NETWORK ROBUST INFERENCE FOR FIXED-EFFECT REGRESSIONS

KENSUKE SAKAMOTO

NOVEMBER 12, 2025

Job Market Paper: Please click HERE for the most recent version

Department of Economics, University of Wisconsin-Madison

ABSTRACT. Fixed effects estimated from network data are central to many economic applications that aim to capture unobserved heterogeneity. Most existing methods assume conditional independence of network observations, an assumption whose plausibility is often application-dependent. This paper studies fixed-effect regressions on network data under a conditional dependence structure in which errors arise from both node- and edge-level shocks that are not fully captured by the fixed effects. We show that the least-squares estimator of the fixed effects can be inconsistent due to a persistent noise term induced by the dependence structure. Leveraging information from the regression residuals, we propose new inference methods for fixed effects that explicitly account for dependence. We also introduce a bias-correction procedure for estimating the sample variance of the fixed effects under dependence. An empirical application using Italian matched employer-employee data demonstrates the practical relevance of our methods and highlights the substantial impact of dependence on inference and variance estimation.

KEYWORDS. Network data, fixed effects, unobserved heterogeneity, dependence, bias correction, matched employer-employee data, limited mobility

E-mail address: ksakamoto2@wisc.edu.

The author thanks Jack Porter, Bruce Hansen, and Harold Chiang for their advice and support. The author is also grateful to Yong Cai for many helpful discussions. The author appreciates comments from Xiaoxia Shi, Kohei Yata, and participants at New York Camp Econometrics XIX, the Bristol Econometrics Study Group Annual Conference (2025), the Celebrating James MacKinnon Conference, the Midwest Econometrics Group Conference (2025), and the 40th Canadian Econometrics Study Group Meeting. The author gratefully acknowledges financial support from the Douglas W. and Sherry A. Caves Dissertation Fellowship.

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.

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

[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).

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) study 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.

This paper is organized as follows. Section 2 introduces the model, estimator, and dependence structure. Section 3 develops the asymptotic theory and inference procedures for fixed effects. Section 4 presents a bias-correction for variance estimation of fixed effects. Section 5 reports Monte Carlo evidence on finite-sample performance. Section 6 provides an empirical application using matched employer-employee data. Section 7 concludes. The Appendix collects proofs and additional discussion.

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.

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

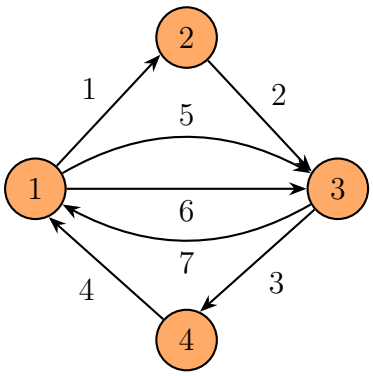
$E_i^s = \{e \in E : s(e) = i\}$ be the set of edges that originate from node i and $E_i^t = \{e \in E : t(e) = i\}$ be the set of edges that terminate at node i. Then, $E_i = E_i^s \cup E_i^t$ is the set of edges incident to node i. We define the out-degree of node i as $d_i^s = |E_i^s|$, the number 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 2.1, 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 2.1. 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. *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 and our empirical application where each edge corresponds to a single worker transition.*

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

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

simplified to

$$\mathbf{y} = \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{2.1}$$

which we refer to as our regression model.

Element-wise, the regression model (2.1) 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 economic applications.

Example 1. *(Two-Period AKM model): An important example of the model* (2.1) *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*

$$w_{g,t} = \phi_g + \alpha_{J(g,t)} + u_{g,t}, \tag{2.2}$$

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 (2.1).

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:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{B}'\mathbf{B})^* \mathbf{B}'\mathbf{y} = \mathbf{L}^* \mathbf{B}'\mathbf{y}, \tag{2.3}$$

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 (2.3):

Assumption 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 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:

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

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

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

where $\hat{\boldsymbol{\alpha}}_{-i}$ 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.

Remark 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).*

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:

$$\epsilon_e = f(U_{s(e)}, U_{t(e)}, V_e); \quad \mathbb{E}[\epsilon_e] = 0, \tag{2.4}$$

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 (2.4) to facilitate the analysis:

Assumption 2. *The error term $\boldsymbol{\epsilon} = (\epsilon_e)_{e=1}^m$ satisfies the following conditions:*

(i) For each $e \in E$, $\mathbb{E}[\epsilon_e] = 0$, $\min_{e \in E} \mathbb{E}[\epsilon_e^2] \geq C_1 > 0$, 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 2 imposes a specific dependence structure on the error terms $\boldsymbol{\epsilon}$. Part (i) ensures that the errors are mean-zero, 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 (2.4) 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 (2.4) under Assumption 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

$$\text{Cov}(\epsilon_e, \epsilon_{e'}) = \begin{cases} \sigma_e^2 \text{ if } e = e'; \\ \sigma_{e,e'} \text{ if } e \neq e' \text{ and } e, e' \in E_i \text{ for some } i \in V; \\ 0 \text{ otherwise} \end{cases}, \tag{2.5}$$

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 (2.4) 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 (2.4) to the AKM model in Example 1:

Example 2. *(Two-Period AKM model, continued). Consider the setting described in Example 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* (2.2)*, 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* (2.4) *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* (2.4)*. 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.*

2.4. **Decomposition.** To isolate the impact of the dependence structure on the fixed-effect estimator $\hat{\boldsymbol{\alpha}}$, 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 (2.4) 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 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 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 distirbuted 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 $\mathbb{E}[V_e] = 1$ and consider the model

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

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

In contrast, Assumption 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 $\hat{\boldsymbol{\alpha}}$ as follows:[3]

$$\hat{\boldsymbol{\alpha}} = \mathbf{L}^*\mathbf{B}'\mathbf{y} = \boldsymbol{\alpha} + \mathbf{L}^*\mathbf{B}'(\mathbf{F}^s\boldsymbol{\tau}^s + \mathbf{F}^t\boldsymbol{\tau}^t) + \mathbf{L}^*\mathbf{B}'(\boldsymbol{\epsilon} - \mathbf{F}^s\boldsymbol{\tau}^s - \mathbf{F}^t\boldsymbol{\tau}^t). \tag{2.6}$$

The first term in (2.6) 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.

2.5. **Variance Bound.** Since the second term and third term in (2.6) are orthogonal to each other, i.e., uncorrelated element-wise, the covariance matrix of $\hat{\boldsymbol{\alpha}}$ is given by:

$$\mathrm{Cov}(\hat{\boldsymbol{\alpha}}) = \mathbf{L}^*\mathbf{B}'\boldsymbol{\Omega}_1\mathbf{B}\mathbf{L}^* + \mathbf{L}^*\mathbf{B}'\boldsymbol{\Omega}_2\mathbf{B}\mathbf{L}^*, \tag{2.7}$$

where

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

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

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

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 (2.7) is bounded by

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

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.

[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

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 2 (ii):

$$\mathbf{\Omega}_2 = \begin{pmatrix} \ddots & & \\ & \mathbf{\Omega}_{(i,j)} & \\ & & \ddots \end{pmatrix}, \quad \mathbf{\Omega}_{(i,j)} = \begin{pmatrix} \sigma^2_{e_k} & \sigma_{e_k,e_l} & \cdots \\ \sigma_{e_k,e_l} & \sigma^2_{e_l} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} - (\mathbb{E}[\tau^s(U_i)^2 + \tau^t(U_i)^2])\boldsymbol{\iota}_{|E_{(i,j)}|}\boldsymbol{\iota}'_{|E_{(i,j)}|},$$

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

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

where $\tilde{\sigma}^2_n \equiv \max_{i,j} \mathbf{A}_{i,j} \max_{e \in E_{(i,j)}} (\sigma^2_e - \mathbb{E}[\tau^s(U_i)^2 + \tau^t(U_i)^2])$. Thus, the second term in (2.7) is bounded by

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

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

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

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 $\sigma^2_\tau \times \lambda_{n,F}$, in addition to the weak dependence structure, as captured by $\tilde{\sigma}^2_n$.

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

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

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 $\mathrm{Cov}(\hat{\boldsymbol{\alpha}})$.

Proposition 1. *Under Assumptions 1-3, the variance of each fixed-effect estimator $\hat{\alpha}_i$ is bounded as follows:*

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

Proposition 1 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., $\tilde{\sigma}_n^2 = \sigma^2$):

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

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 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 is shown to be tight up to a constant factor.

Example 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 $\hat{\alpha}_i$ as follows:

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

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 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

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

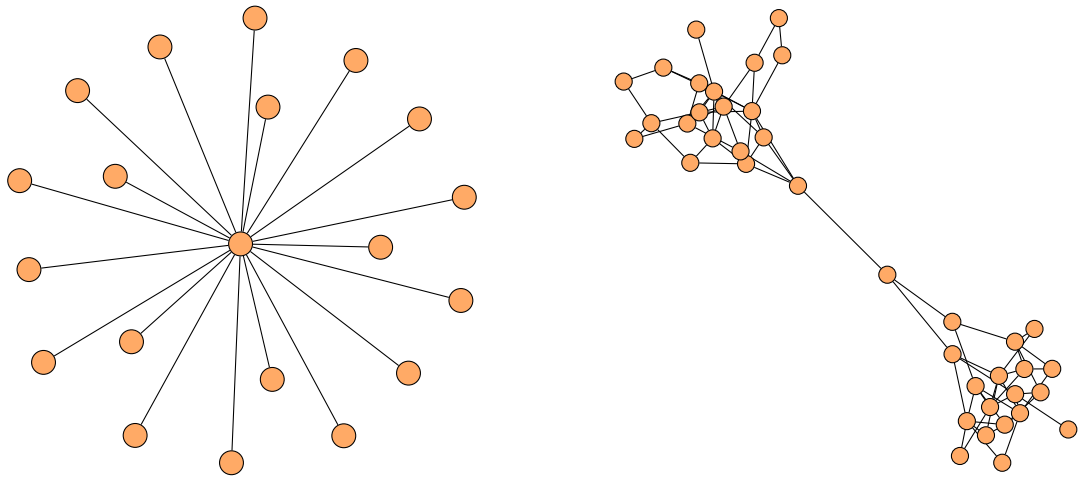
Therefore, the variance bound in Proposition 1 is tight up to a constant factor in this example.

Remark 3. *Note that $\lambda_{n,F}$ and $\lambda_{2,L}$ can vary independently. For instance, in Figure 2.2, 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.*

3. Asymptotic Theory

The variance decomposition in (2.7) and the finite-sample variance bound in Proposition 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

FIGURE 2.2. 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 $\lambda_{2,L}$

bound and does not sharply characterize the estimator's behavior, except for specific graph structures like Example 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.

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:

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

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.

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 $\hat{\alpha}_i$ 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:

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \mathbf{D}^{-1}\mathbf{B}'(\mathbf{F}^s\boldsymbol{\tau}^s + \mathbf{F}^t\boldsymbol{\tau}^t) + \mathbf{D}^{-1}\mathbf{B}'(\boldsymbol{\epsilon} - \mathbf{F}^s\boldsymbol{\tau}^s - \mathbf{F}^t\boldsymbol{\tau}^t) + \boldsymbol{r}. \tag{3.1}$$

The first term on the right-hand side of (3.1) 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 (3.1) 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 4. *The number of edges between each pair of nodes is absolutely bounded, i.e., there exists an absolute constant $C > 0$ such that $\max_{i,j\in V}|E_{(i,j)}| \leq C$ for all $n \in \mathbb{N}$.*

Assumption 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 $\max_{i,j\in V}|E_{(i,j)}| \to \infty$ 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 $\hat{\boldsymbol{\alpha}}$:

Theorem 1. *Under Assumptions 1-4, for each $i \in V$, we have*

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

as $d_i \to \infty$.

Theorem 1 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 extends Theorem 4 in Jochmans and Weidner (2019), which establishes the first-order approximation of $\hat{\alpha}_i$ 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

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

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 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, $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) = 2/(n-1) \to 0$ as $n \to \infty$.*

Example 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 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 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 may fail under such conditions.[6] Indeed, without the dependence structure, $\lambda_{n,F}$ would be replaced 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 4. *Attempting to improve connectivity by aggregating nodes into supernodes or grouping fixed effects into fewer classes, as discussed in Remark 3, may help achieve the negligibility of the remainder term in Theorem 1. For instance, when grouping fixed effects*

[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.

[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 of the remainder term. Thus, even if $\lambda_{n,F}/(d_i\lambda_{2,L}h_i) \not\to 0$, the remainder term may still be negligible depending on the specific graph structure and dependence structure.

into K classes as in Bonhomme et al. (2019), the mobility network likely becomes closer to a complete graph, as in Example 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 can provide a useful approximation for the fixed-effect estimator in such cases after aggregation.

3.2. **Inference.** The first-order approximation in Theorem 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.

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

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

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 2. *Under Assumptions 1-4, for each $i \in V$, we have*

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

as $d_i \to \infty$.

Proposition 2 shows that if node i has sufficient balance between its inflow and outflow of edges, the local average of the residuals $\hat{\epsilon}_e$ 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

$$\hat{\tau}_i \equiv \frac{1}{d_i c_i}\sum_{e\in E_i}\hat{\epsilon}_e = \frac{\tau^t(U_i)+\tau^s(U_i)}{2} + O_p\left(\sqrt{\frac{1}{c_i^2 d_i}}\right). \tag{3.2}$$

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 5. *The convergence rate of $\hat{\tau}_i$ in* (3.2) *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 may not hold, as in Example 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:

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

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

Assumption 5. *For any $u \in \mathcal{U}$, we have*

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

Assumption 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 3. Importantly, this assumption is weaker than requiring symmetry of f in (2.4) 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 5 can hold even when f is not symmetric in $U_{s(e)}$ and $U_{t(e)}$: If

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

with $\mathbb{E}[U_i] = \mathbb{E}[U_i^2] = 1$, then

$$\tau^s(u) = \tau^t(u) = u - 1$$

for all $u \in \mathcal{U}$, even though $(U_{s(e)}, U_{t(e)}) = (2, 0)$ and $(0, 2)$ yield $\epsilon_e = -5$ 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 5.

When Assumption 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 5 to the AKM model discussed earlier:

Example 7. *(Two-Period AKM model, revisited). Consider the two-period AKM model described in Examples 1 and 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 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 5 holds.

Remark 6. *Instead of imposing Assumption 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 7 is an instance of this, with $\tau^t(\cdot) = -\rho\tau^s(\cdot)$. See Appendix C for inference under such settings.*

Under Assumption 5, Proposition 2 implies that we can consistently estimate $\tau(U_i)$ as follows:

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

if $c_i^2 d_i \to \infty$ as $d_i \to \infty$. Moreover, under Assumption 5, Theorem 1 implies that

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

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 $\hat{\tau}_i$ to estimate the distribution of $\tau(U_i)$ and to perform inference on fixed effects.

Remark 7. *Since $\hat{\tau}_i$ is a consistent estimator for $\tau(U_i)$, one might consider correcting the fixed-effect estimator as follows:*

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

For point estimation, this correction would yield a consistent estimator for α_i if the remainder term in Theorem 1 is negligible and i has sufficient balance. Relatedly, Appendix B 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 6 and in Figure 3.1. Second, the correction does not guarantee asymptotic normality, since the remainder term $\boldsymbol{r}$ in (3.1) *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 $\hat{\tau}_i$ 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 6. *$\tau(U_i)$ has a continuous cumulative distribution function F_τ.*

Assumption 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 3.1 illustrates the determination of $\mathcal{C}_n$ using data from the empirical application in Section 6. Let F_τ be the cumulative distribution function of $\tau(U_i)$ and $\hat{F}_{n,\tau}$ be the empirical distribution function of $\hat{\tau}_i$ for $i \in \mathcal{C}_n$, defined as

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

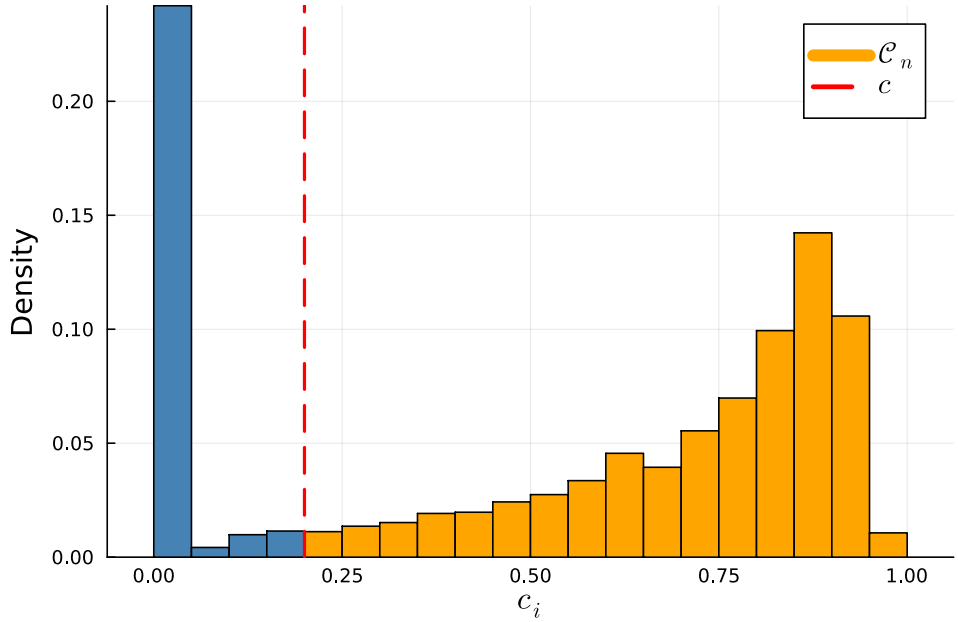
where $\mathbb{I}(\cdot)$ is the indicator function. Also, define the following object to globally control the estimation errors for $|\mathcal{C}_n|$ nodes:

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

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

[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 3.1. 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 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 $c = 0.2$.

Then, we have the following result:

Theorem 2. *Under Assumptions 1-6, if* $|\mathcal{C}_n| \to \infty$ *and* $\eta_n \to 0$ *as* $n \to \infty$*, we have*

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

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 4-6, 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 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

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

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 $\hat{F}_{n,\tau}$. Specifically, for a given confidence level $1-\alpha$, let $\hat{c}_{\alpha/2}$ and $\hat{c}_{1-\alpha/2}$ be the $(\alpha/2)$-th and $(1-\alpha/2)$-th quantiles of $\hat{F}_{n,\tau}$, respectively. Then, the $1-\alpha$ confidence interval for α_i is given by

$$CI_{i,1-\alpha} \equiv \begin{cases} \left[\hat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\hat{c}_{1-\alpha/2}, \hat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\hat{c}_{\alpha/2}\right] & \text{if } d_i^t - d_i^s > 0; \\ \left[\hat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\hat{c}_{\alpha/2}, \hat{\alpha}_i - \left(\frac{d_i^t - d_i^s}{d_i}\right)\hat{c}_{1-\alpha/2}\right] & \text{if } d_i^t - d_i^s < 0. \end{cases} \tag{3.3}$$

Another useful application of Theorem 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:

$$T = \hat{\boldsymbol{\alpha}}_{V_0}' \boldsymbol{M}_{n_0} \hat{\boldsymbol{\alpha}}_{V_0} \tag{3.4}$$

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 $\hat{\boldsymbol{\alpha}}_{V_0}$ 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 $\hat{F}_{n,\tau}$. Let $\hat{c}_{1-\alpha}^T$ denote the $(1-\alpha)$ quantile of the simulated distribution of T under the null. The test rejects H_0 if $T > \hat{c}_{1-\alpha}^T$.

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 $g_n(\hat{\boldsymbol{\alpha}}_{V_0})$ for some known function $g_n(\cdot)$ such that the null distribution can be simulated from $\hat{F}_{n,\tau}$.

The following result establishes the asymptotic validity of the confidence interval in (3.3) and the test based on the statistic T in (3.4):[8]

Proposition 3. *Suppose that the conditions of Theorem 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

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

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

Remark 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*

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

and little information on $\tau(U_i)$ remains in $\hat{\alpha}_i$. The confidence interval in (3.3) 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 7.

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

[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.

variance of the fixed effects $\boldsymbol{\alpha}$ and address the bias in the sample variance estimator that arises from the dependence structure. See Appendix D for estimation of the covariance between two sets of fixed effects.

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

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

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 $\bar{\boldsymbol{\alpha}} = \boldsymbol{\iota}_n'\boldsymbol{\alpha}/n$ is the average of $\boldsymbol{\alpha}$. We can estimate $V_{\boldsymbol{\alpha}}$ by plugging in $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$:

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

It is well known that $\hat{V}_{\boldsymbol{\alpha}}$ is biased upward due to the estimation error in $\hat{\boldsymbol{\alpha}}$ even in cases where dependence is ruled out (Andrews et al., 2008). To see this, from (2.3), we have

$$\begin{aligned} \mathbb{E}[\hat{V}_{\boldsymbol{\alpha}}] &= V_{\boldsymbol{\alpha}} + \frac{\mathbb{E}[\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\boldsymbol{\epsilon}]}{n} \\ &= V_{\boldsymbol{\alpha}} + \frac{\mathrm{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'])}{n}, \end{aligned} \tag{4.1}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Thus, the bias in $\hat{V}_{\boldsymbol{\alpha}}$ is given by the second term in (4.1).

The bias term in (4.1) 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.

4.2. **Bias Correction.** To correct the bias in $\hat{V}_{\boldsymbol{\alpha}}$, 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):

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

where $\hat{\sigma}^2$ 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 (2.7). Thus, under Assumption 5, the bias term in (4.1) can be rewritten as:[9]

$$\mathbb{E}[\tau(U_i)^2] \times \frac{\operatorname{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} + \frac{\operatorname{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\boldsymbol{\Omega}_2)}{n}. \tag{4.2}$$

Here, the first term in (4.2) 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 $\operatorname{tr}(\boldsymbol{M}_n\mathbf{L}^*)$.

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

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

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 not well-connected, both terms will be significant.

Hypothetically, we can estimate the bias term in (4.2) by estimating $\mathbb{E}[\tau(U_i)^2]$ and $\boldsymbol{\Omega}_2$ separately. In particular, we can estimate $\mathbb{E}[\tau^2(U_i)]$ via the following estimator:

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

[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 5. Extensions to asymmetric cases are possible following the discussion in Appendix C.

whose consistency is anticipated by Proposition 2. However, estimating $\mathbf{\Omega}_2$, without imposing additional restrictions on the dependence structure or heteroskedasticity is challenging as $\hat{\boldsymbol{\alpha}}$ 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 (4.2):

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

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

Assumption 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 quantities are $o(1)$ as $n \to \infty$:*

$$\frac{\sqrt{\lambda_{n,F}}}{n} \sum_{i\in V} \frac{1}{d_i}\,, \lambda_{n,F} \left(\frac{1}{n}\sum_{i\in V} \frac{1}{d_i}\right)^{3/2}, \frac{\lambda_{n,F}}{n} \sum_{i\in V} \frac{1}{d_i\, \lambda_{2,L}\, h_i}.$$

Part (i) of Assumption 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 arising from the dependence structure persists (so our bias correction is essential), while the bias from the remainder of the variance–covariance matrix remains relatively small. These conditions hold, for example, in well-connected, homogeneous graphs such as the complete graph of Example 4 and the Erdős–Rényi random graph of Example 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 6.

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

Theorem 3. *Under Assumptions 1-5 and 7, if $|\mathcal{C}_n| \to \infty$ and $\eta_n \to 0$ as $n \to \infty$, we have*

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

as $n \to \infty$.

Note that although Theorem 3 establishes the consistency of the bias-corrected estimator $\hat{V}_{\boldsymbol{\alpha}}^{bc}$, in finite sample, it is typically still biased upward as we have not corrected the second term in (4.2). 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:

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

where

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

The second term in $\hat{V}_{\boldsymbol{\alpha}}^{bc,mod}$ serves as the bias-correction term for the second term in (4.2) 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 $\hat{\sigma}^2$ is not consistent for $\mathbb{E}[\epsilon_e^2]$ in general. However, if the errors are homoskedastic, it can be approximated by

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

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

Remark 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 $\hat{\sigma}^2$ above assumes homoskedasticity. Using a more robust estimator for $\hat{\sigma}^2$, 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 (4.2), 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 (4.2) and are consistent if $\boldsymbol{\epsilon}$ is independent.[10]

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

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

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 (4.2) may be negligible.

However, if the node-level dependence structure is strong with significant $\mathbb{E}[\tau^2(U_i)]$ relative to $\mathbb{E}[\epsilon_e^2]$, then the bias stemming from this dependence structure becomes dominant, as shown in Theorem 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, suffers from the same limitation and will generally be inconsistent when the dependence structure is strong.

Remark 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 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 7, as in the case of complete graphs, thereby increasing the applicability of our bias-correction method.*

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.

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

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 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 3 and 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 $e \in E_{(i,j)}$ 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 (2.1). For each iteration, we compute the least-squares estimator $\hat{\boldsymbol{\alpha}}$ and the empirical distribution of $\tau(U_i)$ using Theorem 2, by setting $\mathcal{C}_n = \{i \in V : c_i > 0.2\}$.

[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.

Table 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 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 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. 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.

5.2. **Results: Inference.** In this exercise, we construct 95% confidence intervals for α_i based on (3.3) 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 $\hat{\alpha}_i$ when $\boldsymbol{\epsilon}$ are independent. Specifically,

TABLE 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.

the standard error of $\hat{\alpha}_i$ in the independent case is given by

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

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

Table 3 summarizes the Monte Carlo simulation results. The confidence intervals based on (3.3) 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 (3.3) 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 $\hat{\alpha}_i$ 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 3. Coverage Probability of the Confidence Intervals

	α_1	d_1	$(d_1^t - d_1^s)/d_1$	$\lambda_{n,F}/(d_1\lambda_{2,L}h_1)$	**95% cov**	**Normal 95% cov**
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 (3.3), and the seventh column reports the coverage probability of the confidence intervals based on Jochmans and Weidner (2019)'s asymptotic normality.

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 $\hat{V}_{\boldsymbol{\alpha}}$, the bias-corrected estimator $\hat{V}_{\boldsymbol{\alpha}}^{bc}$, the rule-of-thumb modified bias-corrected estimator $\hat{V}_{\boldsymbol{\alpha}}^{bc,mod}$, and the Andrews et al. (2008)-type bias-corrected estimator $\hat{V}_{\boldsymbol{\alpha}}^{a}$.[12] We simulate each of these variance estimators 2000 times and compute the mean and standard deviation of the estimated variance.

Table 4 summarizes the results and Figures 5.1 and 5.2 present histograms of the simulated variance estimators for $n = 2500$. The plug-in estimator $\hat{V}_{\boldsymbol{\alpha}}$ exhibits substantial upward bias. The bias-corrected estimator $\hat{V}_{\boldsymbol{\alpha}}^{bc}$ is much closer to the true value $V_{\boldsymbol{\alpha}}$, demonstrating the effectiveness of our bias correction. The rule-of-thumb estimator $\hat{V}_{\boldsymbol{\alpha}}^{bc,mod}$ 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 $\hat{V}_{\boldsymbol{\alpha}}^{a}$ also reduces bias, but its remaining bias is larger than that of $\hat{V}_{\boldsymbol{\alpha}}^{bc}$ and $\hat{V}_{\boldsymbol{\alpha}}^{bc,mod}$ due to

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

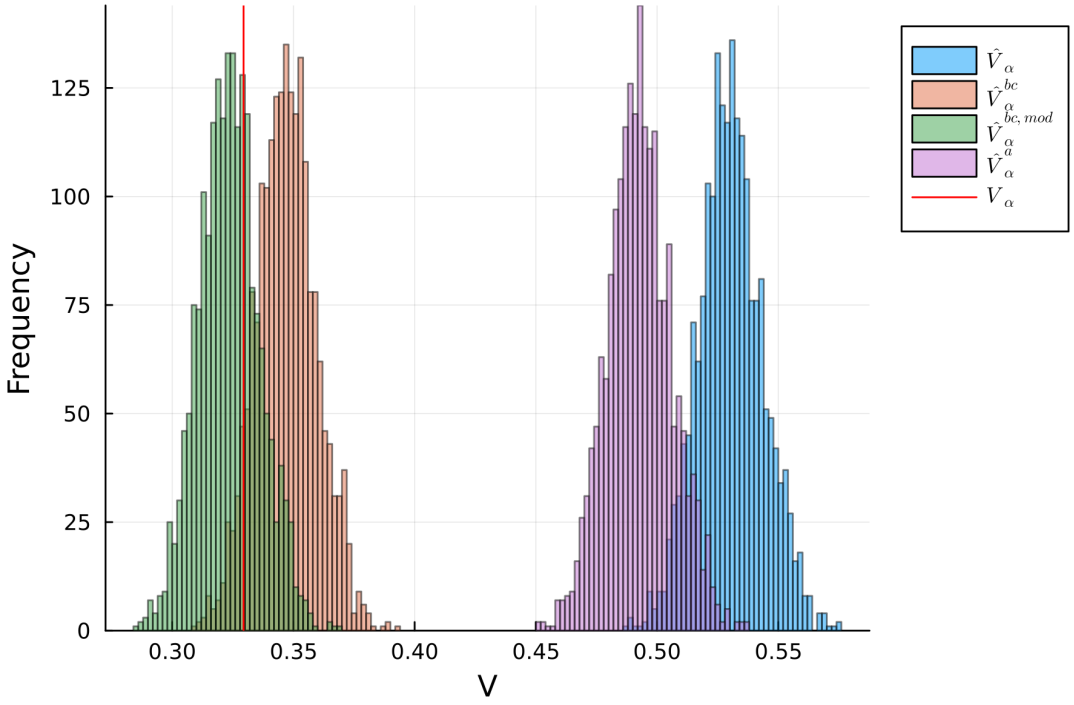
unaccounted dependence. The standard deviations of the variance estimators are generally similar, suggesting that our bias-corrected estimator is preferable in terms of mean squared error.

TABLE 4. Variance Estimation

	$V_{\boldsymbol{\alpha}}$	$\hat{V}_{\boldsymbol{\alpha}}$	$\hat{V}^{bc}_{\boldsymbol{\alpha}}$	$\hat{V}^{bc,mod}_{\boldsymbol{\alpha}}$	$\hat{V}^{a}_{\boldsymbol{\alpha}}$	$sd(\hat{V}_{\boldsymbol{\alpha}})$	$sd(\hat{V}^{bc}_{\boldsymbol{\alpha}})$	$sd(\hat{V}^{bc,mod}_{\boldsymbol{\alpha}})$	$sd(\hat{V}^{a}_{\boldsymbol{\alpha}})$
Panel A: $K = 1$									
$n = 500$	0.334	0.544	0.359	0.328	0.494	0.0301	0.0282	0.0282	0.0298
$n = 1000$	0.340	0.549	0.361	0.333	0.505	0.0215	0.0201	0.0201	0.0214
$n = 2500$	0.329	0.530	0.347	0.323	0.492	0.0129	0.0122	0.0122	0.0128
$n = 5000$	0.336	0.528	0.353	0.33	0.494	0.009	0.0085	0.0085	0.0089
Panel B: $K = 2$									
$n = 500$	0.334	0.609	0.388	0.32	0.492	0.04	0.0391	0.0391	0.0396
$n = 1000$	0.340	0.602	0.388	0.329	0.504	0.0302	0.0293	0.0293	0.0300
$n = 2500$	0.329	0.574	0.356	0.309	0.495	0.0165	0.0164	0.0163	0.0163
$n = 5000$	0.336	0.568	0.362	0.319	0.496	0.0106	0.0109	0.0107	0.0105

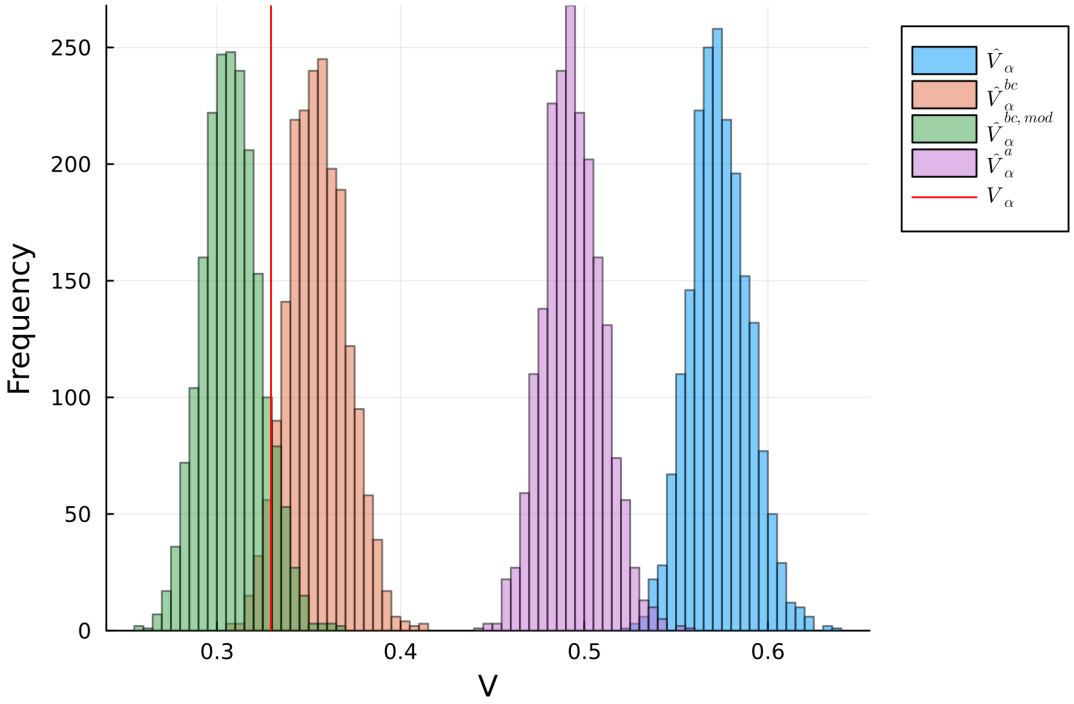
Note: Panel A reports the results for $K = 1$ (Erdős–Rényi model), and Panel B reports the results for $K = 2$. The first column reports the number of nodes n, the second column reports the true variance of $\boldsymbol{\alpha}$, the third to ninth columns report the mean and standard deviation of $\hat{V}_{\boldsymbol{\alpha}},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

FIGURE 5.1. Histogram of the simulated $\hat{V}_{\boldsymbol{\alpha}}$: Erdős–Rényi Model



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

FIGURE 5.2. Histogram of the simulated $\hat{V}_{\boldsymbol{\alpha}}$: Stochastic-Block Model



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

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.

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

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

year. We also exclude workers who work in the public sector and whose wages and days worked are 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; 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 5 presents the degree distribution of the firm-firm network, which is sparse with a mean degree of 18. Table 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 may not be valid, and the finite-sample bias in $\hat{V}^{bc}_{\boldsymbol{\alpha}}$ is likely to be non-negligible.

TABLE 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.

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

[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 5

TABLE 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.

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.

TABLE 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.

Applying the PageRank algorithm to our firm-firm mobility network, we find that the most central firm is firm 931 in our dataset. Table 8 summarizes this firm's key characteristics. Its estimated fixed effect $\hat{\alpha}_{931}$ 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^t_{931} than outflows d^s_{931}, 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 3). This suggests that our inference procedure based on the first-order approximation in Theorem 1 may be valid for this central firm, and we can construct a confidence interval for α_{931}.

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

TABLE 8. Central Firm Information

$\hat{\alpha}_{931}$	d_{931}	d^t_{931}	d^s_{931}	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:

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

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

Note that the mean of the firm fixed effects is -0.017 (see Table 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 $\hat{\alpha}_i$ under the assumption of independent errors, as in the simulation section. The 95% confidence interval for $\hat{\alpha}_{931}$ 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.

6.3. **Results: Variance Estimation.** We estimate the variance of the firm fixed effects $\hat{V}_{\boldsymbol{\alpha}}$ and several bias-corrected estimators: $\hat{V}^{bc}_{\boldsymbol{\alpha}}, \hat{V}^{bc,mod}_{\boldsymbol{\alpha}}$, and $\hat{V}^{a}_{\boldsymbol{\alpha}}$. Table 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 $\hat{V}^{bc}_{\boldsymbol{\alpha}}$ reduces the plug-in estimator $\hat{V}_{\boldsymbol{\alpha}}$ by approximately 28%, which is a larger reduction than that achieved by the Andrews et al. (2008)-type bias-corrected estimator $\hat{V}^{a}_{\boldsymbol{\alpha}}$ (about 23%). The modified bias-corrected estimator $\hat{V}^{bc,mod}_{\boldsymbol{\alpha}}$ 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 9. Variance Component Estimation on the VWH Data

$\hat{V}_{\boldsymbol{\alpha}}$	$\hat{V}^{bc}_{\boldsymbol{\alpha}}$	$\hat{V}^{bc,mod}_{\boldsymbol{\alpha}}$	$\hat{V}^{a}_{\boldsymbol{\alpha}}$
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 $\hat{V}_{\boldsymbol{\alpha}}$, the bias-corrected estimator $\hat{V}^{bc}_{\boldsymbol{\alpha}}$, the modified bias-corrected estimator $\hat{V}^{bc,mod}_{\boldsymbol{\alpha}}$, and the Andrews et al. (2008)-type bias-corrected estimator $\hat{V}^{a}_{\boldsymbol{\alpha}}$.

If we report the two components of the bias in (4.2) separately, we obtain

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

Note that $\hat{\sigma}^2_\tau$, which is an estimator for $\mathbb{E}[\tau(U_i)^2]$, is about 0.015, which is roughly one-quarter of $\hat{\sigma}^2$, 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

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

indicates that the first term is more than three times larger than the second term. Thus, even if $\hat{\sigma}^2_\tau$ 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 D for additional results on covariance estimation between worker and firm fixed effects.

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 D. 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.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

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.

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.

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.

Bonhomme, S. (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.

Bozzo, E. (2013): "The Moore–Penrose inverse of the normalized graph Laplacian," *Linear Algebra and its Applications*, 439, 3038–3043.

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*.

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.

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.

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.

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.

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.

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.

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.

Graham, B. S. (2017): "An Econometric Model of Network Formation With Degree Heterogeneity," *Econometrica*, 85, 1033–1063.

Horn, R. A. and C. R. Johnson (2012): *Matrix Analysis*, Cambridge University Press, 2 ed.

Jackson, C. K. (2013): "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence From Teachers," *The Review of Economics and Statistics*, 95, 1096–1116.

Jochmans, K. and M. Weidner (2019): "Fixed-Effect Regressions on Network Data," *Econometrica*, 87, 1543–1560.

Kallenberg, O. (2002): *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, second ed.

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.

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.

Mas, A. (2025): "Non-Wage Amenities," Tech. rep., Cambridge, MA.

Menzel, K. (2021): "Bootstrap With Cluster-Dependence in Two or More Dimensions," *Econometrica*, 89, 2143–2188.

Newman, M. (2018): *Networks*, Oxford University Press.

Postel-Vinay, F. and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350.

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.

Appendix A. 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:

$$\check{\boldsymbol{\alpha}} = (\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\mathbf{X}}\mathbf{y}, \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{y},$$

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 for $\check{\alpha}_i$.

Algebraically, we have

$$\check{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \mathbf{D}^{-1}\mathbf{B}'\boldsymbol{\epsilon} + \boldsymbol{r} + \tilde{\boldsymbol{r}},$$

where

$$\boldsymbol{r} = \mathbf{D}^{-1}\mathbf{A}(\check{\boldsymbol{\alpha}} - \boldsymbol{\alpha}), \quad \tilde{\boldsymbol{r}} = -\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

We already worked out the first term $\mathbf{D}^{-1}\mathbf{B}'\boldsymbol{\epsilon}$ in the proof of Theorem 1. Thus, we need to show that both $\boldsymbol{r}$ and $\tilde{\boldsymbol{r}}$ are negligible.

First, we show that $\tilde{\boldsymbol{r}}$ is negligible. Note that $\mathbb{E}[\tilde{\boldsymbol{r}}] = 0$. Observe that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\boldsymbol{\epsilon}$$

with $\mathbb{E}[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] = 0$. Then,

$$\begin{aligned}
\mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] &= (\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']\boldsymbol{M}_{\mathbf{B}}\mathbf{X}(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1} \\
&\leq C(\mathbf{X}\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{F}\mathbf{F}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X}'(\mathbf{X}\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1} \\
&\leq \lambda_{n,F} C(\mathbf{X}\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}
\end{aligned}$$

for some absolute constant $C > 0$ under Assumption 1: $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] \leq C\mathbf{F}\mathbf{F}'$. Let

$$\tilde{\rho} = \|(\mathbf{X}'\mathbf{X})(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\|_2,$$

where $\|\cdot\|_2$ is the spectral norm. By the definition of $\tilde{\rho}$, we have $(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1} \leq \tilde{\rho}(\mathbf{X}'\mathbf{X})^{-1}$. Thus, we have

$$\begin{aligned}
\mathbb{E}[\tilde{\boldsymbol{r}}_i^2] &= e_i'\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}\mathbb{E}[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})']\mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i \\
&\leq \lambda_{n,F}Ce_i'\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}(\mathbf{X}'\boldsymbol{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i \\
&\leq \lambda_{n,F}\tilde{\rho}Ce_i'\mathbf{D}^{-1}\mathbf{B}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i \\
&= \frac{\lambda_{n,F}\tilde{\rho}C}{m}\bar{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{x}}_i,
\end{aligned}$$

where $\bar{\boldsymbol{x}}_i = \mathbf{X}'\mathbf{B}\mathbf{D}^{-1}e_i$ 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 $\tilde{\rho} \times \bar{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{x}}_i = O(1)$, we have $\tilde{\boldsymbol{r}}_i = o_p(1)$.

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

- $\max_{i\in V} d_i = o(m)$,
- $\tilde{\rho} \times \bar{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{x}}_i = O(1)$

- $\rho^{-1} = O(1)$,

then Theorem 1 holds for $\check{\boldsymbol{\alpha}}$ as well with the same approximation error rates.

Appendix B. Alternative Estimator

We can consider an alternative estimator for $\boldsymbol{\alpha}$ based on the following transformation.

B.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:

$$\tilde{\boldsymbol{\tau}}^t = (\tau^t(U_i))_{i \in V^{st}}, \quad \tilde{\boldsymbol{\tau}}^s = (\tau^s(U_i))_{i \in V^{st}},$$
$$\check{\boldsymbol{\tau}}^t = (\mathbb{I}\{i \in V^t \setminus V^s\} \cdot \tau^t(U_i))_{i \in V}, \quad \check{\boldsymbol{\tau}}^s = (\mathbb{I}\{i \in V^s \setminus V^t\} \cdot \tau^s(U_i))_{i \in V}.$$

Let $\tilde{\mathbf{F}}$ be the $m \times |V^{st}|$ matrix defined by: for each $e \in E$ and $i \in V^{st}$,

$$\tilde{\mathbf{F}}_{e,i} = \begin{cases} 1 & \text{if } s(e) = i \text{ or } t(e) = i, \\ 0 & \text{otherwise} \end{cases}$$

Finally, let $\tilde{\boldsymbol{\epsilon}} = \boldsymbol{\epsilon} - \mathbf{F}^t\boldsymbol{\tau}^t - \mathbf{F}^s\boldsymbol{\tau}^s$.

Note that the model can be rewritten as

$$\begin{aligned} \mathbf{y} &= \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \\ &= \mathbf{B}\boldsymbol{\alpha} + \mathbf{F}^t\boldsymbol{\tau}^t + \mathbf{F}^s\boldsymbol{\tau}^s + \tilde{\boldsymbol{\epsilon}} \\ &= \mathbf{B}\left(\boldsymbol{\alpha} + \frac{\boldsymbol{\tau}^t}{2} - \frac{\boldsymbol{\tau}^s}{2}\right) + \mathbf{F}\frac{(\boldsymbol{\tau}^t + \boldsymbol{\tau}^s)}{2} + \tilde{\boldsymbol{\epsilon}}, \end{aligned}$$

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 $e \in E$,

$$\left(\mathbf{F}\frac{(\boldsymbol{\tau}^t + \boldsymbol{\tau}^s)}{2}\right)_e = \begin{cases} \frac{\tau^t_{t(e)} + \tau^s_{t(e)} + \tau^t_{s(e)} + \tau^s_{s(e)}}{2} & \text{if } s(e), t(e) \in V^{st} \\ \frac{\tau^t_{t(e)} + \tau^s_{t(e)} + \tau^s_{s(e)}}{2} & \text{if } t(e) \in V^{st},\ s(e) \notin V^{st},\ s(e) \in V^s \\ \frac{\tau^t_{t(e)} + \tau^t_{s(e)} + \tau^s_{s(e)}}{2} & \text{if } s(e) \in V^{st},\ t(e) \notin V^{st},\ t(e) \in V^t \\ \frac{\tau^t_{t(e)} + \tau^s_{s(e)}}{2} & \text{if } t(e), s(e) \notin V^{st},\ t(e) \in V^t,\ s(e) \in V^s \end{cases}$$

This can be decomposed as

$$\left(\mathbf{F}\frac{(\boldsymbol{\tau}^t+\boldsymbol{\tau}^s)}{2}\right)_e = \left(\tilde{\mathbf{F}}\frac{(\tilde{\boldsymbol{\tau}}^t+\tilde{\boldsymbol{\tau}}^s)}{2}\right)_e + \left(\mathbf{B}\frac{(\check{\boldsymbol{\tau}}^t-\check{\boldsymbol{\tau}}^s)}{2}\right)_e,$$

where

$$\left(\tilde{\mathbf{F}}\frac{(\tilde{\boldsymbol{\tau}}^t+\tilde{\boldsymbol{\tau}}^s)}{2}\right)_e = \begin{cases} \frac{\tau^t_{t(e)}+\tau^s_{t(e)}+\tau^t_{s(e)}+\tau^s_{s(e)}}{2} & \text{if } s(e),t(e)\in V^{st} \\ \frac{\tau^t_{t(e)}+\tau^s_{t(e)}}{2} & \text{if } t(e)\in V^{st},\ s(e)\notin V^{st} \\ \frac{\tau^t_{s(e)}+\tau^s_{s(e)}}{2} & \text{if } s(e)\in V^{st},\ t(e)\notin V^{st} \\ 0 & \text{if } t(e),s(e)\notin V^{st} \end{cases}$$

$$\left(\mathbf{B}\frac{(\check{\boldsymbol{\tau}}^t-\check{\boldsymbol{\tau}}^s)}{2}\right)_e = \begin{cases} \frac{\tau^t_{t(e)}}{2} & \text{if } t(e)\in V^t\setminus V^s,\ s(e)\in V^{st} \\ \frac{\tau^s_{s(e)}}{2} & \text{if } s(e)\in V^s\setminus V^t,\ t(e)\in V^{st} \\ \frac{\tau^t_{t(e)}+\tau^s_{s(e)}}{2} & \text{if } t(e),s(e)\notin V^{st},\ t(e)\in V^t,\ s(e)\in V^s \\ 0 & \text{if } t(e),s(e)\in V^{st} \end{cases}$$

Therefore,

$$\frac{\mathbf{F}}{2}(\boldsymbol{\tau}^t+\boldsymbol{\tau}^s) = \tilde{\mathbf{F}}\frac{(\tilde{\boldsymbol{\tau}}^t+\tilde{\boldsymbol{\tau}}^s)}{2} + \mathbf{B}\frac{(\check{\boldsymbol{\tau}}^t-\check{\boldsymbol{\tau}}^s)}{2}.$$

Substituting this into the model, we obtain

$$\mathbf{y} = \mathbf{B}\left(\boldsymbol{\alpha}+\frac{\boldsymbol{\tau}^t+\check{\boldsymbol{\tau}}^t}{2}-\frac{\boldsymbol{\tau}^s+\check{\boldsymbol{\tau}}^s}{2}\right)+\tilde{\mathbf{F}}\frac{(\tilde{\boldsymbol{\tau}}^t+\tilde{\boldsymbol{\tau}}^s)}{2}+\tilde{\boldsymbol{\epsilon}}. \tag{B.1}$$

This is the transformed model where the effects explained by the incidence matrix $\mathbf{B}$ are separated from those not explained by it.

B.2. **Alternative Estimator.** Let $\boldsymbol{M}_{\tilde{\mathbf{F}}} = \mathbf{I}_m - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^+\tilde{\mathbf{F}}'$, which is the projection matrix onto the nullspace of $\tilde{\mathbf{F}}$. Premultiplying both sides of equation (B.1) by $\boldsymbol{M}_{\tilde{\mathbf{F}}}$, we have

$$\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{y} = \boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B}\left(\boldsymbol{\alpha}+\frac{\boldsymbol{\tau}^t+\check{\boldsymbol{\tau}}^t}{2}-\frac{\boldsymbol{\tau}^s+\check{\boldsymbol{\tau}}^s}{2}\right)+\boldsymbol{M}_{\tilde{\mathbf{F}}}\tilde{\boldsymbol{\epsilon}}.$$

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:

$$\begin{aligned}\hat{\boldsymbol{\alpha}}^{alt} &= (\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{y} \\ &= \boldsymbol{\alpha} + (\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B}\left(\frac{\boldsymbol{\tau}^t + \check{\boldsymbol{\tau}}^t}{2} - \frac{\boldsymbol{\tau}^s + \check{\boldsymbol{\tau}}^s}{2}\right) + (\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\tilde{\boldsymbol{\epsilon}}.\end{aligned}$$

An implicit assumption here is that $\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B}$ has rank $n-1$ so that $(\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})^*$ is well-defined. This requires that $rank((\mathbf{B}, \tilde{\mathbf{F}})) = n-1+rank(\tilde{\mathbf{F}})$, which means that the column spaces of $\mathbf{B}$ and $\tilde{\mathbf{F}}$ do not overlap except for the intercept. This assumption holds if the graph contains enough directed cycles so that a column of $\tilde{\mathbf{F}}$ 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.

B.3. **Approximation.** We can derive a first-order approximation for $\hat{\boldsymbol{\alpha}}^{alt}$ similar to Theorem 1. Note that

$$\begin{aligned}\hat{\boldsymbol{\alpha}}^{alt} - \boldsymbol{\alpha} = &\frac{\boldsymbol{\tau}^t + \check{\boldsymbol{\tau}}^t}{2} - \frac{\boldsymbol{\tau}^s + \check{\boldsymbol{\tau}}^s}{2} \\ &+ (\mathbf{I}_n - (\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})\left(\frac{\boldsymbol{\tau}^t + \check{\boldsymbol{\tau}}^t}{2} - \frac{\boldsymbol{\tau}^s + \check{\boldsymbol{\tau}}^s}{2}\right) \\ &+ (\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})^*\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\tilde{\boldsymbol{\epsilon}},\end{aligned}$$

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 $\tilde{\boldsymbol{\epsilon}}$. 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 $\hat{\alpha}_i^{alt}$ can be consistent for α_i, unlike the original estimator $\hat{\alpha}_i$. 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 $\hat{\alpha}_i - \frac{d_i^t - d_i^s}{d_i}\hat{\tau}_i$ in the original estimator.

For the other two terms, once the difference between $(\mathbf{B}'\boldsymbol{M}_{\tilde{\mathbf{F}}}\mathbf{B})^*$ and $(\mathbf{B}'\mathbf{B})^*$ is characterized, we can follow similar steps as in the proof of Proposition 2 and Theorem 1 to derive the approximation with different error rates. Noting that $\tilde{\boldsymbol{\epsilon}}$ is weakly dependent and following Appendix A, we can show that the approximation error rate for the last two terms is given by

$$O_p\left(\sqrt{\frac{1}{\rho d_i}} + \sqrt{\frac{1}{\rho\lambda_{2,L}h_i d_i}} + \sqrt{\frac{\tilde{\rho}}{m}\bar{\boldsymbol{x}}_i'\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{x}}_i}\right),$$

where $\rho, \tilde{\rho}, \bar{\boldsymbol{x}}_i, \boldsymbol{\Omega}$ are defined in Appendix A with $\mathbf{X}$ replaced by $\tilde{\mathbf{F}}$. Since $\boldsymbol{M}_{\tilde{\mathbf{F}}}$ eliminates the strongly dependent component, we do not have $\lambda_{n,F}$ in this rate unlike the rate in Theorem 1. However, ρ^{-1} and $\tilde{\rho}$ may be large depending on the relationship between $\mathbf{B}$ and $\tilde{\mathbf{F}}$, which may make the approximation error rate larger than that in Theorem 1. A more detailed comparison of the two error rates and the choice between the two estimators is left for future research.

Appendix C. 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 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:

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

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.

C.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. Furthermore, $\tau^t = \tau^s$ is a special case of this restriction with $c = 1$.

Under this restriction, the approximations for $\hat{\alpha}_i$ and $\hat{\tau}_i$ become:

$$\hat{\alpha}_i \approx \alpha_i + \frac{cd_i^t - d_i^s}{d_i}\tau(U_i),$$
$$\hat{\tau}_i \approx \frac{(c+1)}{2}\tau(U_i).$$

Write

$$\hat{\alpha}_i = \alpha_i + \frac{cd_i^t - d_i^s}{d_i}\tau(U_i) + r_{i,1}$$
$$\hat{\tau}_i = \frac{c+1}{2}\tau(U_i) + r_{i,2},$$

where $r_{i,1}$ and $r_{i,2}$ are the approximation errors. By the proof of Theorem 1, $r_{i,2}$ 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.

We construct the estimator for c in the following manner. First, we regress $\hat{\alpha}_i$ on $\hat{\tau}_i$ and obtain the following regression coefficient:

$$\hat{\beta}_n \equiv \frac{|\mathcal{C}_n|^{-1}\sum_{i\in\mathcal{C}_n}\hat{\tau}_i\hat{\alpha}_i}{|\mathcal{C}_n|^{-1}\sum_{i\in\mathcal{C}_n}\hat{\tau}_i^2}.$$

By Lemma 3, the denominator converges to $((c+1)/2)^2\mathbb{E}[\tau(U_i)^2]$. The convergence of the numerator is shown as follows:

$$\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\hat{\tau}_i\hat{\alpha}_i = \frac{c+1}{2}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\frac{cd_i^t - d_i^s}{d_i} \times \tau(U_i)^2 \tag{C.1}$$
$$+ \frac{c+1}{2}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\alpha_i\tau(U_i) \tag{C.2}$$
$$+ \frac{c+1}{2}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,1}\tau(U_i) \tag{C.3}$$
$$+ \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,2}\alpha_i \tag{C.4}$$
$$+ \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,2}\frac{cd_i^t - d_i^s}{d_i}\tau(U_i) \tag{C.5}$$
$$+ \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}r_{i,2}r_{i,1}. \tag{C.6}$$

The first term (C.1) 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 (C.2) 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 (C.3) 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

$$\mathbb{E}[(\text{C.3})] = O_p\left(\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\sqrt{\frac{\lambda_{n,F}}{d_i\lambda_{2,L}h_i}}\right) = o(1).$$

Thus, (C.3) $= o_p(1)$ by the Markov inequality. By the similar argument, the fourth to sixth terms (C.4), (C.5), and (C.6) 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

$$\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\hat{\tau}_i\hat{\alpha}_i \to_p \frac{c+1}{2}\mathbb{E}[\tau(U_i)^2] \lim_{n\to\infty}\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}(cd_i^t - d_i^s)/d_i.$$

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

$$\hat{\beta}_n \to_p \frac{2(cd^t - d^s)}{c+1}.$$

Therefore, we can construct a consistent estimator for c as follows:

$$\hat{c}_n = \frac{2d^s + \hat{\beta}_n}{2d^t - \hat{\beta}_n} + O_p\left(\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\sqrt{\frac{\lambda_{n,F}}{d_i\lambda_{2,L}h_i}}\right),$$

as $n \to \infty$.

Hence, by collecting $2(\hat{c}_n + 1)^{-1}\hat{\tau}_i \approx \tau(U_i)$, we can construct the consistent estimator for F_τ and conduct inference on $\hat{\alpha}_i$ 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 $\tilde{\tau}_i = (g(\tau(U_i)) + \tau(U_i))/2$, we have

$$\sum_{i\in\mathcal{C}_n}\mathbb{E}\left[\tilde{\tau}_i^{2m-1}\left(\alpha_i - \tau(U_i) + \frac{2d_i^t}{d_i}g^{-1}(\tilde{\tau}_i)\right)\right] = 0$$

for each integer $m \geq 1$, by g^{-1} being an odd function and the symmetry of $\tilde{\tau}_i$ 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.

C.2. **Inference.** Focus on the case where $\tau^s = \tau$ and $\tau^t = c\tau$ for some constant c. We have shown that $\hat{c}_n$ consistently estimates c. Write $\hat{\tilde{\tau}}_i = 2(\hat{c}_n + 1)^{-1}\hat{\tau}_i$, which is a consistent estimator for $\tau(U_i)$. Then, the distribution F_τ can be estimated by the empirical distribution of $\hat{\tilde{\tau}}_i$, denoted by $\hat{F}_\tau$:

$$\hat{F}_\tau(\cdot) \equiv \frac{1}{|\mathcal{C}_n|} \sum_{i \in \mathcal{C}_n} \mathbb{I}(\hat{\tilde{\tau}}_i \leq \cdot),$$

where $\mathbb{I}(\cdot)$ is the indicator function.

The uniform consistency of $\hat{F}_\tau$ to F_τ over a compact subset of $\mathbb{R}$ follows from modifying the proof of Theorem 2. First, observe that

$$\frac{2}{\hat{c}_n + 1} = \frac{2}{c+1} + r_{n,3}, \quad r_{n,3} = O_p\left(\frac{1}{|\mathcal{C}_n|} \sum_{i \in \mathcal{C}_n} \sqrt{\frac{\lambda_{n,F}}{d_i \lambda_{2,L} h_i}} \right),$$

by the delta method so that

$$\begin{aligned} \hat{\tilde{\tau}}_i &= \frac{2}{\hat{c}_n + 1} \hat{\tau}(U_i) \\ &= \tau(U_i) + \frac{2}{c+1} r_{i,2} + \frac{c+1}{2} \tau(U_i) r_{n,3} + r_{i,2} r_{n,3}. \end{aligned}$$

We can bound $\mathbb{E}[|\hat{\tilde{\tau}}_i - \tau(U_i)|]$ by the rates of convergence of $r_{i,2}$ and $r_{n,3}$. By replacing $\hat{\tau}_i$ with $\hat{\tilde{\tau}}_i$ in the proof of Theorem 2, we have

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

as $n \to \infty$, similarly as before but with slower rates of convergence.

With $\hat{F}_\tau$, we can conduct inference on $\hat{\alpha}_i$ in the same manner as in the symmetric case. For example, a $(1-\alpha)$-level confidence interval for α_i is given by

$$CI_{i,1-\alpha} = \begin{cases} \left[\hat{\alpha}_i - \left(\frac{d_i^t \hat{c}_n - d_i^s}{d_i} \right) \hat{c}_{1-\alpha/2}, \quad \hat{\alpha}_i - \left(\frac{d_i^t \hat{c}_n - d_i^s}{d_i} \right) \hat{c}_{\alpha/2} \right], & \text{if } d_i^t \hat{c}_n - d_i^s > 0, \\ \left[\hat{\alpha}_i - \left(\frac{d_i^t \hat{c}_n - d_i^s}{d_i} \right) \hat{c}_{\alpha/2}, \quad \hat{\alpha}_i - \left(\frac{d_i^t \hat{c}_n - d_i^s}{d_i} \right) \hat{c}_{1-\alpha/2} \right], & \text{if } d_i^t \hat{c}_n - d_i^s < 0, \end{cases}$$

where $\hat{c}_\alpha$ is the α-quantile of $\hat{F}_\tau$.

Appendix D. 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.

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 $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ obtained by the method in the main text, the least-squares estimator for $\boldsymbol{\phi}$ is given by

$$\hat{\boldsymbol{\phi}} = (\mathbf{B}_W' \mathbf{B}_W)^{-1} \mathbf{B}_W' (\boldsymbol{w} - \mathbf{B}_F \hat{\boldsymbol{\alpha}}).$$

The sample covariance between $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\alpha}}$ is given by

$$\begin{aligned} \mathrm{C}_{\boldsymbol{\phi},\boldsymbol{\alpha}} &= \frac{1}{N} \sum_{g \in W} \sum_{t=1}^{2} (\phi_g - \bar{\phi})(\alpha_{J(g,t)} - \bar{\alpha}) \\ &= \frac{1}{N} \boldsymbol{\alpha}' \mathbf{B}_F' \boldsymbol{M}_N \mathbf{B}_W \boldsymbol{\phi}, \end{aligned}$$

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

$$\hat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}} = \frac{1}{N} \hat{\boldsymbol{\alpha}}' \mathbf{B}_F' \boldsymbol{M}_N \mathbf{B}_W \hat{\boldsymbol{\phi}}.$$

To derive the bias correction for $\hat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$, 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 $\hat{\boldsymbol{\phi}}$, we have

$$\begin{aligned}\hat{\boldsymbol{\phi}} &= \boldsymbol{\phi} - (\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) + (\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\boldsymbol{u} \\ &\approx \boldsymbol{\phi} - (\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}.\end{aligned}$$

Thus, the leading bias term in $\hat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$ is given by

$$\begin{aligned}\mathbb{E}[\hat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}] - \mathrm{C}_{\boldsymbol{\phi},\boldsymbol{\alpha}} &\approx -\frac{1}{N}\mathbb{E}[\boldsymbol{\tau}'\mathbf{F}'\mathbf{B}\mathbf{L}^*\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W(\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}] \\ &= -\frac{\mathbb{E}[\tau(U_i)^2]}{N}\mathrm{tr}\left(\mathbf{L}^*\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W(\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\right).\end{aligned}$$

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 $\hat{\sigma}_\tau^2$:

$$\mathrm{C}^{bc}_{\boldsymbol{\phi},\boldsymbol{\alpha}} = \hat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}} + \frac{\hat{\sigma}_\tau^2}{N}\mathrm{tr}\left(\mathbf{L}^*\mathbf{B}_F'\boldsymbol{M}_N\mathbf{B}_W(\mathbf{B}_W'\mathbf{B}_W)^{-1}\mathbf{B}_W'\mathbf{B}_F\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\right).$$

Using the same dataset as in 6, we compare the plug-in estimator and the bias-corrected estimator for the covariance between worker and firm effects. Table 10 reports the results. We find that the bias induced by the dependence is about 30% of the plug-in estimator.

TABLE 10. Variance Component Estimation on the VWH Data

$\hat{\mathbf{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$	$\hat{\mathbf{C}}^{bc}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$
0.014	0.019

Note: This table reports the estimated covariance between worker and firm effects via the plug-in estimator $\hat{\mathrm{C}}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$ and the bias-corrected estimator $\hat{\mathrm{C}}^{bc}_{\boldsymbol{\phi},\boldsymbol{\alpha}}$.

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.

APPENDIX E. 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 5.

E.1. **Inference with non-normal errors.** In Section 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 11. As in the normal case, the coverage probabilities of the confidence intervals based on (3.3) 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.

TABLE 11. Coverage Probability of Confidence Intervals

	α_1	d_1	$(d_1^+ - d_1^-)/d_1$	$\lambda_{n,F}/(d_1\lambda_{2,L}h_1)$	**95% cov**	**Normal 95% cov**
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 (3.3). The last column reports the coverage probability of the confidence intervals based on Jochmans and Weidner (2019)'s asymptotic normality.

E.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 (3.4) 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 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 12 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 12. 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.

APPENDIX F. PROOFS

F.1. **Proof of Theorem 1.**

Proof. We bound each term in the right-hand side of (3.1):

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

We first show the unbiasedness of the fixed-effect estimator:

Lemma 1. *Under the stated assumptions, we have*

$$\mathbb{E}[\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}] = 0.$$

Proof. Observe that

$$\hat{\boldsymbol{\alpha}} = \mathbf{L}^*\mathbf{L}\boldsymbol{\alpha} + \mathbf{L}^*\mathbf{B}'\boldsymbol{\epsilon}.$$

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, 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 2. *Under the stated assumptions, $C\lambda_{n,F} \geq \tilde{\sigma}_n^2$ for some absolute constant $C > 0$.*

Proof. Note that $\mathbf{F}\mathbf{F}'$ is a non-negative, positive-semidefinite matrix and each block corresponding to edges in $E_{(i,j)}$, $(\mathbf{F}\mathbf{F}')_{(i,j)}$ is a principal submatrix of $\mathbf{F}\mathbf{F}'$. Thus, by Corollary 8.1.20 in Horn and Johnson (2012), we have

$$\lambda_{n,F} \geq \lambda_{\max}((\mathbf{F}\mathbf{F}')_{(i,j)}),$$

for each (i,j) and $\lambda_{\max}(\cdot)$ is the largest eigenvalue of a matrix. Notice that

$$(\mathbf{F}\mathbf{F}')_{(i,j)} = 2 \times \boldsymbol{\iota}_{|E_{(i,j)}|}\boldsymbol{\iota}'_{|E_{(i,j)}|}$$

so that $\lambda_{\max}((\mathbf{F}\mathbf{F}')_{(i,j)}) = 2 \times |E_{(i,j)}|$. Therefore,

$$\lambda_{n,F} \geq 2 \max_{(i,j)} |E_{(i,j)}|.$$

By Assumption 1, we have

$$\begin{aligned}\tilde{\sigma}_n^2 &= \max_{(i,j)} \mathbf{A}_{i,j} \times [\max_{e,e'} \in E_{(i,j)}\sigma_{e,e'} - \mathbb{E}[\tau^s(U_i)^2 + \tau^t(U_i)^2]] \\ &\leq C \max_{(i,j)} \mathbf{A}_{i,j} = C \max_{(i,j)} |E_{(i,j)}|,\end{aligned}$$

for some absolute constant $C > 0$. Thus, we have

$$\tilde{\sigma}_n^2 \leq C/2 \times 2 \max_{(i,j)} |E_{(i,j)}| \leq C/2 \times \lambda_{n,F}.$$

This proves the lemma. □

Next, we bound the variance of the remaining term $\boldsymbol{r} = \mathbf{D}^{-1}\mathbf{A}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$. Observe that

$$Var(\hat{\boldsymbol{\alpha}}) \leq C\lambda_{n,F}\mathbf{L}^*$$

for some constant $C > 0$ by Lemma 2. Thus, we have

$$\begin{aligned} Var(\boldsymbol{r}) &= \mathbf{D}^{-1}\mathbf{A}Var(\hat{\boldsymbol{\alpha}})\mathbf{A}'\mathbf{D}^{-1} \\ &\leq C\lambda_{n,F}\mathbf{D}^{-1}\mathbf{A}\mathbf{L}^*\mathbf{A}\mathbf{D}^{-1} \\ &\leq \frac{C\lambda_{n,F}}{\lambda_{2,L}} \times \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}, \end{aligned}$$

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 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:

$$\begin{aligned} Var(\boldsymbol{o}) &= \mathbf{D}^{-1}\mathbf{B}'\boldsymbol{\Omega}_2\mathbf{B}\mathbf{D}^{-1} \\ &\leq \tilde{\sigma}_n^2 \times \mathbf{D}^{-1}\mathbf{B}'\mathbf{B}\mathbf{D}^{-1} \end{aligned}$$

Thus, for each element, we have

$$\begin{aligned} Var(o_i) = e_i' Var(\boldsymbol{o}) e_i &= \tilde{\sigma}_n^2 \times e_i'\mathbf{D}^{-1}\mathbf{B}'\mathbf{B}\mathbf{D}^{-1}e_i \\ &\leq \frac{C \max_{(i,j)} |E_{(i,j)}|}{d_i}. \end{aligned}$$

for some constant $C > 0$ under Assumption 1. 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

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

as $d_i \to \infty$, which completes the proof. □

F.2. **Proof of Theorem 2.**

Proof. Write $\hat{\tau}_i = \tau(U_i) + r_i$ where $r_i = O_p(\sqrt{1/d_i})$ by Proposition 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

$$\sup_{x\in\mathbb{R}} |\hat{F}_{n,\tau}(x) - F_\tau(x)| \le \sup_{x\in\mathbb{R}} |\hat{F}_{n,\tau}(x) - F_{n,\tau}(x)| + \sup_{x\in\mathbb{R}} |F_{n,\tau}(x) - F_\tau(x)| \tag{F.1}$$

The second term in (F.1) 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 (F.1), note that for any $\delta > 0$,

$$\mathbb{I}\{\hat{\tau}_i \le x\} = \mathbb{I}\{\tau(U_i) + r_i \le x\} \le \mathbb{I}\{\tau(U_i) \le x + \delta\} + \mathbb{I}\{|r_i| > \delta\},$$

for each $i \in \mathcal{C}_n$ and $x \in \mathbb{R}$. Since F_τ is non-decreasing, we have

$$\sup_{x\in\mathbb{R}} |\hat{F}_{n,\tau}(x) - F_{n,\tau}(x)| \leq \frac{1}{|\mathcal{C}_n|} \sum_{i\in\mathcal{C}_n} \mathbb{I}\{|r_i| > \delta\} + \sup_{x\in\mathbb{R}} |F_{n,\tau}(x+\delta) - F_{n,\tau}(x-\delta)| \quad \text{(F.2)}$$

The second term in (F.2) is bounded by

$$\sup_{x\in\mathbb{R}} |F_{n,\tau}(x+\delta) - F_{n,\tau}(x-\delta)| \leq \sup_{x\in\mathbb{R}} |F_\tau(x+\delta) - F_\tau(x-\delta)| + 2\sup_{x\in\mathbb{R}} |F_{n,\tau}(x) - F_\tau(x)|.$$

Since F_τ is continuous, the first term is arbitrarily small for sufficiently small δ. For the first term in (F.2), 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

$$\mathbb{P}\left(\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n} \mathbb{I}\{|r_i| > \delta\} > \epsilon\right) \leq \frac{1}{\epsilon|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n} \mathbb{E}[\mathbb{I}\{|r_i| > \delta\}] = O\left(\frac{\sqrt{\eta_n}}{\epsilon\delta}\right) = o(1),$$

as ϵ and δ are arbitrary and $\eta_n \to 0$ by the assumption.

Combining the above, we have,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} |\hat{F}_{n,\tau}(x) - F_\tau(x)| \\
&\leq 3\sup_{x\in\mathbb{R}} |F_{n,\tau}(x) - F_\tau(x)| + \sup_{x\in\mathbb{R}} |F_\tau(x+\delta) - F_\tau(x-\delta)| + \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n} \mathbb{I}\{|r_i| > \delta\} \\
&= o_p(1),
\end{aligned}$$

as $n \to \infty$. This completes the proof. □

F.3. **Proof of Theorem 3.**

Proof. First, we show the consistency of $\hat{\sigma}_\tau^2$:

Lemma 3. *Under the stated assumptions, we have*

$$\hat{\sigma}_\tau^2 \to_p \sigma_\tau^2.$$

Proof. Observe that

$$\hat{\tau}_i = \tau(U_i) + \underbrace{c_i^{-1}\left(\frac{1}{d_i}\sum_{e\in E_i}\hat{\epsilon}_e - c_i\tau(U_i)\right)}_{\equiv\tilde{u}_i}.$$

Then,

$$\hat{\sigma}_\tau^2 = \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tau^2(U_i) \tag{F.3}$$

$$+ \frac{2}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tau(U_i)\tilde{u}_i \tag{F.4}$$

$$+ \frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tilde{u}_i^2. \tag{F.5}$$

The first term (F.3) 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 (F.5) converges to zero in probability. To see this, from the proof of Proposition 2, observe that

$$\mathbb{E}[\tilde{u}_i^2] = O\left(\frac{1}{d_i}\right),$$

since $c_i > c$. Thus, we have

$$\mathbb{E}[(\text{F.5})] = O(\eta_n) = o(1),$$

under the hypothesis that $\eta_n \to 0$ as $n \to \infty$. Then, by Markov's inequality, (F.5) converges to zero in probability.

The second term (F.4) converges to zero in probability by the Cauchy-Schwarz inequality:

$$(\text{F.4}) \leq \sqrt{\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tau^2(U_i)}\sqrt{\frac{1}{|\mathcal{C}_n|}\sum_{i\in\mathcal{C}_n}\tilde{u}_i^2} = O_p(\sqrt{\eta_n}) = o_p(1),$$

as shown above. This completes the proof of Lemma 3. □

Observe that

$$\hat{V}_{\boldsymbol{\alpha}}^{bc} - V_{\boldsymbol{\alpha}} = 2\frac{\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}}{n} \tag{F.6}$$

$$+ 2\frac{(\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n} \tag{F.7}$$

$$+ \frac{(\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'(\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})}{n} \tag{F.8}$$

$$+ \frac{(\mathbb{E}[\tau^2(U_i)] - \hat{\sigma}_\tau^2) \times \mathrm{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} \tag{F.9}$$

$$+ \frac{\boldsymbol{\tau}'\mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n} - \frac{\mathbb{E}[\tau^2(U_i)] \times \mathrm{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} \tag{F.10}$$

The first term (F.6) is mean-zero and its variance is proportional to

$$\begin{aligned}
Var\left(\frac{\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}}{n}\right) &= \frac{1}{n^2}\mathbb{E}[\boldsymbol{\epsilon}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\boldsymbol{\epsilon}] \\
&= \frac{1}{n^2}\mathrm{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']) \\
&\leq \frac{C\lambda_{n,F}}{n^2} \times \boldsymbol{\alpha}'\mathbf{L}^*\boldsymbol{\alpha} \\
&\leq \frac{C\lambda_{n,F}}{n} \times \frac{1}{n\lambda_{2,L}}\sum_{i=1}^n \frac{\alpha_i^2}{d_i} \\
&\leq \frac{C \times C_\alpha \times \lambda_{n,F}}{n} \times \frac{1}{n\lambda_{2,L}}\sum_{i=1}^n \frac{1}{d_i} \\
&= o(1),
\end{aligned}$$

where the first inequality follows from Lemma 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,

$$(\text{F.6}) = o_p(1),$$

by Chebyshev's inequality.

The second term (F.7) is mean-zero and let $\boldsymbol{Q} \equiv \mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'$. Then, we can write

$$(\text{F.7}) = 2\frac{(\boldsymbol{\epsilon} - \mathbf{F}\boldsymbol{\tau})'\boldsymbol{Q}\mathbf{F}\boldsymbol{\tau}}{n}$$

and its variance is given by

$$\begin{aligned}
Var((\text{F.7})) &= \frac{4}{n^2}\mathbb{E}[(\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})'\boldsymbol{Q}\mathbf{F}\boldsymbol{\tau}(\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})'\boldsymbol{Q}\mathbf{F}\boldsymbol{\tau}] \\
&= \frac{4\mathbb{E}[\tau^2(U_i)]}{n^2}\mathrm{tr}(\boldsymbol{Q}\mathbf{F}\mathbf{F}'\boldsymbol{Q}'\boldsymbol{\Omega}_2) \\
&\quad + \frac{4}{n^2}\sum_{e_1,e_2,e_3,e_4\in E}\boldsymbol{Q}_{e_1,e_2}\boldsymbol{Q}_{e_3,e_4}\mathrm{cum}((\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})_{e_1},(\boldsymbol{\epsilon}-\mathbf{F}\boldsymbol{\tau})_{e_2},(\mathbf{F}\boldsymbol{\tau})_{e_3},(\mathbf{F}\boldsymbol{\tau})_{e_4}),
\end{aligned}$$

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

$$\begin{aligned}
\frac{\mathbb{E}[\tau^2(U_i)]}{n^2}\mathrm{tr}(\boldsymbol{Q}\mathbf{F}\mathbf{F}'\boldsymbol{Q}'\boldsymbol{\Omega}_2) &\le \frac{\mathbb{E}[\tau^2(U_i)]}{n^2}\times\tilde{\sigma}_n^2\times\lambda_{n,F}\times\mathrm{tr}(\mathbf{L}^*)^2 \\
&\le \mathbb{E}[\tau^2(U_i)]\times\tilde{\sigma}_n^2\times\lambda_{n,F}\times\left(\frac{1}{n}\sum_{i\in V}d_i^{-1}+\frac{1}{n\lambda_{2,L}}\sum_{i\in V}d_i^{-1}h_i^{-1}\right)^2 \\
&= o(1),
\end{aligned}$$

where the first inequality follows from $\boldsymbol{\Omega}_2 \le \tilde{\sigma}_n\mathbf{I}_m$, $\mathbf{F}\mathbf{F}' \le \lambda_n^2\mathbf{I}_m$,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 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

$$Var((\text{F.7})) = o(1),$$

and by Chebyshev's inequality, we have (F.7) $\to_p 0$.

The third term (F.8)'s mean is given by

$$\begin{aligned}
\mathbb{E}[(\text{F.8})] &= \frac{\text{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\boldsymbol{\Omega}_2)}{n} \\
&\leq \frac{\tilde{\sigma}_n^2 \times \text{tr}(\mathbf{L}^*)}{n} \\
&\leq \frac{\tilde{\sigma}_n^2}{n} \times \left(\sum_{i \in V} \frac{1}{d_i} + \sum_{i \in V} \frac{1}{\lambda_{2,L} h_i d_i} \right) \\
&= o(1),
\end{aligned}$$

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 (F.8) is non-negative ($\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}$ is positive semidefinite), this implies that (F.8) $\to_p 0$.

The fourth term (F.9) converges to zero in probability by Lemma (3) and the fact that

$$\begin{aligned}
\frac{\text{tr}(\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\mathbf{F}')}{n} &\leq \frac{C\lambda_{n,F}}{n} \times \text{tr}(\mathbf{L}^*) \\
&\leq \frac{C\lambda_{n,F}}{n} \times \left(\sum_{i \in V} \frac{1}{d_i} + \sum_{i \in V} \frac{1}{\lambda_{2,L} h_i d_i} \right) \\
&= O(1),
\end{aligned}$$

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 (F.10) 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

$$\begin{aligned}
&Var\left(\frac{\boldsymbol{\tau}'\mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n} \right) \\
&= \frac{\mathbb{E}[\tau^2(U_i)]}{n^2} \times \text{tr}(\boldsymbol{Q}_{\mathbf{F}}^2) + \frac{(\mathbb{E}[\tau^4(U_i)] - 3\mathbb{E}[\tau^2(U_i)]^2)}{n^2} \times \sum_{i \in V} \boldsymbol{Q}_{\mathbf{F},i,i}^2 \\
&\leq \frac{\mathbb{E}[\tau^4(U_i)]}{n^2} \times \text{tr}(\boldsymbol{Q}_{\mathbf{F}}^2),
\end{aligned}$$

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 $\mathbf{B}\mathbf{L}^* = \mathbf{B}\mathbf{D}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{A}\mathbf{L}^*$ 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 $\mathbf{F}\mathbf{F}' \leq \lambda_{n,F}\mathbf{I}_m$, and the last inequality follows from $\mathrm{tr}(\cdot^2) \leq \mathrm{tr}(\cdot)^2$ applied to $\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1} = (\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1/2})(\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1/2})'$ which is positive semidefinite. For the first term on the far right-hand side, we have

$$\begin{aligned}
&\mathrm{tr}(\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{B}'\mathbf{F}\mathbf{F}'\mathbf{B}\mathbf{D}^{-1}) \\
&= \sum_{i\in V}\left(\frac{(d_i^t - d_i^s)^2 + \sum_{j\neq i} r_{i,j}^2}{d_i^2}\right)^2 + \sum_{i\neq j}\left(\frac{\sum_{k\in V} r_{i,k}r_{j,k}}{d_i d_j}\right)^2 \\
&\leq Cn(1 + \sum_{i\in V} d_i^{-2})
\end{aligned}$$

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,

$$\begin{aligned}
&Var\left(\frac{\boldsymbol{\tau}'\mathbf{F}'\mathbf{B}\mathbf{L}^*\boldsymbol{M}_n\mathbf{L}^*\mathbf{B}'\mathbf{F}\boldsymbol{\tau}}{n}\right) \\
&\leq C\left(\frac{1}{n} + \frac{1}{n}\sum_{i\in V} d_i^{-2}\right) + C\left(\frac{\lambda_{n,F}}{\lambda_{2,L}}\frac{1}{n}\sum_{i\in V}\frac{1}{d_i h_i}\right)^2 \\
&= o(1),
\end{aligned}$$

under the stated assumptions. Thus, by Chebyshev's inequality, we have (F.10) $\to_p 0$.

Finally, we combine all the results above to conclude that

$$\hat{V}_{\boldsymbol{\alpha}}^{bc} - V_{\boldsymbol{\alpha}} \to_p 0.$$

This completes the proof. □

F.4. **Proof of Proposition 1.**

Proof. First, observe that

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

where the first inequality holds by $\mathbf{F} = \mathbf{F}^t + \mathbf{F}^s$. Then, the first term in the variance formula (2.7) is bounded by:

$$\mathbf{L}^* \mathbf{B}' \boldsymbol{\Omega}_1 \mathbf{B} \mathbf{L}^* \leq \sigma_\tau^2 \lambda_{n,F} \mathbf{L}^*,$$

where we used the fact that $\mathbf{L}^*$ is a pseudo-inverse of $\mathbf{B}'\mathbf{B}$.

Next, since $\boldsymbol{\Omega}_2 \leq \tilde{\sigma}_n^2 \mathbf{I}_m$, the second term in the variance formula (2.7) is bounded by:

$$\mathbf{L}^* \mathbf{B}' \boldsymbol{\Omega}_2 \mathbf{B} \mathbf{L}^* \leq \tilde{\sigma}_n^2 \mathbf{L}^*.$$

Therefore, the variance of the fixed effect is bounded by:

$$Var(\hat{\boldsymbol{\alpha}}) \leq (\tilde{\sigma}_n^2 + \sigma_\tau^2 \lambda_{n,F}) \times \mathbf{L}^*.$$

Thus, for a unit vector e_i with 1 at the i-th position and 0 elsewhere, we have:

$$Var(\hat{\alpha}_i) = e_i' Var(\hat{\boldsymbol{\alpha}}) e_i \leq (\tilde{\sigma}_n^2 + \sigma_\tau^2 \lambda_{n,F}) \times e_i' \mathbf{L}^* e_i.$$

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,

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

which completes the proof. □

F.5. **Proof of Proposition 2.**

Proof. Observe that from the transformed model (B.1), we can write the residuals as

$$\begin{aligned}\hat{\boldsymbol{\epsilon}} &= \mathbf{y} - \mathbf{B}\hat{\boldsymbol{\alpha}} = (\mathbf{I}_m - \mathbf{B}(\mathbf{B}'\mathbf{B})^*\mathbf{B}')\mathbf{y} = \boldsymbol{M}_{\mathbf{B}}\mathbf{y} \\ &= \boldsymbol{M}_{\mathbf{B}}\tilde{\mathbf{F}}(\tilde{\boldsymbol{\tau}}^t/2 + \tilde{\boldsymbol{\tau}}^s/2) + \boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{\epsilon}}.\end{aligned}$$

Recall that the statistic of interest minus $c_i(\tau^t(U_i) + \tau^s(U_i))/2$ can be written as

$$\begin{aligned}&\frac{1}{d_i}\sum_{e\in E_i}\hat{\epsilon}_e - c_i\frac{\tau^t(U_i) + \tau^s(U_i)}{2} \\ &= \frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{\epsilon}} \qquad \text{(F.11)} \\ &+ \frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\mathbf{F}}(\tilde{\boldsymbol{\tau}}^t/2 + \tilde{\boldsymbol{\tau}}^s/2) - c_i\frac{\tau^t(U_i) + \tau^s(U_i)}{2} \qquad \text{(F.12)}\end{aligned}$$

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 (F.11) converges to zero in probability. By the definition of $\tilde{\boldsymbol{\epsilon}}$, this term has mean zero. Its variance is given by

$$\begin{aligned}\text{Var}\left(\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{\epsilon}}\right) &= \frac{1}{d_i^2}\mathbb{E}\left[1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}'\boldsymbol{M}_{\mathbf{B}}1_{i,m}\right] \\ &\leq \frac{\tilde{\sigma}_n^2}{d_i},\end{aligned}$$

where the inequality follows from $\|\boldsymbol{M}_{\mathbf{B}}\| \leq 1$ (since $\boldsymbol{M}_{\mathbf{B}}$ is idempotent), $\mathbb{E}[\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}'] \leq \tilde{\sigma}_n^2\mathbf{I}_m$, and $1'_{i,m}1_{i,m} = d_i$. Since $d_i \to \infty$, the variance converges to zero. Thus, by Chebyshev's inequality,

$$\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{\epsilon}} = O_p\left(\frac{1}{\sqrt{d_i}}\right) = o_p(1),$$

as $d_i \to \infty$.

Second, we show that (F.12) 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

$$\mathbb{E}\left[\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\mathbf{F}}(\tilde{\boldsymbol{\tau}}^t/2 + \tilde{\boldsymbol{\tau}}^s/2) - c_i\frac{\tau^t(U_i) + \tau^s(U_i)}{2}\middle| U_i\right] = 0$$

where $\tilde{\boldsymbol{f}}_i$ is the i-th column of $\tilde{\mathbf{F}}$, and $c_i = \frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{f}}_i$. Note $c_i \in [0, 1]$ since $\boldsymbol{M}_{\mathbf{B}}$ is a projection matrix and $1'_{i,m}\tilde{\boldsymbol{f}}_i = d_i$.

Since

$$\begin{aligned}&\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\mathbf{F}}(\tilde{\boldsymbol{\tau}}^t/2+\tilde{\boldsymbol{\tau}}^s/2)\\&=\frac{1}{2d_i}\sum_{j\in V^{st}}(\tau^t(U_j)+\tau^s(U_j))\cdot 1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{f}}_j,\end{aligned}$$

the conditional variance given U_i is

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\mathbf{F}}(\tilde{\boldsymbol{\tau}}^t/2+\tilde{\boldsymbol{\tau}}^s/2)\middle|U_i\right]&=\frac{\mathbb{E}[(\tau^t(U_i)+\tau^s(U_i))^2]}{4d_i^2}\sum_{j\neq i\in V^{st}}\left(1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\boldsymbol{f}}_j\right)^2\\&\leq\frac{\mathbb{E}[(\tau^t(U_i)+\tau^s(U_i))^2]}{4d_i^2}\sum_{j\in V^{st}}|E_{(i,j)}|^2\\&\leq\frac{\max_{j\in V^{st}}|E_{(i,j)}|^2\cdot\mathbb{E}[(\tau^t(U_i)+\tau^s(U_i))^2]}{4d_i}\\&=O\left(\frac{1}{d_i}\right)=o(1),\end{aligned}$$

as $d_i\to\infty$, where the first inequality uses $\|\boldsymbol{M}_{\mathbf{B}}\|\leq 1$ and $1'_{i,m}\tilde{\boldsymbol{f}}_j=|E_{(i,j)}|$, and the last equality follows by assumption. By the law of total variance, it follows that

$$\mathrm{Var}\left[\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\mathbf{F}}(\tilde{\boldsymbol{\tau}}^t/2+\tilde{\boldsymbol{\tau}}^s/2)-c_i\frac{\tau^t(U_i)+\tau^s(U_i)}{2}\right]=O\left(\frac{1}{d_i}\right)=o(1),$$

as $d_i\to\infty$. Thus, by Chebyshev's inequality,

$$\frac{1}{d_i}1'_{i,m}\boldsymbol{M}_{\mathbf{B}}\tilde{\mathbf{F}}(\tilde{\boldsymbol{\tau}}^t/2+\tilde{\boldsymbol{\tau}}^s/2)-c_i\frac{\tau^t(U_i)+\tau^s(U_i)}{2}=O_p\left(\frac{1}{\sqrt{d_i}}\right)=o_p(1),$$

as $d_i\to\infty$.

Finally, combining the results from the two steps, we have

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

as $d_i\to\infty$ by the continuous mapping theorem. This completes the proof. □

F.6. **Proof of Proposition 3.**

Proof. **Validity of the confidence interval**: First, we show that $\hat{c}_\alpha\to_p c_\alpha$, 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 $\hat{F}_{n,\tau}$ to F_τ in Theorem 2: Since $\hat{F}_{n,\tau}$ weakly converges to F_τ in probability (Problem 23.1 in van der Vaart, 1998), $\hat{F}^{-1}_{n,\tau}(\cdot)=\inf\{x\in\mathbb{R}:\hat{F}_{n,\tau}(x)\geq\cdot\}$ 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, $F_\tau^{-1}(\alpha) = c_\alpha$ and $\hat{c}_\alpha = \hat{F}_{n,\tau}^{-1}(\alpha) \to_p c_\alpha$.

Next, observe that by Theorem 1 and $((d_i^t - d_i^s)/d_i)^{-1} = O(1)$,

$$\tilde{T} \equiv \left(\frac{d_i^t - d_i^s}{d_i}\right)^{-1} (\hat{\alpha}_i - \alpha_i) \to_p \tau(U_i),$$

as $d_i \to \infty$. Thus, $(\tilde{T}, \hat{c}_{\alpha/2}, \hat{c}_{1-\alpha/2})$ weakly converges to $(\tau(U_i), c_{\alpha/2}, c_{1-\alpha/2})$. By Slutsky's theorem, we have

$$\mathbb{P}(\hat{c}_{\alpha/2} \leq \tilde{T} \leq \hat{c}_{1-\alpha/2}) \to \mathbb{P}(c_{\alpha/2} \leq \tau(U_i) \leq c_{1-\alpha/2}) = F_\tau(c_{1-\alpha/2}) - F_\tau(c_{\alpha/2}) = 1 - \alpha.$$

This shows the asymptotic validity of the confidence interval.

Validity of the test: Observe that under the null hypothesis, by Theorem 1 and the continuous mapping theorem, we have

$$\begin{aligned} T &= \underbrace{((\tau(U_i))_{i\in V_0})' \boldsymbol{C}_n ((\tau(U_i))_{i\in V_0})}_{\equiv g_n((\tau(U_i)_{i\in V_0}))\equiv \tilde{T}} + o_p(1) \\ &\to_d \underbrace{((\tau(U_i))_{i\in V_0})' \boldsymbol{C} ((\tau(U_i))_{i\in V_0})}_{\equiv g((\tau(U_i)_{i\in V_0}))\equiv T_\infty}, \end{aligned}$$

where $\boldsymbol{C} = \mathbf{D}_{V_0}^{st} \boldsymbol{M}_{n_0} \mathbf{D}_{V_0}^{st}$ and $\mathbf{D}_{V_0}^{st} = \lim_{n\to\infty} \mathbf{D}_{n,V_0}^{st}$. Note that the cdf of T_∞ is continuous. Let $\tau_i^{(m)}$ be m-th random draw from $\hat{F}_{n,\tau}$ (with replacement) for $m = 1, ..., M$. Then, the simulated test statistic is given by

$$\tilde{T}^{(m)} = \underbrace{(\tau_i^{(m)})_{i\in V_0})' \boldsymbol{C}_n ((\tau_i^{(m)})_{i\in V_0})}_{\equiv g_n((\tau_i^{(m)})_{i\in V_0})},$$

$\boldsymbol{C}_n = \mathbf{D}_{n,V_0}^{st} \boldsymbol{M}_{n_0} \mathbf{D}_{n,V_0}^{st}$ and $\mathbf{D}_{n,V_0}^{st} = diag((d_i^t - d_i^s)/d_i, i \in V_0)$. Note that $\tilde{T}^{(m)}, m = 1, ..., M$ are independent and identically distributed conditional on $\hat{F}_{n,\tau}$.

Let $\hat{F}_{n,T}^{\otimes n_0}$ be the distribution of $(\tau_i^{(1)})_{i\in V_0}$ given $\hat{F}_{n,\tau}$ and $F_\tau^{\otimes n_0}$ be the distribution of $(\tau(U_i))_{i\in V_0}$. Since n_0 is fixed, by the weak convergence of $\hat{F}_{n,\tau}$ to F_τ in probability, $\hat{F}_{n,T}^{\otimes n_0}$ weakly converges to $F_\tau^{\otimes n_0}$ 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), $\hat{G}_{n,T}$ weakly converges to G_T in probability, where $\hat{G}_{n,T}$ is the distribution of $\tilde{T}^{(1)}$ given $\hat{F}_{n,\tau}$ 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 $\hat{c}_\alpha \to_p c_\alpha$, where $\hat{c}_\alpha$ is the α-quantile of $\hat{G}_{n,T}$ and c_α is the α-quantile of G_T. Thus, $(T, \hat{c}_\alpha)$ weakly converges to

(T_∞, c_α). By Slutsky's theorem, we have

$$\mathbb{P}(T > \hat{c}_{1-\alpha}) \to \mathbb{P}(T_\infty > c_{1-\alpha}) = 1 - G_T(c_{1-\alpha}) = \alpha,$$

which shows the asymptotic validity of the test with the known $\hat{G}_{n,T}$.

Finally, we show that the test remains valid when $\hat{G}_{n,T}$ is estimated by the empirical distribution of $\tilde{T}^{(m)}, m = 1, ..., M$. Let $\hat{G}_{n,T,M}$ be the empirical distribution of $\tilde{T}^{(m)}, m = 1, ..., M$ given $\hat{F}_{n,\tau}$. Let $\hat{c}_{\alpha,M}$ be the α-quantile of $\hat{G}_{n,T,M}$. Observe that, for any $\epsilon > 0$,

$$\begin{aligned}
&|\mathbb{P}(T > \hat{c}_{1-\alpha,M}) - \mathbb{P}(T > c_{1-\alpha})| \\
&= \left|\mathbb{P}(\hat{G}_{n,T,M}(T) > 1-\alpha) - \mathbb{P}(G_T(T) > 1-\alpha)\right| \\
&\leq \mathbb{P}\left(|G_T(T) - 1 + \alpha| \leq \epsilon\right) + \mathbb{P}\left(|G_T(T) - \hat{G}_{n,T,M}(T)| > \epsilon\right) \\
&\leq \mathbb{P}\left(|G_T(T) - 1 + \alpha| \leq \epsilon\right) + \mathbb{P}\left(\sup_{x\in\mathbb{R}} |G_T(x) - \hat{G}_{n,T,M}(x)| > \epsilon\right),
\end{aligned}$$

where the first equality follows from $\hat{G}_{n,T,M}$ and G_T being right-continuous, and the first inequality follows from the fact that $|\mathbb{I}(x \leq a) - \mathbb{I}(x \leq b)| \leq \mathbb{I}(|x - a| \leq \epsilon) + \mathbb{I}(|a - b| > \epsilon)$ 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 $\mathbb{P}(|G_T(T) - 1 + \alpha| \leq \epsilon)$, which can be made arbitrarily small by choosing ϵ small enough. For the second term on the right-hand side, observe that

$$\sup_{x\in\mathbb{R}} |G_T(x) - \hat{G}_{n,T,M}(x)| \leq \sup_{x\in\mathbb{R}} |G_T(x) - \hat{G}_{n,T}(x)| + \sup_{x\in\mathbb{R}} |\hat{G}_{n,T}(x) - \hat{G}_{n,T,M}(x)|.$$

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 $\hat{F}_{n,\tau}$. Thus,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} |G_T(x) - \hat{G}_{n,T,M}(x)| > \epsilon\right) \to 0,$$

as $n, M \to \infty$. Since $\epsilon > 0$ is arbitrary, we have

$$\mathbb{P}(T > \hat{c}_{1-\alpha,M}) \to \mathbb{P}(T > c_{1-\alpha}) = \alpha,$$

as $n, M \to \infty$. This completes the proof. □