilde{\boldsymbol{\epsilon}}.\end{aligned}$$

An implicit assumption here is that $\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B}$ has rank $n-1$ so that $(\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})^{*}$ is well-defined. This requires that $rank((\mathbf{B}, \widetilde{\mathbf{F}})) = n-1+rank(\widetilde{\mathbf{F}})$, which means that the column spaces of $\mathbf{B}$ and $\widetilde{\mathbf{F}}$ do not overlap except for the intercept. This assumption holds if the graph contains enough directed cycles so that a column of $\widetilde{\mathbf{F}}$ cannot be represented as a linear combination of columns of $\mathbf{B}$. This is likely to hold in practice if the graph is dense enough and directions of edges are not too concentrated in one direction.

A.2.3 Approximation

We can derive a first-order approximation for $\widehat{\boldsymbol{\alpha}}^{alt}$ similar to Theorem 1.1. Note that

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{alt}-\boldsymbol{\alpha}=&\frac{\boldsymbol{\tau}^t+\check{\boldsymbol{\tau}}^t}{2}-\frac{\boldsymbol{\tau}^s+\check{\boldsymbol{\tau}}^s}{2}\\&+(\mathbf{I}_n-(\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})\left(\frac{\boldsymbol{\tau}^t+\check{\boldsymbol{\tau}}^t}{2}-\frac{\boldsymbol{\tau}^s+\check{\boldsymbol{\tau}}^s}{2}\right)\\&+(\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\widetilde{\boldsymbol{\epsilon}},\end{aligned}$$

where the first term on the right-hand side is the persistent term, the second term is a misspecification error, and the last term is a linear combination of $\widetilde{\boldsymbol{\epsilon}}$. Each element of the persistent term can be written as

$$\left(\frac{\boldsymbol{\tau}^t+\check{\boldsymbol{\tau}}^t}{2}-\frac{\boldsymbol{\tau}^s+\check{\boldsymbol{\tau}}^s}{2}\right)_i=\begin{cases}\frac{\tau^t(U_i)-\tau^s(U_i)}{2} & \text{if } i\in V^{st}\\ \tau^t(U_i) & \text{if } i\in V^t\setminus V^s\\ -\tau^s(U_i) & \text{if } i\in V^s\setminus V^t\end{cases}$$

Importantly, in the symmetric case where $\tau^t=\tau^s=\tau$, if $i\in V^{st}$, the persistent term is zero. This means that for nodes with both positive in- and out-degrees, the alternative estimator $\widehat{\alpha}_i^{alt}$ can be consistent for α_i, unlike the original estimator $\widehat{\alpha}_i$. This is because the transformation effectively removes the strongly dependent component captured by τ for these nodes. Also, we can view this result as an automatic correction for the persistent term, which is similar to $\widehat{\alpha}_i-\frac{d_i^t-d_i^s}{d_i}\widehat{\tau}_i$ in the original estimator.

For the other two terms, once the difference between $(\mathbf{B}'\boldsymbol{M}_{\widetilde{\mathbf{F}}}\mathbf{B})^*$ and $(\mathbf{B}'\mathbf{B})^*$

is characterized, we can follow similar steps as in the proof of Proposition 1.2 and Theorem 1.1 to derive the approximation with different error rates. Noting that $\widetilde{\boldsymbol{\epsilon}}$ is weakly dependent and following Appendix A.1, we can show that the approximation error rate for the last two terms is given by

$$O_p\left(\sqrt{\frac{1}{\rho d_i}} + \sqrt{\frac{1}{\rho\lambda_{2,L}h_i d_i}} + \sqrt{\frac{\widetilde{\rho}}{m}\overline{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\overline{\boldsymbol{x}}_i}\right),$$

where $\rho, \widetilde{\rho}, \overline{\boldsymbol{x}}_i, \boldsymbol{\Omega}$ are defined in Appendix A.1 with $\mathbf{X}$ replaced by $\widetilde{\mathbf{F}}$. Since $\boldsymbol{M}_{\widetilde{\mathbf{F}}}$ eliminates the strongly dependent component, we do not have $\lambda_{n,F}$ in this rate unlike the rate in Theorem 1.1. However, ρ^{-1} and $\widetilde{\rho}$ may be large depending on the relationship between $\mathbf{B}$ and $\widetilde{\mathbf{F}}$, which may make the approximation error rate larger than that in Theorem 1.1. A more detailed comparison of the two error rates and the choice between the two estimators is left for future research.

A.3 Inference with Asymmetric τ^t and τ^s

In the main text, our analysis with regard to inference and variance estimation focused on the symmetric case where $\tau^t = \tau^s$ is imposed by Assumption 1.5. In this appendix, we discuss potential extensions to the asymmetric case where $\tau^t \neq \tau^s$.

As argued in the main text, the central problem in the asymmetric case is that $\tau^t(U_i)$ and $\tau^s(U_i)$ are not identified separately from the following two approximations:

$$\widehat{\alpha}_i \approx \alpha_i + \frac{d_i^t}{d_i}\tau^t(U_i) - \frac{d_i^s}{d_i}\tau^s(U_i);$$
$$\widehat{\tau}_i \approx \frac{\tau^t(U_i) + \tau^s(U_i)}{2}.$$

In the following, we first discuss possible restrictions on $\tau^t(U_i)$ and $\tau^s(U_i)$ that can help identify them separately, and then we proceed to show that inference can still be conducted under such restrictions.

A.3.1 Restrictions on τ^t and τ^s

One possible restriction is to assume the linear relationship between τ^t and τ^s: for some function τ,

$$\tau^s = \tau, \quad \tau^t = c\tau,$$

where c is an absolute constant. Instead, we can also treat τ^t as a basis function of τ^s. Note also that we do not need to place an intercept restriction on τ^t as $\mathbb{E}[\tau^t(U_i)] = \mathbb{E}[\tau^s(U_i)] = 0$ by Assumption 1.1. Furthermore, $\tau^t = \tau^s$ is a special case of this restriction with $c = 1$.

Under this restriction, the approximations for $\widehat{\alpha}_i$ and $\widehat{\tau}_i$ become:

$$\widehat{\alpha}_i \approx \alpha_i + \frac{cd_i^t - d_i^s}{d_i}\tau(U_i),$$
$$\widehat{\tau}_i \approx \frac{(c+1)}{2}\tau(U_i).$$

Write

$$\widehat{\alpha}_i = \alpha_i + \frac{cd_i^t - d_i^s}{d_i}\tau(U_i) + r_{i,1}$$
$$\widehat{\tau}_i = \frac{c+1}{2}\tau(U_i) + r_{i,2},$$

where $r_{i,1}$ and $r_{i,2}$ are the approximation errors. By the proof of Theorem 1.1, $r_{i,2}$ is $O_p(\sqrt{1/d_i})$. For $r_{i,1}$, it is known that $r_{i,1} = O_p(\sqrt{\lambda_{n,F}/(d_i\lambda_{2,L}h_i)})$ by Theorem 1.1.

We construct the estimator for c in the following manner. First, we regress $\widehat{\alpha}_i$ on $\widehat{\tau}_i$ and obtain the following regression coefficient:

$$\widehat{\beta}_n \equiv \frac{|\mathcal{C}_n|^{-1}\sum_{i\in\mathcal{C}_n}\widehat{\tau}_i\widehat{\alpha}_i}{|\mathcal{C}_n|^{-1}\sum_{i\in\mathcal{C}_n}\widehat{\tau}_i^2}.$$

By Lemma A.3, the denominator converges to $((c+1)/2)^2\mathbb{E}[\tau(U_i)^2]$. The convergence of the numerator is shown as follows:

$$\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\widehat{\tau}_i\widehat{\alpha}_i = \frac{c+1}{2}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\frac{cd_i^t - d_i^s}{d_i}\times\tau(U_i)^2 \tag{A.2}$$
$$+\frac{c+1}{2}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\alpha_i\tau(U_i) \tag{A.3}$$
$$+\frac{c+1}{2}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,1}\tau(U_i) \tag{A.4}$$
$$+\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,2}\alpha_i \tag{A.5}$$
$$+\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,2}\frac{cd_i^t - d_i^s}{d_i}\tau(U_i) \tag{A.6}$$
$$+\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,2}r_{i,1}. \tag{A.7}$$

The first term (A.2) converges to $(c+1)/2 \times \mathbb{E}[\tau(U_i)^2] \lim_{n\to\infty} |\mathcal{C}_n|^{-1} \sum_{i\in\mathcal{C}_n} (cd_i^t - d_i^s)/d_i$ and the second term (A.3) converges to 0 in probability by the law of large numbers as $\mathbb{E}[\tau^2(U_i)] < \infty$, $|(cd_i^t - d_i^s)/d_i| \leq |c| + 1 < \infty$, and $\max_{i\in\mathcal{C}_n} |\alpha_i| = O(1)$. The third term (A.4) converges to 0 in probability since

$$\mathbb{E}[|r_{i,1}\tau(U_i)|] \leq \sqrt{\mathbb{E}[r_{i,1}^2]}\sqrt{\mathbb{E}[\tau(U_i)^2]} = O\left(\sqrt{\frac{\lambda_{n,F}}{d_i\lambda_{2,L}h_i}}\right)$$

by the Cauchy-Schwarz inequality so that

$$\mathbb{E}[(\text{A.4})] = O_p\left(\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\sqrt{\frac{\lambda_{n,F}}{d_i\lambda_{2,L}h_i}}\right) = o(1).$$

Thus, (A.4) $= o_p(1)$ by the Markov inequality. By the similar argument, the fourth to sixth terms (A.5), (A.6), and (A.7) also converge to 0 in probability (if $1/|\mathcal{C}_n| \sum_{i\in\mathcal{C}_n} \sqrt{1/d_i} = o(1)$). Therefore, by the continuous mapping theorem, we have

$$\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n} \widehat{\tau}_i\widehat{\alpha}_i \to_p \frac{c+1}{2}\mathbb{E}[\tau(U_i)^2] \lim_{n\to\infty}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}(cd_i^t - d_i^s)/d_i.$$

Thus, writing $d^t \equiv \lim_{n\to\infty} |\mathcal{C}_n|^{-1} \sum_{i\in\mathcal{C}_n} d_i^t/d_i$ and d^s analogously, we have

$$\widehat{\beta}_n \to_p \frac{2(cd^t - d^s)}{c+1}.$$

Therefore, we can construct a consistent estimator for c as follows:

$$\widehat{c}_n = \frac{2d^s + \widehat{\beta}_n}{2d^t - \widehat{\beta}_n} + O_p\left(\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\sqrt{\frac{\lambda_{n,F}}{d_i\lambda_{2,L}h_i}}\right),$$

as $n \to \infty$.

Hence, by collecting $2(\widehat{c}_n + 1)^{-1}\widehat{\tau}_i \approx \tau(U_i)$, we can construct the consistent estimator for F_τ and conduct inference on $\widehat{\alpha}_i$ similar to the symmetric case.

One could also consider a more general case where τ^t and τ^s are related by a nonlinear function, such as $\tau^t = g(\tau^s)$ for some function g. This generalizes the linear case, since g could be linear, but additional restrictions on the distribution of $\tau(U_i)$ are needed to obtain useful moment conditions for identifying g. For example, suppose g is invertible and its inverse g^{-1} is odd, and $\tau(U_i)$ is symmetrically distributed around 0. Then, defining $\widetilde{\tau}_i = (g(\tau(U_i)) + \tau(U_i))/2$, we have

$$\sum_{i \in \mathcal{C}_n} \mathbb{E}\left[\widetilde{\tau}_i^{2m-1}\left(\alpha_i - \tau(U_i) + \frac{2d_i^t}{d_i} g^{-1}(\widetilde{\tau}_i)\right)\right] = 0$$

for each integer $m \geq 1$, by g^{-1} being an odd function and the symmetry of $\widetilde{\tau}_i$ with respect to $\tau(U_i)$. Estimation of g can proceed by parameterizing g as an odd polynomial function, such as $g(x) = \sum_{k=0}^{K} a_k x^{2k+1}$, and estimating the coefficients a_k using the method of moments up to some finite order K. We do not pursue this direction in this paper, but it is a potential extension of our method.

A.3.2 Inference

Focus on the case where $\tau^s = \tau$ and $\tau^t = c\tau$ for some constant c. We have shown that $\widehat{c}_n$ consistently estimates c. Write $\widehat{\widehat{\tau}}_i = 2(\widehat{c}_n + 1)^{-1}\widehat{\tau}_i$, which is a consistent estimator for $\tau(U_i)$. Then, the distribution F_τ can be estimated by the empirical distribution of $\widehat{\widehat{\tau}}_i$, denoted by $\widehat{F}_\tau$:

$$\widehat{F}_\tau(\cdot) \equiv \frac{1}{|\mathcal{C}_n|} \sum_{i \in \mathcal{C}_n} \mathbb{I}(\widehat{\widehat{\tau}}_i \leq \cdot),$$

where $\mathbb{I}(\cdot)$ is the indicator function.

The uniform consistency of $\widehat{F}_\tau$ to F_τ over a compact subset of $\mathbb{R}$ follows from modifying the proof of Theorem 1.2. First, observe that

$$\frac{2}{\widehat{c}_n + 1} = \frac{2}{c+1} + r_{n,3}, \quad r_{n,3} = O_p\left(\frac{1}{|\mathcal{C}_n|} \sum_{i \in \mathcal{C}_n} \sqrt{\frac{\lambda_{n,F}}{d_i \lambda_{2,L} h_i}} \right),$$

by the delta method so that

$$\begin{aligned} \widehat{\widehat{\tau}}_i &= \frac{2}{\widehat{c}_n + 1} \widehat{\tau}(U_i) \\ &= \tau(U_i) + \frac{2}{c+1} r_{i,2} + \frac{c+1}{2} \tau(U_i) r_{n,3} + r_{i,2} r_{n,3}. \end{aligned}$$

We can bound $\mathbb{E}[|\widehat{\widehat{\tau}}_i - \tau(U_i)|]$ by the rates of convergence of $r_{i,2}$ and $r_{n,3}$. By replacing $\widehat{\tau}_i$ with $\widehat{\widehat{\tau}}_i$ in the proof of Theorem 1.2, we have

$$\sup_{t \in \mathbb{R}} |\widehat{F}_\tau(t) - F_\tau(t)| \to_p 0,$$

as $n \to \infty$, similarly as before but with slower rates of convergence.

With $\widehat{F}_\tau$, we can conduct inference on $\widehat{\alpha}_i$ in the same manner as in the symmetric case. For example, a $(1-\alpha)$-level confidence interval for α_i is given by

$$CI_{i,1-\alpha} = \begin{cases} \left[\widehat{\alpha}_i - \left(\frac{d_i^t \widehat{c}_n - d_i^s}{d_i}\right) \widehat{c}_{1-\alpha/2}, \widehat{\alpha}_i - \left(\frac{d_i^t \widehat{c}_n - d_i^s}{d_i}\right) \widehat{c}_{\alpha/2}\right], & \text{if } d_i^t \widehat{c}_n - d_i^s > 0, \\ \left[\widehat{\alpha}_i - \left(\frac{d_i^t \widehat{c}_n - d_i^s}{d_i}\right) \widehat{c}_{\alpha/2}, \widehat{\alpha}_i - \left(\frac{d_i^t \widehat{c}_n - d_i^s}{d_i}\right) \widehat{c}_{1-\alpha/2}\right], & \text{if } d_i^t \widehat{c}_n - d_i^s < 0, \end{cases}$$

where $\widehat{c}_\alpha$ is the α-quantile of $\widehat{F}_\tau$.

A.4 Bias Correction for Covariance

In this Appendix, we provide a quick discussion on extending the bias correction method for the sample covariance between worker and firm effects based on the two-period AKM model introduced in Example 1.1.

Let W be the set of workers who are employed by firms in V. For simplicity, we focus on the case where workers in W are present in both periods. Let $N = 2|W|$ be the total number of observations in the two periods. Then, the two-period AKM model can be written as

$$\boldsymbol{w} = \mathbf{B}_W \boldsymbol{\phi} + \mathbf{B}_F \boldsymbol{\alpha} + \boldsymbol{u},$$

where $\boldsymbol{w} \in \mathbb{R}^N$ is the vector of workers' log wages in both periods, $\mathbf{B}_W \in \mathbb{R}^{N \times |W|}$ is the incidence matrix for workers, $\mathbf{B}_F \in \mathbb{R}^{N \times |V|}$ is the incidence matrix for firms, $\boldsymbol{\phi} \in \mathbb{R}^{|W|}$ is the vector of worker effects, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is the vector of firm effects, and $\boldsymbol{u} \in \mathbb{R}^N$ is the error term. Given the estimator $\widehat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ obtained by the method in the main text, the least-squares estimator for $\boldsymbol{\phi}$ is given by

$$\widehat{\boldsymbol{\phi}} = (\mathbf{B}_W' \mathbf{B}_W)^{-1} \mathbf{B}_W' (\boldsymbol{w} - \mathbf{B}_F \widehat{\boldsymbol{\alpha}}).$$

The sample covariance between $\widehat{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\alpha}}$ is given by

$$\begin{aligned}\mathrm{C}_{\boldsymbol{\phi},\boldsymbol{\alpha}} &= \frac{1}{N}\sum_{g\in W}\sum_{t=1}^{2}(\phi_g-\overline{\phi})(\alpha_{J(g,t)}-\overline{\alpha})\\ &= \frac{1}{N}\boldsymbol{\alpha}'\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W\boldsymbol{\phi},\end{aligned}$$

where $\boldsymbol{M}_N = \mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N'$ is the demeaning matrix for N observations. Then, the plug-in estimator for $\mathrm{C}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$ is given by

$$\widehat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}} = \frac{1}{N}\widehat{\boldsymbol{\alpha}}'\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W\widehat{\boldsymbol{\phi}}.$$

To derive the bias correction for $\widehat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$, we can follow similar steps as in the main text. Specifically, we focus on the potential leading bias term arising from $\boldsymbol{\tau}$, abstracting from the terms arising from $\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau}$ and $\boldsymbol{u}$ for simplicity. By substituting $\boldsymbol{w} = \mathbf{B}_W\boldsymbol{\phi} + \mathbf{B}_F\boldsymbol{\alpha} + \boldsymbol{u}$ into the expression for $\widehat{\boldsymbol{\phi}}$, we have

$$\begin{aligned}\widehat{\boldsymbol{\phi}} &= \boldsymbol{\phi} - (\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) + (\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\boldsymbol{u}\\ &\approx \boldsymbol{\phi} - (\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}.\end{aligned}$$

Thus, the leading bias term in $\widehat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$ is given by

$$\begin{aligned}&\mathbb{E}[\widehat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}] - \mathrm{C}_{\boldsymbol{\phi},\boldsymbol{\alpha}}\\ &\approx -\frac{1}{N}\mathbb{E}[\boldsymbol{\tau}'\mathbf{F}'\mathbf{B}\mathbf{L}^*\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W(\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}]\\ &= -\frac{\mathbb{E}[\tau(U_i)^2]}{N}\mathrm{tr}\left(\mathbf{L}^*\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W(\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\right).\end{aligned}$$

Therefore, as often pointed out in the literature, the plug-in estimator for the covariance is biased downward. We can estimate this bias term by replacing $\mathbb{E}[\tau(U_i)^2]$ with its consistent estimator $\widehat{\sigma}_\tau^2$:

$$\mathrm{C}_{\boldsymbol{\phi},\boldsymbol{\alpha}}^{bc} = \widehat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}} + \frac{\widehat{\sigma}_\tau^2}{N}\mathrm{tr}\left(\mathbf{L}^*\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W(\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\right).$$

Using the same dataset as in 1.6, we compare the plug-in estimator and the bias-corrected estimator for the covariance between worker and firm effects. Table A.1 reports the results. We find that the bias induced by the dependence is about 30% of the plug-in estimator.

Table A.1: Variance Component Estimation on the VWH Data

$\widehat{\mathbf{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$	$\widehat{\mathbf{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}^{bc}$
0.014	0.019

Note: This table reports the estimated covariance between worker and firm effects via the plug-in estimator $\widehat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$ and the bias-corrected estimator $\widehat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}^{bc}$.

There could be other bias terms arising from $\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau}$ and $\boldsymbol{u}$, which we have abstracted from in this appendix. Establishing the full bias correction and its theoretical properties is left for future research.

A.5 Additional Simulation

In this Appendix, we present additional simulation results to complement the exercises in the main text. Unless otherwise specified, we use the same simulation settings as in Section 1.5.

A.5.1 Inference with non-normal errors

In Section 1.5, the distribution of $\tau(U_i)$ was assumed to be standard normal. Here, we consider the case where $\tau(U_i)$ follows a standard logistic distribution, which has heavier tails than the normal distribution, to verify the validity of the inference procedure to the non-normal case. Specifically, we modify the simulation settings by setting

$$U_i \approx \text{Logistic}(0,1), \quad \tau(U_i) = U_i.$$

The results are reported in Table A.2. As in the normal case, the coverage probabilities of the confidence intervals based on (1.10) are close to the nominal level of 95% for both $K = 1$ and $K = 2$. As before, the normal approximation based on Jochmans and Weidner (2019) is not valid in both cases. This result suggests that our inference procedure is not sensitive to the distribution of $\tau(U_i)$ once the regularity conditions are satisfied.

A.5.2 Testing the joint hypothesis

In the main text, we proposed a Wald-type test for testing $H_0 : \boldsymbol{\alpha}_{V_0} = \boldsymbol{a}$ versus $H_1 : \boldsymbol{\alpha}_{V_0} \neq \boldsymbol{a}$, where $\boldsymbol{a}$ is a vector of constants. Here, we present some preliminary simulation results to check the size and power of the test.

Let $V_0 = \{1, 2, ..., 9, 10\}$ and $\boldsymbol{a} = (0, 0, ..., 0)$. We then generate the data as before except that we set $\alpha_{V_0 \setminus \{1\}} = 0$ and $\alpha_1 \in \{0, 1, 2, 3\}$, corresponding to the null hypothesis and the alternative hypothesis. When computing the test statistic (1.11) and simulate the critical value, we drop the nodes in V_0 if $(d_i^t + d_i^s)/d_i \leq 0.4$ to avoid the cases where the first-order approximation is not

Table A.2: Coverage Probability of Confidence Intervals

	α_1	d_1	$(d_1^t - d_1^s)/d_1$	H_1	**95%**	**Normal 95%**
Panel A: $K = 1$						
$n = 500$	-0.596	119	-0.697	0.02	0.949	0.333
$n = 1000$	0.314	143	-0.986	0.016	0.952	0.2015
$n = 2500$	0.281	161	-0.652	0.014	0.944	0.2905
$n = 5000$	0.539	159	-0.635	0.016	0.94	0.3145
Panel B: $K = 2$						
$n = 500$	-0.594	49	-0.878	3.886	0.942	0.3975
$n = 1000$	0.312	67	-1.0	3.966	0.956	0.3
$n = 2500$	0.278	72	-0.583	4.407	0.944	0.4645
$n = 5000$	0.542	93	-0.613	3.721	0.96	0.418

Panel A reports the results for $K = 1$ and Panel B reports the results for $K = 2$. The first column reports the number of nodes n, the second column reports the true value of α_1, the third and fourth columns report node 1's degree and the coefficient, respectively. The fifth column reports the convergence measure $H_1 = \lambda_{n,F}/(d_1 \lambda_{2,L} h_1)$. The sixth column reports the coverage probability of the confidence interval for α_1 at the 95% level based on (1.10). The last column reports the coverage probability of the confidence intervals based on Jochmans and Weidner (2019)'s asymptotic normality.

valid. If the threshold is closer to 1, the test is likely to be less powerful, but the size is more likely to be controlled.

Table A.3 reports the rejection rates of the joint hypothesis test at the 95% significance level. The results show that the size of the test is controlled at the nominal level of 5% for both $K = 1$ and $K = 2$. As expected, the power of the test increases as α_1 increases. The power tends to be higher for $K = 1$ than for $K = 2$, likely due to the fact that the first-order approximation is more accurate for more well-connected networks. However, there is non-monotonicity in the power of the test as n increases, possibly because the change in the network structure affects the values of d_i^t and d_i^s, which in turn affects which nodes are included in the test and its performance.

Table A.3: Size and Power of the Joint Hypothesis Test

	$H_0 : \alpha_1 = 0$	$H_1 : \alpha_1 = 1$	$H_1 : \alpha_1 = 2$	$H_1 : \alpha_1 = 3$
		Panel A: $K = 1$		
$n = 500$	0.049	0.18	0.626	0.947
$n = 1000$	0.054	0.13	0.429	0.819
$n = 2500$	0.04	0.161	0.58	0.93
$n = 5000$	0.048	0.164	0.57	0.903
		Panel B: $K = 2$		
$n = 500$	0.051	0.138	0.471	0.852
$n = 1000$	0.045	0.143	0.429	0.768
$n = 2500$	0.041	0.161	0.551	0.912
$n = 5000$	0.04	0.169	0.617	0.945

Note: Panel A reports the results for $K = 1$ and Panel B reports the results for $K = 2$. Each column reports the rejection rate of the joint hypothesis test at the 95% significance level. In each case, $\boldsymbol{\alpha}_{V_0 \setminus \{1\}} = 0$ but $\alpha_1 = 0$ for the first column, and $\alpha_1 = 1, 2, 3$ for the second to fourth columns.

A.6 Proofs

A.6.1 Proof of Theorem 1.1

Proof. We bound each term in the right-hand side of (1.8):

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \mathbf{D}^{-1}\mathbf{B}'(\mathbf{F}^s\boldsymbol{\tau}^s + \mathbf{F}^t\boldsymbol{\tau}^t) + \mathbf{D}^{-1}\mathbf{B}'(\boldsymbol{\epsilon} - \mathbf{F}^s\boldsymbol{\tau}^s - \mathbf{F}^t\boldsymbol{\tau}^t) + \boldsymbol{r}$$

We first show the unbiasedness of the fixed-effect estimator:

Lemma A.1. *Under the stated assumptions, we have*

$$\mathbb{E}[\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}] = 0.$$

Proof. Observe that

$$\widehat{\boldsymbol{\alpha}} = \mathbf{L}^*\mathbf{L}\boldsymbol{\alpha} + \mathbf{L}^*\mathbf{B}'\boldsymbol{\epsilon}.$$

Since $\mathbb{E}[\boldsymbol{\epsilon}] = 0$, it suffices to show that $\mathbf{L}^*\mathbf{L}\boldsymbol{\alpha} = \boldsymbol{\alpha}$. By the proof of Theorem 2 and S.1 in Jochmans and Weidner (2019), we have

$$\mathbf{L}^*\mathbf{L} = \mathbf{I}_n - m^{-1}\boldsymbol{\iota}_n\mathbf{d}'.$$

Since $\mathbf{d}'\boldsymbol{\alpha} = 0$ by Assumption 1.1, we have

$$\mathbf{L}^*\mathbf{L}\boldsymbol{\alpha} = \mathbf{I}_n\boldsymbol{\alpha} - m^{-1}\boldsymbol{\iota}_n\mathbf{d}'\boldsymbol{\alpha} = \boldsymbol{\alpha}.$$

This proves the lemma. □

Another useful lemma is the following:

Lemma A.2. *Under the stated assumptions, $C\lambda_{n,F} \geq \widetilde{\sigma}_n^2$ for some absolute constant $C > 0$.*

Proof. Note that $\mathbf{FF}'$ is a non-negative, positive-semidefinite matrix and each block corresponding to edges in $E_{(i,j)}$, $(\mathbf{FF}')_{(i,j)}$ is a principal submatrix of $\mathbf{FF}'$. Thus, by Corollary 8.1.20 in Horn and Johnson (2012), we have

$$\lambda_{n,F} \geq \lambda_{\max}((\mathbf{FF}')_{(i,j)}),$$

for each (i, j) and $\lambda_{\max}(\cdot)$ is the largest eigenvalue of a matrix. Notice that

$$(\mathbf{FF}')_{(i,j)} = 2 \times \boldsymbol{\iota}_{|E_{(i,j)}|}\boldsymbol{\iota}'_{|E_{(i,j)}|}$$

so that $\lambda_{\max}((\mathbf{FF'})_{(i,j)}) = 2 \times |E_{(i,j)}|$. Therefore,

$$\lambda_{n,F} \geq 2 \max_{(i,j)} |E_{(i,j)}|.$$

By Assumption 1.1, we have

$$\begin{aligned}\widetilde{\sigma}_n^2 &= \max_{(i,j)} \mathbf{A}_{i,j} \times [\max_{e,e'} \in E_{(i,j)} \sigma_{e,e'} - \mathbb{E}[\tau^s(U_i)^2 + \tau^t(U_i)^2]] \\ &\leq C \max_{(i,j)} \mathbf{A}_{i,j} = C \max_{(i,j)} |E_{(i,j)}|,\end{aligned}$$

for some absolute constant $C > 0$. Thus, we have

$$\widetilde{\sigma}_n^2 \leq C/2 \times 2 \max_{(i,j)} |E_{(i,j)}| \leq C/2 \times \lambda_{n,F}.$$

This proves the lemma. □

Next, we bound the variance of the remaining term $\boldsymbol{r} = \mathbf{D}^{-1}\mathbf{A}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$. Observe that

$$Var(\widehat{\boldsymbol{\alpha}}) \leq C\lambda_{n,F}\mathbf{L}^*$$

for some constant $C > 0$ by Lemma A.2. Thus, we have

$$\begin{aligned}Var(\boldsymbol{r}) &= \mathbf{D}^{-1}\mathbf{A}Var(\widehat{\boldsymbol{\alpha}})\mathbf{A}'\mathbf{D}^{-1} \\ &\leq C\lambda_{n,F}\mathbf{D}^{-1}\mathbf{A}\mathbf{L}^*\mathbf{A}\mathbf{D}^{-1} \\ &\leq \frac{C\lambda_{n,F}}{\lambda_{2,L}} \times \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1},\end{aligned}$$

where the second inequality follows from $\mathbf{L}^* \leq \mathbf{D}^{-1}/\lambda_{2,L}$. Thus, for each element, we have

$$\begin{aligned} Var(r_i) = e_i' Var(\boldsymbol{r}) e_i &\leq \frac{C\lambda_{n,F}}{\lambda_{2,L}} \times e_i' \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1} e_i. \\ &= \frac{C\lambda_{n,F}}{d_i \lambda_{2,L} h_i}. \end{aligned}$$

Since $\mathbb{E}[r_i] = 0$ by Lemma (A.1), we have

$$r_i = O_p\left(\sqrt{\frac{\lambda_{n,F}}{d_i \lambda_{2,L} h_i}}\right).$$

by Chebyshev's inequality.

Third, we bound the other error term $\boldsymbol{o} \equiv \mathbf{D}^{-1}\mathbf{B}'(\boldsymbol{\epsilon} - \mathbf{F}^t \tau^t - \mathbf{F}^s \boldsymbol{\tau}^s)$. This term is mean-zero and its variance is given by:

$$\begin{aligned} Var(\boldsymbol{o}) &= \mathbf{D}^{-1}\mathbf{B}'\boldsymbol{\Omega}_2\mathbf{B}\mathbf{D}^{-1} \\ &\leq \widetilde{\sigma}_n^2 \times \mathbf{D}^{-1}\mathbf{B}'\mathbf{B}\mathbf{D}^{-1} \end{aligned}$$

Thus, for each element, we have

$$\begin{aligned} Var(o_i) = e_i' Var(\boldsymbol{o}) e_i &= \widetilde{\sigma}_n^2 \times e_i' \mathbf{D}^{-1}\mathbf{B}'\mathbf{B}\mathbf{D}^{-1} e_i \\ &\leq \frac{C \max_{(i,j)} |E_{(i,j)}|}{d_i}. \end{aligned}$$

for some constant $C > 0$ under Assumption 1.1. Thus, we have

$$o_i = O_p\left(\sqrt{\frac{\max_{(i,j)} |E_{(i,j)}|}{d_i}}\right) = o_p(1),$$

by Chebyshev's inequality as $d_i \to \infty$.

Finally, we work on the main term $\boldsymbol{T} \equiv \mathbf{D}^{-1}\mathbf{B}'(\mathbf{F}^t\boldsymbol{\tau}^t + \mathbf{F}^s\boldsymbol{\tau}^s)$. Observe that

$$\boldsymbol{T} = \left(\frac{d_i^t \tau^t(U_i) - d_i^s \tau^s(U_i)}{d_i} + \frac{1}{d_i} \sum_{e \in E_i^t} \tau^s(U_{s(e)}) - \frac{1}{d_i} \sum_{e \in E_i^s} \tau^t(U_{t(e)}) \right)_{i \in V},$$

as pointed out in the main text. Thus, it suffices to show that

$$\frac{1}{d_i} \sum_{e \in E_i^t} \tau^s(U_{s(e)}) = o_p(1), \quad \frac{1}{d_i} \sum_{e \in E_i^s} \tau^t(U_{t(e)}) = o_p(1).$$

These averages are mean-zero and each variance is given by:

$$\frac{d_i^t}{d_i^2} \mathbb{E}[\tau^s(U_j)^2] = o(1), \quad \frac{d_i^s}{d_i^2} \mathbb{E}[\tau^t(U_j)^2] = o(1),$$

as $\tau^s(U_j)$ and $\tau^t(U_j)$ are i.i.d across $j \in V_i$. Thus, the averages converge to zero in probability by Chebyshev's inequality.

By the continuous mapping theorem, we have

$$\widehat{\alpha}_i - \alpha_i = \frac{d_i^t \tau^t(U_i) - d_i^s \tau^s(U_i)}{d_i} + O_p\left(\sqrt{\frac{\lambda_{n,F}}{d_i \lambda_{2,L} h_i}} \right),$$

as $d_i \to \infty$, which completes the proof. □

A.6.2 Proof of Theorem 1.2

Proof. Write $\widehat{\tau}_i = \tau(U_i) + r_i$ where $r_i = O_p(\sqrt{1/d_i})$ by Proposition 1.2 for $i \in \mathcal{C}_n$. Let F_n be the empirical distribution of $\tau(U_i)$ for $i \in \mathcal{C}_n$. Then, observe that

$$\sup_{x\in\mathbb{R}}|\widehat{F}_{n,\tau}(x) - F_\tau(x)| \leq \sup_{x\in\mathbb{R}}|\widehat{F}_{n,\tau}(x) - F_{n,\tau}(x)| + \sup_{x\in\mathbb{R}}|F_{n,\tau}(x) - F_\tau(x)| \quad \text{(A.8)}$$

The second term in (A.8) converges to zero in probability by the Glivenko-Cantelli theorem since $\tau(U_i)$ is i.i.d across i and $|\mathcal{C}_n| \to \infty$.

For the first term in (A.8), note that for any $\delta > 0$,

$$\mathbb{I}\{\widehat{\tau}_i \leq x\} = \mathbb{I}\{\tau(U_i) + r_i \leq x\} \leq \mathbb{I}\{\tau(U_i) \leq x + \delta\} + \mathbb{I}\{|r_i| > \delta\},$$

for each $i \in \mathcal{C}_n$ and $x \in \mathbb{R}$. Since F_τ is non-decreasing, we have

$$\begin{aligned} &\sup_{x\in\mathbb{R}}|\widehat{F}_{n,\tau}(x) - F_{n,\tau}(x)| \\ &\leq \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\mathbb{I}\{|r_i| > \delta\} + \sup_{x\in\mathbb{R}}|F_{n,\tau}(x+\delta) - F_{n,\tau}(x-\delta)| \end{aligned} \quad \text{(A.9)}$$

The second term in (A.9) is bounded by

$$\begin{aligned} &\sup_{x\in\mathbb{R}}|F_{n,\tau}(x+\delta) - F_{n,\tau}(x-\delta)| \\ &\leq \sup_{x\in\mathbb{R}}|F_\tau(x+\delta) - F_\tau(x-\delta)| + 2\sup_{x\in\mathbb{R}}|F_{n,\tau}(x) - F_\tau(x)|. \end{aligned}$$

Since F_τ is continuous, the first term is arbitrarily small for sufficiently small δ. For the first term in (A.9), note that

$$\mathbb{E}[\mathbb{I}\{|r_i| > \delta\}] = \mathbb{P}(|r_i| > \delta) \leq \frac{\mathbb{E}[|r_i|]}{\delta} = O\left(\frac{1}{\sqrt{d_i}\delta}\right).$$

Therefore, by Markov's inequality, for any $\epsilon > 0$, we have

$$\mathbb{P}\left(\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\mathbb{I}\{|r_i|>\delta\}>\epsilon\right)\le\frac{1}{\epsilon|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\mathbb{E}[\mathbb{I}\{|r_i|>\delta\}]=O\left(\frac{\sqrt{\eta_n}}{\epsilon\delta}\right)=o(1),$$

as ϵ and δ are arbitrary and $\eta_n \to 0$ by the assumption.

Combining the above, we have,

$$\begin{aligned}&\sup_{x\in\mathbb{R}}|\widehat{F}_{n,\tau}(x)-F_\tau(x)|\\&\le 3\sup_{x\in\mathbb{R}}|F_{n,\tau}(x)-F_\tau(x)|+\sup_{x\in\mathbb{R}}|F_\tau(x+\delta)-F_\tau(x-\delta)|+\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\mathbb{I}\{|r_i|>\delta\}\\&=o_p(1),\end{aligned}$$

as $n \to \infty$. This completes the proof. □

A.6.3 Proof of Theorem 1.3

Proof. First, we show the consistency of $\widehat{\sigma}_\tau^2$:

Lemma A.3. *Under the stated assumptions, we have*

$$\widehat{\sigma}_\tau^2 \to_p \sigma_\tau^2.$$

Proof. Observe that

$$\widehat{\tau}_i = \tau(U_i) + \underbrace{c_i^{-1}\left(\frac{1}{d_i}\sum_{e\in E_i}\widehat{\epsilon}_e - c_i\tau(U_i)\right)}_{\equiv \widetilde{u}_i}.$$

Then,

$$\widehat{\sigma}_\tau^2 = \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tau^2(U_i) \tag{A.10}$$

$$+ \frac{2}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tau(U_i)\widetilde{u}_i \tag{A.11}$$

$$+ \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\widetilde{u}_i^2. \tag{A.12}$$

The first term (A.10) converges to $\mathbb{E}[\tau^2(U_i)]$ by the law of large numbers as $\tau(U_i)$ is i.i.d across $i \in \mathcal{C}_n$ with a finite second moment.

The third term (A.12) converges to zero in probability. To see this, from the proof of Proposition 1.2, observe that

$$\mathbb{E}[\widetilde{u}_i^2] = O\left(\frac{1}{d_i}\right),$$

since $c_i > c$. Thus, we have

$$\mathbb{E}[(\text{A.12})] = O(\eta_n) = o(1),$$

under the hypothesis that $\eta_n \to 0$ as $n \to \infty$. Then, by Markov's inequality, (A.12) converges to zero in probability.

The second term (A.11) converges to zero in probability by the Cauchy-

Schwarz inequality:

$$(\text{A.11}) \leq \sqrt{\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tau^2(U_i)}\sqrt{\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\widetilde{u}_i^2} = O_p(\sqrt{\eta_n}) = o_p(1),$$

as shown above. This completes the proof of Lemma A.3. □

Observe that

$$\widehat{V}_{\boldsymbol{\alpha}}^{bc} - V_{\boldsymbol{\alpha}} = 2\frac{\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}}{n} \tag{A.13}$$

$$+ 2\frac{(\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n} \tag{A.14}$$

$$+ \frac{(\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'(\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})}{n} \tag{A.15}$$

$$+ \frac{(\mathbb{E}[\tau^2(U_i)] - \widehat{\sigma}_\tau^2)\times \operatorname{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} \tag{A.16}$$

$$+ \frac{\boldsymbol{\tau}'\mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n} - \frac{\mathbb{E}[\tau^2(U_i)]\times \operatorname{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} \tag{A.17}$$

The first term (A.13) is mean-zero and its variance is proportional to

$$\begin{aligned}
Var\left(\frac{\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}}{n}\right) &= \frac{1}{n^2}\mathbb{E}[\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\boldsymbol{\epsilon}] \\
&= \frac{1}{n^2}\operatorname{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']) \\
&\leq \frac{C\lambda_{n,F}}{n^2}\times\boldsymbol{\alpha}'\mathbf{L}^*\boldsymbol{\alpha} \\
&\leq \frac{C\lambda_{n,F}}{n}\times\frac{1}{n\lambda_{2,L}}\sum_{i=1}^n\frac{\alpha_i^2}{d_i} \\
&\leq \frac{C\times C_\alpha\times\lambda_{n,F}}{n}\times\frac{1}{n\lambda_{2,L}}\sum_{i=1}^n\frac{1}{d_i} \\
&= o(1),
\end{aligned}$$

where the first inequality follows from Lemma A.2 for some constant $C > 0$, the second inequality follows from $\mathbf{L}^* \leq \lambda_2^{-1}\mathbf{D}^{-1}$ and $\boldsymbol{\alpha}'\mathbf{L}^*\boldsymbol{\alpha} \leq \sum_{i=1}^n \frac{\alpha_i^2}{\lambda_{2,L} d_i}$, and the third inequality follows from the assumption that $\max_{i\in V}|\alpha_i| \leq C_\alpha$ for some absolute constant $C_\alpha > 0$. Thus,

$$(\text{A.13}) = o_p(1),$$

by Chebyshev's inequality.

The second term (A.14) is mean-zero and let $\boldsymbol{Q} \equiv \mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'$. Then, we can write

$$(\text{A.14}) = 2\frac{(\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})'\boldsymbol{Q}\mathbf{F}\boldsymbol{\tau}}{n}$$

and its variance is given by

$$\begin{aligned}
&Var((\text{A.14}))\\
&= \frac{4}{n^2}\mathbb{E}[(\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})'\boldsymbol{Q}\mathbf{F}\boldsymbol{\tau}(\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})'\boldsymbol{Q}\mathbf{F}\boldsymbol{\tau}]\\
&= \frac{4\mathbb{E}[\tau^2(U_i)]}{n^2}\text{tr}(\boldsymbol{Q}\mathbf{F}\mathbf{F}'\boldsymbol{Q}'\boldsymbol{\Omega}_2)\\
&+ \frac{4}{n^2}\sum_{e_1,e_2,e_3,e_4\in E}\boldsymbol{Q}_{e_1,e_2}\boldsymbol{Q}_{e_3,e_4}\text{cum}((\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})_{e_1}, (\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})_{e_2}, (\mathbf{F}\boldsymbol{\tau})_{e_3}, (\mathbf{F}\boldsymbol{\tau})_{e_4}),
\end{aligned}$$

where

$$cum(x_1, x_2, x_3, x_4) = \mathbb{E}[x_1x_2x_3x_4] - (\mathbb{E}[x_1x_2] + \mathbb{E}[x_1x_3] + \mathbb{E}[x_1x_4]).$$

The first term on the right-hand side is bounded by

$$\begin{aligned}
&\frac{\mathbb{E}[\tau^2(U_i)]}{n^2}\mathrm{tr}(\boldsymbol{Q}\mathbf{F}\mathbf{F}'\boldsymbol{Q}'\boldsymbol{\Omega}_2) \\
&\le \frac{\mathbb{E}[\tau^2(U_i)]}{n^2} \times \widetilde{\sigma}_n^2 \times \lambda_{n,F} \times \mathrm{tr}(\mathbf{L}^*)^2 \\
&\le \mathbb{E}[\tau^2(U_i)] \times \widetilde{\sigma}_n^2 \times \lambda_{n,F} \times \left(\frac{1}{n}\sum_{i\in V} d_i^{-1} + \frac{1}{n\lambda_{2,L}}\sum_{i\in V} d_i^{-1}h_i^{-1}\right)^2 \\
&= o(1),
\end{aligned}$$

where the first inequality follows from $\boldsymbol{\Omega}_2 \le \widetilde{\sigma}_n \mathbf{I}_m$, $\mathbf{F}\mathbf{F}' \le \lambda_n^2 \mathbf{I}_m$, and $\mathrm{tr}(\boldsymbol{Q}\boldsymbol{Q}') = \mathrm{tr}(\mathbf{L}^*\mathbf{L}^*) \le \mathrm{tr}(\mathbf{L}^*)^2$, the second inequality follows from $e_i'\mathbf{L}^* e_i \le 1/d_i + 1/(d_i\lambda_{2,L}h_i)$ for each $i \in V$, and the last equality follows from the hypothesis. The second term on the right-hand side is bounded by

$$\begin{aligned}
&\frac{1}{n^2}\sum_{e_1,e_2,e_3,e_4\in E} \boldsymbol{Q}_{e_1,e_2}\boldsymbol{Q}_{e_3,e_4} \times \mathrm{cum}((\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})_{e_1}, (\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})_{e_2}, (\mathbf{F}\boldsymbol{\tau})_{e_3}, (\mathbf{F}\boldsymbol{\tau})_{e_4}) \\
&\le \frac{C\left(\sum_{e_1,e_2}\boldsymbol{Q}_{e_1,e_2}\right)^2}{n^2} \\
&= \frac{C \times \mathrm{tr}(\boldsymbol{Q}^2)}{n^2} \\
&\le \frac{C \times \mathrm{tr}(\mathbf{L}^*)^2}{n^2} \\
&\le C \times \left(\frac{1}{n}\sum_{i\in V} d_i^{-1} + \frac{1}{n\lambda_{2,L}}\sum_{i\in V} d_i^{-1}h_i^{-1}\right)^2 \\
&= o(1),
\end{aligned}$$

where the first inequality follows from Assumption 1.1 with uniformly bounded ϵ_e, the equality follows from the fact that the square of the Frobenius norm is equal to the trace of the square of the matrix, and the rest is similar to the first term. Thus, we have

$$Var((\text{A.14})) = o(1),$$

and by Chebyshev's inequality, we have (A.14) $\to_p 0$.

The third term (A.15)'s mean is given by

$$\begin{aligned}
\mathbb{E}[(\text{A.15})] &= \frac{\text{tr}(\mathbf{BL}^* \boldsymbol{M}_n \mathbf{L}^* \mathbf{B}' \boldsymbol{\Omega}_2)}{n} \\
&\leq \frac{\widetilde{\sigma}_n^2 \times \text{tr}(\mathbf{L}^*)}{n} \\
&\leq \frac{\widetilde{\sigma}_n^2}{n} \times \left(\sum_{i \in V} \frac{1}{d_i} + \sum_{i \in V} \frac{1}{\lambda_{2,L} h_i d_i} \right) \\
&= o(1),
\end{aligned}$$

where the second inequality follows from $e_i' \mathbf{L}^* e_i \leq 1/d_i + 1/(d_i \lambda_{2,L} h_i)$ for each $i \in V$. Since (A.15) is non-negative ($\mathbf{BL}^* \boldsymbol{M}_n \mathbf{L}^* \mathbf{B}$ is positive semidefinite), this implies that (A.15) $\to_p 0$.

The fourth term (A.16) converges to zero in probability by Lemma A.3 and the fact that

$$\begin{aligned}
\frac{\text{tr}(\mathbf{BL}^* \boldsymbol{M}_n \mathbf{L}^* \mathbf{B}' \mathbf{FF}')}{n} &\leq \frac{C \lambda_{n,F}}{n} \times \text{tr}(\mathbf{L}^*) \\
&\leq \frac{C \lambda_{n,F}}{n} \times \left(\sum_{i \in V} \frac{1}{d_i} + \sum_{i \in V} \frac{1}{\lambda_{2,L} h_i d_i} \right) \\
&= O(1),
\end{aligned}$$

where the second inequality follows from $e_i' \mathbf{L}^* e_i \leq 1/d_i + 1/(d_i \lambda_{2,L} h_i)$ for each

$i \in V$.

The fifth term (A.17) is mean-zero, and letting $\boldsymbol{Q}_{\mathbf{F}} \equiv \mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}$, its variance is given by

$$
\begin{aligned}
&Var\left(\frac{\boldsymbol{\tau}'\mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n}\right)\\
&= \frac{\mathbb{E}[\tau^2(U_i)]}{n^2} \times \mathrm{tr}(\boldsymbol{Q}_{\mathbf{F}}^2) + \frac{(\mathbb{E}[\tau^4(U_i)] - 3\mathbb{E}[\tau^2(U_i)]^2)}{n^2} \times \sum_{i\in V} \boldsymbol{Q}_{\mathbf{F},i,i}^2\\
&\leq \frac{\mathbb{E}[\tau^4(U_i)]}{n^2} \times \mathrm{tr}(\boldsymbol{Q}_{\mathbf{F}}^2),
\end{aligned}
$$

where the inequality follows as $\mathrm{tr}(\boldsymbol{Q}_{\mathbf{F}}^2) \geq \sum_{i\in V} \boldsymbol{Q}_{\mathbf{F},i,i}^2$. We bound $\mathrm{tr}(\boldsymbol{Q}_{\mathbf{F}}^2)$ as follows:

$$
\begin{aligned}
&\mathrm{tr}(\boldsymbol{Q}_{\mathbf{F}}^2)\\
&= \mathrm{tr}(\mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F})\\
&\leq \mathrm{tr}(\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{L}^*\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{L}^*)\\
&\leq 2\mathrm{tr}(\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1})\\
&+ 2\mathrm{tr}(\mathbf{D}^{-1}\mathbf{A}\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{L}^*\mathbf{A}\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{A}\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{L}^*\mathbf{A}\mathbf{D}^{-1})\\
&\leq 2\mathrm{tr}(\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1})\\
&+ \frac{2\lambda_{n,F}^2}{\lambda_{2,L}^2}\mathrm{tr}(\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1})\\
&\leq 2\mathrm{tr}(\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}) + 2\left(\frac{\lambda_{n,F}}{\lambda_{2,L}}\mathrm{tr}(\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1})\right)^2\\
&= 2\mathrm{tr}(\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}) + 2\left(\frac{\lambda_{n,F}}{\lambda_{2,L}}\sum_{i\in V}\frac{1}{d_i h_i}\right)^2
\end{aligned}
$$

where the first inequality follows from the fact that $\mathbf{I}_m - \boldsymbol{M}_n$ is positive semidefinite, the second inequality follows from $\mathbf{BL}^* = \mathbf{BD}^{-1} + \mathbf{BD}^{-1}\mathbf{AL}^*$ and the combination of the Cauchy-Schwarz inequality and the triangle inequality, the third inequality follows from $\mathbf{L}^* \leq \lambda_{2,L}^{-1}\mathbf{D}^{-1}$ and $\mathbf{FF}' \leq \lambda_{n,F}\mathbf{I}_m$, and the last inequality follows from $\mathrm{tr}(\cdot^2) \leq \mathrm{tr}(\cdot)^2$ applied to $\mathbf{D}^{-1}\mathbf{AD}^{-1}\mathbf{AD}^{-1} = (\mathbf{D}^{-1}\mathbf{AD}^{-1/2})(\mathbf{D}^{-1}\mathbf{AD}^{-1/2})'$ which is positive semidefinite. For the first term on the far right-hand side, we have

$$\begin{aligned}
&\mathrm{tr}(\mathbf{D}^{-1}\mathbf{B}'\mathbf{FF}'\mathbf{BD}^{-1}\mathbf{D}^{-1}\mathbf{B}'\mathbf{FF}'\mathbf{BD}^{-1}) \\
&= \sum_{i\in V}\left(\frac{(d_i^t - d_i^s)^2 + \sum_{j\neq i} r_{i,j}^2}{d_i^2}\right)^2 + \sum_{i\neq j}\left(\frac{\sum_{k\in V} r_{i,k}r_{j,k}}{d_i d_j}\right)^2 \\
&\leq Cn(1 + \sum_{i\in V} d_i^{-2})
\end{aligned}$$

for some constant $C > 0$, where $r_{i,j} = \sum_{e\in E_{(i,j)}} \mathbf{B}_{e,i}$ and the inequality follows from $\sum_{j\neq i} r_{i,j}^2 \leq C_A d_i$, $\sum_{k\in V} r_{i,k}r_{j,k} \leq C_A^2 d_i$. Thus,

$$\begin{aligned}
&Var\left(\frac{\boldsymbol{\tau}'\mathbf{F}'\mathbf{BL}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n}\right) \\
&\leq C\left(\frac{1}{n} + \frac{1}{n}\sum_{i\in V} d_i^{-2}\right) + C\left(\frac{\lambda_{n,F}}{\lambda_{2,L}}\frac{1}{n}\sum_{i\in V}\frac{1}{d_i h_i}\right)^2 \\
&= o(1),
\end{aligned}$$

under the stated assumptions. Thus, by Chebyshev's inequality, we have (A.17) $\to_p 0$.

Finally, we combine all the results above to conclude that

$$\widehat{V}_{\boldsymbol{\alpha}}^{bc} - V_{\boldsymbol{\alpha}} \to_p 0.$$

This completes the proof. □

A.6.4 Proof of Proposition 1.1

Proof. First, observe that

$$\mathbf{\Omega}_1 \leq \sigma_\tau^2 \mathbf{F}\mathbf{F}' \leq \sigma_\tau^2 \lambda_{n,F} \mathbf{I}_m,$$

where the first inequality holds by $\mathbf{F} = \mathbf{F}^t + \mathbf{F}^s$. Then, the first term in the variance formula (1.7) is bounded by:

$$\mathbf{L}^* \mathbf{B}' \mathbf{\Omega}_1 \mathbf{B} \mathbf{L}^* \leq \sigma_\tau^2 \lambda_{n,F} \mathbf{L}^*,$$

where we used the fact that $\mathbf{L}^*$ is a pseudo-inverse of $\mathbf{B}'\mathbf{B}$.

Next, since $\mathbf{\Omega}_2 \leq \widetilde{\sigma}_n^2 \mathbf{I}_m$, the second term in the variance formula (1.7) is bounded by:

$$\mathbf{L}^* \mathbf{B}' \mathbf{\Omega}_2 \mathbf{B} \mathbf{L}^* \leq \widetilde{\sigma}_n^2 \mathbf{L}^*.$$

Therefore, the variance of the fixed effect is bounded by:

$$Var(\widehat{\boldsymbol{\alpha}}) \leq (\widetilde{\sigma}_n^2 + \sigma_\tau^2 \lambda_{n,F}) \times \mathbf{L}^*.$$

Thus, for a unit vector e_i with 1 at the i-th position and 0 elsewhere, we have:

$$Var(\widehat{\alpha}_i) = e_i' Var(\widehat{\boldsymbol{\alpha}}) e_i \leq (\widetilde{\sigma}_n^2 + \sigma_\tau^2 \lambda_{n,F}) \times e_i' \mathbf{L}^* e_i.$$

By the proof of Theorem 2 in Jochmans and Weidner (2019), we have

$$e_i'\mathbf{L}^* e_i \leq \frac{1}{d_i}\left(1+\frac{1}{\lambda_{2,L}h_i}\right) - \frac{2}{\sum_{j\in V} d_j}.$$

Thus,

$$Var(\widehat{\alpha}_i) \leq (\widetilde{\sigma}_n^2 + \sigma_\tau^2 \lambda_{n,F}) \times \left\{\frac{1}{d_i}\left(1+\frac{1}{\lambda_{2,L}h_i}\right) - \frac{2}{\sum_{j\in V} d_j}\right\},$$

which completes the proof. □

A.6.5 Proof of Proposition 1.2

Proof. Observe that from the transformed model (A.1), we can write the residuals as

$$\begin{aligned}\widehat{\boldsymbol{\epsilon}} &= \mathbf{y} - \mathbf{B}\widehat{\boldsymbol{\alpha}} = (\mathbf{I}_m - \mathbf{B}(\mathbf{B}'\mathbf{B})^*\mathbf{B}')\mathbf{y} = \boldsymbol{M}_{\mathbf{B}}\mathbf{y} \\ &= \boldsymbol{M}_{\mathbf{B}}\widetilde{\mathbf{F}}(\widetilde{\boldsymbol{\tau}}^t/2 + \widetilde{\boldsymbol{\tau}}^s/2) + \boldsymbol{M}_{\mathbf{B}}\widetilde{\boldsymbol{\epsilon}}.\end{aligned}$$

Recall that the statistic of interest minus $c_i(\tau^t(U_i)+\tau^s(U_i))/2$ can be written as

$$\begin{aligned}&\frac{1}{d_i}\sum_{e\in E_i}\widehat{\epsilon}_e - c_i\frac{\tau^t(U_i)+\tau^s(U_i)}{2} \\ &= \frac{1}{d_i}1_{i,m}'\boldsymbol{M}_{\mathbf{B}}\widetilde{\boldsymbol{\epsilon}} \qquad \text{(A.18)} \\ &+ \frac{1}{d_i}1_{i,m}'\boldsymbol{M}_{\mathbf{B}}\widetilde{\mathbf{F}}(\widetilde{\boldsymbol{\tau}}^t/2 + \widetilde{\boldsymbol{\tau}}^s/2) - c_i\frac{\tau^t(U_i)+\tau^s(U_i)}{2} \qquad \text{(A.19)}\end{aligned}$$

where $1_{i,m}$ is the m-dimensional vector with $1_{i,m,e} = 1$ if $e \in E_i$ and 0 otherwise.

First, we show that (A.18) converges to zero in probability. By the definition of $\widetilde{\boldsymbol{\epsilon}}$, this term has mean zero. Its variance is given by

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\widetilde{\boldsymbol{\epsilon}}\right) &= \frac{1}{d_i^2}\mathbb{E}\left[1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\widetilde{\boldsymbol{\epsilon}}\widetilde{\boldsymbol{\epsilon}}'\boldsymbol{M}_{\mathbf{B}}1_{i,m}\right] \\ &\leq \frac{\widetilde{\sigma}_n^2}{d_i},\end{aligned}$$

where the inequality follows from $\|\boldsymbol{M}_{\mathbf{B}}\| \leq 1$ (since $\boldsymbol{M}_{\mathbf{B}}$ is idempotent), $\mathbb{E}[\widetilde{\boldsymbol{\epsilon}}\widetilde{\boldsymbol{\epsilon}}'] \leq \widetilde{\sigma}_n^2\mathbf{I}_m$, and $1'_{i,m}1_{i,m} = d_i$. Since $d_i \to \infty$, the variance converges to zero. Thus, by Chebyshev's inequality,

$$\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\widetilde{\boldsymbol{\epsilon}} = O_p\left(\frac{1}{\sqrt{d_i}}\right) = o_p(1),$$

as $d_i \to \infty$.

Second, we show that (A.19) converges in probability to zero. For $i \notin V^{st}$, this term is exactly zero since either $1_{i,m} = \mathbf{b}_i$ or $1_{i,m} = -\mathbf{b}_i$, where $\mathbf{b}_i$ is the i-th column of $\mathbf{B}$; as there are no in- or outflows for i, $1_{i,m}$ is orthogonal to $\boldsymbol{M}_{\mathbf{B}}$ and thus the first term vanishes. Also, note that $c_i = 0$ in this case.

For $i \in V^{st}$, since $\mathbb{E}[\tau^o(U_j)|U_i] = 0$ for $j \neq i$ and $o \in \{s,t\}$, we have

$$\mathbb{E}\left[\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\widetilde{\mathbf{F}}(\widetilde{\boldsymbol{\tau}}^t/2 + \widetilde{\boldsymbol{\tau}}^s/2) - c_i\frac{\tau^t(U_i) + \tau^s(U_i)}{2}\middle|U_i\right] = 0$$

where $\widetilde{\boldsymbol{f}}_i$ is the i-th column of $\widetilde{\mathbf{F}}$, and $c_i = \frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\widetilde{\boldsymbol{f}}_i$. Note $c_i \in [0,1]$ since $\boldsymbol{M}_{\mathbf{B}}$ is a projection matrix and $1'_{i,m}\widetilde{\boldsymbol{f}}_i = d_i$.

Since

$$\begin{aligned}
&\frac{1}{d_i} 1'_{i,m} \boldsymbol{M}_{\mathbf{B}} \widetilde{\mathbf{F}} (\widetilde{\boldsymbol{\tau}}^t/2 + \widetilde{\boldsymbol{\tau}}^s/2) \\
&= \frac{1}{2d_i} \sum_{j \in V^{st}} (\tau^t(U_j) + \tau^s(U_j)) \cdot 1'_{i,m} \boldsymbol{M}_{\mathbf{B}} \widetilde{\boldsymbol{f}}_j,
\end{aligned}$$

the conditional variance given U_i is

$$\begin{aligned}
&\text{Var}\left[\frac{1}{d_i} 1'_{i,m} \boldsymbol{M}_{\mathbf{B}} \widetilde{\mathbf{F}} (\widetilde{\boldsymbol{\tau}}^t/2 + \widetilde{\boldsymbol{\tau}}^s/2) \big| U_i\right] \\
&= \frac{\mathbb{E}[(\tau^t(U_i) + \tau^s(U_i))^2]}{4d_i^2} \sum_{j \neq i \in V^{st}} \left(1'_{i,m} \boldsymbol{M}_{\mathbf{B}} \widetilde{\boldsymbol{f}}_j\right)^2 \\
&\leq \frac{\mathbb{E}[(\tau^t(U_i) + \tau^s(U_i))^2]}{4d_i^2} \sum_{j \in V^{st}} |E_{(i,j)}|^2 \\
&\leq \frac{\max_{j \in V^{st}} |E_{(i,j)}|^2 \cdot \mathbb{E}[(\tau^t(U_i) + \tau^s(U_i))^2]}{4d_i} \\
&= O\left(\frac{1}{d_i}\right) = o(1),
\end{aligned}$$

as $d_i \to \infty$, where the first inequality uses $\|\boldsymbol{M}_{\mathbf{B}}\| \leq 1$ and $1'_{i,m} \widetilde{\boldsymbol{f}}_j = |E_{(i,j)}|$, and the last equality follows by assumption. By the law of total variance, it follows that

$$\text{Var}\left[\frac{1}{d_i} 1'_{i,m} \boldsymbol{M}_{\mathbf{B}} \widetilde{\mathbf{F}} (\widetilde{\boldsymbol{\tau}}^t/2 + \widetilde{\boldsymbol{\tau}}^s/2) - c_i \frac{\tau^t(U_i) + \tau^s(U_i)}{2}\right] = O\left(\frac{1}{d_i}\right) = o(1),$$

as $d_i \to \infty$. Thus, by Chebyshev's inequality,

$$\frac{1}{d_i} 1'_{i,m} \boldsymbol{M}_{\mathbf{B}} \widetilde{\mathbf{F}} (\widetilde{\boldsymbol{\tau}}^t/2 + \widetilde{\boldsymbol{\tau}}^s/2) - c_i \frac{\tau^t(U_i) + \tau^s(U_i)}{2} = O_p\left(\frac{1}{\sqrt{d_i}}\right) = o_p(1),$$

as $d_i \to \infty$.

Finally, combining the results from the two steps, we have

$$\frac{1}{d_i}\sum_{e\in E_i}\widehat{\epsilon}_e - c_i(\tau^t(U_i)+\tau^s(U_i))/2 = O_p\left(\frac{1}{\sqrt{d_i}}\right) = o_p(1),$$

as $d_i \to \infty$ by the continuous mapping theorem. This completes the proof. □

A.6.6 Proof of Proposition 1.3

Proof. **Validity of the confidence interval**: First, we show that $\widehat{c}_\alpha \to_p c_\alpha$, where c_α is the α-quantile of F_τ for any $\alpha \in (0,1)$ such that $c_\alpha < \infty$. This is immediate from the uniform convergence of $\widehat{F}_{n,\tau}$ to F_τ in Theorem 1.2: Since $\widehat{F}_{n,\tau}$ weakly converges to F_τ in probability (Problem 23.1 in van der Vaart, 1998), $\widehat{F}^{-1}_{n,\tau}(\cdot) = \inf\{x \in \mathbb{R} : \widehat{F}_{n,\tau}(x) \geq \cdot\}$ weakly converges $F^{-1}_\tau(\cdot) = \inf\{x \in \mathbb{R} : F_\tau(x) \geq \cdot\}$ in probability (Lemma 21.2 in van der Vaart, 1998). Since F_τ is continuous and strictly increasing around c_α, $F^{-1}_\tau(\alpha) = c_\alpha$ and $\widehat{c}_\alpha = \widehat{F}^{-1}_{n,\tau}(\alpha) \to_p c_\alpha$.

Next, observe that by Theorem 1.1 and $((d^t_i - d^s_i)/d_i)^{-1} = O(1)$,

$$\widetilde{T} \equiv \left(\frac{d^t_i - d^s_i}{d_i}\right)^{-1} (\widehat{\alpha}_i - \alpha_i) \to_p \tau(U_i),$$

as $d_i \to \infty$. Thus, $(\widetilde{T}, \widehat{c}_{\alpha/2}, \widehat{c}_{1-\alpha/2})$ weakly converges to $(\tau(U_i), c_{\alpha/2}, c_{1-\alpha/2})$. By Slutsky's theorem, we have

$$\mathbb{P}(\widehat{c}_{\alpha/2} \leq \widetilde{T} \leq \widehat{c}_{1-\alpha/2}) \to \mathbb{P}(c_{\alpha/2} \leq \tau(U_i) \leq c_{1-\alpha/2}) = 1-\alpha.$$

This shows the asymptotic validity of the confidence interval.

Validity of the test: Observe that under the null hypothesis, by Theorem 1.1

and the continuous mapping theorem, we have

$$T = \underbrace{((\tau(U_i))_{i\in V_0})'\boldsymbol{C}_n((\tau(U_i))_{i\in V_0})}_{\equiv g_n((\tau(U_i)_{i\in V_0}))\equiv \widetilde{T}} + o_p(1)$$
$$\rightarrow_d \underbrace{((\tau(U_i))_{i\in V_0})'\boldsymbol{C}((\tau(U_i))_{i\in V_0})}_{\equiv g((\tau(U_i)_{i\in V_0}))\equiv T_\infty},$$

where $\boldsymbol{C} = \mathbf{D}^{st}_{V_0}\boldsymbol{M}_{n_0}\mathbf{D}^{st}_{V_0}$ and $\mathbf{D}^{st}_{V_0} = \lim_{n\to\infty}\mathbf{D}^{st}_{n,V_0}$. Note that the cdf of T_∞ is continuous. Let $\tau_i^{(m)}$ be m-th random draw from $\widehat{F}_{n,\tau}$ (with replacement) for $m = 1, ..., M$. Then, the simulated test statistic is given by

$$\widetilde{T}^{(m)} = \underbrace{(\tau_i^{(m)})_{i\in V_0})'\boldsymbol{C}_n((\tau_i^{(m)})_{i\in V_0})}_{\equiv g_n((\tau_i^{(m)})_{i\in V_0})},$$

$\boldsymbol{C}_n = \mathbf{D}^{st}_{n,V_0}\boldsymbol{M}_{n_0}\mathbf{D}^{st}_{n,V_0}$ and $\mathbf{D}^{st}_{n,V_0} = diag((d_i^t - d_i^s)/d_i, i \in V_0)$. Note that $\widetilde{T}^{(m)}, m = 1, ..., M$ are independent and identically distributed conditional on $\widehat{F}_{n,\tau}$.

Let $\widehat{F}^{\otimes n_0}_{n,T}$ be the distribution of $(\tau_i^{(1)})_{i\in V_0}$ given $\widehat{F}_{n,\tau}$ and $F^{\otimes n_0}_{\tau}$ be the distribution of $(\tau(U_i))_{i\in V_0}$. Since n_0 is fixed, by the weak convergence of $\widehat{F}_{n,\tau}$ to F_τ in probability, $\widehat{F}^{\otimes n_0}_{n,T}$ weakly converges to $F^{\otimes n_0}_{\tau}$ in probability (Theorem 2.8 in Billingsley, 1999). As g_n and g are continuous, by the continuous mapping theorem (Theorem 3.27 in Kallenberg, 2002), $\widehat{G}_{n,T}$ weakly converges to G_T in probability, where $\widehat{G}_{n,T}$ is the distribution of $\widetilde{T}^{(1)}$ given $\widehat{F}_{n,\tau}$ and G_T is the distribution of T_∞. Since G_T is continuous, by the same argument as in the proof of the validity of the confidence interval, we have $\widehat{c}_\alpha \rightarrow_p c_\alpha$, where $\widehat{c}_\alpha$ is the α-quantile of $\widehat{G}_{n,T}$ and c_α is the α-quantile of G_T. Thus, $(T, \widehat{c}_\alpha)$ weakly converges to (T_∞, c_α). By Slutsky's theorem, we have

$$\mathbb{P}(T > \widehat{c}_{1-\alpha}) \to \mathbb{P}(T_\infty > c_{1-\alpha}) = 1 - G_T(c_{1-\alpha}) = \alpha,$$

which shows the asymptotic validity of the test with the known $\widehat{G}_{n,T}$.

Finally, we show that the test remains valid when $\widehat{G}_{n,T}$ is estimated by the empirical distribution of $\widetilde{T}^{(m)}, m = 1, ..., M$. Let $\widehat{G}_{n,T,M}$ be the empirical distribution of $\widetilde{T}^{(m)}, m = 1, ..., M$ given $\widehat{F}_{n,\tau}$. Let $\widehat{c}_{\alpha,M}$ be the α-quantile of $\widehat{G}_{n,T,M}$. Observe that, for any $\epsilon > 0$,

$$\begin{aligned}
&|\mathbb{P}(T > \widehat{c}_{1-\alpha,M}) - \mathbb{P}(T > c_{1-\alpha})| \\
&= \left|\mathbb{P}(\widehat{G}_{n,T,M}(T) > 1-\alpha) - \mathbb{P}(G_T(T) > 1-\alpha)\right| \\
&\le \mathbb{P}\left(|G_T(T) - 1 + \alpha| \le \epsilon\right) + \mathbb{P}\left(|G_T(T) - \widehat{G}_{n,T,M}(T)| > \epsilon\right) \\
&\le \mathbb{P}\left(|G_T(T) - 1 + \alpha| \le \epsilon\right) + \mathbb{P}\left(\sup_{x\in\mathbb{R}} |G_T(x) - \widehat{G}_{n,T,M}(x)| > \epsilon\right),
\end{aligned}$$

where the first equality follows from $\widehat{G}_{n,T,M}$ and G_T being right-continuous, and the first inequality follows from the fact that $|\mathbb{I}(x \le a) - \mathbb{I}(x \le b)| \le \mathbb{I}(|x-a| \le \epsilon) + \mathbb{I}(|a-b| > \epsilon)$ for any $x, a, b \in \mathbb{R}$ and $\epsilon > 0$. By the continuous mapping theorem and the Portmanteau theorem, the first term on the right-hand side converges to $\mathbb{P}(|G_T(T) - 1 + \alpha| \le \epsilon)$, which can be made arbitrarily small by choosing ϵ small enough. For the second term on the right-hand side, observe that

$$\begin{aligned}&\sup_{x\in\mathbb{R}}|G_T(x)-\widehat{G}_{n,T,M}(x)|\\&\leq \sup_{x\in\mathbb{R}}|G_T(x)-\widehat{G}_{n,T}(x)|+\sup_{x\in\mathbb{R}}|\widehat{G}_{n,T}(x)-\widehat{G}_{n,T,M}(x)|.\end{aligned}$$

Since G_T is continuous, the first term on the right-hand side converges to zero in probability (Problem 23.1 in van der Vaart, 1998). The second term on the right-hand side converges to zero almost surely by the Glivenko-Cantelli theorem as $M \to \infty$ conditionally on $\widehat{F}_{n,\tau}$. Thus,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}}|G_T(x)-\widehat{G}_{n,T,M}(x)|>\epsilon\right)\to 0,$$

as $n, M \to \infty$. Since $\epsilon > 0$ is arbitrary, we have

$$\mathbb{P}(T>\widehat{c}_{1-\alpha,M})\to\mathbb{P}(T>c_{1-\alpha})=\alpha,$$

as $n, M \to \infty$. This completes the proof. □

Appendix B

Chapter 2

B.1 Example for $\widetilde{\gamma}_n = \gamma_n^{\text{causal}} + o_p(1)$

In Theorems 2.6 and 2.7, we need some $\widetilde{\gamma}_n$ satisfying $\widetilde{\gamma}_n = \gamma_n^{\text{causal}} + o_p(1)$. By Assumption 2.7 (ii), we, without loss of generality, assume that the first m elements of $Z_{n,i}$ depend on $R_{n,i}$ multiplicatively.[1] We allow general heterogeneous treatment assignment in Assumption 2.1 (iii). We also assume that the researcher knows ρ_n or the population size n. Let $Z_{n,i} = (Z'_{(1:m),n,i}, Z'_{-(1:m),n,i})'$, where $Z_{(1:m),n,i}$ is the first m elements of $Z_{n,i}$ and $Z_{-(1:m),n,i}$ are the remaining elements. Recall that $\widetilde{Z}_{n,i} = Z_{n,i}$ under Assumption 2.7.

Define

$$\widetilde{\gamma}_n = (\widetilde{P}_n^{ZZ})^{-1}\widetilde{P}_n^{ZY}, \tag{B.1}$$

where

$$\widetilde{P}_n^{ZZ} = \frac{1}{N}\sum_{i=1}^{n} R_{n,i}\begin{pmatrix} \rho_n Z_{(1:m),n,i}Z'_{(1:m),n,i} & \rho_n Z_{(1:m),n,i}Z'_{-(1:m),n,i} \\ \rho_n Z_{-(1:m),n,i}Z'_{(1:m),n,i} & Z_{-(1:m),n,i}Z'_{-(1:m),n,i} \end{pmatrix},$$

$$\widetilde{P}_n^{ZY} = \frac{1}{N}\sum_{i=1}^{n} R_{n,i}\begin{pmatrix} \rho_n Z_{(1:m),n,i} \\ Z_{-(1:m),n,i} \end{pmatrix} Y_{n,i}.$$

Note that some elements of $\widetilde{P}_n^{ZZ}$ and $\widetilde{P}_n^{ZY}$ are rescaled by ρ_n from $\widetilde{Q}_n^{ZZ}$ and $\widetilde{Q}_n^{ZY}$. ρ_n can be replaced with its consistent estimator N/n. The consistency

[1] In the usual applications, it is enough to consider the $m = 1$ case.

of $\widetilde{\gamma}_n$ is shown in Lemma B.15.

B.2 Preliminary Results

Remember that for each $i \in \mathcal{N}_n$,

$$T_{n,i} = g(i, \mathbf{D}_n, \mathbf{A}_n);$$
$$\widetilde{T}_{n,i} = \widetilde{g}(i, \mathbf{D}_n, \widetilde{\mathbf{A}}_n);$$
$$X_{n,i} = T_{n,i} - \Lambda_n Z_{n,i};$$
$$\widetilde{X}_{n,i} = \widetilde{T}_{n,i} - \widetilde{\Lambda}_n \widetilde{Z}_{n,i},$$

where

$$\Lambda_n = (\sum_{i=1}^{n} \mathbb{E}[T_{n,i} Z'_{n,i}])(\sum_{i=1}^{n} \mathbb{E}[Z_{n,i} Z'_{n,i}])^{-1};$$
$$\widetilde{\Lambda}_n = (\sum_{i=1}^{n} R_{n,i} \mathbb{E}[\widetilde{T}_{n,i} | \boldsymbol{R}_n] \widetilde{Z}'_{n,i})(\sum_{i=1}^{n} R_{n,i} \widetilde{Z}_{n,i} \widetilde{Z}'_{n,i})^{-1},$$

and

$$\Omega_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\begin{pmatrix} Y_{n,i} \\ X_{n,i} \\ Z_{n,i} \end{pmatrix} \begin{pmatrix} Y_{n,i} \\ X_{n,i} \\ Z_{n,i} \end{pmatrix}'\right] \equiv \begin{pmatrix} \Omega_n^{YY} & \Omega_n^{YX} & \Omega_n^{YZ} \\ \Omega_n^{XY} & \Omega_n^{XX} & \Omega_n^{XZ} \\ \Omega_n^{ZY} & \Omega_n^{ZX} & \Omega_n^{ZZ} \end{pmatrix};$$

$$\widetilde{Q}_n = \frac{1}{N} \sum_{i=1}^{n} R_{n,i} \begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix} \begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}' \equiv \begin{pmatrix} \widetilde{Q}_n^{YY} & \widetilde{Q}_n^{YX} & \widetilde{Q}_n^{YZ} \\ \widetilde{Q}_n^{XY} & \widetilde{Q}_n^{XX} & \widetilde{Q}_n^{XZ} \\ \widetilde{Q}_n^{ZY} & \widetilde{Q}_n^{ZX} & \widetilde{Q}_n^{ZZ} \end{pmatrix};$$

$$\widetilde{\Omega}_n = \frac{1}{N}\sum_{i=1}^{n} R_{n,i}\mathbb{E}\left[\begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}\begin{pmatrix} Y_{n,i} \\ \widetilde{X}_{n,i} \\ \widetilde{Z}_{n,i} \end{pmatrix}' \mid \boldsymbol{R}_n\right] \equiv \begin{pmatrix} \widetilde{\Omega}_n^{YY} & \widetilde{\Omega}_n^{YX} & \widetilde{\Omega}_n^{YZ} \\ \widetilde{\Omega}_n^{XY} & \widetilde{\Omega}_n^{XX} & \widetilde{\Omega}_n^{XZ} \\ \widetilde{\Omega}_n^{ZY} & \widetilde{\Omega}_n^{ZX} & \widetilde{\Omega}_n^{ZZ} \end{pmatrix}.$$

B.2.1 Preliminary Lemmas

We will use the following results from Kojevnikov et al. (2021a). We will only state the conditional version of the results, but also use the unconditional version of the results, which can be understood analogously.

Define

$$\sigma_n^2 = \mathrm{Var}(S_n \mid \boldsymbol{R}_n),$$

where $S_n = \sum_{i\in\mathcal{N}_n} U_{i,n}$.

Condition B.1. *A triangular array $\{U_{n,i}\}$ is conditionally ψ-dependent given $\boldsymbol{R}_n$ with ξ_n satisfying*

- *For some constant $C > 0$,*

$$\psi_{a,b}(f,g) \leq C \times ab(\|f\|_\infty + \mathrm{Lip}(f))(\|g\|_\infty + \mathrm{Lip}(g)).$$

- $\sup_n \max_{s\geq 1} \xi_{n,s} < \infty$ *a.s.*

- *For some $p > 4$, $\sup_{n\geq 1} \max_{i\in\mathcal{N}_n} \mathbb{E}[|U_{n,i}|^p \mid \boldsymbol{R}_n] < \infty$ a.s.*

- *There exists a positive sequence $m_n \to \infty$ such that for $k = 1, 2$,*

$$\frac{n}{\sigma_n^{2+k}} \sum_{s \geq 0} c_n(s, m_n; k) \xi_{n,s}^{1-\frac{2+k}{p}} \xrightarrow{a.s.} 0,$$
$$\frac{n^2 \xi_{n,m_n}^{1-1/p}}{\sigma_n} \xrightarrow{a.s.} 0.$$

- $\mathbb{E}[U_{n,i} \mid \boldsymbol{R}_n] = 0.$

Lemma B.1 (CLT, Theorem 3.2 in Kojevnikov et al., 2021a)**.** *Under Condition B.1,*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{S_n}{\sigma_n} \leq t \mid \boldsymbol{R}_n \right\} - \Phi(t) \right| \xrightarrow{a.s.} 0 \text{ as } n \to \infty,$$

where Φ denotes the distribution function of $\mathcal{N}(0, 1)$.

Lemma B.2 (Linear Transformation, Lemma 2.1 in Kojevnikov et al., 2021a)**.** *For each $n \geq 1$, let $\{a_{n,i}\}_{i \in \mathcal{N}_n}$ be a sequence of $\sigma(\boldsymbol{R}_n)$-measurable vectors such that $\max_{i \in \mathcal{N}_n} \|a_{n,i}\| \leq 1$ a.s. Under the first condition of Condition B.1, the array $a'_{n,i} U_{n,i}$ is conditionally ψ-dependent given $\boldsymbol{R}_n$ with the dependence coefficients $\{\xi_n\}$.*

Condition B.2. *Let $\omega(x) = \mathbb{1}\{|x| \leq 1\}$. There exists $p > 4$ such that*

- $\sup_{n \geq 1} \max_{i \in \mathcal{N}_n} \mathbb{E}[|U_{n,i}|^p \mid \boldsymbol{R}_n] < \infty$ *a.s.*

- $\lim_{n \to \infty} \sum_{s \geq 1} |\omega(s/2K) - 1| \, \delta_n^{\partial}(s, 1) \xi_{n,s}^{1-(2/p)} = 0$ *a.s.*

- $\lim_{n \to \infty} n^{-1} \sum_{s \geq 0} c_n(s, 2K; 2) \, \xi_{n,s}^{1-(4/p)} = 0$ *a.s.*

Lemma B.3 (Variance Consistency, $2K$ Local Case of Proposition 4.1. in Kojevnikov et al., 2021a)**.** *Suppose that Conditions B.1 and B.2 hold. Then as* $n \to \infty$*,*

$$\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i;2K)} U_{n,i}U'_{n,j} - \mathrm{Var}\left(\frac{S_n}{\sqrt{n}} \mid \boldsymbol{R}_n\right)\right\|_F \mid \boldsymbol{R}_n\right] \xrightarrow{a.s.} 0,$$

where $\|\cdot\|_F$ *is the Frobenius norm. By Markov's inequality, we also have*

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i;2K)} U_{n,i}U'_{n,j} - \mathrm{Var}\left(\frac{S_n}{\sqrt{n}} \mid \boldsymbol{R}_n\right) \xrightarrow{p^R} 0.$$

B.2.2 Main Lemmas

Lemma B.4. *Under* $\rho_n n \to \infty$*,*

$$N > 0 \text{ a.s. for large enough } n$$

Proof. Since the result is trivial for $\rho_n = 1$, we focus on the case $\rho_n \in (0,1)$. By the inequality $1 - x \leq e^{-x}$ for $x \in (0,1)$, we have $(1-\rho_n)^n \leq e^{-n\rho_n}$. Thus,

$$\sum_{n=1}^{\infty}\mathbb{P}(N=0) = \sum_{n=1}^{\infty}\mathbb{P}\left(\sum_{i=1}^{n} R_{n,i} = 0\right) = \sum_{n=1}^{\infty}(1-\rho_n)^n \leq \sum_{n=1}^{\infty} e^{-n\rho_n}.$$

$\rho_n n \to \infty$ implies the right-hand side is bounded. By the Borel-Cantelli lemma, we can conclude. □

Lemma B.5. *Under $\rho_n^2 n \to \infty$,*

$$\frac{N}{n\rho_n} \xrightarrow{a.s.} 1$$

as $n \to \infty$.

Proof. Pick any $\varepsilon > 0$. By Hoeffding's inequality with $R_i \in [0,1]$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{N}{n\rho_n} - 1\right| > \varepsilon\right) &= \mathbb{P}\left(|N - n\rho_n| > \varepsilon n\rho_n\right) = \mathbb{P}\left(\left|\sum_{i=1}^{n} R_i - n\rho_n\right| > \varepsilon n\rho_n\right) \\ &\leq 2\exp\left(-\frac{2(\varepsilon n\rho_n)^2}{n}\right) = 2\exp\left(-2\varepsilon^2 n\rho_n^2\right).\end{aligned}$$

$\rho_n^2 n \to \infty$ implies $\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{N}{n\rho_n} - 1\right| > \varepsilon\right)$ is bounded. From the Borel-Cantelli lemma, we can conclude. □

Lemma B.6. *Assume that Assumptions 2.3 and 2.4 hold. Then, for large enough n,*

$$\Lambda_n = L_n, \quad X_{n,i} = T_{n,i} - \mathbb{E}[T_{n,i}|\boldsymbol{R}_n] \quad a.s.,$$

and

$$\widetilde{\Lambda}_n = \widetilde{L}_n, \quad \widetilde{X}_{n,i} = \widetilde{T}_{n,i} - \mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n] \quad a.s.$$

Proof. Observe that $\Lambda_n = L_n$ a.s. for large enough n as

$$\begin{aligned}
\Lambda_n &= \sum_{i=1}^n \mathbb{E}[\mathbb{E}[T_{n,i}|\boldsymbol{R}_n]Z'_{n,i}] \left(\sum_{i=1}^n \mathbb{E}[Z_{n,i}Z'_{n,i}]\right)^{-1} \\
&= L_n \sum_{i=1}^n \mathbb{E}[Z_{n,i}Z'_{n,i}] \left(\sum_{i=1}^n \mathbb{E}[Z_{n,i}Z'_{n,i}]\right)^{-1} \\
&= L_n,
\end{aligned}$$

where Λ_n is well-defined by Assumption 2.3 and the second equality holds by Assumption 2.4. Similarly, $\widetilde{\Lambda}_n = \widetilde{L}_n$ a.s. for large enough n as

$$\begin{aligned}
\widetilde{\Lambda}_n &= \sum_{i=1}^n R_{n,i}\mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n]\widetilde{Z}'_{n,i} \left(\sum_{i=1}^n R_{n,i}\widetilde{Z}_{n,i}\widetilde{Z}'_{n,i}\right)^{-1} \\
&= \widetilde{L}_n \sum_{i=1}^n R_{n,i}\widetilde{Z}_{n,i}\widetilde{Z}_{n,i} \left(\sum_{i=1}^n R_{n,i}\widetilde{Z}_{n,i}\widetilde{Z}'_{n,i}\right)^{-1} \\
&= \widetilde{L}_n,
\end{aligned}$$

where $\widetilde{\Lambda}_n$ is well-defined by Assumption 2.3 and the second equality holds by Assumption 2.4.

Since we define $X_{n,i} = T_{n,i} - \Lambda_n Z_{n,i}$ and $\widetilde{X}_{n,i} = \widetilde{T}_{n,i} - \widetilde{\Lambda}_n \widetilde{Z}_{n,i}$, Assumption 2.4 and the above two displayed qualities imply for large enough n, $X_{n,i} = T_{n,i} - \mathbb{E}[T_{n,i}|\boldsymbol{R}_n]$ a.s. and $\widetilde{X}_{n,i} = \widetilde{T}_{n,i} - \mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n]$ a.s. □

Lemma B.7. *Suppose that $\widetilde{T}_{n,i} = T_{n,i}$ and $\widetilde{Z}_{n,i} = Z_{n,i}$ for all $i \in \mathcal{N}_n$ and $n \in \mathbb{N}$. Under Assumptions 2.3 and 2.4, (i) $\widetilde{\Lambda}_n = \Lambda_n$ a.s. and (ii) $\widetilde{X}_{n,i} = X_{n,i}$ a.s.*

Proof. The results follow directly from Lemma B.6. □

Lemma B.8. *Assume that Assumptions 2.1 to 2.5 hold. The following sequences of triangular arrays are ψ-dependent with $\xi_{n,s} = \mathbb{1}\{s \leq 2K\}$:*

$$X_{n,i}Z'_{n,i},\ X_{n,i}X'_{n,i},\ X_{n,i}Y_{n,i},\ Z_{n,i}Z'_{n,i},\ Z_{n,i}Y_{n,i}.$$

The following sequences of triangular arrays are conditionally ψ-dependent given $\boldsymbol{R}_n$ with $\xi_{n,s} = \mathbb{1}\{s \leq 2K\}$:

$$R_{n,i}\widetilde{X}_{n,i}\widetilde{Z}'_{n,i},\ R_{n,i}\widetilde{X}_{n,i}\widetilde{T}'_{n,i},\ R_{n,i}\widetilde{X}_{n,i}Y_{n,i},\ R_{n,i}\widetilde{Z}_{n,i}\widetilde{Z}'_{n,i},\ R_{n,i}\widetilde{Z}_{n,i}Y_{n,i}.$$

Proof. By Assumption 2.5, we can set $\xi_{n,s} = \mathbb{1}\{s \leq 2K\}$ for $s \geq 1$ since if $d_n(A,B) > 2K$, $f(U_{n,A}) \perp\!\!\!\perp g(U_{n,B})$ for any $f \in \mathcal{L}_{v,a}$ and $g \in \mathcal{L}_{v,b}$ as long as $U_{n,i}$ are based on $\widetilde{T}_{n,i}, T_{n,i}, \widetilde{Z}_{n,i}, Z_{n,i}, \widetilde{Y}_{n,i}, Y_{n,i}$. For large enough n, Lemma B.6 implies $X_{n,i} = T_{n,i} - \mathbb{E}[T_{n,i}|\boldsymbol{R}_n]$ and $\widetilde{X}_{n,i} = \widetilde{T}_{n,i} - \mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n]$ almost surely. Thus, for large enough n, $X_{n,i}$ and $\widetilde{X}_{n,i}$ also have the local dependence with $2K$. By Assumption 2.3, each element is uniformly bounded. Thus, we can set $\psi_{a,b}(f,g) = 2\|f\|_\infty\|g\|_\infty$ for any $f \in \mathcal{L}_{v,a}$ and $g \in \mathcal{L}_{v,b}$. This completes the proof. □

Lemma B.9. *Under Assumption 2.3,*

$$\max_i |\widetilde{\varepsilon}_{n,i}| < \infty \ a.s. \qquad \text{and} \qquad \max_i |\varepsilon_{n,i}| < \infty \ a.s.$$

Proof. Under the uniform boundedness and the invertibility condition (Assumption 2.3), $\|\theta_n^{\text{causal,sample}}\| < \infty$ a.s. and $\|\gamma_n^{\text{causal,sample}}\| < \infty$ a.s. Thus, by the Schwarz Inequality,

$$
\begin{aligned}
|\widetilde{\varepsilon}_{n,i}| &\leq \max_i |Y_{n,i}| + \max_i \|\widetilde{X}_{n,i}\| \times \|\theta_n^{\text{causal,sample}}\| \\
&\quad + \max_i \|\widetilde{Z}_{n,i}\| \times \|\gamma_n^{\text{causal,sample}}\| \\
&< \infty \quad \text{a.s.}
\end{aligned}
$$

for all i. The bound for $|\varepsilon_{n,i}|$ can be derived similarly. □

Lemma B.10. *Under Assumptions 2.1 to 2.6,*

$$
\widetilde{Q}_n - \widetilde{\Omega}_n \xrightarrow{p^R} 0 \quad \textit{and} \quad \widetilde{Q}_n - \widetilde{\Omega}_n \xrightarrow{p} 0.
$$

Proof. Let $W_{n,i} \equiv (Y_{n,i}, \widetilde{X}_{n,i}, \widetilde{Z}_{n,i})'$. Then,

$$
\begin{aligned}
\widetilde{Q}_n - \widetilde{\Omega}_n &= \frac{1}{N} \sum_{i=1}^{n} R_{n,i}(W_{n,i}W_{n,i}' - \mathbb{E}[W_{n,i}W_{n,i}' | \boldsymbol{R}_n]) \\
&= \frac{n\rho_n}{N} \times \frac{1}{n\rho_n} \sum_{i=1}^{n} R_{n,i} \left(W_{n,i}W_{n,i}' - \mathbb{E}[W_{n,i}W_{n,i}' | \boldsymbol{R}_n] \right).
\end{aligned}
$$

Since $(n\rho_n)/N \xrightarrow{a.s.} 1$ (Lemma B.5) implies $(n\rho_n)/N \xrightarrow{p^R} 1$, it suffices to show that

$$
\frac{1}{n\rho_n} \sum_{i=1}^{n} R_{n,i} \left(W_{n,i}W_{n,i}' - \mathbb{E}[W_{n,i}W_{n,i}' | \boldsymbol{R}_n] \right) \xrightarrow{p^R} 0.
$$

We will show it by verifying

$$\mathbb{E}\left[\left(\frac{1}{n\rho_n}\sum_{i=1}^{n}R_{n,i}\left(W_{n,i,(k)}W_{n,i,(\ell)}-\mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)}\mid \boldsymbol{R}_n]\right)\right)^2\mid \boldsymbol{R}_n\right]\xrightarrow{a.s.}0$$

for all $k,\ell=1,\dots,d_{\widetilde{T}}$. Observe that

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{n\rho_n}\sum_{i=1}^{n}R_{n,i}\left(W_{n,i,(k)}W_{n,i,(\ell)}-\mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)}\mid \boldsymbol{R}_n]\right)\right)^2\mid \boldsymbol{R}_n\right]\\
=&\frac{1}{n^2\rho_n^2}\sum_{i=1}^{n}R_{n,i}\mathbb{E}\left[(W_{n,i,(k)}W_{n,i,(\ell)}-\mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)}|\boldsymbol{R}_n])^2\mid \boldsymbol{R}_n\right] && \text{(B.2)}\\
&+\frac{1}{n^2\rho_n^2}\sum_{i\neq j}R_{n,i}R_{n,j}\mathbb{E}[(W_{n,i,(k)}W_{n,i,(\ell)}-\mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)}|\boldsymbol{R}_n])\\
&\qquad\times(W_{n,j,(k)}W_{n,j,(\ell)}-\mathbb{E}[W_{n,j,(k)}W_{n,j,(\ell)}|\boldsymbol{R}_n])\mid \boldsymbol{R}_n] && \text{(B.3)}
\end{aligned}$$

For (B.2), since there is some absolute constant C such that $|W_{n,j,(k)} W_{n,j,(\ell)}|<C$ by Assumption 2.3,

$$(B.2)\leq\frac{1}{n^2\rho_n^2}\sum_{i=1}^{n}(2C)^2=4C^2\times\frac{1}{n\rho_n^2}\to 0$$

where the inequality and the convergence do not depend on $\boldsymbol{R}_n$.

For $(B.3)$, note that if $d_n(i,j)>2K$, then

$$\begin{aligned}
&\mathbb{E}[(W_{n,i,(k)}W_{n,i,(\ell)}-\mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)}|\boldsymbol{R}_n])\\
&\times(W_{n,j,(k)}W_{n,j,(\ell)}-\mathbb{E}[W_{n,j,(k)}W_{n,j,(\ell)}|\boldsymbol{R}_n])\mid \boldsymbol{R}_n]=0
\end{aligned}$$

as $R_{n,i}$ is i.i.d and $(T_{n,i},\widetilde{T}_{n,i})\perp\!\!\!\perp(T_{n,j},\widetilde{T}_{n,j})$ with no overlap in $\mathbf{D}_n$ and $\boldsymbol{R}_n$.

Thus,

$$
\begin{aligned}
&(B.3)\\
&= \frac{1}{n^2\rho_n^2}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i,2K)\setminus\{i\}} R_{n,i}R_{n,j}\mathbb{E}[(W_{n,i,(k)}W_{n,i,(\ell)} - \mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)}|\boldsymbol{R}_n])\\
&\qquad\qquad\qquad \times (W_{n,j,(k)}W_{n,j,(\ell)} - \mathbb{E}[W_{n,j,(k)}W_{n,j,(\ell)}|\boldsymbol{R}_n]) \mid \boldsymbol{R}_n]\\
&\leq 4C^2 \times \frac{1}{n\rho_n^2}\sum_{1\leq s\leq 2K}\delta_n^{\partial}(s;1) \to 0,
\end{aligned}
$$

where the last line holds by Assumption 2.6, and the inequality and the convergence do not depend on $\boldsymbol{R}_n$.

Thus, by Markov's inequality for

$$
\left(\frac{1}{n\rho_n}\sum_{i=1}^{n} R_{n,i}\left(W_{n,i,(k)}W_{n,i,(\ell)} - \mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)} \mid \boldsymbol{R}_n]\right)\right)^2,
$$

we have

$$
\frac{1}{n\rho_n}\sum_{i=1}^{n} R_{n,i}\left(W_{n,i,(k)}W_{n,i,(\ell)} - \mathbb{E}[W_{n,i,(k)}W_{n,i,(\ell)} \mid \boldsymbol{R}_n]\right) \xrightarrow{p^R} 0,
$$

and

$$
\widetilde{Q}_n - \widetilde{\Omega}_n \xrightarrow{p^R} 0.
$$

Unconditional consistency can be shown easily from this result. Since a conditional probability is bounded, the dominated convergence theorem and the law of iterated expectations imply $\widetilde{Q}_n - \widetilde{\Omega}_n \xrightarrow{p} 0$. □

Lemma B.11. *Let $W_{n,i}$ be a scalar random variable satisfying $|W_{n,i}| \leq \overline{W} < \infty$ a.s. We allow $W_{n,i}$ to depend on $\boldsymbol{R}_n$ and $\mathbf{D}_n$, but assume that $W_{n,i} \perp\!\!\!\perp R_{n,i}$ and $W_{n,i} \perp\!\!\!\perp W_{n,j}$ if $d_n(i,j) > 2K$. Then, under Assumptions 2.1 and 2.6,*

$$\frac{1}{N}\sum_{i=1}^{n} R_{n,i}\mathbb{E}[W_{n,i}|\boldsymbol{R}_n] - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[W_{n,i}] \xrightarrow{p} 0.$$

Proof. By Lemma B.5,

$$\frac{1}{N}\sum_{i=1}^{n} R_{n,i}\mathbb{E}[W_{n,i}|\boldsymbol{R}_n] = \frac{1}{n}\sum_{i=1}^{n}\frac{R_{n,i}}{\rho_n}\mathbb{E}[W_{n,i}|\boldsymbol{R}_n] + o_p(1).$$

Thus, it suffices to show that

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{R_{n,i}}{\rho_n}\mathbb{E}[W_{n,i}|\boldsymbol{R}_n] - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[W_{n,i}]\right)^2\right] \to 0 \tag{B.4}$$

The left-hand side of (B.4) is given by

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{R_{n,i}}{\rho_n}\mathbb{E}[W_{i,n}|\boldsymbol{R}_n] - \mathbb{E}[W_{n,i}]\right)^2\right] \tag{B.5}$$

$$+\frac{1}{n^2}\sum_{i\neq j}\mathbb{E}\left[\left(\frac{R_{n,i}}{\rho_n}\mathbb{E}[W_{n,i}|\boldsymbol{R}_n] - \mathbb{E}[W_{n,i}]\right)\left(\frac{R_{n,j}}{\rho_n}\mathbb{E}[W_{n,j}|\boldsymbol{R}_n] - \mathbb{E}[W_{n,j}]\right)\right] \tag{B.6}$$

For (B.5), we have

$$\begin{aligned}(B.5) &\leq \frac{2}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{R_{n,i}}{\rho_n}\right)^2(\mathbb{E}[W_{i,n}|\boldsymbol{R}_n])^2 + (\mathbb{E}[W_{n,i}])^2\right]\\ &\leq \frac{2\overline{W}^2}{n}\left[\mathbb{E}\left[\left(\frac{R_{n,i}}{\rho_n}\right)^2\right] + 1\right],\end{aligned}$$

where the first inequality holds from the inequality $(a-b)^2 \leq 2(a^2+b^2)$ for any $a, b \in \mathbb{R}$ and the second inequality holds by the uniform boundedness. Note that

$$\frac{1}{n}\mathbb{E}\left[\left(\frac{R_{n,i}}{\rho_n}\right)^2\right] = \frac{1}{n\rho_n} = o(1).$$

For (B.6),

$$\begin{aligned}
(B.6) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)\setminus\{i\}} \mathbb{E}\bigg[\left(\frac{R_{n,i}}{\rho_n}\mathbb{E}[W_{n,i}|\boldsymbol{R}_n] - \mathbb{E}[W_{n,i}]\right) \\
&\times \left(\frac{R_{n,j}}{\rho_n}\mathbb{E}[W_{n,j}|\boldsymbol{R}_n] - \mathbb{E}[W_{n,j}]\right)\bigg] \\
&\leq \frac{\overline{W}^2}{n^2}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)\setminus\{i\}} \mathbb{E}\left[\left|\frac{R_{n,i}}{\rho_n}-1\right|\cdot\left|\frac{R_{n,j}}{\rho_n}-1\right|\right] \\
&\leq \frac{\overline{W}^2}{n^2}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)\setminus\{i\}} \mathbb{E}\left[\left(\frac{R_{n,i}}{\rho_n}-1\right)^2\right] \\
&= \left(\frac{1}{\rho_n}-1\right)\frac{\overline{W}^2}{n^2}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)\setminus\{i\}} 1 \\
&= O\left(\frac{1}{n\rho_n}\right)\sum_{1\leq s\leq 2K}\delta_n^{\partial}(s;1) \\
&= o(1),
\end{aligned}$$

where the first equality holds by $W_{n,i} \perp\!\!\!\perp R_{n,j}$, $W_{n,i} \perp\!\!\!\perp W_{n,j}$ if $d_n(i,j) > 2K$, and Assumption 2.1, the first inequality holds by the uniform boundedness, the next inequality holds by the Cauchy-Schwarz inequality, and the last step follows from Assumption 2.6.

Combining the arguments for $(B.5)$ and (B.6), we have shown the convergence (B.4) as $n \to \infty$. □

Lemma B.12. *Let $W_{n,i}$ be a scalar random variable satisfying $|W_{n,i}| \leq \overline{W} < \infty$ a.s. We allow $W_{n,i}$ to depend on $\boldsymbol{R}_n$ and $\mathbf{D}_n$, but assume that $W_{n,i} \perp\!\!\!\perp R_{n,i}$ and $W_{n,i} \perp\!\!\!\perp W_{n,j}$ if $d_n(i,j) > 2K$. Then, under Assumptions 2.1 and 2.6,*

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{n} R_{n,i}\mathbb{E}[R_{n,i}W_{n,i}|\boldsymbol{R}_n] - \frac{1}{n\rho_n}\sum_{i=1}^{n}\mathbb{E}[R_{n,i}W_{n,i}] \\
&= \frac{1}{N}\sum_{i=1}^{n} R_{n,i}\mathbb{E}[W_{n,i}|\boldsymbol{R}_n] - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[W_{n,i}] \\
&\xrightarrow{p} 0.
\end{aligned}$$

Proof. The result follows by the same logic as Lemma B.11. □

Lemma B.13. *Under Assumptions 2.1 to 2.6 and 2.8,*

$$\widetilde{\Sigma}_n^{-1/2}\sum_{i=1}^{n} R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} \xrightarrow{d^R} \mathrm{N}(0, I_{d_{\widetilde{\mathcal{T}}}}).$$

Proof. We use the Cramer-Wold device and verify Condition B.1 for any given $a \in \mathbb{R}^{|\mathcal{T}|}$.

First, we will transform the statistics and verify the zero expectation condition. The orthogonality condition for $\theta_n^{\text{causal,sample}}$ (2.5) implies

$$\sum_{i=1}^{n} R_{n,i}\mathbb{E}\left[\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} \mid \boldsymbol{R}_n\right] = 0. \tag{B.7}$$

Define $U_{n,i} \equiv R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} - \mathbb{E}\left[R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} \mid \boldsymbol{R}_n\right]$. Then, $\widetilde{\Sigma}_n^{-1/2}\sum_{i=1}^{n} R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} = \widetilde{\Sigma}_n^{-1/2}\sum_{i=1}^{n} U_{n,i}$ and we have $\mathbb{E}[U_{n,i} \mid \boldsymbol{R}_n] = 0$ for all i.

By the Cramer-Wold device, it suffices to show that

$$\frac{\sum_{i=1}^{n} a' U_{n,i}}{\sqrt{a' \widetilde{\Sigma}_n a}} \xrightarrow{d^R} \mathrm{N}(0,1)$$

for any $a \in \mathbb{R}^{d_{\tilde{T}}}$ with $a'a = 1$.

By Lemmas B.2 and B.8, $a'U_{n,i}$ is conditionally ψ-dependent with $\xi_{n,s} = \mathbb{1}\{s \leq 2K\}$ given $\boldsymbol{R}_n$. The other conditions are assumed in Assumption 2.8 or automatically satisfied under the local dependence (Assumption 2.5). □

Lemma B.14. *Under Assumptions 2.1 to 2.8,*

$$\Sigma_n^{-1/2} \sum_{i=1}^{n} R_{n,i} X_{n,i} \varepsilon_{n,i} \xrightarrow{d} \mathrm{N}(0, I_{d_T}).$$

Proof. An orthogonality condition for θ_n^{causal} (2.4) implies

$$\sum_{i=1}^{n} \mathbb{E}\left[X_{n,i} \varepsilon_{n,i}\right] = 0. \tag{B.8}$$

By Assumptions 2.2 and 2.7 (i),

$$\begin{aligned} X_{n,i}\varepsilon_{n,i} &= X_{n,i}(Y_{n,i} - X_{n,i}'\theta_n^{\text{causal}} - Z_{n,i}'\gamma_n^{\text{causal}}) \\ &= X_{n,i}T_{n,i}'\theta_{n,i} + X_{n,i}\nu_{n,i} - X_{n,i}X_{n,i}'\theta_n^{\text{causal}} - X_{n,i}Z_{n,i}'\gamma_n^{\text{causal}}. \end{aligned}$$

By Lemma B.6 and Assumption 2.7 (ii), $R_{n,i}$ enters only multiplicatively for $T_{n,i}$ and $X_{n,i} = T_{n,i} - \mathbb{E}[T_{n,i}|\boldsymbol{R}_n]$. By Assumption 2.7 (ii), each element of $Z_{n,i}$ is multiplicatively in $R_{n,i}$. Thus, each element of $X_{n,i}\varepsilon_{n,i}$ is multiplicatively in

$R_{n,i}$ by $R_{n,i}^2 = R_{n,i}$. Combining it with the orthogonality,

$$\sum_{i=1}^{n} \mathbb{E}[R_{n,i}X_{n,i}\varepsilon_{n,i}] = 0.$$

Define $U_{n,i} = R_{n,i}X_{n,i}\varepsilon_{n,i} - \mathbb{E}[R_{n,i}X_{n,i}\varepsilon_{n,i}]$. Then, we have

$$\begin{aligned}\Sigma_n^{-1/2}\sum_{i=1}^{n} R_{n,i}X_{n,i}\varepsilon_{n,i} &= \Sigma_n^{-1/2}\sum_{i=1}^{n} U_{n,i} + \Sigma_n^{-1/2}\sum_{i=1}^{n}\mathbb{E}[R_{n,i}X_{n,i}\varepsilon_{n,i}] \\ &= \Sigma_n^{-1/2}\sum_{i=1}^{n} U_{n,i},\end{aligned}$$

and $\mathbb{E}[U_{n,i}] = 0$.

The remaining parts of the proof are similar to Lemma B.13. □

Lemma B.15. *Under Assumptions 2.1 to 2.6,*

$$\widehat{\gamma}_n - \gamma_n^{\text{causal,sample}} \xrightarrow{p^R} 0.$$

If we assume Assumption 2.7 additionally,

$$\widetilde{\gamma}_n - \gamma_n^{\text{causal}} \xrightarrow{p} 0,$$

where $\widetilde{\gamma}_n$ *is defined in (B.1).*

Proof. We can show $\widehat{\gamma}_n - \gamma_n^{\text{causal,sample}} \xrightarrow{p^R} 0$ by Lemma B.10 as the proof for Theorem 2.2.

Next, we show $\widetilde{\gamma}_n - \gamma_n^{\text{causal}} \xrightarrow{p} 0$. By Lemma B.10, $\widetilde{P}_n^{ZZ} \xrightarrow{p^R} \mathbb{E}[\widetilde{P}_n^{ZZ}|\boldsymbol{R}_n]$ and $\widetilde{P}_n^{ZY} \xrightarrow{p^R} \mathbb{E}[\widetilde{P}_n^{ZY}|\boldsymbol{R}_n]$. By Lemma B.11 and Lemma B.12, $\mathbb{E}[\widetilde{P}_n^{ZZ}|\boldsymbol{R}_n] \xrightarrow{p} \Omega_n^{ZZ}$

and $\mathbb{E}[\widetilde{P}_n^{ZY}|\boldsymbol{R}_n] \xrightarrow{p} \Omega_n^{ZY}$. Thus, we can conclude by the continuous mapping theorem. □

B.3 Proofs

B.3.1 Proof of Theorem 2.1

Proof. Lemma B.6 implies that

$$\Omega_n^{XZ} = 0 = \mathbb{E}[(T_{n,i} - \mathbb{E}[T_{n,i}|\boldsymbol{R}_n])Z'_{n,i}] = 0$$

for large enough n. Similarly,

$$\widetilde{\Omega}_n^{XZ} = \mathbb{E}[\widetilde{X}_{n,i}\widetilde{Z}'_{n,i}|\boldsymbol{R}_n] = \mathbb{E}[(\widetilde{T}_{n,i} - \mathbb{E}[\widetilde{T}_{n,i}|\boldsymbol{R}_n])\widetilde{Z}'_{n,i}|\boldsymbol{R}_n] = 0 \quad \text{a.s.}$$

for large enough n since $\widetilde{Z}_{n,i}$ is measurable with respect to $\sigma(\boldsymbol{R}_n)$.

Therefore, for large enough n,

$$\theta_n^{\text{causal}} = \left(\Omega_n^{XX}\right)^{-1} \Omega_n^{XY},$$

and

$$\theta_n^{\text{causal,sample}} = \left(\widetilde{\Omega}_n^{XX}\right)^{-1} \widetilde{\Omega}_n^{XY} \quad \text{a.s.}$$

They are well-defined under Assumption 2.3. Then, it suffices to show that for large enough n,

$$\mathbb{E}[X_{n,i}Y_{n,i}] = \mathbb{E}[X_{n,i}X'_{n,i}]\theta_{n,i},$$

and

$$\mathbb{E}[\widetilde{X}_{n,i}Y_{n,i}|\boldsymbol{R}_n] = \mathbb{E}[\widetilde{X}_{n,i}X'_{n,i}|\boldsymbol{R}_n]\theta_{n,i} \quad \text{a.s.}$$

The following transformations hold for large enough n:

$$\begin{aligned}
\mathbb{E}[X_{n,i}Y_{n,i}] &= \mathbb{E}[X_{n,i}T'_{n,i}]\theta_{n,i} + \mathbb{E}[X_{n,i}]\nu_{n,i} \\
&= \mathbb{E}[X_{n,i}X'_{n,i}]\theta_{n,i} + \mathbb{E}[X_{n,i}(T_{n,i} - X_{n,i})']\theta_{n,i} \\
&= \mathbb{E}[X_{n,i}X'_{n,i}]\theta_{n,i} + \mathbb{E}[X_{n,i}]Z'_{n,i}\Lambda'_n\theta_{n,i} \\
&= \mathbb{E}[X_{n,i}X'_{n,i}]\theta_{n,i},
\end{aligned}$$

where the first equality holds by Assumption 2.2, the second and the last equalities follow by $\mathbb{E}[X_{n,i}] = 0$, which is implied by Lemma B.6, and the third equality follows by the definition of $X_{n,i}$. Similarly, the following transformations hold almost surely for large enough n:

$$\begin{aligned}
\mathbb{E}[\widetilde{X}_{n,i}Y_{n,i}] &= \mathbb{E}[\widetilde{X}_{n,i}T'_{n,i}|\boldsymbol{R}_n]\theta_{n,i} + \mathbb{E}[\widetilde{X}_{n,i}|\boldsymbol{R}_n]\nu_{n,i} \\
&= \mathbb{E}[\widetilde{X}_{n,i}X_{n,i}|\boldsymbol{R}_n]\theta_{n,i} + \mathbb{E}[\widetilde{X}_{n,i}(T_{n,i} - X_{n,i})'|\boldsymbol{R}_n]\theta_{n,i} \\
&= \mathbb{E}[\widetilde{X}_{n,i}X'_{n,i}|\boldsymbol{R}_n]\theta_{n,i} + \mathbb{E}[\widetilde{X}_{n,i}|\boldsymbol{R}_n]Z'_{n,i}\Lambda'_n\theta_{n,i} \\
&= \mathbb{E}[\widetilde{X}_{n,i}X'_{n,i}|\boldsymbol{R}_n]\theta_{n,i},
\end{aligned}$$

where we used $\mathbb{E}[\widetilde{X}_{n,i}|\boldsymbol{R}_n] = 0$. This completes the proof. $\square$

B.3.2 Proof of Corollary 2.1

Proof. By the population version of the Frisch-Waugh-Lovell theorem,

$$\theta_{n,(k)}^{\text{causal}} = \frac{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} Y_{n,i}]}{\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)}^2]}$$

By the linearity of the model (Assumption 2.2), the numerator can be transformed as

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} Y_{n,i}] \\
&= \sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} T_{n,i}'] \theta_{n,i} + \sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)}] \nu_{n,i} \\
&= \sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} (X_{n,i} + \mathbb{E}[T_{n,i} \mid \boldsymbol{R}_n])'] \theta_{n,i} \\
&= \sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} X_{n,i,(k)}] \theta_{n,i,(k)} + \sum_{i=1}^{n} \mathbb{E}[U_{n,i,(k)} X_{n,i,(-k)}'] \theta_{n,i,(-k)},
\end{aligned}$$

where the second equality holds as $\mathbb{E}[U_{n,i,(k)}] = 0$, which is implied by $\mathbb{E}[X_{n,i}] = 0$, a consequence of Lemma B.6 and the last equality follows from the law of iterated expectations and $\mathbb{E}[X_{n,i} \mid \boldsymbol{R}_n] = 0$.

Similarly,

$$\theta_{n,(k)}^{\text{causal,sample}} = \frac{\sum_{i=1}^{n} \mathbb{E}[\widetilde{U}_{n,i,(k)} Y_{n,i} | \boldsymbol{R}_n]}{\sum_{i=1}^{n} \mathbb{E}[\widetilde{U}_{n,i,(k)}^2 | \boldsymbol{R}_n]}.$$

The numerator is given by

$$\begin{aligned}\sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}Y_{n,i}|\boldsymbol{R}_n] &= \sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}T'_{n,i}|\boldsymbol{R}_n]\theta_{n,i} \\ &= \sum_{i=1}^{n} R_{n,i}\sum_{l=1}^{d_T}\mathbb{E}[\widetilde{U}_{n,i,(k)}X_{n,i,(l)}|\boldsymbol{R}_n]\theta_{n,i,(l)}.\end{aligned}$$

Under $d_T = d_{\widetilde{T}}$, the last equation can be simplified further to

$$\sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n]\theta_{n,i,(k)} + \sum_{i=1}^{n}\sum_{l\neq k} R_{n,i}\mathbb{E}[\widetilde{U}_{n,i,(k)}X_{n,i,(l)}|\boldsymbol{R}_n]\theta_{n,i,(l)}$$

as above. This completes the proof. □

B.3.3 Proof of Corollary 2.2

Proof. By Lemma B.6,

$$\begin{aligned}&\mathbb{E}[\widetilde{X}_{n,i,(k)}X_{n,i,(l)}|\boldsymbol{R}_n] \\ =&\mathbb{E}[(\widetilde{T}_{n,i,(k)} - \mathbb{E}[\widetilde{T}_{n,i,(k)}|\boldsymbol{R}_n])(T_{n,i,(l)} - \mathbb{E}[T_{n,i,(l)}])|\boldsymbol{R}_n] \\ =&\mathbb{E}[(\widetilde{T}_{n,i,(k)} - \mathbb{E}[\widetilde{T}_{n,i,(k)}|\boldsymbol{R}_n])(T_{n,i,(l)} - \mathbb{E}[T_{n,i,(l)}|\boldsymbol{R}_n] \\ &+ \mathbb{E}[T_{n,i,(l)}|\boldsymbol{R}_n] - \mathbb{E}[T_{n,i,(l)}])|\boldsymbol{R}_n] \\ =&\mathbb{E}[(\widetilde{T}_{n,i,(k)} - \mathbb{E}[\widetilde{T}_{n,i,(k)}|\boldsymbol{R}_n])(T_{n,i,(l)} - \mathbb{E}[T_{n,i,(l)}|\boldsymbol{R}_n])|\boldsymbol{R}_n] \\ =&\operatorname{Cov}(\widetilde{T}_{n,i,(k)}, T_{n,i,(l)}|\boldsymbol{R}_n).\end{aligned}$$

Also, by the law of iterated expectations,

$$\begin{aligned}\mathbb{E}[X_{n,i,(k)}X_{n,i,(l)}] &= \mathbb{E}[\mathbb{E}[X_{n,i,(k)}X_{n,i,(l)}|\boldsymbol{R}_n]]\\ &= \mathbb{E}[\mathbb{E}[(T_{n,i,(k)} - \mathbb{E}[T_{n,i,(k)}|\boldsymbol{R}_n])(T_{n,i,(l)} - \mathbb{E}[T_{n,i,(l)}|\boldsymbol{R}_n])]|\boldsymbol{R}_n]]\\ &= \mathbb{E}[\mathrm{Cov}(T_{n,i,(k)}, T_{n,i,(l)}|\boldsymbol{R}_n)].\end{aligned}$$

By Theorem 2.1 and the above equivalences, the no contamination result follows if the covariance condition is satisfied.

Moreover, the numerator of $\theta^{\text{causal,sample}}_{n,(k)}$ is

$$\begin{aligned}&\sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{X}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n]\theta_{n,i,(k)}\\ &+\sum_{i=1}^{n}\sum_{l\in\{1,\cdots,d_T\}\setminus\{k\}} R_{n,i}\mathbb{E}[\widetilde{X}_{n,i,(k)}X_{n,i,(l)}|\boldsymbol{R}_n]\theta_{n,i,(l)}\\ =&\sum_{i=1}^{n} R_{n,i}\mathbb{E}[\widetilde{X}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n]\theta_{n,i,(k)},\end{aligned}$$

and $R_{n,i}\mathbb{E}[\widetilde{X}_{n,i,(k)}X_{n,i,(k)}|\boldsymbol{R}_n] \geq 0$ if we assume that $\mathrm{Cov}(\widetilde{T}_{n,i,(k)}, T_{n,i,(k)}|\boldsymbol{R}_n) \geq 0$. □

B.3.4 Proof of Theorem 2.2

Proof. By Lemma B.6, we have $\mathbb{E}[\widetilde{X}_{n,i}\widetilde{Z}_{n,i}|\boldsymbol{R}_n] = 0$ a.s. for large enough n. Thus, $\widetilde{Q}_n^{ZX}, \widetilde{Q}_n^{XZ} \xrightarrow{a.s.} 0$. Since

$$\begin{pmatrix}\widehat{\theta}_n\\ \widehat{\gamma}_n\end{pmatrix} = \begin{pmatrix}\widetilde{Q}_n^{XX} & \widetilde{Q}_n^{XZ}\\ \widetilde{Q}_n^{ZX} & \widetilde{Q}_n^{ZZ}\end{pmatrix}^{-1}\begin{pmatrix}\widetilde{Q}_n^{XY}\\ \widetilde{Q}_n^{ZY}\end{pmatrix},$$

Lemma B.10 implies that

$$\widehat{\theta}_n - \theta_n^{\text{causal,sample}} = \widehat{\theta}_n - (\widetilde{\Omega}_n^{XX})^{-1}\widetilde{\Omega}_n^{XY} + o_{p^R}(1) \xrightarrow{p^R} 0,$$

which further implies

$$\widehat{\theta}_n - \theta_n^{\text{causal,sample}} \xrightarrow{p} 0.$$

□

B.3.5 Proof of Theorem 2.3

Proof. Since we have already shown Theorem 2.2, it suffices to prove

$$\theta_n^{\text{causal,sample}} - \theta_n^{\text{causal}} \xrightarrow{p} 0.$$

By Lemma B.6, we have $\mathbb{E}[X_{n,i}Z_{n,i}|\boldsymbol{R}_n] = 0$ a.s. and $\mathbb{E}[X_{n,i}Z_{n,i}] = 0$ for large enough n. Thus, for large enough n, $\theta_n^{\text{causal,sample}} = (\widetilde{\Omega}_n^{XX})^{-1}\widetilde{\Omega}_n^{XY}$ a.s. and $\theta_n^{\text{causal}} = (\Omega_n^{XX})^{-1}\Omega_n^{XY}$. Without loss of generality, assume that the first element of $T_{n,i}$ depends on $R_{n,i}D_{n,i}^*$. By Assumption 2.7 (i) and (iii), we can treat the first element of $\theta_{n,(1)}^{\text{causal,sample}}$ and $\theta_{n,(1)}^{\text{causal,sample}}$ the other elements separately as $\theta_{n,(1)}^{\text{causal}} = (\Omega_{n,(1,1)}^{XX})^{-1}\Omega_{n,(1,1)}^{XY}$, $\theta_{n,(-1)}^{\text{causal}} = (\Omega_{n,(-1,-1)}^{XX})^{-1}\Omega_{n,(-1,-1)}^{XY}$, $\theta_{n,(1)}^{\text{causal,sample}} = (\widetilde{\Omega}_{n,(1,1)}^{XX})^{-1}\widetilde{\Omega}_{n,(1,1)}^{XY}$, and $\theta_{n,(-1)}^{\text{causal,sample}} = (\widetilde{\Omega}_{n,(-1,-1)}^{XX})^{-1}\widetilde{\Omega}_{n,(-1,-1)}^{XY}$, where $\Omega_{n,(1,1)}$ is the $(1,1)$ element of Ω_n and $\Omega_{n,(-1,-1)}$ is the submatrix of Ω_n except for its first row and first column. $\widetilde{\Omega}_{n,(1,1)}$ and $\widetilde{\Omega}_{n,(-1,-1)}$ are defined analogously. By Lemma B.11,

$$\theta^{\text{causal,sample}}_{n,(-1)} - \theta^{\text{causal}}_{n,(-1)} = (\widetilde{\Omega}^{XX}_{n,(-1,-1)})^{-1}\widetilde{\Omega}^{XY}_{n,(-1,-1)} - (\Omega^{XX}_{n,(-1,-1)})^{-1}\Omega^{XY}_{n,(-1,-1)} \xrightarrow{p} 0.$$

By Lemma B.12,

$$\theta^{\text{causal,sample}}_{n,(1)} - \theta^{\text{causal}}_{n,(1)} = (\widetilde{\Omega}^{XX}_{n,(1,1)})^{-1}\widetilde{\Omega}^{XY}_{n,(1,1)} - ((1/\rho_n)\Omega^{XX}_{n,(1,1)})^{-1}(1/\rho_n)\Omega^{XY}_{n,(1,1)} \xrightarrow{p} 0.$$

We can conclude by stacking them. $\square$

B.3.6 Proof of Theorem 2.4

Proof. We have $\widetilde{\Omega}^{XZ}_n \xrightarrow{a.s.} 0$ and $(n\rho_n)/N \xrightarrow{a.s.} 1$ under the invertibility and the moment conditions. Thus,

$$\begin{aligned}
&\sqrt{n\rho_n}\begin{pmatrix} \widehat{\theta}_n - \theta^{\text{causal,sample}}_n \\ \widehat{\gamma}_n - \gamma^{\text{causal,sample}}_n \end{pmatrix} \\
&= \begin{pmatrix} \widetilde{Q}^{XX}_n & \widetilde{Q}^{XZ}_n \\ \widetilde{Q}^{ZX}_n & \widetilde{Q}^{ZZ}_n \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sqrt{n\rho_n}}{N}\sum_{i=1}^n R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} \\ \frac{\sqrt{n\rho_n}}{N}\sum_{i=1}^n R_{n,i}\widetilde{Z}_{n,i}\widetilde{\varepsilon}_{n,i} \end{pmatrix} \\
&= \left[\begin{pmatrix} \widetilde{Q}^{XX}_n & O \\ O & \widetilde{Q}^{ZZ}_n \end{pmatrix}^{-1} + o_{p^R}(1)\right] \begin{pmatrix} (1+o_{p^R}(1))\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^n R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} \\ (1+o_{p^R}(1))\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^n R_{n,i}\widetilde{Z}_{n,i}\widetilde{\varepsilon}_{n,i} \end{pmatrix},
\end{aligned}$$

and it suffices to show[2]

$$\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^{n} R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} = O_{p^R}(1), \tag{B.9}$$

$$\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^{n} R_{n,i}\widetilde{Z}_{n,i}\widetilde{\varepsilon}_{n,i} = O_{p^R}(1), \tag{B.10}$$

$$\frac{1}{\sqrt{n\rho_n}}\widetilde{\Sigma}_n^{-1/2} = O_{\text{a.s.}}(1) \tag{B.11}$$

since these conditions imply that

$$\begin{aligned}&\widetilde{\Sigma}_n^{-1/2}\widetilde{Q}_n^{XX}\left(\widehat{\theta}_n - \theta_n^{\text{causal,sample}}\right)\\ =&\frac{1}{\sqrt{n\rho_n}}\widetilde{\Sigma}_n^{-1/2}\widetilde{Q}_n^{XX}\left(\widetilde{Q}_n^{XX}\right)^{-1}\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^{n} R_{n,i}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i} + o_{p^R}(1),\end{aligned}$$

and we can conclude the convergence in conditional distribution with Lemma B.13. The dominated convergence theorem and the law of iterated expectations imply the unconditional result.

We show (B.9)-(B.11). By Chebyshev's inequality, it suffices to show that

[2] A random variable X_n is $O_{p^R}(1)$ if for any $\varepsilon > 0$, there exist some constant $M_\varepsilon < \infty$ such that

$$\mathbb{P}\left(\|X_n\| > M_\varepsilon \mid \boldsymbol{R}_n\right) < \varepsilon \quad \text{a.s.}$$

for large enough n.

its conditional variance is almost surely bounded.

$$
\begin{aligned}
&\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{R_{n,i}}{\sqrt{\rho_n}}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i}\mid \boldsymbol{R}_n\right)\\
=&\frac{1}{n}\sum_{i=1}^{n}\operatorname{Var}\left(\frac{R_{n,i}}{\sqrt{\rho_n}}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i}\mid \boldsymbol{R}_n\right)\\
&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)\setminus\{i\}}\operatorname{Cov}\left(\frac{R_{n,i}}{\sqrt{\rho_n}}\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i},\frac{R_{n,j}}{\sqrt{\rho_n}}\widetilde{X}_{n,j}\widetilde{\varepsilon}_{n,j}\mid \boldsymbol{R}_n\right)\\
\leq&\frac{1}{n}\sum_{i=1}^{n}\frac{R_{n,i}}{\rho_n}\mathbb{E}\left[\widetilde{X}_{n,i}\widetilde{X}_{n,i}'\widetilde{\varepsilon}_{n,i}^2\mid \boldsymbol{R}_n\right] \qquad \text{(B.12)}\\
&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)\setminus\{i\}}\frac{R_{n,i}R_{n,j}}{\rho_n}\Bigg(\mathbb{E}\left[\widetilde{X}_{n,i}\widetilde{X}_{n,j}'\widetilde{\varepsilon}_{n,i}\widetilde{\varepsilon}_{n,j}\mid \boldsymbol{R}_n\right]\\
&-\mathbb{E}\left[\widetilde{X}_{n,i}\widetilde{\varepsilon}_{n,i}\mid \boldsymbol{R}_n\right]\mathbb{E}\left[\widetilde{X}_{n,j}\widetilde{\varepsilon}_{n,j}\mid \boldsymbol{R}_n\right]'\Bigg). \qquad \text{(B.13)}
\end{aligned}
$$

Each element of the first term (B.12) is almost surely bounded by

$$
\left(\frac{N}{n\rho_n}\right)\cdot\max_i\|\widetilde{X}_{n,i}\|^2\cdot\max_i|\widetilde{\varepsilon}_{n,i}|^2.
$$

Thus, the first term (B.12) is $O_{\text{a.s.}}(1)$ by Assumption 2.3 and Lemmas B.5 and B.9. The second term (B.13) is also $O_{\text{a.s.}}(1)$ by a similar argument as the first term and Assumption 2.6. Hence, (B.9) is $O_{p^R}(1)$. Similarly, we can show that (B.10) is $O_{p^R}(1)$. (B.11) is also $O_{\text{a.s.}}(1)$ by the invertibility assumption (Assumption 2.3). □

B.3.7 Proof of Theorem 2.5

Proof. Since $\widetilde{X}_{n,i} = X_{n,i}$ and $\widetilde{Z}_{n,i} = Z_{n,i}$,

$$
\begin{aligned}
&\sqrt{n\rho_n} \begin{pmatrix} \widehat{\theta}_n - \theta_n^{\text{causal}} \\ \widehat{\gamma}_n - \gamma_n^{\text{causal}} \end{pmatrix} \\
&= \begin{pmatrix} \widetilde{Q}_n^{XX} & \widetilde{Q}_n^{XZ} \\ \widetilde{Q}_n^{ZX} & \widetilde{Q}_n^{ZZ} \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} \frac{\sqrt{n\rho_n}}{N} \sum_{i=1}^n R_{n,i} X_{n,i} (Y_{n,i} - X'_{n,i}\theta_n^{\text{causal}} - Z'_{n,i}\gamma_n^{\text{causal}}) \\ \frac{\sqrt{n\rho_n}}{N} \sum_{i=1}^n R_{n,i} Z_{n,i} (Y_{n,i} - X'_{n,i}\theta_n^{\text{causal}} - Z'_{n,i}\gamma_n^{\text{causal}}) \end{pmatrix} \\
&= \left[\begin{pmatrix} \widetilde{Q}_n^{XX} & O \\ O & \widetilde{Q}_n^{ZZ} \end{pmatrix}^{-1} + o_p(1) \right] \\
&\qquad \times \begin{pmatrix} (1 + o_p(1)) \frac{1}{\sqrt{n\rho_n}} \sum_{i=1}^n R_{n,i} X_{n,i} \varepsilon_{n,i} \\ (1 + o_p(1)) \frac{1}{\sqrt{n\rho_n}} \sum_{i=1}^n R_{n,i} Z_{n,i} \varepsilon_{n,i} \end{pmatrix}.
\end{aligned}
$$

By an argument similar to the proof of Theorem 2.4, we can show that

$$\frac{1}{\sqrt{n\rho_n}} \sum_{i=1}^n R_{n,i} X_{n,i} \varepsilon_{n,i} = O_p(1), \tag{B.14}$$

$$\frac{1}{\sqrt{n\rho_n}} \sum_{i=1}^n R_{n,i} Z_{n,i} \varepsilon_{n,i} = O_p(1), \tag{B.15}$$

$$\frac{1}{\sqrt{n\rho_n}} \Sigma_n^{-1/2} = O_p(1). \tag{B.16}$$

Thus, (B.14) to (B.16) imply that

$$\begin{aligned}
&\Sigma_n^{-1/2}\widetilde{Q}_n^{XX}\left(\widehat{\theta}_n-\theta_n^{\text{causal}}\right)\\
&=\frac{1}{\sqrt{n\rho_n}}\Sigma_n^{-1/2}\widetilde{Q}_n^{XX}\left(\widetilde{Q}_n^{XX}\right)^{-1}\frac{1}{\sqrt{n\rho_n}}\sum_{i=1}^{n}R_{n,i}X_{n,i}\varepsilon_{n,i}+o_p(1),
\end{aligned}$$

and we can conclude with Lemma B.14. □

B.3.8 Proof of Theorem 2.6

Proof. **[Proof for $\frac{1}{n\rho_n}\widetilde{\Sigma}_n$]**

Let

$$\begin{aligned}
&\frac{1}{n\rho_n}\widetilde{\Sigma}_n^{\dagger}\\
&\qquad=\frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)}R_{n,i}R_{n,j}\\
&\qquad\times\left(\widetilde{\Psi}_{n,i}-\mathbb{E}\left[\widetilde{\Psi}_{n,i}\mid\boldsymbol{R}_n\right]\right)\times\left(\widetilde{\Psi}_{n,j}-\mathbb{E}\left[\widetilde{\Psi}_{n,j}\mid\boldsymbol{R}_n\right]\right)'.
\end{aligned}$$

Under the boundedness (Assumption 2.3) and the local dependence (Assumption 2.5), Condition B.2 is automatically satisfied. Then, Lemma B.3 implies that

$$\frac{1}{n\rho_n}\widetilde{\Sigma}_n^{\dagger}=\frac{1}{n\rho_n}\widetilde{\Sigma}_n+o_{p^R}(1).$$

Hence, it suffices to show that

$$\frac{1}{N}\widehat{\Sigma}_n=\frac{1}{n\rho_n}\widetilde{\Sigma}_n^{\dagger}+\widetilde{B}_n+o_{p^R}(1).$$

Here, $\max_i|\widehat{\varepsilon}_{n,i}-\widetilde{\varepsilon}_{n,i}|=o_p^R(1)$ by Assumption 2.3, Theorem 2.2, and

Lemma B.15. Also,

$$\frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\widetilde{X}_{n,i}\widetilde{X}'_{n,j}\widehat{\varepsilon}_{n,i}\widehat{\varepsilon}_{n,j} = O_{\text{a.s.}}(1)$$

by Assumptions 2.3 and 2.6, $\rho \in (0,1]$, and Lemma B.9. Thus, we can show that

$$\begin{aligned}
\frac{1}{N}\widehat{\Sigma}_n =& \frac{1}{N}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\widehat{\Psi}_{n,i}\widehat{\Psi}'_{n,j} \\
=& \frac{1}{N}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\widetilde{X}_{n,i}\widetilde{X}'_{n,j}\widehat{\varepsilon}_{n,i}\widehat{\varepsilon}_{n,j} \\
=& \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\widetilde{X}_{n,i}\widetilde{X}'_{n,j}\widetilde{\varepsilon}_{n,i}\widetilde{\varepsilon}_{n,j} + o_{p^R}(1), \qquad \text{(B.17)}
\end{aligned}$$

where the last equality holds by Lemma B.5.

Then,

$$\begin{aligned}
(B.17) =& \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\widetilde{\mathcal{N}}_n(i,2K)} R_{n,i}R_{n,j}\widetilde{\Psi}_{n,i}\widetilde{\Psi}'_{n,j} + o_{p^R}(1) \\
=& \frac{1}{n\rho_n}\widetilde{\Sigma}_n^{\dagger} + \widetilde{B}_n + o_{p^R}(1) \qquad \text{(B.18)} \\
&+ \frac{2}{n\rho_n}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{n,i}R_{n,j} \\
&\times \left(\widetilde{\Psi}_{n,i} - \mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\right)\mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]' \mathbb{1}\{\widetilde{d}_n(i,j) \le 2K\}, \qquad \text{(B.19)}
\end{aligned}$$

thus, it suffices to show that the remainder term $(B.19) = o_{p^R}(1)$.

We will show it element-wise. Take the (k,k')-element of (B.19). Let

$$\widetilde{\varphi}_i = \sum_{j=1}^{n} R_{n,j}\mathbb{E}\left[\widetilde{\Psi}_{n,j,(k')} \mid \boldsymbol{R}_n\right] \mathbb{1}\{\widetilde{d}_n(i,j) \leq 2K\}.$$

Then,

$$\begin{aligned}
&\mathbb{E}\left[\left|(k,k')\text{-element of } (B.19)\right| \mid \boldsymbol{R}_n\right] \\
=&\mathbb{E}\left[\left|\frac{2}{n\rho_n}\sum_{i=1}^{n} R_{n,i}\left(\widetilde{\Psi}_{n,i} - \mathbb{E}\left[\widetilde{\Psi}_{n,i,(k)} \mid \boldsymbol{R}_n\right]\right)\widetilde{\varphi}_i\right| \mid \boldsymbol{R}_n\right] \\
\leq&\mathbb{E}\left[\left(\frac{2}{n\rho_n}\sum_{i=1}^{n} R_{n,i}\left(\widetilde{\Psi}_{n,i} - \mathbb{E}\left[\widetilde{\Psi}_{n,i,(k)} \mid \boldsymbol{R}_n\right]\right)\widetilde{\varphi}_i\right)^2 \mid \boldsymbol{R}_n\right]^{1/2} \\
\leq&\frac{2}{\rho_n}\left(\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(\widetilde{\Psi}_{n,i,(k)} \mid \boldsymbol{R}_n\right)\widetilde{\varphi}_i^2\right. \\
&\left. + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\left|\mathrm{Cov}\left(\widetilde{\Psi}_{n,i,(k)}, \widetilde{\Psi}_{n,j,(k)} \mid \boldsymbol{R}_n\right)\right| \times |\widetilde{\varphi}_i\widetilde{\varphi}_j|\right)^{1/2},
\end{aligned}$$

where the first inequality follows from Jensen's inequality.

By Assumption 2.3 and Lemma B.9, $\widetilde{\Psi}_{n,i,(k)}$ is uniformly bounded, thus $\max_i \mathrm{Var}\left(\widetilde{\Psi}_{n,i,(k)} \mid \boldsymbol{R}_n\right) = O_{\text{a.s.}}(1)$ and $\widetilde{\varphi}_i^2 \leq C \times (\sum_{j=1}^{n}\mathbb{1}\{\widetilde{d}_n(i,j) \leq 2K\})^2 \leq C \times |\mathcal{N}_n(i;2K)|^2$ for some constant $C > 0$. Hence, $\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(\widetilde{\Psi}_{n,i,(k)} \mid \boldsymbol{R}_n\right)$ $\widetilde{\varphi}_i^2 \leq C'\delta_n(2K,2)/n$ for some constant $C' > 0$. By Assumption 2.9 (i), $\delta_n(2K,2)/n \to 0$ as $n \to \infty$.

By Lemma B.8, $\widetilde{\Psi}_{n,i,(k)}$ is conditionally ψ-dependent with $\xi_{n,s} = \mathbb{1}\{s \leq 2K\}$ given $\boldsymbol{R}_n$, thus $\left|\mathrm{Cov}\left(\widetilde{\Psi}_{n,i,(k)}, \widetilde{\Psi}_{n,j,(k)} \mid \boldsymbol{R}_n\right)\right| \leq C''\sum_{s=1}^{\infty}\mathbb{1}\{s \leq 2K\} \times$

$\mathbb{1}\{d_n(i,j)=s\}$ for some constant $C''>0$. Thus,

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\left|\mathrm{Cov}\left(\widetilde{\Psi}_{n,i,(k)},\widetilde{\Psi}_{n,j,(k)}\mid \boldsymbol{R}_n\right)\right|\times|\widetilde{\varphi}_i\widetilde{\varphi}_j|\\ \leq&\frac{C''}{n^2}\sum_{s=1}^{2K}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{1}\{d_n(i,j)=s\}\sum_{i'\in\mathcal{N}(i,2K)}\sum_{j'\in\mathcal{N}(j,2K)}1\\ \leq&\frac{C'''}{n^2}\sum_{s=1}^{2K}|\mathcal{J}_n(s,2K)|\end{aligned}$$

for some constant $C'''>0$. By Assumption 2.9 (ii), $\sum_{s=1}^{2K}\mathcal{J}_n(s,2K)/n^2\to 0$ as $n\to\infty$.

Therefore, we have shown that

$$\mathbb{E}\left[\left|(k,k')\text{-element of }(B.19)\right|\mid \boldsymbol{R}_n\right]=o_{\text{a.s.}}(1).$$

By Markov's inequality, we can conclude that the remainder term $(B.19)=o_{p^R}(1)$.

[Proof for $\frac{1}{n\rho_n}\Sigma_n$] Let

$$\frac{1}{n\rho_n}\Sigma_n^{\dagger}=\frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i,2K)}\left(R_{n,i}\Psi_{n,i}-\rho_n\mathbb{E}[\Psi_{n,i}]\right)\left(R_{n,j}\Psi_{n,j}-\rho_n\mathbb{E}[\Psi_{n,j}]\right)'.$$

We can show that

$$\begin{aligned}
\frac{1}{N}\widehat{\Sigma}_n =& \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_n(i,2K)} R_{n,i}\Psi_{n,i}R_{n,j}\Psi'_{n,j} + o_p(1) \\
=& \frac{1}{n\rho_n}\Sigma_n^{\dagger} + \widehat{B}_n + o_p(1) \\
&+ \frac{2}{n\rho_n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(R_{n,i}\Psi_{n,i} - \rho_n\mathbb{E}[\Psi_{n,i}]\right)\rho_n\mathbb{E}[\Psi_{n,j}]\mathbb{1}\{d_n(i,j)\leq 2K\} \\
=& \frac{1}{n\rho_n}\Sigma_n^{\dagger} + \widehat{B}_n + o_p(1),
\end{aligned}$$

where the first equality follows by the similar arguments as we derive (B.17) and by Lemma B.7, the second equality is a simple transformation, and the last equality holds by the similar arguments for the remainder term (B.19). We can conclude by applying Lemma B.3 to $(n\rho_n)^{-1}\Sigma_n^{\dagger}$. □

B.3.9 Proof of Theorem 2.7

Proof. Let $\frac{1}{N}\widehat{\Sigma}_n^{-} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{n,i}R_{n,j}\widehat{\Psi}_{n,i}\widehat{\Psi}'_{n,j}\widetilde{K}^{-}_{n,i,j}$. Since $\widetilde{K}_n^{+} = \widetilde{K}_n + \widetilde{K}_n^{-}$, we have

$$\frac{1}{N}\widehat{\Sigma}_n^{+} = \frac{1}{N}\widehat{\Sigma}_n + \frac{1}{N}\widehat{\Sigma}_n^{-}. \tag{B.20}$$

[Proof for $\frac{1}{n\rho_n}\widetilde{\Sigma}_n$]

Theorem 2.6 implies

$$
\begin{aligned}
&\frac{1}{N}\widehat{\Sigma}_n \\
&= \frac{1}{n\rho_n}\widetilde{\Sigma}_n + \widetilde{B}_n + o_{p^R}(1) \\
&= \frac{1}{n\rho_n}\widetilde{\Sigma}_n \\
&+ \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{n,i}R_{n,j}\mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]'\left(\widetilde{K}^{+}_{n,i,j} - \widetilde{K}^{-}_{n,i,j}\right) \\
&+ o_{p^R}(1).
\end{aligned}
$$

By the same logic as in the proof of Theorem 2.6 after replacing $\mathbb{1}\{\widetilde{d}_n(i,j) \leq 2K\}$ by $\widetilde{K}^{-}_{n,i,j}$ and Assumption 2.9 by Assumption 2.10, we can show that

$$
\begin{aligned}
&\frac{1}{N}\widehat{\Sigma}^{-}_n \\
&= \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{n,i}R_{n,j}\mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]'\widetilde{K}^{-}_{n,i,j} \\
&+ \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{n,i}R_{n,j}\mathbb{E}\Bigg[\left(\widetilde{\Psi}_{n,i} - \mathbb{E}\left[\widetilde{\Psi}_{n,i} \mid \boldsymbol{R}_n\right]\right) \\
&\times \left(\widetilde{\Psi}_{n,j} - \mathbb{E}\left[\widetilde{\Psi}_{n,j} \mid \boldsymbol{R}_n\right]\right)' \mid \boldsymbol{R}_n\Bigg]\widetilde{K}^{-}_{n,i,j} \\
&+ o_{p^R}(1).
\end{aligned}
$$

We get the conclusion by substituting these results into (B.20).

[Proof for $\frac{1}{n\rho_n}\Sigma_n$]

The proof is similar. By the same logic as in the proof of Theorem 2.6,

$$\begin{aligned}
&\frac{1}{N}\widehat{\Sigma}_n^{-} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_n\mathbb{E}\left[\Psi_{n,i}\right]\mathbb{E}\left[\Psi_{n,j}\right]' K_{n,i,j}^{-} \\
&+ \frac{1}{n\rho_n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[\left(R_{n,i}\Psi_{n,i} - \rho_n\mathbb{E}\left[\Psi_{n,i}\right]\right)\left(R_{n,j}\Psi_{n,j} - \rho_n\mathbb{E}\left[\Psi_{n,j}\right]\right)'\right] K_{n,i,j}^{-} \\
&+ o_p(1).
\end{aligned}$$

We get the conclusion by combining it with the result of Theorem 2.6. □

B.4 Additional Simulation Results

We consider the following exposure mapping:

$$T_{n,i} = \left(R_{n,i}D_{n,i}^{*}, \sum_{j\neq i} A_{n,i,j}R_{n,j}D_{n,j}^{*}\right) =: (D_{n,i}, \mathrm{net}_{n,i}).$$

We set $\widetilde{T}_{n,i} = T_{n,i}$. Note that, since $D_{n,i} \perp\!\!\!\perp \mathrm{net}_{n,i}$, no contamination bias would arise. Our focus here is to evaluate our inference procedure based on the asymptotic approximation in this correctly specified model.

We follow the same implementation procedure as in the simulation exercise in Section 2.5, except for the definition of $T_{n,i}$ and $\widetilde{T}_{n,i}$, and $\theta_{n,i,(1)} \sim$ Exponential(1/3) and $\theta_{n,i,(2)} = \frac{\sum_{j\neq i} A_{n,i,j}}{\max_k \sum_{j\neq k} A_{n,k,k}}$. Here, the average direct effect is 1/3 and the average spillover effect is about 2/9.

In Table B.1, we report the results of this simulation when we vary $\rho_n \in \{0.1, 0.5, 1.0\}$. Since the population size (the number of nodes) is 1770, the sample size varies from about 177 to 1770. In each panel, the first three rows report the averages of the population and sample-level causal estimands and the OLS estimator. The fourth to sixth rows report the averages of the EHW standard errors and the averages of our proposed standard errors in Equation (2.13). The seventh and eighth rows report the average absolute deviations of the estimator from the causal estimands. The last four rows report the coverage probabilities of the 95% confidence intervals constructed using the EHW standard errors and those based on (2.13) for the two causal estimands.

The first three rows in Table B.1 show that the estimator closely approximates both estimands, as expected from our asymptotic theory (Theorems 2.2 and 2.3). The difference between θ_n^{causal} and $\theta_n^{\text{causal,sample}}$ is negligible because $T_{n,i} = \widetilde{T}_{n,i}$. We also observe that while the direct effect estimands $\theta_{(1)}^{\text{causal}}$ and $\theta_{(1)}^{\text{causal,sample}}$ are close to the average direct effect of 1/3, the spillover effect estimands $\theta_{(2)}^{\text{causal}}$ and $\theta_{(2)}^{\text{causal,sample}}$ are larger than the average spillover effect of 2/9. This occurs because the spillover effect estimands place greater weight on nodes with more connections, who tend to have larger spillover effects, resulting in an upward bias. The seventh and eighth rows, showing the average absolute deviations of the estimator from the estimands, also confirm that the estimator closely approximates the estimands, especially as ρ_n increases and the sample size becomes larger.

The fourth to sixth rows show that our proposed standard errors based on (2.13) tend to be larger than the EHW standard errors, especially as ρ_n increases. This is because (i) the EHW standard errors do not account for the network dependence structure, and the observed network becomes denser as ρ_n increases, and (ii) our standard errors are designed to be conservative, as established in Theorem 2.7. When ρ_n is small, the difference between the two types of standard errors is less pronounced because (i) the observed network is sparser and the dependence structure is less important, and (ii) the sample-to-population ratio approaches the infinite population case, where the standard model-based inference is valid. Additionally, we observe that our proposed standard errors based on (2.13) for θ_n^{causal} tend to be slightly larger than those for $\theta_n^{\text{causal,sample}}$, reflecting the additional adjustment for sampling variation in the former.

The last two rows in Table B.1 show that the coverage rates based on our proposed method (2.13) are reasonably close to the nominal 95% target. We observe under-coverage for $\theta_{n,(2)}^{\text{causal}}$ and $\theta_{n,(2)}^{\text{causal,sample}}$ when ρ_n is small, likely due to the small sample size and limited variation in the net variable in sparse networks. In contrast, the coverage rates for $\theta_{n,(2)}^{\text{causal}}$ and $\theta_{n,(2)}^{\text{causal,sample}}$ based on the EHW standard errors are substantially below the nominal level as ρ_n increases. This is because the EHW standard errors ignore the network dependence structure and finite population bias, which likely leads to over-rejection of the null hypothesis.

Overall, our simulation exercise shows that as long as the model is correctly specified and relevant network information is observed, reliable inference for the causal estimands is possible even when not everyone in the population is sampled. Since exhaustive network collection can be costly in practice, our results provide a rationale for collecting network data based on sampled units, which is less costly.

Table B.1: Simulation Results: $T_{n,i} = \widetilde{T}_{n,i}$ case

	$\rho = 0.1$		$\rho = 0.5$		$\rho = 1.0$	
	D	net	D	net	D	net
θ^{causal}	0.348	0.312	0.348	0.312	0.348	0.312
$\theta^{\text{causal,sample}}$	0.346	0.311	0.349	0.312	0.348	0.312
$\widehat{\theta}$	0.347	0.285	0.350	0.305	0.350	0.307
SE EHW	0.214	0.265	0.126	0.093	0.100	0.058
SE mod, θ^{causal}	0.214	0.263	0.132	0.109	0.110	0.083
SE mod, $\theta^{\text{causal,sample}}$	0.214	0.263	0.132	0.110	0.110	0.085
$\lvert\widehat{\theta} - \theta^{\text{causal}}\rvert$	0.172	0.233	0.097	0.093	0.084	0.067
$\lvert\widehat{\theta} - \theta^{\text{causal,sample}}\rvert$	0.172	0.232	0.096	0.092	0.084	0.066
Cov. EHW, θ^{causal}	0.945	0.920	0.953	0.879	0.936	0.816
Cov. EHW, $\theta^{\text{causal,sample}}$	0.948	0.914	0.954	0.886	0.940	0.822
Cov. mod, θ^{causal}	0.941	0.904	0.963	0.931	0.962	0.956
Cov. mod, $\theta^{\text{causal,sample}}$	0.946	0.907	0.963	0.928	0.966	0.959

Note: This table reports simulation results for selected values $\rho = 0.1, 0.5, 1.0$. The first three rows report the averages of the population-level causal estimand θ^{causal}, the sample-level causal estimand $\theta^{\text{causal,sample}}$, and the OLS estimator $\widehat{\theta}$. The next three rows report the average standard errors: the Eicker–Huber–White (EHW) standard errors and the proposed standard errors based on (2.13). The following two rows report the average absolute deviations of the estimator from the two causal estimands. The final four rows report the coverage probabilities of the 95% confidence intervals constructed using the EHW standard errors and the proposed standard errors for both causal estimands.

B.5 Survey of OLS usage in network experiment applications

In this section, we summarize our survey of the usage of OLS in network experiment applications in economics, as introduced in the second paragraph of the introduction. Our survey provides an overview of the prevalence of OLS in estimating spillover effects in network experiments.

We considered papers published from April 2010 through April 2025 in the following journals: American Economic Review, Econometrica, Quarterly Journal of Economics, Journal of Political Economy, Review of Economic Studies, American Economic Journal: Applied Economics, and Journal of

Development Economics. We searched for articles that included both "networks" and either "field experiments" or "randomized trial" as keywords on the Web of Science platform. This search yielded 52 papers, as listed in Table B.2. We then reviewed each paper to determine whether it conducted a network experiment and estimated spillover effects using regression. Among these, 29 papers ran regressions to estimate spillover effects; all 29 used the OLS estimator, while only two papers (Coutts, 2022 and Fafchamps and Vicente, 2013) mentioned propensity scores or used related estimators.

Table B.2: Survey of OLS usage in network experiment applications

Citation	Field/Lab Exp w/ Network?	Regression for Causal Effects?	Estimator(s) Used
Evsyukova et al. (2024)	Yes	Yes	OLS, Causal Forest
Batista et al. (2025)	No	Yes	OLS
Karing (2024)	No	Yes	OLS, Logit
Chegere et al. (2024)	Yes	Yes	OLS
Deutschmann et al. (2024)	Yes	Yes	OLS
Barsbai et al. (2024)	No	Yes	OLS
Banerjee et al. (2024)	No	Yes	OLS, IV
Colonnelli et al. (2024)	No	Yes	OLS, DiD
Hernandez-Agramonte et al. (2024)	No	Yes	OLS, IPW, Logit
Borusyak and Hull (2023b)	No	Yes	OLS, 2SLS

Continued on next page

Citation	Field/Lab Exp w/ Network?	Regression?	Estimator(s) Used
Banerjee et al. (2023)	Yes	Yes	OLS
Soldani et al. (2023)	Yes	Yes	OLS
Bobonis et al. (2022)	No	Yes	OLS, IV
Alan et al. (2022)	Yes	Yes	OLS
Coutts (2022)	Yes	Yes	Propensity score matching, OLS
Leung (2022b)	No (method)	-	-
Bjorkegren and Karaca (2022)	Yes	No, Structural	OLS
Beaman et al. (2021b)	Yes	Yes	OLS
Hess et al. (2021)	Yes	Yes	OLS
Meghir et al. (2022)	No	Yes	OLS
Carter et al. (2021b)	Yes	Yes	OLS,
Hardy and McCasland (2021)	Yes	Yes	OLS
Breza et al. (2020)	No (method)	-	-
Abel et al. (2020)	No	Yes	OLS
Afridi et al. (2020)	Yes	Yes	OLS
Drago et al. (2020)	Yes	Yes	OLS
BenYishay et al. (2020)	Yes	Yes	OLS
Cai (2020)	No	Yes	OLS, Propensity Score Matching
Banerjee et al. (2019)	Yes	Yes	OLS

Continued on next page

Citation	Field/Lab Exp w/ Network?	Regression?	Estimator(s) Used
Kandpal and Baylis (2019)	No (natural experiment)	Yes	OLS, IV
Benyishay and Mobarak (2019)	Yes	Yes	OLS
Boltz et al. (2019)	Yes	Yes	OLS, Logit
Breza and Chandrasekhar (2019)	Yes	Yes	OLS
Flory (2018)	Yes	Yes	OLS
Chandrasekhar et al. (2018)	Yes	Yes	OLS
Cai and Szeidl (2018)	Yes	Yes	OLS
Di Falco et al. (2018)	Yes	Yes	OLS
Gine and Mansuri (2018)	No (cluster)	Yes	OLS, IV
Kessler (2017)	No	Yes	OLS,
Cruz et al. (2017)	No	Yes	OLS, IV,
Barnhardt et al. (2017)	Yes	Yes	OLS
Belloni et al. (2017)	No (method)	-	-
Pallais and Sands (2016)	No	Yes	OLS
Alatas et al. (2016)	Yes	Yes	OLS
Nagavarapu and Sekhri (2016)	No	Yes	OLS
Levine et al. (2016)	No	Yes	OLS
Jakiela and Ozier (2016)	Yes	Yes	OLS
Cai et al. (2015b)	Yes	Yes	OLS

Continued on next page

Citation	Field/Lab Exp w/ Network?	Regression?	Estimator(s) Used
Callen and Long (2015)	No	Yes	OLS
Fafchamps and Vicente (2013)	Yes	Yes	OLS, Propensity score matching
Robinson (2012)	No	Yes	OLS
Godlonton and Thornton (2012)	Yes	Yes	OLS

Notes: The first column lists the citation of the paper. The second column indicates whether the paper uses a field or lab experiment with a network structure. The third column indicates whether the paper uses regression to estimate causal effects, and the fourth column lists the specific estimator(s) used in the regression analysis. Methodological papers are marked with "No (method)" in the second column and do not have the third and fourth columns filled in.

Appendix C

Chapter 3

C.1 Proofs of Theorems and Propositions

C.1.1 Proof of Theorem 3.1

Proof. First, consider the infeasible version of β_n, where $\widehat{\gamma}_n$ is replaced by the true γ:

$$\widetilde{\beta}_n = \beta + S_{WW}^{-1} S_{W\lambda} + S_{WW}^{-1} S_{W\nu},$$

where $S_{WW}, S_{W\lambda}$, and $S_{W\nu}$ are the same as $\widehat{S}_{WW}, \widehat{S}_{W\lambda}$, and $\widehat{S}_{W\nu}$ except $\widehat{\gamma}_n$ replaced by γ. We use the following lemmas

Lemma C.1. *Suppose Assumptions 3.1-3.9 hold. Then,*

$$S_{WW} \to_p \Sigma_{WW},$$

as $n \to \infty$.

Lemma C.2. *Suppose Assumptions 3.1-3.9 hold. Fix some* $h \in [0, \infty)$*. If* $Nh_n^{2k+3} \to h$*, then*

$$\sqrt{Nh_n} S_{W\lambda} \to_p \sqrt{h} \Sigma_{W\lambda},$$

as $n \to \infty$. *If* $Nh_n^{2k+3} \to \infty$ *and* $nh_n^{2k} \to \infty$, *then*

$$h_n^{-(k+1)} S_{W\lambda} \to_p \Sigma_{W\lambda},$$

as $n \to \infty$.

Lemma C.3. *Suppose Assumptions 3.1-3.9 hold. Fix an arbitrary non-zero vector* $c \in \mathbb{R}^{q_w}$ *and some constant* $h \in [0, \infty)$. *Let* $c_W = \Sigma_{WW}^{-1} c$. *If* $c_W' \Sigma_{W\nu,1} c_W > 0$ *and* $Nh_n^{2k+3} \to h$, *then*

$$\sqrt{n} c' S_{WW}^{-1} S_{W\nu} \to_d \mathcal{N}(0, c_W' \Sigma_{W\nu,1} c_W),$$

as $n \to \infty$. *If* $c_W' \Sigma_{W\nu,1} c_W = 0$ *and* $Nh_n^{2k+3} \to h$, *then*

$$\sqrt{Nh_n} c' S_{WW}^{-1} S_{W\nu} \to_d \mathcal{N}(0, c_W' \Sigma_{W\nu,2} c_W),$$

as $n \to \infty$.

By combining Lemmas C.1-C.3, the statement of Theorem 3.1 follows for $\widetilde{\beta}_n$. The following lemmas are used to show the negligibility of $\widehat{\beta}_n - \widetilde{\beta}_n$:

Lemma C.4. *Suppose Assumptions 3.1-3.10 hold. Fix some constant* $h \in [0, \infty)$. *If* $Nh_n^{2k+3} \to h$, *then,*

$$\widehat{S}_{WW} = S_{WW} + o_p(1).$$

Lemma C.5. *Suppose Assumptions 3.1-3.10 hold. Fix some constant* $h \in [0, \infty)$*. If* $Nh_n^{2k+3} \to h$*, then*

$$\widehat{S}_{W\lambda} = S_{W\lambda} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

Lemma C.6. *Suppose Assumptions 3.1-3.10 hold. Fix some constant* $h \in [0, \infty)$*. If* $Nh_n^{2k+3} \to h$*, then*

$$\widehat{S}_{W\nu} = S_{W\nu} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

By combining Lemmas C.4-C.6, we have

$$\widehat{\beta}_n - \beta = \widetilde{\beta}_n - \beta + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

Thus, the normalization $r_n \in \{\sqrt{n}, \sqrt{Nh_n}, h_n^{-(k+1)}\}$ corresponding to each case results in

$$r_n(\widehat{\beta}_n - \beta) = r_n(\widetilde{\beta}_n - \beta) + o_p(1).$$

Since $\widetilde{\beta}_n$ satisfies the statement of Theorem 3.1, this completes the proof. □

C.1.2 Proof of Proposition 3.1

Proof. We show the claim by the following steps.

C.1.2.1 Step 1: $\widehat{\Sigma}_{W\nu,2} \to_p \Sigma_{W\nu,2}$

By expanding $K^2(\Delta R'_{ij}\widehat{\gamma}_n/h_n)$ around $\Delta R'_{ij}\gamma$, we get

$$\begin{aligned}&K^2(\Delta R'_{ij}\widehat{\gamma}_n/h_n)\\&= K^2(\Delta R'_{ij}\gamma/h_n) + 2\Delta R'_{ij}(\widehat{\gamma}_n - \gamma)/h_n k(c^*_{ij,n}/h_n)K(c^*_{12n}/h_n),\end{aligned}$$

where $c^*_{ij,n}$ is between $\Delta R'_{ij}\gamma$ and $\Delta R'_{ij}\widehat{\gamma}_n$ and $k(\cdot)$ is the derivative of $K(\cdot)$. Then,

$$\begin{aligned}\widehat{\Sigma}_{W\nu,2} &= \underbrace{\frac{1}{Nh_n}\sum_{i<j} K^2(\Delta R'_{12}\gamma/h_n)d_{ij}\Delta W_{ij}\Delta W'_{ij}\Delta\widehat{\varepsilon}^2_{ij}}_{D_{p1,1}}\\&\quad+ \underbrace{\frac{2}{Nh_n^2}\sum_{i<j}\Delta R'_{ij}(\widehat{\gamma}_n - \gamma)k(c^*_{ijn}/h_n)K(c^*_{ijn}/h_n)\Delta W_{ij}\Delta W'_{ij}\Delta\widehat{\varepsilon}^2_{ij}}_{D_{p1,2}}.\end{aligned}$$

C.1.3 Sub-Step 1: $D_{p1,1} \to_p \Sigma_{W\nu,2}$

Observe that

$$\begin{aligned}\Delta\widehat{\varepsilon}^2_{ij} - \nu^2_{ij} &= \left(\Delta W'_{ij}(\beta - \widehat{\beta}_n)\right)^2 + \lambda^2_{ij} + 2\Delta W'_{ij}(\beta - \widehat{\beta}_n)\lambda_{ij}\\&\quad+ 2\Delta W'_{ij}(\beta - \widehat{\beta}_n)\nu_{ij} + 2\lambda_{ij}\nu_{ij}.\end{aligned}$$

Thus,

$$
\begin{aligned}
D_{p1,1} &= \frac{1}{Nh_n}\sum_{i<j} K^2(\Delta R_{ij}'\gamma/h_n)d_{ij}\Delta W_{ij}\Delta W_{ij}'\nu_{ij}^2 \\
&+ \frac{1}{Nh_n}\sum_{i<j} K^2(\Delta R_{ij}'\gamma/h_n)d_{ij}\Delta W_{ij}\Delta W_{ij}'\left(\Delta W_{ij}'(\beta-\widehat{\beta}_n)\right)^2 \\
&+ \frac{1}{Nh_n}\sum_{i<j} K^2(\Delta R_{ij}'\gamma/h_n)d_{ij}\Delta W_{ij}\Delta W_{ij}'\lambda_{ij}^2 \\
&+ \frac{2}{Nh_n}\sum_{i<j} K^2(\Delta R_{ij}'\gamma/h_n)d_{ij}\Delta W_{ij}\Delta W_{ij}'\Delta W_{ij}'(\beta-\widehat{\beta}_n)\lambda_{ij} \\
&+ \frac{2}{Nh_n}\sum_{i<j} K^2(\Delta R_{ij}'\gamma/h_n)d_{ij}\Delta W_{ij}\Delta W_{ij}'\Delta W_{ij}'(\beta-\widehat{\beta}_n)\nu_{ij} \\
&+ \frac{2}{Nh_n}\sum_{i<j} K^2(\Delta R_{ij}'\gamma/h_n)d_{ij}\Delta W_{ij}\Delta W_{ij}'\lambda_{ij}\nu_{ij}.
\end{aligned}
$$

We call each term by $D_{p1,1}^i$ for $i = 1, ..., 6$ that is corresponding to each row.

The first term $D_{p1,1}^1$ converges to $\Sigma_{W\nu}$. Its expectation coincides with $\Sigma_{W\nu}$ in the limit as

$$
\begin{aligned}
\mathbb{E}[D_{p1,1}^1] &= \frac{1}{h_n}\int \mathbb{E}[d_{12}\Delta W_{12}\Delta W_{12}'\nu_{12}^2|\Delta R_{12}'\gamma = r]K^2(r/h_n)f_{R\gamma}(r)dr \\
&= \int \mathbb{E}[d_{12}\Delta W_{12}\Delta W_{12}'\nu_{12}^2|\Delta R_{12}'\gamma = rh_n]K^2(r)f_{R\gamma}(rh_n)dr \\
&= \Sigma_{W\nu,2} + o(1),
\end{aligned}
$$

where the last line holds by the dominated convergence theorem under Assumptions 3.4, 3.6, and 3.8. For the variance, denoting each summand by $D_{p1,1,ij}^1$

and for any vector a with $\|a\| = 1$, we have

$$\begin{aligned}\mathrm{Var}[\|D^1_{p1,1}\|] \leq \frac{1}{Nh_n^2}\mathbb{E}\left[\|D^1_{p1,1,12}\|^2\right] \\ + \frac{2(n-2)}{Nh_n^2}\mathbb{E}[\|D^1_{p1,1,12}\| \times \|D^1_{p1,1,13}\|].\end{aligned}$$

The first term in the right hand side is $O(1/(Nh_n))$ because,

$$\begin{aligned}\mathbb{E}\left[\|D^1_{p1,1,12}\|^2\right] &\leq \int \mathbb{E}[\|\Delta W_{12}\|^4 \nu_{12}^4 | \Delta R_{12}'\gamma = r] K^2(r/h_n) f_{R\gamma}(r) dr \\ &= h_n \int \mathbb{E}[\|\Delta W_{12}\|^4 \nu_{12}^4 | \Delta R_{12}'\gamma = rh_n] K^2(r) f_{R\gamma}(rh_n) dr \\ &= O(h_n),\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6, and 3.8. The second term on the right-hand side is $O(1/n)$ because

$$\begin{aligned}&\mathbb{E}[\|D^1_{p1,1,12}\| \times \|D^1_{p1,1,13}\|] \\ &\leq \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R_{12}'\gamma = r_1, \xi_1, U_1] \\ &\times \mathbb{E}[\|\Delta W_{13}\|^2 \nu_{13}^2 | \Delta R_{13}'\gamma = r_1, \xi_1, U_1] \\ &\times K^2(r_1/h_n) K^2(r_2/h_n) f_{R\gamma|\xi_1,U_1}(r_1) f_{R\gamma|\xi_1,U_1}(r_2) dr_1 dr_2\Big] \\ &= h_n^2 \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R_{12}'\gamma = r_1 h_n, \xi_1, U_1] \\ &\times \mathbb{E}[\|\Delta W_{13}\|^2 \nu_{13}^2 | \Delta R_{13}'\gamma = r_1 h_n, \xi_1, U_1] \\ &\times K^2(r_1) K^2(r_2) f_{R\gamma|\xi_1,U_1}(r_1 h_n) f_{R\gamma|\xi_1,U_1}(r_1 h_n) dr_1 dr_2\Big] \\ &= O(h_n^2),\end{aligned}$$

where the first line follows from Assumptions 3.4, 3.6, and 3.8. Thus,

$$\mathrm{Var}[\|D^1_{p1,1}\|] = O\left(\frac{1}{Nh_n}\right) + O\left(\frac{1}{n}\right) = o(1).$$

This implies that $D^1_{p1,1} \to_p \Sigma_{W\nu,2}$ as $n \to \infty$.

The second term $D^2_{p1,1}$ converges to 0. Observe that, as K is bounded by Assumption 3.8, for some absolute constant $C > 0$,

$$\|D^2_{p1,1}\| \leq \frac{\|\beta - \widehat{\beta}_n\|^2}{h_n} \times \frac{C}{N} \sum_{i<j} \|\Delta W_{ij}\|^4.$$

Since $\mathbb{E}[\|\Delta W_{ij}\|^4] < \infty$ by Assumption 3.7, we can apply the law of large numbers for U-statistics (Hoeffding, 1961) to $C/N \sum_{i<j} \|\Delta W_{ij}\|^4$, which is $O_p(1)$. Also, since $\|\beta - \widehat{\beta}_n\| = O_p(1/\sqrt{n})$ (which is the worst-case rate for the specified h_n by Theorem 3.1), we have $\|\beta - \widehat{\beta}_n\|^2/h_n = O_p(1/(nh_n^2)) = o_p(1)$ as $nh_n^2 \approx n \times n^{-2/(2k+3)} = n^{(2k+1)/(2k+3)}$ diverges. Thus,

$$\|D^2_{p1,1}\| = o_p(1) \times O_p(1) = o_p(1),$$

and $D^2_{p1,1} \to_p 0$ as $n \to \infty$.

The third term $D^3_{p1,1}$ converges to 0. Observe that,

$$\begin{aligned}
&\mathbb{E}[\|D^3_{p1,1}\|] \\
&\leq h_n^{-1} \int \mathbb{E}[\|\Delta W_{12}^2 \Lambda_{12}^2 | \Delta R_{12}'\gamma = r] r^2 K^2(r/h_n) f_{R\gamma}(r) dr \\
&= h_n^2 \mathbb{E}[\|\Delta W_{12}^2 \Lambda_{12}^2 | \Delta R_{12}'\gamma = rh_n] r^2 K^2(r) f_{R\gamma}(rh_n) dr \\
&= O(h_n^2),
\end{aligned}$$

where the last line follows from Assumptions 3.4, 3.6, and 3.8. Thus, $\mathbb{E}[\|D^3_{p1,1}\|] = o(1)$. Observe that, by writing each summand of $D^3_{p1,1}$ as $D^3_{p1,1,ij}$,

$$\mathrm{Var}[\|D^3_{p1,1}\|] \leq \frac{1}{Nh_n^2}\mathbb{E}[\|D^3_{p1,1,12}\|^2] + \frac{2(n-2)}{Nh_n^2}\mathbb{E}[\|D^3_{p1,1,12}\| \times \|D^3_{p1,1,13}\|].$$

The first term on the right hand is $O(h_n^3/N)$ because

$$\begin{aligned}
\mathbb{E}[\|D^3_{p,1,12}\|^2] &\leq \int \mathbb{E}[\|\Delta W_{12}\|^4 \Lambda_{12}^4 | \Delta R'_{12}\gamma = r] r^4 K^2(r/h_n) f_{R\gamma}(r) dr \\
&= h_n^5 \mathbb{E}[\|\Delta W_{12}\|^4 \Lambda_{12}^4 | \Delta R'_{12}\gamma = rh_n] r^4 K^2(r) f_{R\gamma}(rh_n) dr \\
&= O(h_n^5),
\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6, and 3.8. The second term on the right hand side is $O(h_n^4/n)$ because

$$\begin{aligned}
&\mathbb{E}[\|D^3_{p1,1,12}\| \times \|D^3_{p1,1,13}\|] \\
&\leq \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|^2 \Lambda_{12}^2 | \Delta R'_{12}\gamma = r_1, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\|^2 \Lambda_{13}^2 | \Delta R'_{13}\gamma = r_1, \xi_1, U_1] \\
&\times r_1^2 r_2^2 K^2(r_1/h_n) K^2(r_2/h_n) f_{R\gamma|\xi_1,U_1}(r_1) f_{R\gamma|\xi_1,U_1}(r_2) dr_1 dr_2\Big] \\
&= h_n^6 \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|^2 \Lambda_{12}^2 | \Delta R'_{12}\gamma = r_1 h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\|^2 \Lambda_{13}^2 | \Delta R'_{13}\gamma = r_2 h_n, \xi_1, U_1] \\
&\times r_1^2 r_2^2 K^2(r_1) K^2(r_2) f_{R\gamma|\xi_1,U_1}(r_1 h_n) f_{R\gamma|\xi_1,U_1}(r_2 h_n) dr_1 dr_2\Big] \\
&= O(h_n^6),
\end{aligned}$$

where the last line follows from Assumptions 3.4, 3.6, and 3.8. Hence, we have

$$\mathrm{Var}[\|D^3_{p1,1}\|] = O\left(\frac{h_n^3}{N}\right) + O\left(\frac{h_n^4}{n}\right) = o(1).$$

This implies that $D^3_{p1,1} \to_p 0$ as $n \to \infty$.

The fourth term $D^4_{p1,1}$ converges to 0. Observe that, since K is bounded by Assumption 3.8 and $\|\gamma\| < \infty$, for some constant $C > 0$,

$$\|D^4_{p1,1}\| \leq \frac{C\|\beta - \widehat{\beta}_n\|}{h_n} \times \frac{1}{N}\sum_{i<j} \|\Delta W_{ij}\|^3 \|\Delta R_{ij}\| |\Lambda_{12}|$$

The sum part converges to the expectation of summand by the law of large numbers for U-statistics (Hoeffding, 1961) as $\mathbb{E}[\|\Delta W_{12}\|^3 \|\Delta R_{12}\| |\Lambda_{12}|] < \infty$ is bounded by Cauchy-Schwartz and Assumption 3.7. Thus, this part is $O_p(1)$. Also note that

$$\frac{\|\beta - \widehat{\beta}_n\|}{h_n} = O_p\left(\frac{1}{\sqrt{nh_n^2}}\right) = o_p(1)$$

by Assumption 3.9 and $\sqrt{nh_n^2} \approx \sqrt{n(2k-1)/(2k+3)}$ diverges. Hence,

$$\|D^4_{p1,1}\| = o_p(1).$$

This shows that $D^4_{p1,1} \to_p 0$ as $n \to \infty$.

The fifth term $D^5_{p1,1}$ converges to 0. Observe that, since K is bounded by

Assumption 3.8, for some constant $C > 0$

$$\|D_{p1,1}^5\| \leq \frac{C}{N}\sum_{i<j}\|\Delta W_{ij}\|^3|\nu_{ij}| \times \frac{\|\beta - \widehat{\beta}_n\|}{h_n}.$$

The sum part is $O_p(1)$ because

$$\mathbb{E}[\|\Delta W_{12}\|^3|\nu_{12}|] < \infty,$$

by Assumption 3.7 and

$$\begin{aligned}&\text{Var}\left[\frac{1}{N}\sum\|\Delta W_{ij}\||\nu_{ij}|\right]\\ &\leq \frac{\mathbb{E}[\|\Delta W_{12}\|^6\nu_{12}^2]}{N} + \frac{2(n-2)}{N}\mathbb{E}[\|\Delta W_{12}\|^3\Delta W_{13}\|^3|\nu_{12}||\nu_{13}|] = o(1),\end{aligned}$$

as these two moments are bounded by Assumption 3.7. Thus,

$$\|D_{p1,1}^5\| = o_p(1),$$

by the previous calculation for the term involving $\widehat{\beta}_n - \beta$. This shows that $D_{p1,1}^5 \to_p 0$ as $n \to \infty$.

The sixth term $D_{p1,1}^6$ converges to 0. Its expectation is exactly 0 by the conditional mean independence of ν_{ij}. Also, by repeating the similar calculation as $\text{Var}[\|D_{p1,1}^2\|]$ (by replacing ν_{ij}^2 by $\lambda_{ij}\nu_{ij}$), we have

$$\text{Var}[\|D_{p1,1}^6\|] = O\left(\frac{1}{N}\right) + O\left(\frac{h_n^2}{n}\right) = o(1).$$

This shows that $\widetilde{D}_{1,6} \to_p 0$ as $n \to \infty$.

C.1.4 Sub-Step 2: $D_{p1,2} \to_p 0$

As before, we can decompose $D_{p1,2}$ into $D^i_{p1,2}$ for $i = 1, ..., 6$. Unlike in $D_{p1,1}$, we can no longer have the moments scaled by h_n^α because the middle values $c^*_{ij,n}$ are in the kernels. Thus, by the previous calculation for $D_{p1,1}$, the $D^i_{p1,2}$ that involves ν^2_{ij},λ^2_{ij}, or $\lambda_{ij}\nu_{ij}$ will have the slowest convergence rate. So, it suffices to show that those terms converge to 0 in probability.

Pick up such $D^i_{p1,2}$ with ν^2_{ij}, which is $D^1_{p1,2}$ and given by

$$D^1_{p1,2} = \frac{2}{Nh_n^2} \sum_{i<j} d_{ij} \Delta W_{ij} \Delta W'_{ij} \Delta R'_{ij} \nu^2_{ij} k(c^*_{ij,n}/h_n) K(c^*_{ij,n})(\widehat{\gamma}_n - \gamma)$$

Observe that, for some constant $C > 0$

$$\|D^1_{p1,2}\| \leq \frac{C}{N} \sum_{i<j} \|\Delta W_{ij}\|^2 \|\Delta R_{ij}\| \nu^2_{ij} \times \frac{\|\widehat{\gamma}_n - \gamma\|}{h_n^2}$$

The sum part is $O_p(1)$ because

$$\mathbb{E}[\|\Delta W_{12}\|^2 \|\Delta R_{12}\| \nu^2_{12}] < \infty,$$

by Assumption 3.7, and

$$\begin{aligned}
&\text{Var}\left[\frac{1}{N} \sum_{i<j} \|\Delta W_{ij}\|^2 \|\Delta R_{ij}\| \nu^2_{ij}\right] \\
&\leq \frac{\mathbb{E}[\|\Delta W_{12}\|^4 \|\Delta R_{12}\|^2 \nu^4_{12}]}{N} \\
&+ \frac{2(n-2)}{N} \mathbb{E}[\|\Delta W_{12}\|^2 \|\Delta W_{13}\|^2 \|\Delta R_{12}\| \|\Delta R_{13}\| \nu^2_{12} \nu^2_{13}] \\
&= o(1),
\end{aligned}$$

as these moments are bounded by Assumption 3.7. The term involving $\widehat{\gamma}_n$ is $o_p(1)$ because

$$\frac{\|\widehat{\gamma}_n - \gamma\|}{h_n^2} = \frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|}{\sqrt{Nh_n^5}} = o_p(1),$$

by Assumption 3.10 and $Nh_n^5 = Nh_n^{2k+3} \times h_n^{-2k+2}$ diverges for $k \geq 2$. Hence,

$$\|D_{p1,2}^1\| = O_p(1) \times o_p(1) = o_p(1).$$

This shows that $D_{p1,2}^1 \to_p 0$ as $n \to \infty$. Thus, by the above argument, it follows that $D_{p1,2} \to_p 0$ as $n \to \infty$.

These two sub-steps conclude that

$$\widehat{\Sigma}_{W\nu,2} \to_p \Sigma_{W\nu,2},$$

as $n \to \infty$. This finishes Step 1.

C.1.4.1 Step 2: $\widehat{\Sigma}_{W\nu,1} \to_p \Sigma_{W\nu,1}$

Define

$$S_{ij} \equiv 2d_{ij}K_{h_n}(\Delta R_{ij}'\gamma)\Delta W_{ij}\Delta\widehat{\epsilon}_{ij},$$

and let $\widetilde{\Sigma}_{W\nu,1}$ be $\widehat{\Sigma}_{W\nu,1}$ with $\widehat{S}_{ij}$ replaced by S_{ij}. First, we use the following result:

Lemma C.7. *Suppose that Assumptions 3.1-3.10 hold. If* $h_n = hN^{-1/(2k+3)}$

for some $h > 0$, we have

$$\widetilde{\Sigma}_{W\nu,1} \to_p \Sigma_{W\nu,1},$$

as $n \to \infty$.

Then, it is enough to show that $\widehat{\Sigma}_{W\nu,1}$ is well approximated by $\widetilde{\Sigma}_{W\nu,1}$:

Lemma C.8. *Suppose that Assumptions 3.1-3.10 hold. If $h_n = hN^{-1/(2k+3)}$ for some $h > 0$, we have*

$$\|\widehat{\Sigma}_{W\nu,1} - \widetilde{\Sigma}_{W\nu,1}\| = o_p(1).$$

Lemmas C.7 and C.8 imply that

$$\|\widehat{\Sigma}_{W\nu,1} - \Sigma_{W\nu,1}\| \leq \|\widehat{\Sigma}_{W\nu,1} - \widetilde{\Sigma}_{W\nu,1}\| + \|\widetilde{\Sigma}_{W\nu,1} - \Sigma_{W\nu,1}\| = o_p(1),$$

which shows the consistency of $\widehat{\Sigma}_{W\nu,1}$ for $\Sigma_{W\nu,1}$. This finishes Step 2.

C.1.4.2 Step 3: $c_W'\Sigma_{W\nu,1}c_W = 0$ case

Observe that, by some algebra,

$$
\begin{aligned}
& nh_n c'\widehat{S}_{WW}^{-1}\widehat{\Sigma}_{W\nu,1}\widehat{S}_{WW}^{-1}c \\
& = nh_n c'(\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1})\widehat{\Sigma}_{W\nu,1}(\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1})c \\
& + nh_n c'\Sigma_{WW}^{-1}\widehat{\Sigma}_{W\nu,1}(\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1})c + nh_n c'(\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1})\widehat{\Sigma}_{W\nu,1}\Sigma_{WW}^{-1}c \\
& + nh_n c'\Sigma_{WW}^{-1}\widehat{\Sigma}_{W\nu,1}\Sigma_{WW}^{-1}c.
\end{aligned}
$$

We show the negligibility of the first line in the right hand side of this decomposition. By the proof of Lemma C.1, we have that

$$\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1} = o_p(n^{-\alpha/2})$$

for any $\alpha \in (0,1)$. Thus,

$$
\begin{aligned}
& nh_n c'(\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1})\widehat{\Sigma}_{W\nu,1}(\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) \\
& = n^{1-\alpha}h_n o_p(1)\widehat{\Sigma}_{W\nu,1}o_p(1) \\
& = o_p(1),
\end{aligned}
$$

for $\alpha \in [(2k+1)/(2k+3), 1)$ as $n^{1-\alpha}h_n = n^{(2k+1-\alpha(2k+3))/(2k+3)} = o(1)$ and $\widehat{\Sigma}_{W\nu,1} = O_p(1)$ by the above Step 2.

The remaining terms are shown to be negligible by applying the following lemmas:

Lemma C.9. *Suppose that Assumptions 3.1-3.10 hold. If* $c_W'\Sigma_{W\nu,1}c_W = 0$

and $h_n = hN^{-1/(2k+3)}$ *for some* $h \in (0, \infty)$, *we have*

$$n^{1-\alpha/2} h_n \widehat{\Sigma}_{W\nu,1} c_W \to_p 0,$$
$$n^{1-\alpha/2} h_n c_W' \widehat{\Sigma}_{W\nu,1} \to_p 0,$$

as $n \to \infty$ *for* $\alpha \in [6/(2k+3), 1)$.

Lemma C.10. *Suppose that Assumptions 3.1-3.10 hold. If* $c_W' \Sigma_{W\nu,1} c_W = 0$ *and* $h_n = hN^{-1/(2k+3)}$ *for some* $h \in (0, \infty)$, *we have*

$$nh_n c_W' \widehat{\Sigma}_{W\nu,1} c_W \to_p 0,$$

as $n \to \infty$.

Then, by Lemmas C.9 and C.10, the last two lines are shown to be

$$\begin{aligned}
&nh_n c' \Sigma_{WW}^{-1} \widehat{\Sigma}_{W\nu,1} (\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) c + nh_n c' (\widehat{S}_{WW}^{-1} - \Sigma_{WW}^{-1}) \widehat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c \\
&+ 3nh_n c' \Sigma_{WW}^{-1} \widehat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c \\
&= n^{1-\alpha/2} h_n c' \Sigma_{WW}^{-1} \widehat{\Sigma}_{W\nu,1} o_p(1) + c' o_p(1) n^{1-\alpha/2} h_n \widehat{\Sigma}_{W\nu,1} \Sigma_{WW}^{-1} c + o_p(1) \\
&= o_p(1).
\end{aligned}$$

Hence,

$$nh_n c' \widehat{S}_{WW}^{-1} \widehat{\Sigma}_{W\nu,1} \widehat{S}_{WW} c = o_p(1).$$

Steps 1-3 finish the proof of Proposition 3.1. □

C.1.5 Proof of Proposition 3.2

Proof. Since

$$h_{n,\delta}^{-(k+1)}(\widehat{\beta}_{n,\delta} - \widehat{\beta}_n) = h_{n,\delta}^{-(k+1)}(\widehat{\beta}_{n,\delta} - \beta) - h_{n,\delta}^{-(k+1)}(\widehat{\beta}_n - \beta),$$

where the first term on the right hand side converges to $\Sigma_{WW}^{-1}\Sigma_{W\lambda}$ by Theorem 3.1 as $Nh_{n,\delta}^{2k+3} \to \infty$, it suffices to show that

$$h_{n,\delta}^{-(k+1)}(\widehat{\beta}_n - \beta) = o_p(1).$$

Take an arbitrary non-zero vector $c \in \mathbb{R}^{q_w}$. Since $\widehat{\beta}_n$ is calculated based on $h_n = hN^{-1/(2k+3)}$ such that $Nh_n^{2k+3} \to h$, by Theorem 3.1,

$$h_{n,\delta}^{-(k+1)}c'(\widehat{\beta}_n - \beta) = \frac{1}{\sqrt{nh_{n,\delta}^{2(k+1)}}} \times \underbrace{\sqrt{n}c'(\widehat{\beta}_n - \beta)}_{=O_p(1)} = o_p(1)$$

since

$$nh_{n,\delta}^{2(k+1)} \approx n \times n^{-4\delta(k+1)/(2k+3)} = n^{\frac{2k+3-4\delta(k+1)}{2k+3}}$$

diverges for $\delta \in (0, \frac{2k+3}{4k+4})$, which is assumed by the hypothesis. Since c is arbitrary, $h_{n,\delta}^{-(k+1)}(\widehat{\beta}_n - \beta) = o_p(1)$, which completes the proof. $\square$

C.2 Proofs of Lemmas

C.2.1 Proof of Lemma C.1

Proof. Write each summand of S_{WW} as $S_{WW,ij}$. Since it suffices to show the element-wise convergence of S_{WW} to Σ_{WW}, we use a unit vector $e \in \mathbb{R}^{q_w}$ with the arbitrary element being 1 and 0 elsewhere. Observe that

$$\begin{aligned}
&\mathbb{E}[e' S_{WW} e] = \mathbb{E}[e' S_{WW,ij} e] \\
&= \frac{1}{h_n} \int \mathbb{E}[d_{12} e' \Delta W_{12} \Delta W_{12}' e | \Delta R_{12}' \gamma = r] K(r/h_n) f_{R\gamma}(r) dr \\
&= \int \mathbb{E}[d_{12} e' \Delta W_{12} \Delta W_{12}' e | \Delta R_{12}' \gamma = r h_n] K(r) f_{R\gamma}(r h_n) dr \\
&= e' \Sigma_{WW} e + o_p(1),
\end{aligned}$$

where the last line holds from the dominated convergence theorem under Assumptions 3.4, 3.6 and 3.8. Since $S_{WW,ij}$ and $S_{WW,kl}$ are independent if $i \neq k, l$ and $j \neq k, l$ by Assumption 3.1, observe that

$$\mathrm{Var}(e' S_{WW} e) = \frac{\mathrm{Var}(S_{WW,12})}{N} + \frac{2(n-2)}{N} \mathrm{Cov}(e' S_{WW,12} e, e' S_{WW,13} e).$$

For the variance, we have

$$\begin{aligned}
\mathrm{Var}(S_{WW,12}) &\leq \mathbb{E}[(e' S_{WW,12} e)^2] \\
&\leq \frac{1}{h_n^2} \int \mathbb{E}[\|\Delta W_{12}\|^4 | \Delta R_{12}' \gamma = r] K^2(r/h_n) f_{R\gamma}(r) dr \\
&= \frac{1}{h_n} \int \mathbb{E}[\|\Delta W_{12}\|^4 | \Delta R_{12}' \gamma = r h_n] K^2(r) f_{R\gamma}(r h_n) dr \\
&= O\left(\frac{1}{h_n}\right),
\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6 and 3.8. For the covariance, by the conditional independence of ΔW_{12} and ΔW_{13}, we have

$$
\begin{aligned}
&\mathrm{Cov}(e'S_{WW,12}e, e'S_{WW,13}e) \\
&\leq \mathbb{E}[|e'S_{WW,12}e \times e'S_{WW,13}e|] \\
&\leq \frac{1}{h_n^2} \int \mathbb{E}[\|\Delta W_{12}'\|^2 \|\Delta W_{13}\|^2 | \Delta R_{12}'\gamma = r_1, \Delta R_{13}'\gamma = r_2] \\
&\times |K(r_1/h_n)||K(r_2/h_n)| f_{R\gamma,2}(r_1, r_2) dr_1 dr_2 \\
&= \int \mathbb{E}[\|\Delta W_{12}'\|^2 \|\Delta W_{13}\|^2 | \Delta R_{12}'\gamma = h_n r_1, \Delta R_{13}'\gamma = h_n r_2] \\
&\times |K(r_1)||K(r_2)| f_{R\gamma,2}(h_n r_1, h_n r_2) dr_1 dr_2 \\
&= O(1),
\end{aligned}
$$

where the last line holds from Assumptions 3.4, 3.6 and 3.8. Thus,

$$
\mathrm{Var}(e'S_{WW}e) = O\left(\frac{1}{Nh_n}\right) + O\left(\frac{1}{n}\right) = o(1).
$$

By Chebychev's inequality, $e'S_{WW}e \to_p e'\Sigma_{WW}e$ as $n \to \infty$. Since e is arbitrary, this completes the proof.

□

C.2.2 Proof of Lemma C.2

Proof. Write each summand of $S_{W\lambda}$ as $S_{W\lambda,ij}$. We use a unit vector $e \in \mathbb{R}^{q_w}$ with an arbitrary element being 1 and 0 elsewhere. Observe that, for large enough n,

$$
\begin{aligned}
&\mathbb{E}[e'S_{W\lambda}] = \mathbb{E}[e'S_{W\lambda,ij}] \\
&= \frac{1}{h_n}\int \mathbb{E}[d_{12}e'\Delta W_{12}\lambda_{12}|\Delta R_{12}'\gamma = r]K(r/h_n)f_{R\gamma}(r)dr \\
&= h_n \int e'g(rh_n)rK(r)dr \\
&= \frac{h_n^{k+1}}{k!}\int \left(e'\frac{\partial^k g(rh_n)}{\partial r^k} + o(1)\right) r^{k+1}K(r)dr \\
&= h_n^{k+1}e'\Sigma_{W\lambda} + o(h_n^{k+1}),
\end{aligned}
$$

where the second line holds from $\lambda_{12} = \Lambda_{12} \times \Delta R_{12}'\gamma$, the third line holds from Assumption 3.8 eliminating $\int s^i K(s)$ for $i = 1, ..., k$, and the last line holds from the dominated convergence theorem under Assumptions 3.6 and 3.8. Observe that

$$
\mathrm{Var}[e'S_{W\lambda}] = \frac{\mathrm{Var}[e'S_{W\lambda,12}]}{N} + \frac{2(n-2)}{N}\,\mathrm{Cov}[e'S_{W\lambda,12}, e'S_{W\lambda,13}].
$$

For the variance, we have

$$
\begin{aligned}
\mathrm{Var}[e'S_{W\lambda,12}] &\leq \mathbb{E}[(e'S_{W\lambda,12})^2] \\
&\leq \frac{1}{h_n^2}\int \mathbb{E}[\|\Delta W_{12}\|^2\lambda_{12}^2|\Delta R_{12}'\gamma = r]K^2(r/h_n)f_{R\gamma}(r)dr \\
&= h_n \int \mathbb{E}[\|\Delta W_{12}\|^2\Lambda_{12}^2|\Delta R_{12}'\gamma = rh_n]r^2K^2(r)f_{R\gamma}(rh_n)dr \\
&= O(h_n),
\end{aligned}
$$

where the last line holds from Cauchy-Schwartz and Assumptions 3.4, 3.6, and 3.8. For the covariance, we have

$$
\begin{aligned}
&\mathrm{Cov}[e'S_{W\lambda,12}, e'S_{W\lambda,13}] \\
&\leq \mathbb{E}[e'S_{W\lambda,12} \times e'S_{W\lambda,13}] \\
&= \frac{1}{h_n^2}\mathbb{E}\Big[\int \mathbb{E}[d_{12}e'\Delta W_{12}\Lambda_{12}|\Delta R_{12}'\gamma = r_1, \xi_1, U_1] \\
&\times \mathbb{E}[d_{13}e'\Delta W_{13}\Lambda_{13}|\Delta R_{13}'\gamma = r_1, \xi_1, U_1] \\
&\times r_1 r_2 K(r_1/h_n)K(r_2/h_n) f_{R\gamma,\xi_1,U_1}(r_1) f_{R\gamma,\xi_1,U_1}(r_2) dr_1 dr_2\Big] \\
&\leq h_n^2\mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\||\Lambda_{12}||\Delta R_{12}'\gamma = r_1 h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\||\Lambda_{13}||\Delta R_{13}'\gamma = r_2 h_n, \xi_1, U_1] \\
&\times r_1 r_2 K(r_1)K(r_2) f_{R\gamma,\xi_1,U_1}(r_1 h_n) f_{R\gamma,\xi_1,U_1}(r_2 h_n) dr_1 dr_2\Big] \\
&= O(h_n^2),
\end{aligned}
$$

where the last line holds from Cauchy-Schwartz and Assumptions 3.4, 3.6, and 3.8. Thus,

$$\mathrm{Var}[e'S_{W\lambda}] = O\left(\frac{h_n}{N}\right) + O\left(\frac{h_n^2}{n}\right) = O\left(\frac{h_n^2}{n}\right),$$

since $O(h_n/N) = O(h_n^2/n) \times O(1/(nh_n)) = o(h_n^2/n)$ under Assumption 3.9.

If $Nh_n^{2k+3} \to h$ for some $0 < h < \infty$, note that

$$\sqrt{Nh_n}\mathbb{E}[e'S_{W\lambda}] = \sqrt{Nh_n^{2k+3}}e'\Sigma_{W\lambda} + o(\sqrt{Nh_n^{2k+3}}) \to \sqrt{h}\Sigma_{W\lambda},$$

as $n \to \infty$. Also,

$$\mathrm{Var}[\sqrt{Nh_n}e'S_{W\lambda}] = O\left(\frac{Nh_n^3}{n}\right) = O(nh_n) \times O(h_n^2) = o(1),$$

by Assumption 3.9. Thus, by Chebyshev's inequality, we have

$$\sqrt{Nh_n}e'S_{W\lambda} \to_p \sqrt{h}e'\Sigma_{W\lambda},$$

as $n \to \infty$.

If $Nh_n^{2k+3} \to \infty$, note that

$$h_n^{-(k+1)}\mathbb{E}[e'S_{W\lambda}] = e'\Sigma_{W\lambda} + o(1) \to e'\Sigma_{W\lambda},$$

as $n \to \infty$. Also,

$$\mathrm{Var}[h_n^{-(k+1)}e'S_{W\lambda}] = O\left(\frac{h_n^2}{nh_n^{2k+2}}\right) = o(1),$$

as $nh_n^{2k} \to \infty$ by the hypothesis. Thus, by Chebyshev's inequality, we have

$$h_n^{-(k+1)}e'S_{W\lambda} \to_p e'\Sigma_{W\lambda},$$

as $n \to \infty$. Since e is arbitrary, this completes the proof. □

C.2.3 Proof of Lemma C.3

Proof. The proof is done in the following steps:

C.2.3.1 Step 0: Decomposition

Observe that

$$c' S_{WW}^{-1} S_{W\nu} = c'(S_{WW}^{-1} - \Sigma_{WW}^{-1}) S_{W\nu} + c' \Sigma_{WW}^{-1} S_{W\nu}$$

In Steps 1-2, we verify the asymptotic normality of $S_{W\nu}$, with the worst-case convergence rate being $\sqrt{n}$. Given that result, the first term on the right-hand side is shown to be negligible even when normalized by $\sqrt{Nh_n}$:

$$\begin{aligned} \sqrt{Nh_n} c'(S_{WW}^{-1} - \Sigma_{WW}^{-1}) S_{W\nu} &= \sqrt{Nh_n} o_p(n^{-\alpha/2}) O_p(1/\sqrt{n}) \\ &= \sqrt{n^{1-\alpha} h_n} o_p(1) = o_p(1) \end{aligned}$$

because by Lemma C.1, $S_{WW}^{-1} - \Sigma_{WW}^{-1} = o_p(n^{-\alpha/2})$ for any $\alpha \in (0,1)$ and $n^{1-\alpha} h_n = o(1)$ for sufficiently large α under the hypothesis. Thus,

$$c' S_{WW}^{-1} S_{W\nu} = c' \Sigma_{WW}^{-1} S_{W\nu} + o_p(1/\sqrt{Nh_n}),$$

and it suffices to establish the asymptotic normality of $S_{W\nu}$. Write $c = c_W$ for short. Observe that, since $\mathbb{E}[S_{W\nu}] = 0$ by the definition of ν_{ij}, $c' S_{W\nu}$ can be decomposed as

$$c' S_{W\nu} = \underbrace{\frac{1}{n} \sum_{i=1}^{n} L_{i,W\nu}}_{L_{W\nu}} + \underbrace{\frac{1}{N} \sum_{i<j} P_{ij,W\nu}}_{P_{W\nu}} + \underbrace{\frac{1}{N} \sum_{i<j} Q_{ij,W\nu}}_{Q_{W\nu}}$$

where

$$\begin{aligned}
L_{i,W\nu} &\equiv 2\mathbb{E}[d_{ij}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_i, U_i] \\
P_{ij,W\nu} &= \mathbb{E}[d_{ij}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_i, U_i, \xi_j, U_j] \\
&\quad - \mathbb{E}[d_{ij}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_i, U_i] \\
&\quad - \mathbb{E}[d_{ij}\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_j, U_j] \\
Q_{ij,W\nu} &= d_{ij}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma) \\
&\quad - \mathbb{E}[d_{ij}c'\Delta W_{ij}\nu_{ij}K_{h_n}(\Delta R'_{ij}\gamma)|\xi_i, U_i, \xi_j, U_j].
\end{aligned}$$

By design, we have that $\text{Cov}[L_{i,W\nu}, L_{j,W\nu}] = \text{Cov}[L_{i,W\nu}, P_{kl,W\nu}] = \text{Cov}[L_{i,W\nu}$ $, Q_{kl,W\nu}] = 0$, $\text{Cov}[P_{ij,W\nu}, P_{kl,W\nu}] = \text{Cov}[P_{ij,W\nu}, Q_{kl,W\nu}] = 0$, and $\text{Cov}[Q_{ij,W\nu}$ $, Q_{kl,W\nu}] = 0$ for any $i \neq j$, $k \neq l$, and $ij \neq kl$. We show the asymptotic normality of $c'S_{W\nu}$ in the following.

C.2.3.2 Step 1: Asymptotic Normality of $L_{W\nu}$

Define V_L by

$$V_L = \sqrt{n}L_{W\nu} = \sum_{i=1} \underbrace{\frac{L_{i,W\nu}}{\sqrt{n}}}_{V_{i,L}}$$

Note that $\mathbb{E}[V_{i,W\nu}] = 0$ by the mean independence of ν_{ij}. Observe that

$$
\begin{aligned}
&\text{Var}[V_L] = \text{Var}[L_{i,W\nu}] \\
&= 4\mathbb{E}\left[\mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R_{12}'\gamma)|\xi_1, U_1]^2\right] \\
&= 4\mathbb{E}[d_{12}d_{13}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}K_{h_n}(\Delta R_{12}'\gamma)K_{h_n}(\Delta R_{13}'\gamma)] \\
&= \frac{4}{h_n^2}\int \mathbb{E}[d_{12}d_{13}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}|\Delta R_{12}'\gamma = r_1, \Delta R_{13}'\gamma = r_2] \\
&\times K(r_1/h_n)K(r_2/h_n)f_{R\gamma,2}(r_1, r_2)dr_1dr_2 \\
&= \int \mathbb{E}[d_{12}d_{13}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}|\Delta R_{12}'\gamma = r_1h_n, \Delta R_{13}'\gamma = r_2h_n] \\
&\times K(r_1)K(r_2)f_{R\gamma,2}(r_1h_n, r_2h_n)dr_1dr_2 \times 4 \\
&= c'\Sigma_{W\nu,1}c + o_p(1),
\end{aligned}
$$

where the last line holds from the dominated convergence theorem under Assumptions 3.4, 3.6, and 3.8. Furthermore, note that

$$
\begin{aligned}
&\mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R_{12}'\gamma)|\xi_1, U_1] \\
&= \frac{1}{h_n}\int \mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}|\Delta R_{12}'\gamma = r, \xi_1, U_1]K(r/h_n)f_{R\gamma|\xi_1,U_1}(r)dr \\
&= \int \mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}|\Delta R_{12}'\gamma = r, \xi_1, U_1]K(r/h_n)f_{R\gamma|\xi_1,U_1}(r)dr \\
&= O(1),
\end{aligned}
$$

almost surely for sufficiently large n by Assumptions 3.4, 3.6, and 3.8. Thus, we have

$$
\sum_{i=1}^{n}\mathbb{E}[|V_{i,L}|^3] = n \times O\left(\frac{1}{n\sqrt{n}}\right) = o(1).
$$

If $\Sigma_{W\nu,1}$ is positive definite, we have $c'\Sigma_{W\nu,1}c > 0$. Thus, by Lyapunov CLT, we have

$$V_L/\sqrt{\text{Var}[V_L]} \to_d \mathcal{N}(0,1),$$

as $n \to \infty$. Thus, $V_L = \sqrt{n}L_{W\nu} \to_d \mathcal{N}(0, c'\Sigma_{W\nu,1}c)$ as $n \to \infty$.

If $c'\Sigma_{W\nu,1}c = 0$, observe that, for some constant $C > 0$

$$\begin{aligned}
&\text{Var}[L_{i,W\nu}] \\
&= 4\mathbb{E}\Bigg[\bigg(\int h_n^{-1}\mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}|\Delta R_{12}'\gamma = r, \xi_1, U_1] \\
&\times K(r/h_n)f_{R\gamma|\xi_1,U_1}(r)dr\bigg)^2\Bigg] \\
&= 4\mathbb{E}\left[\left(\int g_{\xi_1,U_1}(rh_n)K(r)dr\right)^2\right] \\
&\leq 4\mathbb{E}\left[\left(g_{\xi_1,U_1}(0) + Ch_n^{k+1}\right)^2\right] \\
&= O(h_n^{2(k+1)}),
\end{aligned}$$

where the first inequality holds from the Taylor expansion of $g_{\xi_1,U_1}(rh_n)$ and

K eliminating $\int r^i K(r)dr = 0$ for $i = 1, ..., k$, and the last line holds since

$$\begin{aligned}
&\mathbb{E}\left[\left(g_{\xi_1,U_1}(0)\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}f_{R\gamma|\xi_1,U_1}(0)|\Delta R_{12}'\gamma = 0, \xi_1, U_1]\right)^2\right] \\
&= \mathbb{E}\Bigg[\mathbb{E}[d_{12}d_{13}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13} \\
&\times f_{R\gamma,2}(0,0)|\Delta R_{12}'\gamma = \Delta R_{13}'\gamma = 0, \xi_1, U_1]\Bigg] \\
&= f_{R\gamma,2}(0,0)\mathbb{E}[d_{12}d_{13}c'\Delta W_{12}c'\Delta W_{13}\nu_{12}\nu_{13}|\Delta R_{12}'\gamma = \Delta R_{13}'\gamma = 0] \\
&= c'\Sigma_{W\nu,1}c/4 = 0,
\end{aligned}$$

and $\mathbb{E}[g_{\xi_1,U_1}] = 0$ by the conditional mean independence of ν_{12}. Thus,

$$\text{Var}[\sqrt{Nh_n}L_{W\nu}] = nh_n \times O(h_n^{2(k+1)}) = o(1),$$

since $nh_n^{2k+3} \approx Nh_n^{2k+3}/n \to 0$ by the hypothesis. Hence, by Chebyshev's inequality,

$$\sqrt{Nh_n}L_{W\nu} \to_p 0,$$

as $n \to \infty$ when $\Sigma_{W\nu,1} = 0$.

C.2.3.3 Step 2: Asymptotic Normality of $P_{W\nu}$

Notice that $P_{ij,W\nu}$ is a degenerate U-statistic of order 2. We use the CLT for degenrate U-statistics by Hall (1984). Note that $P_{ij,W\nu}$ is symmetric in i and j, $\mathbb{E}[P_{ij,W\nu}^2] < \infty$ by Assumption 3.6, and $\mathbb{E}[P_{ij,W\nu}|\xi_i, U_i] = \mathbb{E}[P_{ij,W\nu}|\xi_j, U_j] = 0$

by the definition of $P_{ij,W\nu}$. Also, we can verify that

$$\mathbb{E}\left[\left(\mathbb{E}\left[P_{12,W\nu} \times P_{13,W\nu}|\xi_2, U_2, \xi_3, U_3\right]\right)^2\right] / \left[\mathbb{E}[P_{12,W\nu}^2]\right]^2 \to 0$$
$$\frac{1}{n} \times \mathbb{E}\left[P_{12,W\nu}^4\right] / \left[\mathbb{E}[P_{12,W\nu}^2]\right]^2 \to 0.$$

To see this, first note that,

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}\left[P_{12,W\nu} \times P_{13,W\nu}|\xi_2, U_2, \xi_3, U_3\right]\right)^2\right] \\
&= \mathbb{E}\Big[\Big(\mathbb{E}\Big[\mathbb{E}[d_{12}\nu_{12}|\xi_1, U_1, \xi_2, U_2]\mathbb{E}[d_{13}\nu_{13}|\xi_1, U_1, \xi_3, U_3] \\
&\times c'\Delta W_{12} c'\Delta W_{13} \nu_{12}\nu_{13} K_{h_n}(\Delta R_{12}'\gamma) K_{h_n}(\Delta R_{13}'\gamma)|\xi_2, U_2, \xi_3, U_3\Big]\Big)^2\Big] + O(1) \\
&= \mathbb{E}\Big[\Big(h_n^{-2} \int \mathbb{E}[\mathbb{E}[d_{12}\nu_{12}|\xi_1, U_1, \xi_2, U_2]\mathbb{E}[d_{13}\nu_{13}|\xi_1, U_1, \xi_3, U_3] \\
&\times c'\Delta W_{12} c'\Delta W_{13} \nu_{12}\nu_{13}|\Delta R_{12}'\gamma = r_1, \Delta R_{13}'\gamma = r_2, \xi_2, U_2, \xi_3, U_3] \\
&\times K(r_1/h_n) K(r_2/h_n) f_{R\gamma,2|\xi_2,U_2,\xi_3,U_3}(r_1, r_2) dr_1 dr_2\Big)^2\Big] + O(1) \\
&= \mathbb{E}\Big[\Big(\int \mathbb{E}[\mathbb{E}[d_{12}\nu_{12}|\xi_1, U_1, \xi_2, U_2]\mathbb{E}[d_{13}\nu_{13}|\xi_1, U_1, \xi_3, U_3] \\
&\times c'\Delta W_{12} c'\Delta W_{13} \nu_{12}\nu_{13}|\Delta R_{12}'\gamma = h_n r_1, \Delta R_{13}'\gamma = h_n r_2, \xi_2, U_2, \xi_3, U_3] \\
&\times K(r_1) K(r_2) f_{R\gamma,2|\xi_2,U_2,\xi_3,U_3}(h_n r_1, h_n r_2) dr_1 dr_2\Big)^2\Big] + O(1) \\
&= O(1),
\end{aligned}$$

where the first equality holds from the calculation in Step 1, and the last line holds from Assumptions 3.4, 3.6, and 3.8. Next, note that

$$
\begin{aligned}
\mathbb{E}[P_{12,W\nu}^2] \le & \mathbb{E}[(d_{12}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R_{12}'\gamma))^2] \\
&= h_n^{-2}\int \mathbb{E}[d_{12}(c'\Delta W_{12})^2\nu_{12}^2|\Delta R_{12}'\gamma = r]K^2(r/h_n)f_{R\gamma}(r)dr \\
&= h_n^{-1}\int \mathbb{E}[d_{12}(c'\Delta W_{12})^2\nu_{12}^2|\Delta R_{12}'\gamma = h_n r]K^2(r)f_{R\gamma}(h_n r)dr \\
&= O(h_n^{-1}),
\end{aligned}
$$

where the last line holds from Assumptions 3.4, 3.6, and 3.8. Similarly, we can verify that $\mathbb{E}[P_{12,W\nu}^4] = O(h_n^{-3})$. Thus, we have

$$
\begin{aligned}
&\mathbb{E}\left[(\mathbb{E}\left[P_{12,W\nu}\times P_{13,W\nu}|\xi_2,U_2,\xi_3,U_3\right])^2\right] / \left[\mathbb{E}[P_{12,W\nu}^2]\right]^2 \\
&= O(1)/O(h_n^{-1}) = O(h_n) \to 0, \\
&\frac{1}{n}\times\mathbb{E}\left[P_{12,W\nu}^4\right] / \left[\mathbb{E}[P_{12,W\nu}^2]\right]^2 = O(h_n^{-3}n^{-1})/O(h_n) = O(n^{-1}h_n^{-2}) \to 0,
\end{aligned}
$$

by the hypothesis on the bandwidth.

Thus, we can apply Theorem 1 in Hall (1984) to have

$$
\sqrt{Nh_n}P_{W\nu} \to_d \mathcal{N}(0, c'\Sigma_P c),
$$

as $n \to \infty$, where

$$
\Sigma_P = \lim_{n\to\infty} h_n \mathbb{E}[P_{12,W\nu}^2].
$$

To complete the characterization of Σ_P, from Step 1, we have

$$
\begin{aligned}
&(Nh_n)^{-1}\mathbb{E}[P_{12,W\nu}^2] \\
&= h_n\mathbb{E}\left[\left(\mathbb{E}\left[d_{12}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R_{12}'\gamma)|\xi_1,U_1,\xi_2,U_2\right]\right)^2\right] + o(1)
\end{aligned}
$$

so that

$$
\Sigma_P = \lim_{n\to\infty} h_n\mathbb{E}\left[\left(\mathbb{E}\left[d_{12}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R_{12}'\gamma)|\xi_1,U_1,\xi_2,U_2\right]\right)^2\right].
$$

C.2.3.4 Step 3: Asymptotic Normality of $Q_{W\nu}$

We use the CLT for martingale differences (Theorem 5.24 and Corollary 5.26 in White (2001)). Define $V_{n,t}(1 \leq t \leq N)$, a triangular array, as

$$
\begin{aligned}
V_{n,1} &= \frac{1}{N}Q_{12,W\nu}, \\
V_{n,2} &= \frac{1}{N}Q_{13,W\nu}, \\
&\vdots \\
V_{n,n-1} &= \frac{1}{N}Q_{1n,W\nu}, \\
&\vdots \\
V_{n,N} &= \frac{1}{N}Q_{n-1n,W\nu}.
\end{aligned}
$$

Notice that $Q_{ij,W\nu}$ is independent of $Q_{km,W\nu}$ if $i \neq k, m$ and $j \neq k, m$. Also, $Q_{ij,W\nu}$ is conditionally independent of $Q_{km,W\nu}$ even if $i = k$ or m, or $j = k$ or m; Note that $(\epsilon_{ijt}, \eta_{ijt})_{t=1,2}$ and $(\epsilon_{imt}, \eta_{imt})_{t=1,2}$ are conditionally independent given ξ_i, U_i by Assumption 3.1. Since $(\xi_i, U_i), i = 1, ...n$ is i.i.d., this implies that, for $1 < t \leq N$ such that ij corresponds to t,

$$\begin{aligned}\mathbb{E}[V_{n,t}|\{V_{n,s}; s<t\}] &= \mathbb{E}\left[\mathbb{E}[V_{n,t}|\{V_{n,s}; s<t\}, \{\xi_i, U_i\}_{i=1}^n|]\{V_{n,s}; s<t\}\right] \\ &= \mathbb{E}\left[\mathbb{E}[V_{n,t}|\xi_i, U_i, \xi_j, U_j]|\{V_{n,s}; s<t\}\right] \\ &= 0,\end{aligned}$$

as $\mathbb{E}[V_{n,t}|\xi_i, U_i, \xi_j, U_j] = \mathbb{E}[Q_{ij,W\nu}|\xi_i, U_i, \xi_j, U_j] = 0$ by construction. Thus, letting $\mathcal{F}_t \equiv \sigma(V_s|1 \leq s \leq t)$ be a sigma algebra generated by $V_1, ..., V_{t-1}$ ($\mathcal{F}_1$ is set to be a trivial σ-algebra) and $\mathbb{F} \equiv (\mathcal{F}_t)_{1\leq t\leq N}$ be a filtration, we have

$$\mathbb{E}[V_{n,t}|\mathcal{F}_{t-1}] = 0$$

for $1 \leq t \leq N$. Also, for each t, for some constant $C > 0$,

$$\begin{aligned}&\mathbb{E}[|V_{n,t}|] \leq \frac{3}{N}\mathbb{E}[|\Delta W_{12}||\nu_{12}||K_{h_n}(\Delta R_{12}'\gamma)|] \\ &\leq \frac{C}{Nh_n}\mathbb{E}[\|\Delta W_{12}\|^2]^{1/2}\mathbb{E}[\nu_{12}^2]^{1/2} < \infty,\end{aligned}$$

by Assumptions 3.7 and 3.8. This shows that $\{V_{n,t}\}$ is a martingale difference sequence.

Let $V_n = \sum_{t=1}^N V_{n,t}$. Define the variance of this sequence by

$$v_n^2 = \mathrm{Var}\left[\sum_{t=1}^N V_{n,t}\right] = N\,\mathrm{Var}[V_{n,1}] = \frac{1}{N}\mathbb{E}[Q_{12,W\nu}^2].$$

We can calculate that

$$\mathbb{E}[Q_{12,W\nu}^2] = \mathbb{E}\left[d_{12}(c'\Delta W_{12})^2\nu_{12}^2 K^2(\Delta R_{12}'\gamma)\right]$$
$$- \mathbb{E}\left[\mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}K_{h_n}(\Delta R_{12}'\gamma)|\xi_1, U_1, \xi_2, U_2]^2\right].$$

Observe that

$$\begin{aligned}
&\mathbb{E}\left[d_{12}(c'\Delta W_{12})^2\nu_{12}^2 K^2(\Delta R_{12}'\gamma)\right] \\
&= \frac{1}{h_n^2}\int \mathbb{E}[d_{12}(c'\Delta W_{12})^2\nu_{12}^2|\Delta R_{12}'\gamma = r]K(r/h_n)f_{R\gamma}(r)dr \\
&= \frac{1}{h_n}\int \mathbb{E}[d_{12}(c'\Delta W_{12})^2\nu_{12}^2|\Delta R_{12}'\gamma = rh_n]K(r)f_{R\gamma}(rh_n)dr \\
&= \frac{1}{h_n}c'\Sigma_{W\nu,2}c + o\left(\frac{1}{h_n}\right),
\end{aligned}$$

where the last line holds from the dominated convergence theorem under Assumptions 3.4, 3.6, and 3.8. Then,

$$\mathbb{E}[Q_{12,W\nu}^2] = \frac{1}{h_n}c'\Sigma_{W\nu,2}c - \frac{1}{h_n}c'\Sigma_P c + o\left(\frac{1}{h_n}\right).$$

Hence, we have

$$v_n^2 = \frac{1}{Nh_n}c'\Sigma_{W\nu,2}c - \frac{1}{Nh_n}c'\Sigma_P c + o\left(\frac{1}{Nh_n}\right).$$

The CLT for martingale differences holds if we can show the following two conditions:

$$\sum_{t=1}^{N}\mathbb{E}\left[\left(\frac{V_{n,t}}{v_n}\right)^{2+\delta}\right] \to 0 \text{ (Lyapunov)},$$

for some $\delta > 0$ as $n \to \infty$ and

$$\sum_{t=1}^{N} \left(\frac{V_{n,t}}{v_n} \right)^2 \to_p 1 \text{ (Stability)},$$

as $n \to \infty$. If these conditions are met, we can apply Theorem 5.24 and Corollary 5.26 in White (2001) to show that

$$\frac{V_n}{v_n} \to_d \mathcal{N}(0,1),$$

as $n \to \infty$. Since $\sqrt{Nh_n v_n} \to \sqrt{c'(\Sigma_{W\nu,2} - \Sigma_P)c}$, by Slutsky's lemma,

$$\sqrt{Nh_n} V_n \to_d \mathcal{N}(0, c'(\Sigma_{W\nu,2} - \Sigma_P)c),$$

which is equivalent to

$$\sqrt{Nh_n} Q_{W\nu} \to_d \mathcal{N}(0, c'(\Sigma_{W\nu,2} - \Sigma_P)c),$$

as $n \to \infty$.

For Lyapunov's condition, observe that for some constant $C > 0$,

$$\begin{aligned}
\mathbb{E}\left[|V_{n,1}|^3\right] &\leq \frac{C}{(Nh_n)^3} \int \mathbb{E}[\|\Delta W_{12}\|^3 |\nu_{12}|^3 | \Delta R_{12}'\gamma = r] |K(r/h_n)|^3 f_{R\gamma}(r) dr \\
&= \frac{Ch_n}{(Nh_n)^3} \int \mathbb{E}[\|\Delta W_{12}\|^3 |\nu_{12}|^3 | \Delta R_{12}'\gamma = rh_n] |K(r)|^3 f_{R\gamma}(rh_n) dr \\
&= O\left(\frac{1}{N^3 h_n^2} \right),
\end{aligned}$$

where the first inequality follows from Jensen's inequality, the last line follows from Cauchy-Schwartz and Assumptions 3.4, 3.6, and 3.8. Since $v_n =$

$O(1/\sqrt{Nh_n})$, we have

$$\sum_{t=1}^{N} \mathbb{E}\left[\left|\frac{V_{n,t}}{v_n}\right|^3\right] = NO\left(\frac{\sqrt{Nh_n}}{N^3 h_n^2}\right) = O\left(\frac{1}{(Nh_n)^{3/2}}\right) = o(1)$$

by Assumption 3.9. Thus Lyapunov's condition holds.

For the stability condition, we can alternatively show that

$$\frac{1}{v_n^2}\sum_{t=1}^{N}\left(V_{n,t}^2 - \mathbb{E}[V_{n,t}^2]\right) \to_p 0,$$

as $n \to \infty$. Note that

$$\frac{1}{v_n^2}\sum_{t=1}^{N}\left(V_{n,t}^2 - \mathbb{E}[V_{n,t}^2]\right) = \frac{1}{Nv_n^2}\left(\frac{1}{N}\sum_{i<j} Q_{ij,W\nu}^2 - \mathbb{E}[Q_{12,W\nu}^2]\right).$$

Since $Nv_n^2 = O(1/h_n)$, we need to show that the remaining term is $o_p(1/h_n)$. Since $Q_{ij,W\nu}$ is independent fron $Q_{km,W\nu}$ if there is no common node,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{N}\sum_{i<j} Q_{ij,W\nu}^2 - \mathbb{E}[Q_{12,W\nu}^2]\right)^2\right] \\
&= \frac{\mathrm{Var}[Q_{12,W\nu}^2]}{N} + \frac{2(n-2)}{N}\,\mathrm{Cov}[Q_{12,W\nu}^2, Q_{13,W\nu}^2] \\
&\leq \frac{\mathbb{E}[Q_{12,W\nu}^4]}{N} + \frac{2(n-2)}{N}\mathbb{E}[Q_{12,W\nu}^2 \times Q_{13,W\nu}^2].
\end{aligned}$$

The first term in the far right-hand side is bounded as follows: For some constant $C > 0$,

$$\frac{\mathbb{E}[Q^4_{12,W\nu}]}{N} \leq \frac{C}{Nh_n^4}\int \mathbb{E}[\|\Delta W_{12}\|^4\nu^4_{12}|\Delta R'_{12}\gamma = r]K^4(r)f_{R\gamma}(r)dr$$
$$= \frac{C}{Nh_n^3}\int \mathbb{E}[\|\Delta W_{12}\|^4\nu^4_{12}|\Delta R'_{12}\gamma = rh_n]K^4(r)f_{R\gamma}(rh_n)dr$$
$$= O\left(\frac{1}{Nh_n^3}\right),$$

where the first inequality follows from Jensen's inequality, and the last line follows from Cauchy-Schwartz and Assumptions 3.4, 3.6, and 3.8. The second term on the far right-hand side is bounded as follows: For some constant $C > 0$,

$$\frac{2(n-2)}{N}\mathbb{E}[Q^2_{12,W\nu} \times Q^2_{13,W\nu}]$$
$$\leq \frac{C(n-2)}{Nh_n^4}\mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|^2\nu^2_{12}|\Delta R'_{12}\gamma = r_1, \xi_1, U_1]$$
$$\times \mathbb{E}[\|\Delta W_{13}\|^2\nu^2_{13}|\Delta R'_{13}\gamma = r_2, \xi_1, U_1]$$
$$\times K^2(r_1/h_n)K^2(r_2/h_n)f_{R\gamma|\xi_1,U_1}(r_1)f_{R\gamma|\xi_1,U_1}(r_2)dr_1dr_2\Big]$$
$$= \frac{C(n-2)}{Nh_n^2}\mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|^2\nu^2_{12}|\Delta R'_{12}\gamma = r_1h_n, \xi_1, U_1]$$
$$\times \mathbb{E}[\|\Delta W_{13}\|^2\nu^2_{13}|\Delta R'_{13}\gamma = r_2h_n, \xi_1, U_1]$$
$$\times K^2(r_1)K^2(r_2)f_{R\gamma|\xi_1,U_1}(r_1h_n)f_{R\gamma|\xi_1,U_1}(r_2h_n)dr_1dr_2\Big]$$
$$= O\left(\frac{1}{nh_n^2}\right)$$

Thus,

$$h_n\mathbb{E}\left[\left(\frac{1}{N}\sum_{i<j}Q^2_{ij,W\nu} - \mathbb{E}[Q^2_{12,W\nu}]\right)^2\right] = O\left(\frac{1}{Nh_n^2}\right) + O\left(\frac{1}{nh_n}\right) = o(1),$$

and by Markov's inequality,

$$\sqrt{h_n}\left(\frac{1}{N}\sum_{i<j} Q_{ij,W\nu}^2 - \mathbb{E}[Q_{12,W\nu}^2]\right) = o_p(1).$$

Then,

$$\frac{1}{v_n^2}\sum_{t=1}^{N}\left(V_{n,t}^2 - \mathbb{E}[V_{n,t}^2]\right) = O(h_n)\times o_p\left(\frac{1}{\sqrt{h_n}}\right) = o_p(\sqrt{h_n}) = o_p(1),$$

which shows the stability condition.

C.2.3.5 Step 3: Conclusion

By Steps 0-3, we have established that if $c_W'\Sigma_{W\nu,1}c_W > 0$,

$$\begin{aligned}
&\sqrt{n}c'S_{WW}^{-1}S_{W\nu}\\
&= \underbrace{\sqrt{n}L_{W\nu}}_{\to_d\mathcal{N}(0,c_W'\Sigma_{W\nu,1}c_W)} + \underbrace{\frac{\sqrt{n}}{\sqrt{Nh_n}}}_{\to 0}\times\underbrace{\sqrt{Nh_n}(P_{W\nu}+Q_{W\nu})}_{\to_d\mathcal{N}(0,c_W'\Sigma_{W\nu,2}c_W)} + o_p(\sqrt{n}/\sqrt{Nh_n})\\
&\to_d \mathcal{N}(0,c_W'\Sigma_{W\nu,1}c_W),
\end{aligned}$$

as $n\to\infty$ by Assumption 3.9, and if $c_W'\Sigma_{W\nu,1}c_W = 0$,

$$\begin{aligned}
&\sqrt{Nh_n}c'S_{WW}^{-1}S_{W\nu}\\
&= \underbrace{\sqrt{Nh_n}L_{W\nu}}_{\to_p 0} + \underbrace{\sqrt{Nh_n}(P_{W\nu}+Q_{W\nu})}_{\to_d\mathcal{N}(0,c_W'\Sigma_{W\nu,2}c_W)} + o_p(1)\\
&\to_d \mathcal{N}(0,c_W'\Sigma_{W\nu,2}c_W),
\end{aligned}$$

as $n\to\infty$. This completes the proof. □

C.2.4 Proof of Lemma C.4

Proof. By expanding $K(\Delta R'_{ij}\widehat{\gamma}_n/h_n)$ around $\Delta R'_{ij}\gamma$, we have

$$\widehat{S}_{WW} = S_{WW} + \frac{1}{Nh_n^2}\sum_{i<j} d_{ij}\Delta W_{ij}\Delta W'_{ij}\Delta R'_{ij}(\widehat{\gamma}_n - \gamma)k(c^*_{ij,n})$$

where $c^*_{ij,n}$ is in between $\Delta R'_{ij}\gamma$ and $\Delta R'_{ij}\widehat{\gamma}_n$ and k is the first derivative of K. Thus, for some constant $C > 0$,

$$\|\widehat{S}_{WW} - S_{WW}\| \leq \underbrace{\frac{C}{N}\sum_{i<j}\|\Delta W_{ij}\|^2\|\Delta R_{ij}\|}_{D_{4,1}} \times h_n^{-2}\|\widehat{\gamma}_n - \gamma\|$$

Notice that $\|\Delta W_{ij}\| = \|w(X_{i1}, X_{j1}) - w(X_{i2}, X_{j2})\|, \|R_{ij}\| = \|r(Z_{i1}, Z_{j1} - r(Z_{i2}, Z_{j2})\|$ are symmetric in i and j by the symmetry of w, r, and $\|\cdot\|$ so that $D_{4,1}$ is a second-order U-statistics. Also,

$$\mathbb{E}[\|\Delta W_{12}\|^2\|\Delta R_{ij}\|] < \infty.$$

by Cauchy-Schwartz with Assumption 3.7. Thus, we can apply the law of large numbers for U-statistics (Hoeffding, 1961):

$$D_{4,1} = O_p(1).$$

By the hypothesis and Assumption 3.10,

$$
\begin{aligned}
h_n^{-2}\|\widehat{\gamma}_n - \gamma\| &= \frac{\|\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma)\|}{\sqrt{Nh_n^5}} \\
&= \frac{\|\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma)\|}{\sqrt{Nh_n^{2k+3}}} \times \sqrt{h_n^{2k-2}} \\
&= o_p(1),
\end{aligned}
$$

as $\sqrt{Nh_n^{2k+3}}$ is either diverging or $O(1)$, $\sqrt{h^{2k-2}} = o(1)$ for $k \geq 2$. Thus,

$$
\|\widehat{S}_{WW} - S_{WW}\| = O_p(1) \times o_p(1) = o_p(1).
$$

This shows that $\widehat{S}_{WW} = S_{WW} + o_p(1)$. □

C.2.5 Proof of Lemma C.5

Proof. Expanding $K(\Delta R_{ij}'\widehat{\gamma}_n/h_n)$ around $\Delta R_{ij}'\gamma$, for some constant $C > 0$, we have

$$
\begin{aligned}
&\sqrt{Nh_n}\|\widehat{S}_{W\lambda} - S_{W\lambda}\| \\
&\leq \underbrace{\frac{1}{Nh_n^2}\sum_{i<j}\|\Delta W_{ij}\|\|\Delta R_{ij}\||\lambda_{ij}||k(\Delta R_{ij}'\gamma/h_n)|}_{D_{5,1}} \times \sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\| \\
&+ \underbrace{\frac{1}{Nh_n^2}\sum_{i<j}\|\Delta W_{ij}\|\|\Delta R_{ij}\|^2|\lambda_{ij}||k(\Delta R_{ij}'\gamma/h_n)|}_{D_{5,2}} \times \frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|}{2h_n} \\
&+ C\underbrace{\frac{1}{N}\sum_{i<j}\|\Delta W_{ij}\|\|\Delta R_{ij}\|^3|\lambda_{ij}|}_{D_{5,3}} \times \frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|^3}{6h_n^4}
\end{aligned}
$$

We follow the following steps to bound the right hand side.

C.2.5.1 Step 1 $D_{5,1}$ and $D_{5,2}$

Observe that

$$\begin{aligned}\mathbb{E}[D_{5,1}] &= \frac{1}{h_n^2}\int \mathbb{E}[\|\Delta W_{12}\|\|\Delta R_{12}\||\Lambda_{12}||\Delta R_{12}'\gamma = r]|r||k(r/h_n)|f_{R\gamma}(r)dr \\ &= \int \mathbb{E}[\|\Delta W_{12}\|\|\Delta R_{12}\||\Lambda_{12}||\Delta R_{12}'\gamma = rh_n]|r||k(r)|f_{R\gamma}(rh_n)dr \\ &= O(1),\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6 and 3.8. Also, writing each summand of $D_{5,1}$ by $D_{5,1,ij}$, we have

$$\begin{aligned}\mathrm{Var}[D_{5,1}] &= \frac{1}{Nh_n^4}\,\mathrm{Var}[D_{5,1,12}] + \frac{2(n-2)}{Nh_n^4}\,\mathrm{Cov}[D_{5,1,12}, D_{5,1,13}] \\ &\leq \frac{1}{Nh_n^4}\mathbb{E}[D_{5,1,12}^2] + \frac{2(n-2)}{Nh_n^4}\mathbb{E}[D_{5,1,12} \times D_{5,1,13}].\end{aligned}$$

The first term on the far right side is $O(1/(Nh_n^2))$ because

$$\begin{aligned}\mathbb{E}[D_{5,1,12}^2] &= \int \mathbb{E}[\|\Delta W_{12}\|^2\|\Delta R_{12}\|^2\Lambda_{12}^2|\Delta R_{12}'\gamma = r]r^2k(r/h_n)^2f_{R\gamma}(r)dr \\ &= h_n^2\int \mathbb{E}[\|\Delta W_{12}\|^2\|\Delta R_{12}\|^2\Lambda_{12}^2|\Delta R_{12}'\gamma = rh_n]r^2k(r)^2f_{R\gamma}(rh_n)dr \\ &= O(h_n^2),\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6, and 3.8. The second term on the far right side is $O(1/n)$ because

$$\begin{aligned}
&\mathbb{E}[D_{5,1,12} \times D_{5,1,13}] \\
&= \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\| \|\Delta R_{12}\| |\Lambda_{12}| | \Delta R_{12}'\gamma = r_1, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\| \|\Delta R_{13}\| |\Lambda_{13}| | \Delta R_{13}'\gamma = r_2, \xi_1, U_1] \\
&\times |r_1||r_2||k(r_1/h_n)||k(r_2/h_n)| f_{R\gamma|\xi_1,U_1}(r_1) f_{R\gamma|\xi_1,U_1}(r_2) dr_1 dr_2 \Big] \\
&= h_n^4 \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\| \|\Delta R_{12}\| |\Lambda_{12}| | \Delta R_{12}'\gamma = r_1 h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\| \|\Delta R_{13}\| |\Lambda_{13}| | \Delta R_{13}'\gamma = r_2 h_n, \xi_1, U_1] \\
&\times |r_1||r_2||k(r_1)||k(r_2)| f_{R\gamma|\xi_1,U_1}(r_1 h_n) f_{R\gamma|\xi_1,U_1}(r_2 h_n) dr_1 dr_2 \Big] \\
&= O(h_n^4),
\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6, and 3.8. Thus,

$$\mathrm{Var}[D_{5,1}] = O\left(\frac{1}{Nh_n^2}\right) + O\left(\frac{1}{n}\right) = o(1),$$

since $Nh_n^2 = Nh^{2k+3} \times h_n^{1-2k}$ diverges by the hypothesis. Thus,

$$D_{5,1} = O_p(1).$$

By a similar calculation, we have

$$D_{5,2} = O_p(1).$$

C.2.5.2 Step 2: $D_{5,3}$

First, observe that

$$\mathbb{E}[D_{5,3}] = \mathbb{E}[\|\Delta W_{12}\|\|\Delta R_{12}\|^4|\Lambda_{12}|] < \infty$$

by Hölder's inequality with Assumption 3.7. Note that by construction, Λ_{ij} is written as a function of ξ_i and ξ_j with symmetry with respect to i and j, which implies that $D_{5,3}$ is a second-order U-statistics. Since each summand is non-negative and has a finite mean, we can apply the law of large numbers for U-statistics (Hoeffding, 1961) to show that

$$D_{5,3} = O_p(1).$$

C.2.5.3 Step 3: Conclusion

Finally, by Assumption 3.10 and the hypothesis,

$$\begin{aligned}
&\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\| = o_p(1), \\
&\frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|^2}{2h_n} = \frac{\|\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma)\|^2}{2\sqrt{Nh_n^3}} = o_p(1), \\
&\frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|^3}{6h_n^4} = \frac{\|\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma)\|^3}{6Nh_n^5} = o_p(1).
\end{aligned}$$

Thus,

$$\sqrt{Nh_n}\|\widehat{S}_{W\lambda} - S_{W\lambda}\| = o_p(1).$$

This implies that

$$\widehat{S}_{W\lambda} = S_{W\lambda} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

This completes the proof. □

C.2.6 Proof of Lemma C.6

Proof. By expanding $K(\Delta R_{ij}'\widehat{\gamma}_n/h_n)$ around $\Delta R_{ij}'\gamma$, we have

$$\begin{aligned}
&\sqrt{Nh_n}(\widehat{S}_{W\nu} - S_{W\nu}) \\
&= \underbrace{\frac{1}{Nh_n^2}\sum_{i<j} d_{ij}\Delta W_{ij}\Delta R_{ij}'\nu_{ij}k(\Delta R_{ij}'\gamma/h_n)}_{D_{6,1}}\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma) \\
&+ (\widehat{\gamma}_n - \gamma)'\underbrace{\frac{1}{Nh_n^2}\sum_{i<j} d_{ij}\Delta W_{ij}\Delta R_{ij}\Delta R_{ij}'\nu_{ij}k(\Delta R_{ij}'\gamma/h_n)}_{D_{6,2}}\sqrt{Nh_n}\frac{(\widehat{\gamma}_n - \gamma)}{h_n} \\
&+ \underbrace{\frac{\sqrt{Nh_n}}{Nh_n^4}\sum_{i<j} d_{ij}\Delta W_{ij}\nu_{ij}k(c_{ij,n}^*/h_n)\left(\Delta R_{ij}'(\widehat{\gamma}_n - \gamma)\right)^3}_{D_{6,3}}
\end{aligned}$$

We bound each component by the following steps.

C.2.6.1 Step 1: $D_{6,1}$ and $D_{6,2}$

Note that $\mathbb{E}[D_{6,1}] = \mathbb{E}[D_{6,2}] = 0$ by the conditional mean independence of ν_{ij}. Write $D_{6,1,ij}$ as each summand of $D_{6,1}$. Observe that, by the similar calculation as above,

$$\text{Var}[\|D_{6,1}\|] \leq \frac{1}{Nh_n^4}\mathbb{E}[\|D_{6,1,12}\|^2] + \frac{2(n-2)}{Nh_n^4}\mathbb{E}[\|D_{6,1,12}\| \times \|D_{6,1,13}\|].$$

The first term on the right hand side is $O(1/(Nh_n^3))$ since

$$\begin{aligned}
\mathbb{E}[\|D_{6,1,12}\|^2] &\leq \int \mathbb{E}[\|\Delta W_{12}\|^2\|\Delta R_{12}\|^2\nu_{12}^2|\Delta R_{12}'\gamma = r]k(r/h_n)^2 f_{R\gamma}(r)dr \\
&= h_n \int \mathbb{E}[\|\Delta W_{12}\|^2\|\Delta R_{12}\|^2\nu_{12}^2|\Delta R_{12}'\gamma = rh_n]k(r)^2 f_{R\gamma}(rh_n)dr \\
&= O(h_n),
\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6, and 3.8. The second term on the right hand side is $O(1/(nh_n^2))$ since

$$\begin{aligned}
&\mathbb{E}[\|D_{6,1,12}\| \times \|D_{6,1,13}\|] \\
&\leq \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|\|\Delta R_{12}\||\nu|_{12}|\Delta R_{12}'\gamma = r_1, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\|\|\Delta R_{13}\||\nu|_{13}|\Delta R_{13}'\gamma = r_2, \xi_1, U_1] \\
&\times |k(r_1/h_n)||k(r_2/h_n)|f_{R\gamma|\xi_1,U_1}(r_1)f_{R\gamma|\xi_1,U_1}(r_2)dr_1dr_2\Big] \\
&= h_n^2\mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\|\|\Delta R_{12}\||\nu|_{12}|\Delta R_{12}'\gamma = r_1h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\|\|\Delta R_{13}\||\nu|_{13}|\Delta R_{13}'\gamma = r_2h_n, \xi_1, U_1] \\
&\times |k(r_1)||k(r_2)|f_{R\gamma|\xi_1,U_1}(r_1h_n)f_{R\gamma|\xi_1,U_1}(r_2h_n)dr_1dr_2\Big] \\
&= O(h_n^2),
\end{aligned}$$

where the last line holds from Assumptions 3.4, 3.6, and 3.8. Hence,

$$\mathrm{Var}[\|D_{6,1}\|] = O\left(\frac{1}{Nh_n^3}\right) + O\left(\frac{1}{nh_n^2}\right) = o(1),$$

since both $Nh_n^3 = Nh_n^{2k+3} \times h_n^{-2k}$ and $nh_n^2 \approx \sqrt{Nh_n^4} = \sqrt{Nh_n^{2k+3}} \times \sqrt{h_n^{-2k+1}}$ diverge under the hypothesis. This shows that

$$D_{6,1} = o_p(1).$$

A similar calculation shows that

$$D_{6,2} = o_p(1),$$

as well.

C.2.6.2 Step 2: $D_{6,3}$

Observe that, for some constant $C > 0$

$$\|D_{6,3}\| \leq C \underbrace{\frac{1}{N}\sum_{i<j} \|\Delta W_{ij}\| \|\Delta R_{ij}\|^3 |\nu_{ij}|}_{D_{6,4}} \times \frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|^3}{Nh_n^4}.$$

Observe that

$$\mathbb{E}[D_{6,4}] = \mathbb{E}[\|\Delta W_{12}\| \|\Delta R_{12}\|^3 |\nu_{12}|] < \infty,$$

by Cauchy-Schwartz with Assumption 3.6. Also, by writing each summand of $D_{6,4}$ as $D_{6,4,ij}$, we have

$$\mathrm{Var}[D_{6,4}] \leq \frac{\mathbb{E}[D_{6,4,12}^2]}{N} + \frac{2(n-2)}{N}\mathbb{E}[D_{6,4,12} \times D_{6,4,13}].$$

Since

$$\mathbb{E}[D_{6,4,12}^2] = \mathbb{E}[\|\Delta W_{12}\|^2 \|\Delta R_{12}\|^6 \nu_{12}^2] < \infty$$
$$\mathbb{E}[D_{6,4,12} \times D_{6,4,13}] = \mathbb{E}[\|\Delta W_{12}\| \|\Delta W_{13}\| \|\Delta R_{12}\|^3 \|\Delta R_{13}\|^3 |\nu_{12}||\nu_{13}|] < \infty$$

by Hölder's inequality with Assumption 3.7,

$$\mathrm{Var}[D_{6,4}] = O\left(\frac{1}{N}\right) + O\left(\frac{1}{n}\right) = o(1).$$

This shows that

$$D_{6,4} = O_p(1).$$

Hence, by the previous calculation for the term involving $\widehat{\gamma}_n - \gamma$,

$$\|D_{6,3}\| = O_p(1) \times o_p(1) = o_p(1).$$

C.2.6.3 Step 3: Conclusion

By the above steps and the hypothesis on $\widehat{\gamma}_n - \gamma$,

$$\sqrt{Nh_n}\|\widehat{S}_{W\nu} - S_{W\nu}\| = o_p(1).$$

This implies that

$$\widehat{S}_{W\nu} = S_{W\nu} + o_p\left(\frac{1}{\sqrt{Nh_n}}\right).$$

This completes the proof. □

C.2.7 Proof of Lemma C.7

Proof. Define

$$S_{ij,1} \equiv 2d_{ij}K_{h_n}(\Delta R_{ij}'\gamma)\Delta W_{ij}\nu_{ij},$$
$$S_{ij,2} \equiv 2d_{ij}K_{h_n}(\Delta R_{ij}'\gamma)\Delta W_{ij}\lambda_{ij},$$
$$S_{ij,3} \equiv 2d_{ij}K_{h_n}(\Delta R_{ij}'\gamma)\Delta W_{ij}\Delta W_{ij}'(\beta - \widehat{\beta}_n).$$

Since

$$\Delta\widehat{\epsilon}_{ij} = \Delta W_{ij}'(\beta - \widehat{\beta}_n) + \lambda_{ij} + \nu_{ij},$$

we have

$$S_{ij} = S_{ij,1} + S_{ij,2} + S_{ij,3}.$$

Thus,

$$\widetilde{\Sigma}_{W\nu,1} = \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \underbrace{\frac{1}{3}(S_{ij,1}S'_{ik,1} + S_{ij,1}S'_{jk,1} + S_{ik,1}S_{jk,1})}_{D_{7,ijk}}}_{D_7} + \mathcal{O}_7,$$

where $\mathcal{O}_7$ is the remainder term.

We first show that $c'D_7c \to_p c'\Sigma_{W\nu,1}c$. Note that

$$\begin{aligned}
&\mathbb{E}[c'D_7c] \\
&= \mathbb{E}[c'S_{12,1}S_{13,1}c] \\
&= \frac{4}{h_n^2}\int \mathbb{E}[d_{12}d_{13}c'\Delta W_{12}\Delta W'_{13}c\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = s_1, \Delta R'_{13}\gamma = s_2] \\
&\times K(s_1/h_n)K(s_2/h_n)f_{R\gamma,2}(s_1, s_2)ds_1ds_2 \\
&= 4\int \mathbb{E}[d_{12}d_{13}c'\Delta W_{12}\Delta W'_{13}c\nu_{12}\nu_{13}|\Delta R'_{12}\gamma = s_1h_n, \Delta R'_{13}\gamma = s_2h_n] \\
&\times K(s_1)K(s_2)f_{R\gamma,2}(s_1h_n, s_2h_n)ds_1ds_2 \\
&\to c'\Sigma_{W\nu,1}c,
\end{aligned}$$

as $n \to \infty$ by the dominated convergence theorem under Assumptions 3.4, 3.6, and 3.8. Define the third order U-statistics

$$U_{n,1} = \binom{n}{3}^{-1} \sum_{i<j<k} p_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k),$$

where $\boldsymbol{\xi}_i = (\xi_i, U_i)$ and

$$p_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k) = \mathbb{E}[c'D_{7,ijk}c|\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k]$$

By the calculation of Graham et al. (2019) in Appendix B,

$$\begin{aligned}
&\mathbb{E}[(c'D_7c - U_{n,1})^2] \\
&= \binom{n}{3}^{-1} \mathbb{E}\left[(c'D_{7,123}c - \mathbb{E}[c'D_{7,123}c|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3])^2\right] \\
&+ \binom{n}{3}^{-2} \times 3\binom{n}{2}\binom{n-2}{2} \times \mathbb{E}\left[c'D_{7,123}c - \mathbb{E}[c'D_{7,123}c|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3]\right] \\
&\times \mathbb{E}\left[c'D_{7,124}c - \mathbb{E}[c'D_{7,124}c|\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_4]\right] \\
&= O\left(\frac{\mathbb{E}[(c'D_{7,123}c)^2]}{n^3}\right).
\end{aligned}$$

Observe that

$$\begin{aligned}
&\mathbb{E}[(c'D_{7,123}c)^2] \\
&= \frac{1}{9}\left(3\mathbb{E}[(c'S_{12,1}c \times c'S_{13,1}c)^2] + 6\mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{23,1}c]\right) \\
&= O\left(\frac{1}{h_n^2}\right),
\end{aligned}$$

since for some positive constant $C > 0$,

$$
\begin{aligned}
&\mathbb{E}[(c'S_{12,1}c \times c'S_{13,1}c)^2] \\
&\leq \frac{C}{h_n^4}\int \mathbb{E}[\|\Delta W_{12}\|^2\|\Delta W_{13}\|^2\nu_{12}^2\nu_{13}^2|\Delta R_{12}'\gamma = s_1, \Delta R_{13}'\gamma = s_2] \\
&\times K^2(s_1/h_n)K^2(s_2/h_n)f_{R\gamma,2}(s_1, s_2)ds_1ds_2 \\
&= \frac{C}{h_n}\int \mathbb{E}[\|\Delta W_{12}\|^2\|\Delta W_{13}\|^2\nu_{12}^2\nu_{13}^2|\Delta R_{12}'\gamma = s_1h_n, \Delta R_{13}'\gamma = s_2h_n] \\
&\times K^2(s_1)K^2(s_2)f_{R\gamma,2}(s_1h_n, s_2h_n)ds_1ds_2 \\
&= O\left(\frac{1}{h_n^2}\right),
\end{aligned}
$$

as $n \to \infty$ with the last line coming from Assumption 3.4, 3.6, and 3.8, and,

$$
\begin{aligned}
&\mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{23,1}c] \\
&= \mathbb{E}[(c'S_{12,1}c)^2 \times \mathbb{E}[c'S_{13,1}c|\xi_1, U_1] \times \mathbb{E}[c'S_{23,1}c|\xi_2, U_2]] \\
&= \mathbb{E}[\mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c|\xi_1, U_1] \times \mathbb{E}[c'S_{23,1}|\xi_2, U_2]] \\
&= \mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c] \times \mathbb{E}[c'S_{12,1}c] \\
&\leq O(1) \times \frac{C}{h_n^3}\int \mathbb{E}[\|\Delta W_{12}\|^2\nu_{12}^2|\Delta R_{12}'\gamma = s_1, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\||\nu_{13}||\Delta R_{13}'\gamma = s_2, \xi_1, U_1] \\
&\times K^2(s_1/h_n)K(s_2/h_n)f_{R\gamma|\xi_1,U_1}(s_1)f_{R\gamma|\xi_1,U_1}(s_2)ds_1ds_2ds_3 \\
&= O(1) \times \frac{C}{h_n}\int \mathbb{E}[\|\Delta W_{12}\|^2\nu_{12}^2|\Delta R_{12}'\gamma = s_1h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\||\nu_{13}||\Delta R_{13}'\gamma = s_2h_n, \xi_1, U_1] \\
&\times K^2(s_1)K(s_2)f_{R\gamma|\xi_1,U_1}(s_1h_n)f_{R\gamma|\xi_1,U_1}(s_2h_n)ds_1ds_2ds_3 \\
&= O\left(\frac{1}{h_n}\right),
\end{aligned}
$$

where the first to third lines follow from the conditional independence of $S_{ij,1}$, the random sampling of ξ_i, and the conditional independence and exchangeability of U_i under Assumption 3.1, and the last line follows from Assumptions 3.4, 3.6, and 3.8. Observe that, by conditional independence of $S_{ij,1}$ and $S_{ik,1}$ given ξ_i, U_i and $S_{ij,1} = S_{ji,1}$, one can show that

$$
\begin{aligned}
&\mathbb{E}[c'D_{7,123}c \times c'D_{7,124}c] \\
&= \frac{1}{9}\big\{2\mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{14,1}c] \\
&+ 2\mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c] \times \mathbb{E}[c'S_{13,1}c] + 5\mathbb{E}[c'S_{12,1}c \times c'S_{13,1}c]^2\big\} \\
&= \left(\frac{1}{h_n}\right),
\end{aligned}
$$

where the last line holds since

$$
\begin{aligned}
&\mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c \times c'S_{14,1}c] \\
&\leq \frac{C}{h_n^4}\mathbb{E}\bigg[\int \mathbb{E}[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R_{12}'\gamma = s_1, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{12}\| \nu_{12} | \Delta R_{12}'\gamma = s_2, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{14}\| \nu_{14} | \Delta R_{14}'\gamma = s_3, \xi_1, U_1] \times K^2(s_1/h_n) K(s_2/h_n) K(s_3/h_n) \\
&\times f_{R\gamma|\xi_1,U_1}(s_1) f_{R\gamma|\xi_1,U_1}(s_2) f_{R\gamma|\xi_1,U_1}(s_3) ds_1 ds_2 ds_3\bigg] \\
&= \frac{C}{h_n}\mathbb{E}\bigg[\int \mathbb{E}[\|\Delta W_{12}\|^2 \nu_{12}^2 | \Delta R_{12}'\gamma = s_1 h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{12}\| \nu_{12} | \Delta R_{12}'\gamma = s_2 h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{14}\| \nu_{14} | \Delta R_{14}'\gamma = s_3 h_n, \xi_1, U_1] \times K^2(s_1) K(s_2) K(s_3) \\
&\times f_{R\gamma|\xi_1,U_1}(s_1 h_n) f_{R\gamma|\xi_1,U_1}(s_2 h_n) f_{R\gamma|\xi_1,U_1}(s_3 h_n) ds_1 ds_2 ds_3\bigg] \\
&= O\left(\frac{1}{h_n}\right),
\end{aligned}
$$

where the last line holds by Assumptions 3.4, 3.6, 3.8,

$$
\begin{aligned}
&\mathbb{E}[(c'S_{12,1}c)^2 \times c'S_{13,1}c] \\
&\leq \frac{C}{h_n^3}\mathbb{E}\bigg[\int \mathbb{E}[\|\Delta W_{12}\|^2\nu_{12}^2|\Delta R_{12}'\gamma = s_1, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\||\nu_{13}||\Delta R_{13}'\gamma = s_2, \xi_1, U_1] \\
&\times K^2(s_1/h_n)K(s_2/h_n)f_{R\gamma|\xi_1,U_1}(s_1)f_{R\gamma|\xi_1,U_1}(s_2)ds_1ds_2\bigg] \\
&\leq \frac{C}{h_n}\mathbb{E}\bigg[\int \mathbb{E}[\|\Delta W_{12}\|^2\nu_{12}^2|\Delta R_{12}'\gamma = s_1h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\||\nu_{13}||\Delta R_{13}'\gamma = s_2h_n, \xi_1, U_1] \\
&\times K^2(s_1)K(s_2)f_{R\gamma|\xi_1,U_1}(s_1h_n)f_{R\gamma|\xi_1,U_1}(s_2h_n)ds_1ds_2\bigg] \\
&= O\left(\frac{1}{h_n}\right),
\end{aligned}
$$

where the last equality holds from Assumptions 3.4, 3.6, and 3.8, and

$$
\begin{aligned}
&\mathbb{E}[c'S_{12}c \times c'S_{13}c] \\
&\leq \frac{C}{h_n^2}\mathbb{E}\bigg[\int \mathbb{E}[\|\Delta W_{12}\||\nu_{12}||\Delta R_{12}'\gamma = s_1, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\||\nu_{13}||\Delta R_{13}'\gamma = s_2, \xi_1, U_1] \\
&\times K(s_1/h_n)K(s_2/h_n)f_{R\gamma|\xi_1,U_1}(s_1)f_{R\gamma|\xi_1,U_1}(s_2)ds_1ds_2\bigg] \\
&\leq C\mathbb{E}\bigg[\int \mathbb{E}[\|\Delta W_{12}\||\nu_{12}||\Delta R_{12}'\gamma = s_1h_n, \xi_1, U_1] \\
&\times \mathbb{E}[\|\Delta W_{13}\||\nu_{13}||\Delta R_{13}'\gamma = s_2h_n, \xi_1, U_1] \\
&\times K(s_1)K(s_2)f_{R\gamma|\xi_1,U_1}(s_1h_n)f_{R\gamma|\xi_1,U_1}(s_2h_n)ds_1ds_2\bigg] \\
&= O(1),
\end{aligned}
$$

where the last equality holds from Assumptions 3.4, 3.6, and 3.8. Thus,

$$\mathbb{E}\left[(c'D_7c - U_{n,1})^2\right] = O\left(\frac{1}{n^3h_n^2}\right) = o(1).$$

Thus, $c'D_1$ is well approximated by U_n. Also, since $nh_n^2 \to \infty$ with the stated assumption on h_n,

$$\mathbb{E}\left[\left(p_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k)\right)^2\right] = O\left(\mathbb{E}[(c'D_{7,123}c)^2]\right) = O\left(\frac{n}{nh_n^2}\right) = o(1) \times O(n),$$

and by Lemma A.3 of Ahn and Powell (1993), we have

$$U_n = \mathbb{E}[U_{n,1}] + o_p(1).$$

This shows that

$$\begin{aligned} c'D_7c &= \mathbb{E}[U_{n,1}] + \underbrace{c'D_7c - U_{n,1}}_{=o_p(1)} + \underbrace{U_{n,1} - \mathbb{E}[U_{n,1}]}_{=o_p(1)} \\ &= \mathbb{E}[c'D_7c] + o_p(1) \\ &= c'\Sigma_{W\nu,1}c + o_p(1). \end{aligned}$$

This completes $c'D_7c \to_p c'\Sigma_{W\nu,1}c$ as $n \to \infty$.

The remainder term $\mathcal{O}_7$ with each term involving either $S_{ij,2}$ or (and) $S_{ij,3}$ is of smaller order than D_7 since $S_{ij,2}$ and $S_{ij,3}$ involve $\|\widehat{\beta}_n - \beta\| = O_p(1/\sqrt{n})$ and $\lambda_{ij} \approx h_n$ for large n; By computing in a similar way as before, we can establish that $\mathbb{E}[|c'\mathcal{O}_7c|] = o(1)$ and $\mathrm{Var}[c'\mathcal{O}_7c] = o(1)$ so that $|c'\mathcal{O}_7c| \to_p 0$.

Hence,

$$|c'\widetilde{\Sigma}_{\nu,1}c - c'\Sigma_{W\nu,1}c| \leq |c'D_1c - c'\Sigma_{W\nu,1}c| + |c'\mathcal{O}_7c| = o_p(1),$$

which completes the proof for Lemma C.7. □

C.2.8 Proof of Lemma C.8

Proof. Define

$$\widehat{S}_{ij,1} = \frac{2}{h_n^2} d_{ij} k\left(\frac{c^*_{ij,n}}{h_n}\right) \Delta W_{ij} \Delta R'_{ij} \Delta\widehat{\epsilon}_{ij}$$

where $c^*_{ij,n}$ is in between $\Delta R'_{ij}\gamma$ and $\Delta R'_{ij}\widehat{\gamma}_n$. In the following argument, we treat $\Delta\widehat{\epsilon}_{ij}$ in $\widehat{S}_{ij,1}$ as ν_{ij} because only the existence of higher moments is important and bounding the terms involving ν_{ij} suffices. By the expression for $\Delta\widehat{\epsilon}_{ij}$, we have

$$\widehat{S}_{ij} = S_{ij,1} + S_{ij,2} + S_{ij,3} + \widehat{S}_{ij,1}(\widehat{\gamma}_n - \gamma).$$

By the proof of C.7, we know that

$$\binom{n}{3}^{-1} \sum_{i<j<k} \frac{1}{3}(S_{ij,p}S'_{ik,p} + S_{ij,p}S'_{jk,p} + S_{ik,p}S'_{jk,p}) = o_p(1),$$

for $p = 2, 3$.

$$
\begin{aligned}
&\|\widehat{\Sigma}_{W\nu,1} - \widetilde{\Sigma}_{W\nu,1}\| \\
&\leq \sum_{p=1}^{3} \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} D^p_{8,1,ijk}}_{D^p_{8,1}} \frac{\|\widehat{\gamma}_n - \gamma\|}{h_n^2} \\
&+ \underbrace{\sum_{p=1}^{3} \binom{n}{3}^{-1} \sum_{i<j<k} D^p_{8,2,ijk}}_{D_{8,2}} \frac{\|\widehat{\gamma}_n - \gamma\|}{h_n^2} \\
&+ \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} D^p_{8,3,ijk}}_{D_{8,3}} \frac{\|\widehat{\gamma}_n - \gamma\|^2}{h_n^4},
\end{aligned}
$$

where

$$
\begin{aligned}
D^p_{8,1,ijk} &= \frac{h_n^2}{3}(\|S_{ij,p}\|\|\widehat{S}_{ik,1}\| + \|S_{ij,p}\|\|\widehat{S}_{jk,1}\| + \|S_{ik,p}\|\|\widehat{S}_{jk,1}\|), \\
D^p_{8,2,ijk} &= \frac{h_n^2}{3}(\|\widehat{S}_{ij,1}\|\|S_{ik,p}\| + \|\widehat{S}_{ij,1}\|\|S_{jk,p}\| + \|\widehat{S}_{ik,1}\|\|S_{jk,p}\|), \\
D^p_{8,3,ijk} &= \frac{h_n^4}{3}(\|\widehat{S}_{ij,1}\|\|\widehat{S}_{ik,1}\| + \|\widehat{S}_{ij,1}\|\|\widehat{S}_{jk,1}\| + \|\widehat{S}_{ik,1}\|\|\widehat{S}_{jk,1}\|),
\end{aligned}
$$

for each $p = 1, 2, 3$ and $i < j < k$.

For $D^p_{8,1}$, it suffices to bound $D^1_{8,1}$ as the similar calculation applies to the other terms. By Assumption 3.8, for some constant $C > 0$,

$$
\begin{aligned}
\|S_{ij,1}\| &\leq \frac{2}{h_n} \underbrace{\|\Delta W_{ij}\||\nu_{ij}||K(\Delta R'_{ij}\gamma/h_n)|}_{g_{ij,1}}, \\
\|\widehat{S}_{ij,1}\| &\leq \frac{C}{h_n^2} \underbrace{\|\Delta W_{ij}\|\|\Delta R_{ij}\||\nu_{ij}|}_{g_{ij,2}}.
\end{aligned}
$$

Thus,

$$D^1_{8,1} \leq C\binom{n}{3}^{-1} \sum_{i<j<k} \underbrace{(g_{ij,1}g_{ik,2} + g_{ij,1}g_{jk,2} + g_{ik,1}g_{jk,2})}_{g_{ijk,12}}.$$

Observe that

$$\begin{aligned}
\mathbb{E}[D^1_{8,1}] &\leq \frac{1}{h_n}\mathbb{E}[g_{12,1}g_{13,2}] \\
&= \frac{1}{h_n}\mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\||\nu_{12}||\Delta R'_{12}\gamma = r, \xi_1, U_1]K(r/h_n)f_{R\gamma|\xi_1,U_1}(r)dr \\
&\times \mathbb{E}[\|\Delta W_{13}\|\|\Delta R_{13}\||\nu_{13}||\xi_1, U_1]\Big] \\
&= \mathbb{E}\Big[\int \mathbb{E}[\|\Delta W_{12}\||\nu_{12}||\Delta R'_{12}\gamma = rh_n, \xi_1, U_1]K(r)f_{R\gamma|\xi_1,U_1}(rh_n)dr \\
&\times \mathbb{E}[\|\Delta W_{13}\|\|\Delta R_{13}\||\nu_{13}||\xi_1, U_1]\Big] \\
&= O(1),
\end{aligned}$$

where the last equality follows from Assumptions 3.4, 3.6, 3.7, and 3.8. For variance, the leading term involves covariances between variables with one common node, which has $n \times \binom{n-1}{4}$ elements (up to some constant scale):

$$\begin{aligned}
&\mathrm{Var}[D^1_{8,1}] \\
&= O\left(\frac{1}{h_n^2} \times \binom{n}{3}^{-2} \times n \times \binom{n-1}{4} \times \mathbb{E}[D^1_{8,1,123} \times D^1_{8,1,145}]\right) \\
&= O_p\left(\frac{1}{nh_n^2}\right) = o(1),
\end{aligned}$$

since $nh_n^2 \approx n^{(2k-1)/(2k+3)}$ diverges for $k \geq 2$ and

$$\begin{aligned}
\mathbb{E}[D^1_{8,1,123} \times D^1_{8,1,145}] &\leq \mathbb{E}[g_{123,12} \times g_{145,12}] \\
&\leq \mathbb{E}[g^2_{123,12}] \\
&\leq C\mathbb{E}[\|\Delta W_{12}\|^2 \|\Delta W_{13}\|^2 \|\Delta R_{13}\|^2 \nu_{12}^2 \nu_{13}^2] \\
&= O(1),
\end{aligned}$$

by Cauchy-Schwartz under Assumption 3.7. Thus, $D_{8,1} = O_p(1)$. Similarly, $D_{8,2} = O_p(1)$ and $D_{8,3} = O_p(1)$ hold.

Notice that

$$\frac{\|\widehat{\gamma}_n - \gamma\|}{h_n^2} = \frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|}{\sqrt{Nh_n^5}} = o_p(1),$$

since $1/\sqrt{Nh_n^5}$ diverges by the hypothesis, and similarly,

$$\frac{\|\widehat{\gamma}_n - \gamma\|^2}{h_n^4} = \frac{(\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|)^2}{Nh_n^5} = o_p(1).$$

Hence,

$$\|\widehat{\Sigma}_{W\nu,1} - \widetilde{\Sigma}_{W\nu,1}\| \leq O_p(1) \times o_p(1) + o_p(1) = o_p(1),$$

which completes the proof for Lemma C.8. □

C.2.9 Proof of Lemma C.9

Proof. We only show $nh_n c_W' \widehat{\Sigma} = o_p(1)$ as the other case follows by taking transpose. We also write c_W as c for short. The statement is proved by showing that Lemmas C.7 and C.8 hold even after re-scaled by $n^{1-\alpha/2}h_n$. First, we show that $c'\Sigma_{W\nu,1}c = 0$ implies that $c'\Sigma_{W\nu,1} = 0$.

C.2.9.1 Step 1: The implication of $c'\Sigma_{W\nu,1}c = 0$

Note that

$$\begin{aligned}\Sigma_{W\nu,1} = f_{R\gamma,2}(0,0)Pr(d_{12}d_{13} = 1|\Delta R_{12}'\gamma = \Delta R_{13}'\gamma = 0)\\ \times \mathbb{E}[\Delta W_{12}\Delta W_{13}'\nu_{12}\nu_{13}|\Delta R_{12}'\gamma = \Delta R_{13}'\gamma = 0].\end{aligned}$$

By Assumption 3.4, $f_{R\gamma,2}(0,0) > 0$. Also, by the conditional independence of $(d_{12}, \Delta R_{12}'\gamma)$ and $(d_{13}, \Delta R_{13}'\gamma)$ given ξ_1, U_1 under Assumption 3.1,

$$\begin{aligned}&Pr(d_{12}d_{13} = 1|\Delta R_{12}'\gamma = \Delta R_{13}'\gamma = 0)\\ &= \mathbb{E}\left[\left(Pr(d_{12} = 1|\Delta R_{12}'\gamma = 0, \xi_1, U_1)\right)^2 |\Delta R_{12}'\gamma = 0\right].\end{aligned}$$

Since $Pr(d_{12} = 1|\Delta R_{12}'\gamma = 0) > 0$ is implied by Assumption 3.3, it must be that

$$Pr(d_{12}d_{13} = 1|\Delta R_{12}'\gamma = \Delta R_{13}'\gamma = 0) > 0,$$

as otherwise $Pr(d_{12} = 1|\Delta R_{12}'\gamma = 0, \xi_1, U_1)$ is constant at 0, which contradicts with the locally positive probability of $d_{12} = 1$. Thus, $c'\Sigma_{W\nu,1}c = 0$ is equivalent to

$$\begin{aligned}
&\mathbb{E}[\Delta c' W_{12} \Delta W_{13}' c \nu_{12} \nu_{13} | \Delta R_{12}' \gamma = \Delta R_{13}' \gamma = 0] \\
&= \mathbb{E}\left[\left(\mathbb{E}[c' \Delta W_{12} \nu_{12} | \Delta R_{12}' \gamma = 0, \xi_1, U_1] \right)^2 | \Delta R_{12}' \gamma = 0 \right] \\
&= 0,
\end{aligned}$$

which, in turn, is equivalent to (by the mean independence of ν_{12}),

$$\mathbb{E}[c' \Delta W_{12} \nu_{12} | \Delta R_{12}' \gamma = 0, \xi_1, U_1] = 0,$$

almost surely. Thus, $c' \Sigma_{W\nu,1} c = 0$ implies that

$$\begin{aligned}
&c' \Sigma_{W\nu,1} \\
&= f_{R\gamma,2}(0,0) Pr(d_{12} d_{13} = 1 | \Delta R_{12}' \gamma = \Delta R_{13}' \gamma = 0) \\
&\times \mathbb{E}\Big[\mathbb{E}[c' \Delta W_{12} \nu_{12} | \Delta R_{12}' \gamma = 0, \xi_1, U_1] \\
&\times \mathbb{E}[\Delta W_{13}' \nu_{13} | \Delta R_{13}' \gamma, \xi_1, U_1] \big| \Delta R_{12}' \gamma = \Delta R_{13}' \gamma = 0 \Big] \\
&= 0.
\end{aligned}$$

C.2.9.2 Step 2: $nh_n c' \widetilde{\Sigma}_{W\nu,1} = o_p(1)$

Remember that

$$\widetilde{\Sigma}_{W\nu,1} = D_7 + \mathcal{O}_7.$$

For D_7, from the calculation in the proof of Lemma C.3, we have that

$$\mathbb{E}[c'D_7] = \mathbb{E}[c'S_{12,1}S'_{13,1}] = O(h_n^k).$$

, which shows $nh_n\mathbb{E}[c'D_7] = O(nh_n^{k+1}) = o(1)$ under the hypothesis. For any non-zero vector $a \in \mathbb{R}^{q_w}$, redefining U_n and $U_{n,1}$ with the kernel $p_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k) = \mathbb{E}[c'D_{7,ijk}a|\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k]$, we can repeat the calculation in the proof of Lemma C.7 to get

$$\mathbb{E}[(nh_nc'D_7a - nh_nU_{n,1})^2] = n^2h_n^2 \times O\left(\frac{1}{n^3h_n^2}\right) = o(1).$$

Also, $\mathbb{E}[nh_np_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k)] = \mathbb{E}[nh_nc'D_7] = o(1)$ and

$$\mathbb{E}[(nh_np_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_l))^2] = O(n),$$

so that by Lemma A.3 of Ahn and Powell (1993),

$$nh_nU_n = nh_n\mathbb{E}[U_{n,1}] + o_p(1) = o_p(1).$$

This shows that, since a is arbitrary,

$$nh_nc'D_7 = o_p(1).$$

For the remainder term $\mathcal{O}_7$, this should again be of smaller order than $nh_nc'D_7$ since $\beta - \widehat{\beta}_n = O_p(1/\sqrt{n})$ and $\lambda_{ij} = \Delta R'_{ij}\gamma\Lambda_{ij}$ is locally $O(h_n^{k+1})$ under the smoothing kernel and smoothness conditions on the density. For example, one of the elements in $\mathcal{O}_7$ is given by

$$
\begin{aligned}
& nh_n \binom{n}{3} \sum_{i<j<k} \frac{1}{3}(S_{ij,2}S'_{ik,2} + S_{ij,2}S'_{jk,2} + S_{ik,2}S_{jk,2}) \\
& \leq \binom{n}{3} \sum_{i<j<k} \frac{4}{3h_n^2} \Bigg(|\Delta W_{ij}\|^2 \|\Delta W_{ik}\|^2 |K(\Delta R'_{ij}\gamma/h_n)||K(\Delta R'_{ik}\gamma/h_n)| \\
& + \|\Delta W_{ij}\|^2 \|\Delta W_{jk}\|^2 |K(\Delta R'_{ij}\gamma/h_n)||K(\Delta R_{jk}/h_n)| \\
& + \|\Delta W'_{ik}\|^2 \|\Delta W_{jk}\|^2 |K(\Delta R_{ik}/h_n)||K(\Delta R_{jk}/h_n)| \Bigg) nh_n \|\beta - \widehat{\beta}_n\|^2 \\
& = O_p(h_n) = o_p(1),
\end{aligned}
$$

where the last line can be shown by the same calculation as before to show $O_p(1)$ for the summation part and $\|\beta - \widehat{\beta}_n\|^2 = O_p(1/n)$ from Theorem 3.1. Similarly, we can show the negligibility of the elements in $\mathcal{O}_7$. This finishes the step 2.

C.2.9.3 Step 3: $n^{1-\alpha/2}h_n\|\widehat{\Sigma}_{W\nu,1} - \widetilde{\Sigma}_{W\nu,1}\| = o_p(1)$

By the proof of Lemma C.8, we have

$$
n^{1-\alpha/2}h_n\|\widehat{\Sigma}_{W\nu,1} - \widetilde{\Sigma}_{W\nu,1}\| \leq O_p(1)\left(\frac{n^{1-\alpha/2}\|\widehat{\gamma}_n - \gamma\|}{h_n} + \frac{n^{1-\alpha/2}\|\widehat{\gamma}_n - \gamma\|^2}{h_n^3}\right).
$$

Observe that

$$
\begin{aligned}
\frac{n^{1-\alpha/2}\|\widehat{\gamma}_n - \gamma\|}{h_n} &= O\left(\frac{\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|}{\sqrt{n^\alpha h_n^3}}\right) = o_p(1), \\
\frac{n^{1-\alpha/2}\|\widehat{\gamma}_n - \gamma\|^2}{h_n^3} &= O\left(\frac{\|\sqrt{Nh_n}(\widehat{\gamma}_n - \gamma)\|^2}{\sqrt{n^{2+\alpha} h_n^7}}\right) = o_p(1),
\end{aligned}
$$

by Assumption 3.10 and

$$n^{\alpha}h_n^3 \approx n^{(\alpha(2k+3)-6)/(2k+3)} \to \infty,$$
$$n^{2+\alpha}h_n^7 \approx n^{((2k+3)\alpha+4k-8)/(2k+3)} \to \infty,$$

for $\alpha \in [6/(2k+3), 1)$. Hence,

$$n^{1-\alpha/2}h_n\|\widehat{\Sigma}_{W\nu,1} - \widetilde{\Sigma}_{W\nu,1}\| = o_p(1).$$

Steps 1-3 complete the proof of Lemma C.9. □

C.2.10 Proof of Lemma C.10

Proof. Write c_W as c for short. It suffices to show that $nh_n\|c'\widehat{\Sigma}_{W\nu,1}c - c'\widetilde{\Sigma}_{W\nu,1}c\| = o_p(1)$ as we already show, in the proof of Lemma C.9, that $nh_n c'\widetilde{\Sigma}_{W\nu,1} = o_p(1)$, which implies $nh_n c'\widetilde{\Sigma}_{W\nu,1}c = o_p(1)$. To save the space, in the following argument, we treat $\Delta\widehat{\epsilon}_{ij}$ as ν_{ij}; The other terms are similarly bounded using the properties of $\beta - \widehat{\beta}_n$ and λ_{ij}.

Re-define

$$S_{ij} = \frac{2}{h_n} d_{ij} K\left(\frac{\Delta R_{ij}'\gamma}{h_n}\right) c'\Delta W_{ij}\nu_{ij}$$
$$\widehat{S}_{ij} = \frac{2}{h_n} d_{ij} K\left(\frac{\Delta R_{ij}'\widehat{\gamma}_n}{h_n}\right) c'\Delta W_{ij}\nu_{ij},$$
$$\widehat{S}_{ij,1} = \frac{2}{h_n^2} d_{ij} k\left(\frac{\Delta R_{ij}'\gamma}{h_n}\right) c'\Delta W_{ij}\Delta R_{ij}'\nu_{ij}$$
$$\widehat{S}_{ij,2} = \frac{1}{h_n^3} d_{ij} k\left(\frac{c_{ij,n}^*}{h_n}\right) \Delta R_{ij} c'\Delta W_{ij}\Delta R_{ij}'\nu_{ij},$$

where $c^*_{ij,n}$ is in between $\Delta R'_{ij}\gamma$ and $\Delta R'_{ij}\widehat{\gamma}_n$. We have that

$$\widehat{S}_{ij} = S_{ij} + \widehat{S}_{ij,1}(\widehat{\gamma}_n - \gamma) + (\widehat{\gamma}_n - \gamma)'\widehat{S}_{ij,2}(\widehat{\gamma}_n - \gamma).$$

Note that, for some constant $C > 0$, $\|\widehat{S}_{ij,1}\| \leq Ch_n^{-2} g_{ij,2}$ as before and

$$\|\widehat{S}_{ij,2}\| \leq \frac{C}{h_n^3} \underbrace{\|\Delta W_{ij}\| \|\Delta R_{ij}\|^2 |\nu_{ij}|}_{g_{ij,3}}.$$

Observe that

$$
\begin{aligned}
&c'\widehat{\Sigma}_{W\nu,1}c - c'\widetilde{\Sigma}_{W\nu,1}c \\
&\le \frac{(\widehat{\gamma}_n - \gamma)'}{\sqrt{h_n}} \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \underbrace{\frac{\sqrt{h_n}}{3}(S_{ij}\widehat{S}'_{ik,1} + S_{ij}\widehat{S}'_{jk,1} + S_{ik}\widehat{S}'_{jk,1})}_{D_{10,1,ijk}}}_{D_{10,1}} \\
&+ \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \frac{\sqrt{h_n}}{3}(\widehat{S}_{ij,1}S_{ik} + \widehat{S}_{ij,1}S_{jk} + \widehat{S}_{ik,1}S_{jk})}_{D_{10,2}} \frac{(\widehat{\gamma}_n - \gamma)}{\sqrt{h_n}} \\
&+ \frac{(\widehat{\gamma}_n - \gamma)'}{h_n^{3/2}} \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \frac{h_n^3}{3}(S_{ij}\widehat{S}'_{ik,2} + S_{ij}\widehat{S}'_{jk,2} + S_{ik}\widehat{S}'_{jk,2})}_{D_{10,3}} \frac{(\widehat{\gamma}_n - \gamma)}{h_n^{3/2}} \\
&+ \frac{(\widehat{\gamma}_n - \gamma)'}{h_n^{3/2}} \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \frac{h_n^3}{3}(\widehat{S}_{ij,2}S_{ik} + \widehat{S}_{ij,2}S_{jk} + \widehat{S}_{ik,2}S_{jk})}_{D_{10,4}} \frac{(\widehat{\gamma}_n - \gamma)}{h_n^{3/2}} \\
&+ C^2 \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \frac{1}{3}(g_{ij,2}g_{ik,3} + g_{ij,2}g_{jk,3} + g_{ik,2}g_{jk,3})}_{D_{10,5}} \frac{\|\widehat{\gamma}_n - \gamma\|^3}{h_n^5} \\
&+ C^2 \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \frac{1}{3}(g_{ij,3}g_{ik,2} + g_{ij,3}g_{jk,2} + g_{jk,3}g_{ik,2})}_{D_{10,6}} \frac{\|\widehat{\gamma}_n - \gamma\|^3}{h_n^5} \\
&+ \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} D_{10,7,ijk}}_{D_{10,7}} \frac{\|\widehat{\gamma}_n - \gamma\|^2}{h_n^2} \\
&+ C^2 \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} \frac{1}{3}(g_{ij,3}g_{ik,3} + g_{ij,3}g_{jk,3} + g_{jk,3}g_{ik,3})}_{D_{10,8}} \frac{\|\widehat{\gamma}_n - \gamma\|^4}{h_n^6},
\end{aligned}
$$

where

$$D_{10,7,ijk} = \frac{h_n^2}{3}(\|\widehat{S}_{ij,1}\|\|\widehat{S}_{ik,1}\| + \|\widehat{S}_{ij,1}\|\|\widehat{S}_{jk,1}\| + \|\widehat{S}_{jk,1}\|\|\widehat{S}_{ik,1}\|)$$

for $i < j < k$.

First we stochastically bound $D_{10,1}$ and $D_{10,2}$. For any vector $a \in \mathbb{R}^{q_r}$ and some constant $C > 0$,

$$\begin{aligned}
&\mathbb{E}[a'D_{10,1}] \\
&= \frac{\sqrt{h_n}}{h_n^3}\mathbb{E}\left[\mathbb{E}[c'S_{12}|\xi_1, U_1]\mathbb{E}[a'\widehat{S}_{13}c|\xi_1, U_1]\right] \\
&\leq \frac{1}{h_n^{3/2}}\left\{\mathbb{E}\left[\int \mathbb{E}[d_{12}c'\Delta W\nu_{12}|\Delta R_{12} = s_1 h_n, \xi_1, U_1]^2\right.\right. \\
&\left.\left.\times K(s_1) f_{R\gamma|\xi_1,U_1}(s_1 h_n) ds_1\right]\right\}^{1/2} \\
&\times C\mathbb{E}[\|\Delta W_{12}\|^2\|\Delta R_{13}\|^2\nu_{13}^2]^{1/2} \\
&= O(h_n^{(2k-1)/2}) = o(1),
\end{aligned}$$

where the first line hold from the conditional independence, the inequality follows from Cauchy-Schwartz and Assumption 3.8, and the final line holds from the implication from $c'\Sigma_{W\nu,1}c = 0$ and Assumptions 3.4, 3.6, 3.7, and 3.8. Repeating the calculation in the proof for Lemma C.7 and adjusting for covariances with one index in common,

$$\begin{aligned}
\mathrm{Var}[a'D_{10,1}] &= O\left(\frac{1}{n^2 h_n^3}\right) + O\left(\frac{h_n \mathbb{E}[S_{12}a'\widehat{S}'_{13,1}S_{14}a'\widehat{S}'_{15,1}]}{n}\right) \\
&= O\left(\frac{1}{nh_n}\right) = o(1),
\end{aligned}$$

where the last equality holds from the stated assumption on h_n for $k \geq 1$ and

$$
\begin{aligned}
&\mathbb{E}[S_{12}a'\widehat{S}'_{13,1}S_{14}a'\widehat{S}'_{15,1}] \\
&= \frac{16}{h_n^2}\mathbb{E}\Big[\mathbb{E}[d_{12}c'\Delta W_{12}\nu_{12}|\Delta R'_{12}\gamma = h_n s_1, \boldsymbol{\xi}_1] \\
&\times \mathbb{E}[d_{13}c'\Delta W_{13}a'\Delta R_{13}\nu_{13}|\Delta R'_{13}\gamma = h_n s_2, \boldsymbol{\xi}_1] \\
&\times \mathbb{E}[d_{14}c'\Delta W_{14}\nu_{14}|\Delta R'_{14}\gamma = h_n s_3, \boldsymbol{\xi}_1] \\
&\times \mathbb{E}[d_{15}c'\Delta W_{15}a'\Delta R_{15}\nu_{15}|\Delta R'_{15}\gamma = h_n s_4, \boldsymbol{\xi}_1] \\
&\times K(s_1)k(s_2)K(s_3)k(s_4)\Pi_{i=1}^4 f_{R\gamma|\xi_1,U_1}(s_i h_n)\Big] \\
&= O\left(\frac{1}{h_n^2}\right)
\end{aligned}
$$

by Assumptions 3.4, 3.6, and 3.8. Thus, $D_{1,1} = O_p(h_n^{(2k-1)/2} + (nh_n)^{-1/2})$. Similarly, we have $D_{1,2} = O_p(h_n^{(2k-1)/2} + (nh_n)^{-1/2})$.

Next, we stochastically bound $D_{10,3}$ and $D_{10,4}$. For any finite $a, b \in \mathbb{R}^{q_r}$, for some $C > 0$,

$$
\begin{aligned}
&\mathbb{E}[a'D_{10,3}b] \\
&\leq \frac{C}{h_n}\mathbb{E}[|S_{12}|g_{13,3}] \\
&= C\mathbb{E}\Big[\int \mathbb{E}[d_{12}|c'\Delta W_{12}||\nu_{12}||\Delta R'_{12}\gamma = h_n s_1, \boldsymbol{\xi}_1] \\
&\times \mathbb{E}[g_{13}|\boldsymbol{\xi}_1] \\
&\times K(s_1)f_{R\gamma|\xi_1,U_1}(s_1 h_1)f_{R\gamma|\xi_1,U_1}(s_2)\Big] \\
&= O(1),
\end{aligned}
$$

where the last equality holds from Assumptions 3.4, 3.6, 3.7 and 3.8. The variance is calculated similarly as before:

$$\mathrm{Var}[a' D_{10,3} b] = O\left(\frac{1}{n^2 h_n}\right) + O\left(\frac{h_n}{n}\right) = o(1).$$

Thus, $D_{10,3} = O_p(1)$. Similarly, $D_{10,4} = O_p(1)$.

$D_{10,5}, D_{10,6}$, and $D_{10,8}$ are all $O_p(1)$ by the similar computation as in Lemma C.7.

$D_{10.7}$ is stochastically bounded as follows. Let $s_{10,7,c}(\cdot,\cdot)$ be defined as

$$\begin{aligned}
&s_{10,7,c}(s_1, s_2) \\
&= \mathbb{E}[|c'\Delta W_{12}||c'\Delta W_{13}|\|\Delta R_{12}\|\Delta R_{13}\||\nu_{12}||\nu_{13}||\Delta R_{12}'\gamma = s_1, \Delta R_{13}'\gamma = s_2]
\end{aligned}$$

Observe that

$$\begin{aligned}
&\mathbb{E}[D_{10,7}] \\
&= h_n^2 \mathbb{E}[\|\widehat{S}_{12,1}\|\|\widehat{S}_{13,1}\|] \\
&\leq 4 \int s_{10,7,c}(h_n s_1, h_n s_2) k(s_1) k(s_2) f_{R\gamma,2}(h_n s_1, h_n s_2) ds_1 ds_2 \\
&= O(1),
\end{aligned}$$

where the last line holds from Assumption 3.4, 3.6, and 3.8. The variance is calculated similarly as before:

$$\mathrm{Var}[D_{10,7}] = O\left(\frac{1}{n^2 h_n^2}\right) + O\left(\frac{1}{n}\right) = o(1).$$

Thus, $D_{10,7} = O_p(1)$.

Finally, the above implies

$$
\begin{aligned}
&nh_n|c'\widehat{\Sigma}_{W\nu,1}c - c'\widetilde{\Sigma}_{W\nu,1}c| \\
&\leq O_p(h_n^{(2k-1)/2} + (nh_n)^{-1/2}) \times \frac{nh_n\|\widehat{\gamma}_n - \gamma\|}{\sqrt{h_n}} \\
&+ O_p(1) \times \left(\frac{nh_n\|\widehat{\gamma}_n - \gamma\|^2}{h_n^3} + \frac{nh_n\|\widehat{\gamma}_n - \gamma\|^3}{h_n^5} + \frac{nh_n\|\widehat{\gamma}_n - \gamma\|^4}{h_n^6}\right) \\
&= o_p(1).
\end{aligned}
$$

To see this, first note that

$$
\begin{aligned}
&O_p(h_n^{(2k-1)/2} + (nh_n)^{-1/2}) \times \frac{nh_n\|\widehat{\gamma}_n - \gamma\|}{\sqrt{h_n}} \\
&= O\left((h_n^{(2k-2)/2} + (nh_n^2)^{-1/2})\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|\right) = o_p(1),
\end{aligned}
$$

by Assumption 3.10 and $nh_n^2 = O(n^{(2k-1)/(2k+3)})$ diverging for $k \geq 2$. Also,

$$
\begin{aligned}
\frac{nh_n\|\widehat{\gamma}_n - \gamma\|^2}{h_n^3} &= O\left(\frac{(\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|)^2}{nh_n^3}\right) = o_p(1), \\
\frac{nh_n\|\widehat{\gamma}_n - \gamma\|^3}{h_n^5} &= O\left(\frac{(\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|)^3}{n^2h_n^{11/2}}\right) = o_p(1), \\
\frac{nh_n\|\widehat{\gamma}_n - \gamma\|^4}{h_n^6} &= O\left(\frac{(\sqrt{Nh_n}\|\widehat{\gamma}_n - \gamma\|)^4}{n^3h_n^7}\right) = o_p(1),
\end{aligned}
$$

since

$$
\begin{aligned}
nh_n^3 &= O(n^{(2k-3)/(2k+3)}), \\
n^2h_n^{11/2} &= O(n^{(4k-5)/(2k+3)}), \\
n^3h_n^7 &= O(n^{(6k-5)/(2K+3)}),
\end{aligned}
$$

all diverging for $k \geq 2$. This completes the proof for Lemma C.10. $\square$

Bibliography

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020): "Sampling-based versus Design-based Uncertainty in Regression Analysis," *Econometrica*, 88, 265–296.

——— (2023): "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics*, 138, 1–35.

Abel, M., R. Burger, and P. Piraino (2020): "The Value of Reference Letters: Experimental Evidence from South Africa," *American Economic Journal: Applied Economics*, 12, 40–71.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Afridi, F., A. Dhillon, S. X. Li, and S. Sharma (2020): "Using social connections and financial incentives to solve coordination failure: A quasi-field experiment in India's manufacturing sector," *Journal of Development Economics*, 144.

Ahn, H. and J. L. Powell (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics*, 58, 3–29.

Alan, S., G. Corekcioglu, and M. Sutter (2022): "Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention," *Quarterly Journal of Economics*, 138, 151–203.

Alatas, V., A. Banerjee, A. G. Chandrasekhar, R. Hanna, and B. A. Olken (2016): "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia," *American Economic Review*, 106, 1663–1704.

Alviarez, V., K. Head, and T. Mayer (2025): "Global giants and local stars: How changes in brand ownership affect competition," *Am. Econ. J. Microecon.*, 17, 389–432.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008): "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 673–697.

Angrist, J. D. (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66, 249–288.

Aronow, P. M. and C. Samii (2017): "Estimating average causal effects under general interference, with application to a social network experiment," *Annals of Applied Statistics*, 11, 1912–1947.

Auerbach, E. (2022): "Identification and Estimation of a Partially Linear Regression Model Using Network Data," *Econometrica*, 90, 347–365.

Bacher-Hicks, A. and C. Koedel (2023): "Chapter 2 - Estimation and interpretation of teacher value added in research applications," Elsevier, vol. 6 of *Handbook of the Economics of Education*, 93–134.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104, 1551–96.

Banerjee, A., E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan (2023): "Changes in Social Network Structure in Response to Exposure to Formal Credit Markets," *Review of Economic Studies*.

Banerjee, A., E. Breza, A. G. Chandrasekhar, and B. Golub (2024): "When Less Is More: Experimental Evidence on Information Delivery During India's Demonetisation," *Review of Economic Studies*, 91, 1884–1922.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The diffusion of microfinance," *Science*, 341, 1236498.

——— (2019): "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials," *Review of Economic Studies*, 86, 2453–2490.

BARNHARDT, S., E. FIELD, AND R. PANDE (2017): "Moving to Opportunity or Isolation? Network Effects of a Randomized Housing Lottery in Urban India," *American Economic Journal: Applied Economics*, 9, 1–32.

BARSBAI, T., V. LICUANAN, A. STEINMAYR, E. TIONGSON, AND D. YANG (2024): "Information and immigrant settlement," *Journal of Development Economics*, 170.

BATISTA, C., D. M. COSTA, P. FREITAS, G. LIMA, AND A. B. REIS (2025): "What matters for the decision to study abroad? A lab-in-the-field experiment in Cape Verde," *Journal of Development Economics*, 173.

BEAMAN, L., A. BENYISHAY, J. MAGRUDER, AND A. M. MOBARAK (2021a): "Can network theory-based targeting increase technology adoption?" *American Economic Review*, 111, 1918–1943.

——— (2021b): "Can Network Theory-Based Targeting Increase Technology Adoption?" *American Economic Review*, 111, 1918–1943.

BELLONI, A., V. CHERNOZHUKOV, I. FERNANDEZ-VAL, AND C. HANSEN (2017): "Program Evaluation and Causal Inference with High-Dimensional Data," *Econometrica*, 85, 233–298.

BENYISHAY, A., M. JONES, F. KONDYLIS, AND A. M. MOBARAK (2020): "Gender gaps in technology diffusion," *Journal of Development Economics*, 143.

BENYISHAY, A. AND A. M. MOBARAK (2019): "Social Learning and Incentives for Experimentation and Communication," *Review of Economic Studies*, 86, 976–1009.

BERNARD, A. B., E. DHYNE, G. MAGERMAN, K. MANOVA, AND A. MOXNES (2022): "The Origins of Firm Heterogeneity: A Production Network Approach," *Journal of Political Economy*, 130, 1765–1804.

BILLINGSLEY, P. (1999): *Convergence of Probability Measures*, Hoboken, NJ, USA: John Wiley & Sons, Inc.

BJORKEGREN, D. AND B. C. KARACA (2022): "Network adoption subsidies: A digital evaluation of a rural mobile phone program in Rwanda," *Journal of Development Economics*, 154.

Bobonis, G. J., P. J. Gertler, M. Gonzalez-Navarro, and S. Nichter (2022): "Vulnerability and Clientelism," *American Economic Review*, 112, 3627–3659.

Boltz, M., K. Marazyan, and P. Villar (2019): "Income Hiding and Informal Redistribution: A Lab-in-the-Field Experiment in Senegal," *Journal of Development Economics*, 137, 78–92.

Bonhomme, S. (2020): "Econometric analysis of bipartite networks," *The Econometric Analysis of Network Data*.

——— (2021): "Teams: Heterogeneity, Sorting, and Complementarity," ArXiv:2102.01802v1 [econ.EM].

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023): "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics*, 41, 291–322.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87, 699–739.

Borusyak, K. and P. Hull (2023a): "Nonrandom exposure to exogenous shocks," *Econometrica*, 91, 2155–2185.

——— (2023b): "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 91, 2155–2185.

——— (2024): "Negative weights are no concern in design-based specifications," *AEA Papers and Proceedings*, 114, 597–600.

Bozzo, E. (2013): "The Moore–Penrose inverse of the normalized graph Laplacian," *Linear Algebra and its Applications*, 439, 3038–3043.

Breza, E. and A. G. Chandrasekhar (2019): "Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment," *Econometrica*, 87, 175–216.

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2020): "Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data," *American Economic Review*, 110, 2454–2484.

Cai, J., A. de Janvry, and E. Sadoulet (2015a): "Social networks and the decision to insure," *American Economic Journal: Applied Economics*, 7, 81–108.

Cai, J., A. De Janvry, and E. Sadoulet (2015b): "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 7, 81–108.

Cai, J. and A. Szeidl (2018): "Interfirm Relationships and Business Performance," *Quarterly Journal of Economics*, 133, 1229–1282.

Cai, S. (2020): "Migration under liquidity constraints: Evidence from randomized credit access in China," *Journal of Development Economics*, 142, 10th AFD-World-Bank-Development Conference, CERDI, Clermont Ferrand, FRANCE, JUN 30-JUL 01, 2017.

Callen, M. and J. D. Long (2015): "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan," *American Economic Review*, 105, 354–381.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference With Multiway Clustering," *Journal of Business & Economic Statistics*, 29, 238–249.

Cameron, A. C. and D. Miller (2014): "Robust Inference for Dyadic Data," *Unpublished manuscript.*

Candelaria, L. E. (2020): "A Semiparametric Network Formation Model with Unobserved Linear Heterogeneity," .

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

Carter, M., R. Laajaj, and D. Yang (2021a): "Subsidies and the African Green Revolution: Direct Effects and Social Network Spillovers of Randomized Input Subsidies in Mozambique," *American Economic Journal: Applied Economics*, 13, 206–29.

——— (2021b): "Subsidies and the African Green Revolution: Direct Effects and Social Network Spillovers of Randomized Input Subsidies in Mozambique," *American Economic Journal: Applied Economics*, 13, 206–229.

Cattaneo, M. D., Y. Feng, and W. G. Underwood (2024): "Uniform Inference for Kernel Density Estimators with Dyadic Data," *Journal of the American Statistical Association*, 119, 2695–2708.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 47, 225–238.

Chandrasekhar, A. and R. Lewis (2016): "Econometrics of sampled networks," *Unpublished manuscript*.

Chandrasekhar, A. G., C. Kinnan, and H. Larreguy (2018): "Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field," *American Economic Journal: Applied Economics*, 10, 43–78.

Chegere, M. J., P. Falco, and A. Menzel (2024): "Social ties at work and effort choice: Experimental evidence from Tanzania," *Journal of Development Economics*, 171.

Chiang, H. D., B. E. Hansen, and Y. Sasaki (2024): "Standard Errors for Two-Way Clustering with Serially Correlated Time Effects," *The Review of Economics and Statistics*, 1–40.

Colonnelli, E., B. Li, and E. Liu (2024): "Investing with the Government: A Field Experiment in China," *Journal of Political Economy*, 132, 248–294.

Conley, T. G. and C. R. Taber (2011): "Inference with "Difference in Differences" with a Small Number of Policy Changes," *The Review of Economics and Statistics*, 93, 113–125.

Conley, T. G. and C. R. Udry (2010): "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 100, 35–69.

Coutts, A. (2022): "Identifying communication spillovers in lab-in-the-field experiments," *Journal of Development Economics*, 157.

Cruz, C., J. Labonne, and P. Querubin (2017): "Politician Family Networks and Electoral Outcomes: Evidence from the Philippines," *American Economic Review*, 107, 3006–3037.

Cvetković, D., P. Rowlinson, and S. Simić (2007): "Eigenvalue Bounds for the Signless Laplacian," *Publications de l'Institut Mathématique*, 81(95), 11–27.

Davezies, L., X. D'Haultfœuille, and Y. Guyonvarch (2021): "Empirical process results for exchangeable arrays," *The Annals of Statistics*, 49, 845 – 862.

Deng, S., S. Ling, and T. Strohmer (2021): "Strong Consistency, Graph Laplacians, and the Stochastic Block Model," *Journal of Machine Learning Research*, 22, 1–44.

Deutschmann, J. W., M. Lipscomb, L. Schechter, and J. Zhu (2024): "Spillovers without Social Interactions in Urban Sanitation," *American Economic Journal: Applied Economics*, 16, 482–515.

Di Addario, S., P. Kline, R. Saggio, and M. Sølvsten (2023): "It ain't where you're from, it's where you're at: Hiring origins, firm heterogeneity, and wages," *Journal of Econometrics*, 233, 340–374.

Di Falco, S., F. Feri, P. Pin, and X. Vollenweider (2018): "Ties that bind: Network redistributive pressure and economic decisions in village economies," *Journal of Development Economics*, 131, 123–131.

Dizon, F., E. Gong, and K. Jones (2020): "The Effect of Promoting Savings on Informal Risk Sharing," *Journal of Human Resources*, 55, 963–998.

Drago, F., F. Mengel, and C. Traxler (2020): "Compliance Behavior in Networks: Evidence from a Field Experiment," *American Economic Journal: Applied Economics*, 12, 96–133.

Dubé, J. P., A. Hortaçsu, and J. Joo (2021): "Random-coefficients logit demand estimation with zero-valued market shares," *Marketing Science*, 40.

Engbom, N. and C. Moser (2022): "Earnings Inequality and the Minimum Wage: Evidence from Brazil," *American Economic Review*, 112, 3803–47.

Engbom, N., C. Moser, and J. Sauermann (2023): "Firm pay dynamics," *Journal of Econometrics*, 233, 396–423.

Evsyukova, Y., F. Rusche, and W. Mill (2024): “Linkedout? A Field Experiment on Discrimination in Job Network Formation,” *Quarterly Journal of Economics*, 140, 283–334.

Fafchamps, M. and F. Gubert (2007): “The formation of risk sharing networks,” *Journal of Development Economics*, 83.

Fafchamps, M. and P. C. Vicente (2013): “Political violence and social networks: Experimental evidence from a Nigerian election,” *Journal of Development Economics*, 101, 27–48.

Fernando, A. N. (2021): “Seeking the treated: The impact of mobile extension on farmer information exchange in India,” *Journal of Development Economics*, 153, 102713.

Ferreras Garrucho, D. and T. Zylkin (2023): *penppml: Penalized Poisson Pseudo Maximum Likelihood Regression*, r package version 0.2.3.

Finkelstein, A., M. Gentzkow, and H. Williams (2016): “Sources of Geographic Variation in Health Care: Evidence From Patient Migration,” *The Quarterly Journal of Economics*, 131, 1681–1726.

——— (2021): “Place-Based Drivers of Mortality: Evidence from Migration,” *American Economic Review*, 111, 2697–2735.

Flory, J. A. (2018): “Formal finance and informal safety nets of the poor: Evidence from a savings field experiment,” *Journal of Development Economics*, 135, 517–533.

Gao, M. and P. Ding (2023): “Causal inference in network experiments: regression-based analysis and design-based properties,” *arXiv preprint arXiv:2309.07476*.

Gine, X. and G. Mansuri (2018): “Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan,” *American Economic Journal: Applied Economics*, 10, 207–235.

Godlonton, S. and R. Thornton (2012): “Peer effects in learning HIV results,” *Journal of Development Economics*, 97, 118–129.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2022): “Contamination bias in linear regressions,” Tech. rep., National Bureau of Economic Research.

Graham, B. S. (2017): "An Econometric Model of Network Formation With Degree Heterogeneity," *Econometrica*, 85, 1033–1063.

——— (2020): "Dyadic regression," *The Econometric Analysis of Network Data*.

Graham, B. S., F. Niu, and J. L. Powell (2019): "Kernel Density Estimation for Undirected Dyadic Data," .

——— (2021): "Minimax Risk and Uniform Convergence Rates for Nonparametric Dyadic Regression," .

Griffith, A. (2022): "Name your friends, but only five? the importance of censoring in peer effects estimates using social network data," *Journal of Labor Economics*, 40, 779–805.

Hall, P. (1984): "Central limit theorem for integrated square error of multivariate nonparametric density estimators," *Journal of Multivariate Analysis*, 14, 1–16.

Hardy, M. and J. McCasland (2021): "It takes two: Experimental evidence on the determinants of technology diffusion," *Journal of Development Economics*, 149.

Head, K. and T. Mayer (2014): "Gravity Equations: Workhorse,Toolkit, and Cookbook," *Handbook of International Economics*, 4, 131–195.

Heckman, J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Helpman, E., M. Melitz, and Y. Rubinstein (2008): "Estimating trade flows: Trading partners and trading volumes," *Quarterly Journal of Economics*, 123, 441–487.

Hernandez-Agramonte, J. M., O. Namen, E. Naslund-Hadley, and M. L. Biehl (2024): "Supporting early childhood development remotely: Experimental evidence from SMS messages," *Journal of Development Economics*, 166.

Hess, S., D. Jaimovich, and M. Schuendeln (2021): "Development Projects and Economic Networks: Lessons from Rural Gambia," *Review of Economic Studies*, 88, 1347–1384.

Hoeffding, W. (1961): "The strong law of large numbers for U-statistics," .

Horn, R. A. and C. R. Johnson (2012): *Matrix Analysis*, Cambridge University Press, 2 ed.

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60.

Hoshino, T. and T. Yanagi (2024): "Causal Inference with Noncompliance and Unknown Interference," *Journal of the American Statistical Association*, 119, 2869–2880.

Hsieh, C.-S., Y.-C. Hsu, S. I. Ko, J. Kovarik, and T. D. Logan (2024): "Non-representative sampled networks: Estimation of network structural properties by weighting," *Journal of Econometrics*, 240, 105689.

Jackson, C. K. (2013): "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence From Teachers," *The Review of Economics and Statistics*, 95, 1096–1116.

Jakiela, P. and O. Ozier (2016): "Does Africa Need a Rotten Kin Theorem? Experimental Evidence from Village Economies," *Review of Economic Studies*, 83, 231–268.

Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina (2014): "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking?" *Econometrica*, 82, 463–505.

Jochmans, K. (2023): "Peer effects and endogenous social interactions," *Journal of Econometrics*, 235, 1203–1214.

Jochmans, K. and M. Weidner (2019): "Fixed-Effect Regressions on Network Data," *Econometrica*, 87, 1543–1560.

Johnsson, I. and H. R. Moon (2021): "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach," *The Review of Economics and Statistics*, 103, 328–345.

Kallenberg, O. (2002): *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, second ed.

KANDPAL, E. AND K. BAYLIS (2019): "The social lives of married women: Peer effects in female autonomy and investments in children," *Journal of Development Economics*, 140, 26–43.

KARING, A. (2024): "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone," *Quarterly Journal of Economics*, 139, 2083–2133.

KESSLER, J. B. (2017): "Announcements of Support and Public Good Provision," *American Economic Review*, 107, 3760–3787.

KLINE, P. (2024): "Chapter 2 - Firm wage effects," Elsevier, vol. 5 of *Handbook of Labor Economics*, 115–181.

KLINE, P., R. SAGGIO, AND M. SØLVSTEN (2020): "Leave-Out Estimation of Variance Components," *Econometrica*, 88, 1859–1898.

KOJEVNIKOV, D., V. MARMER, AND K. SONG (2021a): "Limit theorems for network dependent random variables," *Journal of Econometrics*, 222, 882–908.

——— (2021b): "Limit theorems for network dependent random variables," *Journal of Econometrics*, 222, 882–908.

KOLACZYK, E. D. AND G. CSÁRDI (2014): *Statistical analysis of network data with R*, vol. 65, Springer.

KYRIAZIDOU, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65.

LACHOWSKA, M., A. MAS, R. SAGGIO, AND S. A. WOODBURY (2023): "Do firm effects drift? Evidence from Washington administrative data," *Journal of Econometrics*, 233, 375–395.

LEUNG, M. P. (2020): "Treatment and spillover effects under network interference," *Review of Economics and Statistics*, 102, 368–380.

——— (2022a): "Causal inference under approximate neighborhood interference," *Econometrica*, 90, 267–293.

——— (2022b): "Causal Inference Under Approximate Neighborhood Interference," *Econometrica*, 90, 267–293.

Levine, D., R. Polimeni, and I. Ramage (2016): "Insuring health or insuring wealth? An experimental evaluation of health insurance in rural Cambodia," *Journal of Development Economics*, 119, 1–15.

Lewbel, A., X. Qu, and X. Tang (2023): "Ignoring measurement errors in social networks," *The Econometrics Journal*, utad028.

Manski, C. F. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

Mas, A. (2025): "Non-Wage Amenities," Tech. rep., Cambridge, MA.

Meghir, C., A. M. Mobarak, C. Mommaerts, and M. Morten (2022): "Migration and Informal Insurance: Evidence from a Randomized Controlled Trial and a Structural Model," *Review of Economic Studies*, 89, 452–480.

Menzel, K. (2021): "Bootstrap With Cluster-Dependence in Two or More Dimensions," *Econometrica*, 89, 2143–2188.

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

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018): "Commuting, migration, and local employment elasticities," *American Economic Review*, 108.

Moretti, E. and D. J. Wilson (2017): "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists," *American Economic Review*, 107, 1858–1903.

Nagavarapu, S. and S. Sekhri (2016): "Informal monitoring and enforcement mechanisms in public service delivery: Evidence from the public distribution system in India," *Journal of Development Economics*, 121, 63–78.

Newman, M. (2018): *Networks*, Oxford University Press.

Pallais, A. and E. G. Sands (2016): "Why the Referential Treatment? Evidence from Field Experiments on Referrals," *Journal of Political Economy*, 124, 1793–1828.

Postel-Vinay, F. and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350.

Robinson, J. (2012): "Limited Insurance within the Household: Evidence from a Field Experiment in Kenya," *American Economic Journal: Applied Economics*, 4, 140–164.

Sassi, Y. S. (2023): "The Ordinary Least Eigenvalues Estimator," ArXiv:2304.12554v1 [econ.EM].

Silva, J. M. S. and S. Tenreyro (2006): "The log of gravity," *Review of Economics and Statistics*, 88.

Soldani, E., N. Hildebrandt, Y. Nyarko, and G. Romagnoli (2023): "Price information, inter-village networks, and "bargaining spillovers": Experimental evidence from Ghana," *Journal of Development Economics*, 164.

Sorkin, I. (2018): "Ranking Firms Using Revealed Preference*," *The Quarterly Journal of Economics*, 133, 1331–1393.

Tabord-Meehan, M. (2019): "Inference With Dyadic Data: Asymptotic Behavior of the Dyadic-Robust t-Statistic," *Journal of Business and Economic Statistics*, 37.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Verdier, V. (2020): "Estimation and Inference for Linear Models with Two-Way Fixed Effects and Sparsely Matched Data," *The Review of Economics and Statistics*, 102, 1–16.

Viviano, D. (2024): "Policy targeting under network interference," *Review of Economic Studies*, rdae041.

White, H. (2001): *Asymptotic Theory for Econometricians*, Academic Press.

Xu, R. and J. M. Wooldridge (2022): "A design-based approach to spatial correlation," *arXiv preprint arXiv:2211.14354*.

Zeleneev, A. (2020): "Identification and Estimation of Network Models with Nonparametric Unobserved Heterogeneity," https://www.princeton.edu/~zeleneev/azeleneev_jmp.pdf.