Inference in Linear Dyadic Data Models with Network Spillovers

Nathan Canen* **and Ko Sugiura**†

Department of Economics, University of Houston, Science Building 3581 Cullen Boulevard Suite 230, Houston, 77204-5019, Texas, United States. Tel.: (713) 743-3800, Fax: (713) 743-3798

Abstract

When using dyadic data (i.e., data indexed by pairs of units, such as trade flow data between two countries), researchers typically assume a linear model, estimate it using Ordinary Least Squares and conduct inference using "dyadic-robust" variance estimators. The latter assumes that dyads are uncorrelated if they do not share a common unit (e.g., if one country does not appear in both pairs of trade flow data). We show that this assumption does not hold in many empirical applications because indirect links may exist due to network connections, e.g., different country-pairs may have correlated trade outcomes due to sharing common trading partner links. Hence, as we show in Monte Carlo simulations, "dyadic-robust"' estimators can be severely biased in such situations. We develop a consistent variance estimator appropriate for such contexts by leveraging results in network econometrics. Our estimator has good finite sample properties in numerical simulations. We then illustrate our message with an application to voting behavior by seating neighbors in the European Parliament.

Key words: Dyadic data, Networks, Inference, Cross-sectional dependence, Congressional Voting

---

*E-mail: ncanen@uh.edu (Corresponding author)

†E-mail: ksugiura@uh.edu

1 Introduction

Dyadic data is categorized by the dependence between two sets of sampled units (dyads). For example, exports between the U.S. and Canada depend on both countries (and, plausibly, their characteristics), while arrangements between a seller and a buyer show similar dependence. This contrasts to classical data in the social sciences that only depends on a single unit of observation (e.g., the GDP of the U.S.).

The empirical relevance of dyadic data is showcased by its widespread use, which has increased over the past two decades (Graham (2020b) provides an extensive review). For example, applications are found in international trade (export-import outcomes across countries, Anderson and van Wincoop (2003)), macroeconomics (exchange rate determination, Lustig and Richmond (2020)), development (gift giving and insurance between two members in a village, Fafchamps and Gubert (2007)), political economy (correlation in voting behavior in Parliament across seating neighbors, Harmon, Fisman, and Kamenica (2019)), among many others. In all of these examples, applied researchers model the dependence between dyadic outcomes and observable characteristics using a linear model which they then estimate using Ordinary Least Squares (OLS). However, inference on such estimators for the linear parameters is more complex.

The main approach in applied work for the latter has been the use of the so-called "dyadic-robust" estimators (e.g., Cameron, Gelbach, and Miller (2011), Aronow, Samii, and Assenova (2015), Tabord-Meehan (2019), among others). This builds on the widely used assumption in dyadic data that the error term for dyad (i, j) and for dyad (j, k) can only be correlated if they share a unit (see Aronow et al. (2015), Tabord-Meehan (2019) for a discussion, and Cameron and Miller (2014) for a review). In the context of international trade, this implies that Canada-U.S. and France-U.S. trade may be correlated because the U.S. is present in both, but there is no correlation between Canada-U.S. exports and France-Germany ones (conditional on observable characteristics).

In this paper, we first show that such an assumption does not hold in many models using dyadic data where dyads may be indirectly connected along a network.[1] Figure 1 (A) presents a simple example in the context of international trade. That is, it is completely possible that Canada-US and France-Germany trades might be correlated along unobservables because they have many indirect

[1]This is a concrete class of applied examples where the assumption fails. The possibility that cross-sectional dependence in dyadic data might be more extensive than assumed has been pointed out by Cameron and Miller (2014).

connections (in the figure, through Canada trading with China which then trades with France).[2] This implies that the use of dyadic-robust variance estimators may be inappropriate for inference in linear models with dyadic data when spillovers exist along a network.

We make this point in two ways. First, we prove that existing “dyadic-robust" variance estimators are biased for the true asymptotic variance when there are network spillovers (i.e., when dyads may be correlated along a network even when they do not share a common unit). This result is verified in Monte Carlo simulations with standard set-ups, where we show that such variance estimators lead to a bias of approximately 14% and its associated confidence interval undercovers the true parameter by over 9 percentage points.[3] This performance worsens with the density of the underlying network (e.g., in political economy and international trade, networks are often dense, see Cameron and Miller (2014)) and with sample size, as these features increase the role of indirect spillovers. This issue may be prevalent in many applied examples: in the case of international trade (Cameron et al. (2011) and Cameron and Miller (2014)), dyads may be correlated even if they do not share a member country because there are many indirect connections among them. In political economy, those seating far apart from one another in Parliament may still have correlated behavior through the indirect links of whom they are seating next to sits with. In Macroeconomics, exchange rates may depend on other countries beyond the dyads.

To deal with these issues, we develop a consistent variance estimator that can account for such network spillovers even with dyadic data. This generalizes existing results for dyadic-robust variance estimators (e.g., Aronow et al. (2015)) to explicitly account for the network structure of spillovers. To do so, we leverage recent results in inference in such cross-sectional data, such as Kojevnikov, Marmer, and Song (2021).[4] We prove that our proposed variance estimator is consistent for the true variance of the OLS estimator in linear models with dyadic data when the cross-sectional dependence follows an observed network.

Finally, we illustrate the extent to which neglecting network spillovers with dyadic data may

[2]Such spillovers could be rationalized as individual-level unobserved heterogeneity: e.g., an unmeasured export orientation of an economy and unmodeled tastes for consumption in the literature of international trade (see Graham (2020a) and references therein for detail).

[3]These numbers are obtained from the simulation results in which the underlying network is modestly complex (the Barabási and Albert's (1999) random network with parameter $\nu = 3$), the sample size is large ($N = 5000$) and the spillovers are significant ($\gamma = 0.8$). See Section 5 for details.

[4]We carefully review the existing literature on the econometrics of networks, including Vainora (2020), Kojevnikov et al. (2021), Leung (2021, 2022), and Leung and Moon (2021), and multiway clustering (e.g., Chiang, Kato, Ma, and Sasaki (2021)), in Appendix B.

bias inference results. Beyond the Monte Carlo simulations described above, we revisit the application in Harmon et al. (2019) of voting in the European Parliament. The authors study whether random seating arrangements (based on naming conventions) induce neighboring politicians to agree with one another in policy votes. The outcome, whether politicians i and j vote the same way on a policy, is dyadic in nature. However, i and j's votes may be correlated even if they are not neighbors: for instance, i and j may sit on either side of common neighbors k and m, who influence them both. This chain of influences is sufficient to induce strong correlation across non-dyads. We show that neglecting such higher order spillovers has significant empirical consequences: their estimated variance using Aronow et al. (2015) is roughly 22% smaller than using our consistent estimator accounting for such spillovers; and the estimate based on the Eicker-Huber-White estimator ignoring spillovers is approximately 73% smaller compared to the one delivered by our estimator.

The paper is organized as follows. In the next section, we define the linear model and additional notation necessary for our results. Section 3 presents results on the consistency and asymptotic normality of the OLS estimator in this environment, necessary to prove the consistency of our variance estimator. The latter is our main result, proved in Section 4. Section 5 presents Monte Carlo simulation results comparing the finite sample performance of our estimator to the Eicker-Huber-White estimator and the dyadic-robust one, while Section 6 presents the empirical application. Section 7 concludes. Throughout the paper, as appropriate, we compare our results to those in the literature. In Appendix, we provide all the proofs, additional Monte Carlo exercises and additional details on the empirical application as well as an extended review of related literature.

2 Set-up

2.1 Sampling Scheme

Assume that we are given a cross section of $N \in \mathbb{N}$ individuals locating along a network that can be interpreted as consumers, firms, and other observation units depending on the context. Amongst $\binom{N}{N-1}$ possible dyads, the ones that are present in the network over the N individuals are called active dyads (e.g., importer-exporter, employer-employee relationships). Let each active dyad between the units be indexed by $m \in \mathcal{M}_N \subseteq \mathbb{N}$ where $\mathcal{M}_N$ denotes an index set of the active dyads among N individuals with $M := |\mathcal{M}_N| \in \mathbb{N}$. That is, the dyad for two units (e.g., firms, individuals) i and j is denoted as some m.

To make our point clear, it is quintessential to distinguish between a pair of dyads who share a member (i.e., who are directly linked) and a pair of dyads who are directly or indirectly linked.

Definition 2.1 (Adjacent & Connected Dyads)**.** Two active dyads m and m' are said to be *adjacent* if they have an individual in common, and they are called *connected* if there are other pairs of adjacent dyads between them.

In Figure 1 (A), U.S.-Canada trade is adjacent to Canada-China, and connected with, though not adjacent to, France-Germany. We emphasize that the adjacency relationship constitutes a network structure among active dyads, and networks over individuals can thus be transformed to ones over active dyads: for example, Figure 1 (B) provides a network over pairs of trading countries (i.e., active dyads).[5] We define the geodesic distance between two connected dyads m and m' to be the smallest number of adjacent dyads between them. Note that adjacent dyads are a special case of connected dyads with the geodesic distance equal one.

Suppose for any $N \in \mathbb{N}$ that each dyad m is endowed with a triplet of the dyad-specific variables, forming a triangular array $\{(y_{M,m}, x_{M,m}, \varepsilon_{M,m})\}_{m \in \mathcal{M}_N}$ with respect to M, where $y_{M,m} \in \mathbb{R}$ is an observable outcome, $x_{M,m} \in \mathbb{R}^K$ is a K-dimensional vector of observable characteristics with $K \in \mathbb{N}$, and $\varepsilon_{M,m} \in \mathbb{R}$ is a random error term that is not observable to the econometrician. Here, it is assumed that the network structure among individuals is exogenously determined and so is the network among dyads, which is furthermore assumed to be conditionally independent of $\{\varepsilon_{M,m}\}_{m \in \mathcal{M}_N}$. This means that we only consider exogenous network formation.

It is worth noting that there are two different units at play: the number of sampling unit (i.e., individuals), N, and the number of dyads, M. The latter constitutes the basis of statistical analysis. The asymptotic theories of this paper are stated in the scenario that the number of sampling unit goes to infinity, i.e., a large N asymptotics. To bridge these two units, we assume that the number of active dyads also goes to infinity as the number of sampling unit is taken to infinity. This is formalized by the following assumption.

Assumption 2.1 (Denseness)**.** There exists a positive constant c such that $M \geq cN$ whenever N is sufficiently large.

Assumption 2.1 requires the number of active dyads to grow at least at some constant rate relative to the number of individuals, eliminating the possibility of extremely sparse networks. This is similar in spirit to Assumption 2.3 of Tabord-Meehan (2019), in which the minimum degree is assumed to grow at some constant rate relative to the number of individuals. Note that since the

[5] This corresponds to thinking about the line graph of the original graph over individuals.

average degree of the network across individuals is $\frac{2M}{N}$, Assumption 2.1 gives its lower bound to be $2c$. For example, when $c = \frac{1}{2}$, it requires that each individual is linked with, on average, at least one other individual. This is in stark contrast with the notion of bounded degree in which the maximum degree in a network is bounded even when $N \to \infty$ (de Paula, Richards-Shubik, & Tamer, 2018; Penrose & Yukich, 2003).

Remark 2.1. Assumption 2.1 is somewhat milder than Assumption 2.3 of Tabord-Meehan (2019) In fact, Assumption 2.3 of Tabord-Meehan (2019) does not allow any individual to be isolated,[6] while such situations are not precluded in Assumption 2.1, as it merely constrains the average degree to be larger than some constant.

Remark 2.2. Assumption 2.1 marks a crucial distinction between our analysis and standard panel data analysis. In panel data, the econometrician chooses an appropriate asymptotic framework (e.g., a large N (sample size) and/or large T (time) asymptotics) depending on the purpose of the analysis, and there is no necessary relationship between N and T. On the other hand, our analysis under Assumption 2.1 focuses on the situation where an increase in N is likely to create a corresponding increase in M.

2.2 The Linear Model

2.2.1 Setup & Identification

The cross-sectional model of interest takes the form of the linear network-regression model: for any $N \in \mathbb{N}$,

$$y_{M,m} = x'_{M,m}\beta + \varepsilon_{M,m} \quad \forall m \in \mathcal{M}_N, \tag{1}$$

where

$$Cov(\varepsilon_{M,m}, \varepsilon_{M,m'} \mid X_M) = 0 \quad \text{unless } m \text{ and } m' \text{ are connected,} \tag{2}$$

and β is a $K \times 1$ vector of the regression coefficients and X_M denotes the $M \times K$ matrix that records the observed dyad-specific characteristics, i.e., $X_M := [x_{M,1}, \ldots, x_{M,M}]'$. We note that equation (2) allows for there to be spillovers across the error terms even when dyads m and m' are not adjacent (i.e., when they are connected through indirect links). This specification has been of substantial empirical interest (see, e.g., Harmon et al. (2019) and Lustig and Richmond (2020)).

[6] An individual on a network is said to be *isolated* if that individual is not linked to anyone else on the same network.

Let us turn to dyadic models commonly studied in the econometrics literature. To do so, it is convenient to write $m = d(i,j)$ to denote that dyad m is the dyad related to units i and j.

Remark 2.3 (Dyadic-Regression Models)**.** Aronow et al. (2015) and Tabord-Meehan (2019) consider a linear dyadic-regression model of the form:

$$y_{M,d(i,j)} = x'_{M,d(i,j)}\beta + \varepsilon_{M,d(i,j)} \quad \forall i,j \in \mathbb{N}, \tag{3}$$

where

$$Cov\left(\varepsilon_{M,d(i,j)}, \varepsilon_{M,d(k,l)} \mid x_{M,d(i,j)}, x_{M,d(k,l)}\right) = 0 \quad \text{unless} \quad \{i,j\} \cap \{k,l\} \neq \emptyset, \tag{4}$$

with $m = d(i,j)$ representing the dyad between i and j. The dyadic-regression model (3) and (4) can be embedded in our network-regression model (1) by restricting the correlation structure (2) to be

$$Cov(\varepsilon_{M,m}, \varepsilon_{M,m'} \mid X_M) = 0 \quad \text{unless } m \text{ and } m' \text{ are adjacent.} \tag{5}$$

It is clear from comparing the expression above to our assumption in (2) that our assumed covariance structure is more general than those in the cited papers.

Towards identification of β in (1), we impose the following assumptions, all of which are standard in the literature.

Assumption 2.2 (Identification Condition)**.** For each $N \in \mathbb{N}$, the following conditions are satisfied:

(i) $\sup_{m \in \mathcal{M}_N} E\left[|\varepsilon_{M,m}|^2\right]$ exists and is finite;

(ii) $\sup_{m \in \mathcal{M}_N} E\left[\|x_{M,m}\|_2^2\right]$ exists and is finite;

(iii) $E\left[x_{M,m}x'_{M,m}\right]$ exists with finite elements and positive definite for all $m \in \mathcal{M}_N$; and

(iv) $E\left[\varepsilon_{M,m} \mid X_M\right] = 0$ for all $m \in \mathcal{M}_N$,

where $\|\cdot\|_2$ is meant to be the Euclidean norm.

Assumption 2.2 (i) and (ii) are standard and jointly imply the finite existence of the second moment of $y_{M,m}$ for all $m \in \mathcal{M}_N$, which in turn implies the finite existence of the cross moment of $y_{M,m}$ and $x_{M,m}$ for all $m \in \mathcal{M}_N$. The third and fourth assumptions are also standard in the context of the linear regression models and require no multicolinearity and strict exogeneity, respectively.[7]

Assumption 2.2 is used to prove identification of the linear parameters in equation (1):

[7]For identification purposes, both of these assumptions can be relaxed to

Proposition 2.1 (Identification)**.** *Under Assumption 2.2, the regression parameter β in (1) is identified.*

Proof. See Appendix D.1. $\square$

2.2.2 Estimation

Stacking the dyad-specific observations, the model, in matrix notation, is given by

$$y_M = X_M \beta + \varepsilon_M, \tag{6}$$

where y_M is the $M \times 1$ vector of the observed dyad-specific outcomes, i.e., $y_M := [y_{M,1}, \ldots, y_{M,M}]'$, and ε_M represents the $M \times 1$ vector of unobserved dyad-specific error terms, i.e., $\varepsilon_M := [\varepsilon_{M,1}, \ldots, \varepsilon_{M,M}]'$. Note that the error terms have a non-homoskedastic structure as specified in (2).

Towards estimation, we assume the following rank condition.

Assumption 2.3 (Full Rank Condition)**.** The matrix $X_M' X_M$ is positive-definite with finite elements whenever M is sufficiently large.

This assumption is an algebraic requirement and guarantees the invertibility of $X'X$. The intuition and interpretation carry over from Assumption 2.2 (iii). Under Assumption 2.3, the OLS estimator of β is given by

$$\begin{aligned} \hat{\beta} &= \left(X_M' X_M\right)^{-1} X_M' y_M \\ &= \left(\sum_{j \in \mathcal{M}_N} x_{M,j} x_{M,j}'\right)^{-1} \sum_{m \in \mathcal{M}_N} x_{M,m} y_{M,m}. \end{aligned} \tag{7}$$

From this expression, we can write

$$\hat{\beta} - \beta = \left(\sum_{j \in \mathcal{M}_N} x_{M,j} x_{M,j}'\right)^{-1} \sum_{m \in \mathcal{M}_N} x_{M,m} \varepsilon_{M,m}. \tag{8}$$

In light of Assumption 2.2 (iv), it is straightforward to verify that $\hat{\beta}$ is unbiased for β. However, a consistency result is by no means trivial due to the dependence along the network which induces a complex form of cross-sectional correlations, hindering the blind application of the standard theory for independently and identically distributed random vectors. This is clarified in Section 2.3.

In the remainder of this subsection, we omit the subscript M purely for notational convenience.

(iii)' $E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} x_{M,m}'\right]$ exists with finite elements and positive definite;

(iv)' $E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} \varepsilon_{M,m}\right] = 0$,

respectively. We maintain Assumptions 2.2 (iii) and (iv) solely to facilitate exposition.

2.2.3 Inference

Inference about β is based on a normal approximation of the distribution of $\hat{\beta}$ around β. In practice, the hypothesis testing is conducted using the expression:

$$\left(\widehat{Var}(\hat{\beta})\right)^{-\frac{1}{2}}(\hat{\beta}-\beta), \tag{9}$$

where $\widehat{Var}(\hat{\beta})$ is a consistent estimator of the asymptotic variance of $\hat{\beta}$. Our main result in Section 4 is providing such an appropriate estimator, which takes the form:

$$\widehat{Var}(\hat{\beta}) := \left(\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N}\kappa_{m,m'}h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}' x_m x_{m'}'\right)\left(\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}, \tag{10}$$

where $\kappa_{m,m'}$ is a kernel function that will be defined formally in Section 4; $h_{m,m'}$ represents an indicator function that takes one if dyads m and m' are connected and zero otherwise; and $\hat{\varepsilon}_m := y_m - x_m'\hat{\beta}$. This paper derives conditions under which $\widehat{Var}(\hat{\beta})$ is consistent for the asymptotic variance of $\hat{\beta}$. Before doing so, let us compare the variance estimator (10) and a dyadic-robust variance estimator.

Example 2.1 (Dyadic-Robust Variance Estimator)**.** An increasing number of applied researchers, such as Harmon et al. (2019) and Lustig and Richmond (2020), estimate model (1) under the specification (2). To conduct inference, these works employ dyadic-robust variance estimators proposed by Aronow et al. (2015) and Tabord-Meehan (2019) which are given by:

$$\widehat{Var}(\hat{\beta}) := \left(\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N}\mathbb{1}_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'} x_m x_{m'}'\right)\left(\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}, \tag{11}$$

where $\mathbb{1}_{m,m'}$ equals one if dyads m and m' are adjacent and zero otherwise. On the other hand, our estimator (10) yields

$$\widehat{Var}(\hat{\beta}) := \left(\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N}h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'} x_m x_{m'}'\right)\left(\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}, \tag{12}$$

where we assume, for the sake of simplicity, that $\kappa_{m,m'} = 1$ for all $m, m' \in \mathcal{M}_N$. Note that the use of the dyadic-robust variance estimator sets cases in which two dyads are not adjacent but connected to zero. This means that the structure of the variance estimator (11) may not be coherent to the model specification (2). However, our estimator accounting for network spillovers accommodates the correlation across both adjacent and connected dyads.[8] This suggests that the dyadic-robust variance estimator may be inconsistent when there is cross-sectional dependence along a network: i.e., when non-adjacent dyads can still affect the correlation structure and outcomes of dyad m.[9] This conjecture is formally proven in Corollary 4.1 and illustrated in Monte Carlo simulations in Section 5.

In order to validate the test statistic (9) for inference, this paper establishes a consistency and asymptotic normality results of $\hat{\beta}$ and develops a consistent network-robust estimator for the asymptotic variance.

2.3 Network Dependent Processes

Let $Y_{M,m}$ be a random vector defined as

$$Y_{M,m} := \left(\frac{1}{M} \sum_{j \in \mathcal{M}_N} x_{M,j} x'_{M,j} \right)^{-1} x_{M,m} \varepsilon_{M,m}, \tag{13}$$

[10]

and denote $\mathcal{C}_M := \{x_m\}_{m \in \mathcal{M}_N}$.[11] From equation (8) we can write

$$\hat{\beta} - \beta = \frac{1}{M} \sum_{m \in \mathcal{M}_N} Y_{M,m}. \tag{14}$$

Our interest lies in proving the asymptotic properties of $\hat{\beta}$ taking advantage of the expression (14). However, the presence of $\varepsilon_{M,m}$ in $Y_{M,m}$, which is allowed to be correlated along the network over active dyads, renders our approach nonstandard. The canonical results about independently and identically distributed ($i.i.d.$) random variables are by no means applicable due to the complex form of dependence inherent to $\varepsilon_{M,m}$. The theory of dyadic data merely accounts for a special case where the correlations between connected dyads except adjacent dyads are muted. Moreover, the asymptotic theory of spatially-correlated and time-series data builds on the assumption that

[8]See Definition 2.1 and the subsequent discussion.

[9]Standard clustering estimators (e.g., at the individual level) may be inappropriate as each agent in a network has a complex structure of connections. If the network model features positive spillovers, then the dyadic-robust variance estimator will probably underestimate the true variance, leading to conservative hypothesis testing. Meanwhile, it is likely to overstate the true variance when there are negative spillovers. We expand on this point in our numerical simulations.

[10]By construction, the collection of $Y_{M,m}$'s constitutes a triangular array of random vectors.

[11]For the case of stochastic networks, it is defined to include information about the network topology as well as the collection of the dyad-specific attributes $\{x_m\}_{m \in \mathcal{M}_N}$.

the index set is a metrizable space,[12] an assumption that may not necessarily be true in the network context. In this light, the right hand side of (14) cannot simply be embedded in other widely-studied variants of dependent random vectors.

In this paper, we embrace the dependence of the $Y_{M,m}$'s along the network among active dyads (hereby, referred to as the "network") by exploiting the concept of network dependent random variables recently developed by Kojevnikov et al. (2021), in which the correlations of the random variables given a common shock are explicitly dictated by their distance on the network. Compared to the other existing frameworks, the network-dependent random variables circumvent the aforementioned limitations, so that the network-regression model (1) and (2) become amenable to formal statistical inference.

To formally define our notion of network dependence, we first introduce notation intentionally set close to those of Kojevnikov et al. (2021). By construction, $Y_{M,m}$ resides in the K-dimensional Euclidean space $\mathbb{R}^K$. First, let $\rho_M(m, m')$ denote the geodesic distance between dyads m and m',[13] and A and B be sets of dyads whose cardinalities are a and b, respectively. Define the distance between these two sets as

$$\rho_M(A, B) \coloneqq \min_{m \in A} \min_{m' \in B} \rho_M(m, m'). \tag{15}$$

We denote by $\mathcal{P}_M(a, b; s)$ the collection of pairs of A and B with distance s. Next we consider a collection of bounded Lipschitz functions $f : \mathbb{R}^{K\times c} \to \mathbb{R}$ and denote it by $\mathcal{L}_{K,c}$. Let $\mathcal{L}_K$ be a collection that gathers $\mathcal{L}_{K,c}$ for all $c \in \mathbb{N}$.

Now we are finally in a position to formally define the notion of conditional ψ-dependence.

Definition 2.2 (Conditional ψ-Dependence given $\{\mathcal{C}_M\}$: Kojevnikov et al. (2021), Definition 2.2)**.** A triangular array $\{Y_{M,m} \in \mathbb{R}^K : M \geq 1, m \in \{1, \ldots, M\}\}$ is called conditionally ψ-dependent given $\{\mathcal{C}_M\}$, if for each $M \in \mathbb{N}$, there exist a $\mathcal{C}_M$-measurable sequence $\theta_M \coloneqq \{\theta_{M,s}\}_{s\geq 0}$ with $\theta_{M,0} = 1$, and a collection of nonrandom function $(\psi_{a,b})_{a,b\in\mathbb{N}}$ where $\psi_{a,b} : \mathcal{L}_{K,a} \times \mathcal{L}_{K,b} \to [0, \infty)$, such that for all $(A, B) \in \mathcal{P}_M(a, b; s)$ with $s > 0$ and all $f \in \mathcal{L}_{K,a}$ and $g \in \mathcal{L}_{K,b}$,

$$|Cov\left(f(Y_{M,A}), g(Y_{M,B}) \mid \mathcal{C}_M\right)| \leq \psi_{a,b}(f, g)\theta_{M,s} \quad a.s.$$

We call $\{\theta_{M,s}\}$ the dependence coefficients of $\{Y_{M,m}\}$. Intuitively, $\psi_{a,b}(f, g)$ dictates the magnitude of covariation of $Y_{M,A}$ and $Y_{M,B}$ as well as those of f and g. Meanwhile $\theta_{M,s}$ captures the

[12] Typically the literature assumes a Euclidean space. See Conley (1999), Ibragimov and Müller (2010) and references therein.

[13] Recall that we define the geodesic distance between two connected dyads m and m' to be the smallest number of adjacent dyads between them.

strength of covariation net of the effect of scaling, analogous to correlation coefficients. As the minimum distance between A and B, s, grows, the dependence between $Y_{M,A}$ and $Y_{M,B}$, $\{\theta_{M,s}\}$, tends to zero.[14] We maintain Assumption 2.1 of Kojevnikov et al. (2021) on the upper bounds of both $\psi_{a,b}(f,g)$ and $\theta_{M,s}$ throughout the paper.[15]

We base our theory on Definition 2.2 in favor of conditioning on $\mathcal{C}_M$, which allows us to focus only on the dependence of the random vectors $Y_{M,m}$ intrinsic to the error terms $\varepsilon_{M,m}$ (i.e., after conditioning out those from covariates.) This proves to be of practical importance when $Y_{M,m}$ consists of a number of other covariates with the error term being the sole source of the systematic dependence structured over the network.

In Appendix B, we carefully compare our framework to others in the literature.

3 Preliminary Results

This section establishes the asymptotic properties tailored for our network dependent random variables.

3.1 Notations

As will become transparent shortly, asymptotic theories for $\hat{\beta}$ rest on tradeoffs between the correlation of the network-dependent random vectors (i.e., the dependence coefficients) and the denseness of the underlying network. To measure the denseness, we first define two concepts of neighborhoods: for each $m \in \mathcal{M}_N$ and $s \in \mathbb{N}$,

$$\mathcal{M}_N(m;s) \coloneqq \{m' \in \mathcal{M}_N : x_m(m,m') \leq s\},$$
$$\mathcal{M}_N^{\partial}(m;s) \coloneqq \{m' \in \mathcal{M}_N : x_m(m,m') = s\}.$$

The first set $\mathcal{M}_N(m;s)$ collects all the m's neighbors whose distance from m is no more than s, whilst the other set $\mathcal{M}_N^{\partial}(m;s)$ registers all the m's neighbors whose distance from m is exactly s. In what follows, we refer to $\mathcal{M}_N(m;s)$ as the neighborhood of m within distance s, and call $\mathcal{M}_N^{\partial}(m;s)$ the neighborhood shell of m at distance s.

[14] Note that $\{\theta_{M,s}\}$ are random because Definition 2.2 is a conditional statement.

[15] Details of the mathematical environment are delegated to Appendix A.

Next, we define two types of density measures of a network: for $k, r > 0$,

$$\Delta_M(s, r; k) := \frac{1}{M} \sum_{m \in \mathcal{M}_N} \max_{m' \in \mathcal{M}_N^{\partial}(m;s)} |\mathcal{M}_N(m; r) \backslash \mathcal{M}_N(m'; s-1)|^k,$$
$$\delta_M^{\partial}(s; k) := \frac{1}{M} \sum_{m \in \mathcal{M}_N} |\mathcal{M}_N^{\partial}(m; s)|^k,$$

where it is assumed that $\mathcal{M}_N(m'; -1) = \emptyset$. The measure $\Delta_M(s, r; k)$ gauges the denseness of a network in terms of the average size of a version of the neighborhood. The other measure $\delta_M^{\partial}(s; k)$ expresses the denseness of a network as the average size of the neighborhood shell. Moreover, we consider a composite of these two density measures:

$$c_M(s, m; k) := \inf_{\alpha > 1} \left(\Delta_M(s, r; k\alpha)\right)^{\frac{1}{\alpha}} \left(\delta_M^{\partial}\left(s; \frac{\alpha}{\alpha - 1}\right)\right)^{\frac{\alpha-1}{\alpha}}.$$

Notice that any restriction on $c_M(s, m; k)$ indirectly regulates the limiting behaviors of $\Delta_M(s, r; k)$ and $\delta_M^{\partial}(s; k)$. In fact, Kojevnikov et al. (2021) show that controlling the asymptotic behavior of $c_M(s, m; k)$ is sufficient for the Law of Large Numbers (LLN) and Central Limit Theorem (CLT) of the network dependent random variables (Condition ND).

Lastly, conditions in this section are stated with respect to each element of $Y_{M,m}$ for notational simplicity. We refer to the u-th entry of the network-dependent random vector $Y_{M,m}$ as $Y_{M,m}^u$ for $u \in \{1, \ldots, K\}$.

3.2 Limit Theorems for Network Dependent Processes

3.2.1 Consistency

Proving that $\hat{\beta}$ is consistent for β requires two additional conditions.

Assumption 3.1 (Conditional Finite Moment of ε_m)**.** There exists $\eta > 0$ such that

$$\sup_{N \geq 1} \max_{m \in \mathcal{M}_N} E\left[|\varepsilon_{M,m}|^{1+\eta} \mid \mathcal{C}_M\right] < \infty.$$

Assumption 3.2. $\frac{1}{M}\sum_{s \geq 1} \delta_M^{\partial}(s; 1)\theta_{M,s} \xrightarrow{a.s.} 0$ as $M \to \infty$.

Assumption 3.1 requires that the errors are not too large once conditioned on common shocks. This assumption is generally stronger than Assumption 2.2 (i) as it requires the finiteness of a conditional moment. Under Assumption 2.3, this assumption implies Assumption 3.1 of Kojevnikov et al. (2021) for each $Y_{M,m}^u$ with $u \in \{1, \ldots, K\}$.

Assumption 3.2 controls the tradeoff between the denseness of the underlying network and the covariability of the random vectors. If the network becomes dense, then the dependence of the associated random variables has to decay much faster. Such tradeoffs can readily be envisioned, among many others, in the context of international trade: as a country diversifies its trading partners, the country's trade flows as a whole become less susceptible to the shock to a particular one of its trading counterparts. Moreover, this is also consistent with the typical assumption in the spatial econometrics literature upon which the empirical analyses of the aforementioned examples (i.e., international trade, legislative voting, and exchange rates) are based, because it embodies the idea that spillovers decay as they propagate farther (a "fading" memory). See, e.g., Kelejian and Prucha (2010). For instance, Acemoglu, García-Jimeno, and Robinson (2015) assume that network spillovers are zero if agents are sufficiently distantly connected on a geographical network.

Theorem 3.1 (Consistency of $\hat{\beta}$)**.** *Suppose that Assumptions 2.1, 2.2, 2.3, 3.1 and 3.2 hold. Then,*

$$\hat{\beta} \xrightarrow{p} \beta$$

as $N \to \infty$.

Proof. See Appendix D.2. □

Theorem 3.1 establishes the consistency of $\hat{\beta}$ under the scenario where the number of sampling units N goes to infinity. When Assumption 2.1 is dropped, the result continues to hold in terms of the number of active dyads M.

3.2.2 Asymptotic Normality

Before we can prove the consistency of our variance estimator, we must derive the asymptotic distribution of $\hat{\beta}$. To do so, we study the asymptotic distribution of the following sum of network dependent random vectors

$$S_M := \sum_{m \in \mathcal{M}_N} Y_{M,m},$$

which is present in $\hat{\beta}$ in equation (14). Let S_M^u be the u-th entry of S_M for $u \in \{1, \ldots, K\}$. We consider two types of variances: namely, unconditional and conditional variances. Denote the unconditional variance of S_M^u by

$$\tau_M^2 := Var\left(S_M^u\right),$$

and the conditional variance of S^u_M given $\mathcal{C}_M$ by

$$\sigma^2_M \coloneqq Var\left(S^u_M \mid \mathcal{C}_M\right).$$

Since S^u_M is not a sum of independent variables, τ^2_M and σ^2_M cannot be expressed as a sum of variances. We thus need to explicitly take into account covariance between the random variables $\{Y^u_{M,m}\}_{m\in\mathcal{M}_N}$. We study the the CLT for the normalized sum of $Y^u_{M,m}$ given by $\frac{S^u_M}{\tau_M}$.

As usual, the CLT for this normalized sum requires us to strengthen conditions (3.1) and (3.2) to control the variability of higher moments of our random variables. This is done in Assumptions A.2 and A.3 in Appendix A. These assumptions are written in terms of conditional expectations, whereas we are interested in the unconditional distribution of $\frac{S_M}{\tau_M}$. The following assumption bridges the conditional variance σ_M and the unconditional variance τ_M.

Assumption 3.3 (Growth Rates of Variances)**.** There exists a sequence of (possibly random) positive numbers, $\{\pi_{N,M}\}_{N>0}$, such that

$$\frac{\sigma^2_M}{\pi_{N,M}\tau^2_M} \xrightarrow{a.s.} 1 \qquad \text{as} \quad N\to\infty.$$

In Assumption 3.3, the relative growth rate of the conditional and unconditional variances are regulated by $\pi_{N,M}$. $\pi_{N,M}$ reflects the effects of the randomness conveyed in $\mathcal{C}_M$. Moreover, $\pi_{N,M}$ can be thought of capturing the ratio between N and M, and plays a pivotal role in determining the convergence rates (see Remark 3.1 and 4.2). In the remainder of this paper, we assume that the rates $\{\pi_{N,M}\}_{N>0}$ are known to the econometrician (see Remark 4.3 for details on its empirical implementation).

Under these assumptions, we obtain the asymptotic distribution of $\hat{\beta}$:

Theorem 3.2 (Asymptotic Normality of a Sum of the Network-Dependent Random Variables)**.** *Suppose that Assumptions 2.1, A.2, A.3 and 3.3 hold. Then,*

$$\sqrt{N}\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}(0, AVar(\hat{\beta})) \qquad as \quad N\to\infty, \tag{16}$$

where $AVar(\hat{\beta})$ is given by:

$$AVar(\hat{\beta}) \coloneqq \lim_{N\to\infty}\frac{N\pi_{N,M}}{M^2}E\left[\left(\sum_{k\in\mathcal{M}_N}x_{M,k}x'_{M,k}\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N}\varepsilon_{M,m}\varepsilon'_{M,m'}x_{M,m}x'_{M,m'}\right)\left(\sum_{k\in\mathcal{M}_N}x_{M,k}x'_{M,k}\right)^{-1}\right], \tag{17}$$

which is assumed to have finite elements and be positive semidefinite, where $\pi_{N,M}$ is defined in Assumption 3.3.

Proof. See Appendix D.3. $\square$

Theorem 3.2 demonstrates the asymptotic normality of $\hat{\beta}$ in terms of N, the number of sampling units. If Assumption 2.1 is omitted, the theorem holds as long as M, the number of active dyads, goes to infinity. Since dyadic dependence is a more stringent form of correlations than networks, the claim of this theorem remains valid for the case of dyadic dependence (5). As an immediate consequence, Theorem 3.2 further implies an asymptotic normality for linear dyadic-regression models (1) and (5), a result comparable to Lemma 1 of Aronow et al. (2015) and Proposition 3.1 and 3.2 of Tabord-Meehan (2019).

Remark 3.1. The requirement on the behavior of $AVar(\hat{\beta})$ mirrors Assumptions 2.4, 2.5 and 2.6 of Tabord-Meehan (2019): the expression (17) boils down to his Assumption 2.4, if it is well-defined with $\pi_{N,M} = \frac{M}{N}$; it reduces to his Assumption 2.5, if it is compatible with $\pi_{N,M} = \frac{M}{N^2}$; and it coincides with Assumption 2.6, if it is maintained with $\pi_{N,M} = \frac{M}{N^{r+1}}$ for $r \in [0,1]$. Moreover, if $AVar(\hat{\beta})$ is well-defined for $\pi_{N,M} = 1$, the expression (17) coincides with the one that appears in Lemma 1 of Aronow et al. (2015).

4 Consistent Estimation of the Variance of β Under Network Spillovers

This section studies consistent estimation of $AVar(\hat{\beta})$. In light of Assumption 2.2 (iv), $Y_{M,m}$ are centered, i.e., $E\left[Y_{M,m}\right] = 0$ for each $m \in \mathcal{M}_N$. Define $Var(\hat{\beta})$ as

$$Var(\hat{\beta}) := \frac{N\pi_{N,M}}{M^2} \sum_{s \geq 0} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} E\left[Y_{M,m} Y'_{M,m'}\right]. \tag{18}$$

Clearly, $AVar(\hat{\beta}) = \lim_{N\to\infty} Var(\hat{\beta})$. The estimator developed in this section is a type of a kernel estimator. Define a kernel function $\omega : \bar{\mathbb{R}} \to [-1,1]$ such that $\omega(0) = 1$, $\omega(z) = 0$ whenever $|z| > 1$, and $\omega(z) = \omega(-z)$ for all $z \in \bar{\mathbb{R}}$. We denote by b_M the bandwidth, or the lag truncation. The feasible variance estimator of our interest (10) can be written as

$$\widehat{Var}(\hat{\beta}) = \frac{1}{M^2} \sum_{s \geq 0} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \omega_M(s) \hat{Y}_{M,m} \hat{Y}'_{M,m'}, \tag{19}$$

with

$$\omega_M(s) := \omega\left(\frac{s}{b_M}\right) \qquad \forall s \geq 0,$$

and

$$
\hat{Y}_{M,m} := \left(\frac{1}{M} \sum_{j \in \mathcal{M}_N} x_{M,j} x'_{M,j} \right)^{-1} x_{M,m} \hat{\varepsilon}_{M,m},
$$

where $\hat{\varepsilon}_{M,m} := y_{M,m} - x'_{M,m} \hat{\beta}$.

Assumption 4.1 (Kojevnikov et al. (2021), Assumption 4.1)**.** There exists $p > 4$ such that

(i) $\sup_{N \geq 1} \max_{m \in \mathcal{M}_N} E\left[|\varepsilon_m|^p \mid \mathcal{C}_M \right] < \infty \quad a.s.$;

(ii) $\lim_{M \to \infty} \sum_{s \geq 1} |\omega_M(s) - 1| \delta_M^{\partial}(s) \theta_{M,s}^{1-\frac{2}{p}} = 0 \quad a.s.$; and

(iii) $\lim_{M \to \infty} \frac{1}{M} \sum_{s \geq 0} c_M(s, b_M; 2) \theta_{M,s}^{1-\frac{4}{p}} = 0 \quad a.s.$

Assumption 4.1 (i) is essentially the same as Assumption A.2. Assumption (ii) posits a tradeoff between the kernel function, the denseness of a network and the dependence coefficients. Specifically, the kernel function ω_M is required to converge to one sufficiently fast. Kojevnikov et al. (2021) demonstrate primitive conditions under which this requirement is fulfilled (Proposition 4.2). Assumption (iii) indicates a condition between the bandwidth, the denseness and the dependence coefficients. Kojevnikov et al. (2021) provide a preferred choice of the bandwidth that satisfies this condition. For a fixed $p > 4$, both (ii) and (iii) preserve the same interpretation as Assumption 3.2: namely, the correlation must decay much faster relative to the growth of the density of the network, or vice versa.

To prove our main result, we need three additional regularity assumptions.

Assumption 4.2. For all $N \geq 1$, $\{x_{M,m}\}_{m \in \mathcal{M}_N}$ have uniformly bounded support.

Assumption 4.2 is employed to guarantee the uniform boundedness of arbitrary moments of $x_{M,m}$.

Remark 4.1. As pointed out in Tabord-Meehan (2019), the bounded support assumption can be relaxed at the cost of imposing another condition pertain to higher-order moments. For instance, the boundedness of the 16th order moment of $x_{M,m}$ is sufficient for the subsequent results.

Assumption 4.3.

(i) $\lim_{M \to \infty} \frac{1}{M} \sum_{s \geq 0} \delta_M^{\partial}(s; 1) = 0$;

(ii) $\lim_{M \to \infty} \frac{1}{M} \sum_{s \geq 0} c_M(s, b_M; 2) = 0$.

Both requirements of Assumption 4.3 restrict the denseness of the network among dyads: the requirement (i) means that the average density with respect to the neighborhood shell vanishes as M goes to infinity, and the requirement (ii) embodies the same idea in terms of the composite measure of the denseness (see Section 3.1). These conditions rule out the situation where the network becomes progressively dense. Note that while Assumption 4.1 constrains the denseness of the network relative to the correlations of the associated random variables, Assumption 4.3 only refers to the network configuration.

Assumption 4.4. $\sup_{N\geq 1} \frac{N\pi_{N,M}}{M} < \infty$.

Under Assumption 2.1, this condition amounts to requiring the boundedness of $\pi_{N,M}$. In light of Definition 3.3, this is typically satisfied when the variance due to $\mathcal{C}_M$, i.e., conditioning variables, does not diverge.

The following theorem is the main theoretical contribution of this paper.

Theorem 4.1 (Consistency of the Network-HAC Variance Estimator)**.** *Suppose that Assumptions 4.1, 4.2, 4.3 and 4.4 hold. Then,*

$$\left\| N\pi_{N,M}\widehat{Var}(\hat{\beta}) - Var(\hat{\beta}) \right\|_F \xrightarrow{p} 0 \qquad as \quad N \to \infty,$$

where $\|\cdot\|_F$ indicates the Frobenius norm.[16]

Proof. See Appendix D.4. □

Theorem 4.1 establishes the consistency of our proposed variance estimator accounting for network spillovers across dyads in the sense of the Frobenius norm. By the equivalence of matrix norms, the consistency remains true in the sense of the spectral norm and the nuclear norm. When the network structure is specified to show dyadic dependence (5), this result can thus be understood as a consistency result of a dyadic-robust variance estimator, as anticipated in Remark 2.3. In this sense, our result further adds to the literature of dyadic-robust variance estimators such as Aronow et al. (2015) and Tabord-Meehan (2019), by providing an alternative approach under a different set of assumptions.

[16]The Frobenius norm of a real matrix $A := (a_{ij})_{i,j}$, dented by $\|A\|_F$, is given by

$$\|A\|_F := \sqrt{\mathrm{tr}(A^T A)} = \sqrt{\sum_i \sum_j a_{ij}^2}.$$

Remark 4.2. The convergence rate of $\widehat{Var}(\hat{\beta})$ rests on $\pi_{N,M}$. If $\pi_{N,M} = 1$, the estimator converges to the true variance at the standard rate of $\sqrt{N}$ analogous to Proposition 1 of Aronow et al. (2015). Under the condition that $\pi_{N,M} = \frac{M}{N}$, the convergence rate turns to be $\sqrt{M}$ in a similar vein to Proposition 3.3 of Tabord-Meehan (2019), while that of Proposition 3.4 and 3.5 of Tabord-Meehan (2019) can be associated with the cases when $\pi_{N,M} = \frac{M}{N^2}$ and $\pi_{N,M} = \frac{M}{N^{r+1}}$, respectively.

Remark 4.3. In Section 5 (Monte Carlo simulation) and 6 (empirical application), we set $\pi_{N,M} = 1$ because the primary goal of these sections lies in contrasting the widely-used dyadic-robust variance estimator ignoring higher order moments with the network-robust variance estimator accounting for correlations across indirect links – a comparison raised in Example 2.1. In the observance of Remark 4.2, this modeling choice helps us to explicitly compare these two variance estimators within the same asymptotic approximation (i.e., the same rate of convergence). Evaluating the impacts of incorrectly choosing the value for this parameter and developing a methodology for estimating it are beyond the scope of this paper and thus left for future work.

It follows from Theorem 4.1 that the dyadic-robust variance estimator (11) is inconsistent for the true variance when the underlying network involves a nonnegligible degree of far-away correlations.

Corollary 4.1 (Inconsistency of Dyadic-Robust Variance Estimators in the Presence of Network Spillovers Across Dyads)**.** *Suppose that the assumptions required in Theorem 4.1 hold. Assume, in addition, that*

$$\inf_{N\geq 1} \frac{1}{M} \sum_{s\geq 2} \sum_{m\in\mathcal{M}_N} \sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} \left\| E\left[\varepsilon_{M,m}\varepsilon_{M,m'}x_{M,m}x'_{M,m'}\right]\right\|_F > 0. \tag{20}$$

Then, the dyadic-robust estimator (11) applied to the network-regression model (1) and (2) is inconsistent.

Proof. See Appendix D.4. □

The added assumption (20) in Corollary 4.1 pertains to both the network configuration of active dyads and the regression variables. It represents a setting where the spillovers from far-away neighbors are not negligible even when N is large. If, on the other hand, far-away neighbors in the network only have a negligible effect on the cross-sectional dependence of our model, the dyadic-robust variance estimator remains a good approximation for the asymptotic variance of linear dyadic data models with network spillovers across dyads. This insight is further investigated through numerical simulations in Section 5.

5 Monte Carlo Simulation

This section explores the finite sample properties of the asymptotic inference of the proposed variance estimator by means of Monte Carlo simulations.

5.1 Simulation Design

We compare three types of variance estimators across different specifications and network configurations. We use the Eicker-Huber-White estimator as a benchmark, the dyadic-robust estimator of Tabord-Meehan (2019) as a comparison accounting for the dyadic nature of the data, and our proposed estimator which is robust to network spillovers across dyads.

We first generate networks on which random variables are assigned. We follow Canen, Schwartz, and Song (2020) among others by employing two models of random graph formations. They are referred to as Specifications 1 and 2. Specification 1 uses the Barabási and Albert's (1999) model of preferential attachment, with the fixed number of edges $\nu \in \{1, 2, 3\}$ being established by each new node.[17] Specification 2 is based on the Erdös-Renyi random graph (Erdös & Rényi, 1959, 1960) with probability $p = \frac{\lambda}{N}$ for N denoting the number of nodes and $\lambda \in \{1, 2, 3\}$ being a parameter that governs the probability relative to the node size.

For each of the randomly generated networks, we consider a set of agents (denoted by i, j, etc) populated along the network. We can construct the network of active dyads from this simulated network among individual agents: there exists an edge between two active dyads only if they share a common agent (i.e., adjacent). The simulation data is generated from the following simple network-linear regression:

$$y_m = x_m \beta + \varepsilon_m, \text{[18]}$$

with $m := d(i, j)$ representing the dyad between agent i and j.

The dyad-specific regressor x_m is defined as $x_m := |z_i - z_j|$, where both z_i and z_j are drawn independently from $\mathcal{N}(0, 1)$. The regression coefficient is fixed to be $\beta = 1$ across specifications.

The dyad-specific error term ε_m is constructed to exhibit non-zero correlation with $\varepsilon_{m'}$ as long as dyads m and m' are connected (i.e., in the network terminology, there exists a path in the simu-

[17]In generating the Barabási-Albert random graphs, we follow Canen et al. (2020) by choosing the seed to be the Erdös-Renyi random graph with the number of nodes equal the smallest integer above $5\sqrt{N}$, where N denotes the number of nodes.

[18]To simplify notation, we drop the M subscript, delivering the triangular array structure implicit.

lated network), while the strength of the correlation is assumed to decay as they are more distant. To that end, we draw $\varepsilon_m := \sum_{m'} \gamma_{m,m'}\eta_{m,m'}$, where $\gamma_{m,m'}$ equals γ^s if the distance between m and m' is s, and 0 otherwise, for $\gamma \in [0, 1]$[19] and $s \in \{1, \ldots, S\}$ with S being the maximum geodesic distance that the spillover propagates to. Each $\eta_{m,m'}$ is drawn i.i.d. from $\mathcal{N}(0, 1)$. If $\gamma = 1$, then spillover effects are the same no matter how far the agents are apart, i.e., the spillover effects do not decay. If $\gamma = 0$, there are no spillover effects, so the dyadic-robust variance estimator should be consistent. The case of $S = 2$ corresponds to a situation where up to friends of friends may matter for spillovers.

We consider three specifications for each type of network. In the main text, we set $S = 2$ and $\gamma = 0.8$. The results for $S = 2$ with $\gamma = 0.2$ are given in Appendix C.3, and the ones for $S = 1$ with $\gamma = 0.8$ are in Appendix C.4. Throughout the experiments, we assume $\pi_{N,M} = 1$ to facilitate comparison across the three variance estimators under the same asymptotic approximation (see Remark 4.3). For simplicity, the kernel function is set to be one. The nominal size is set at $\alpha = 0.05$ throughout the 5000 iterations of the Monte Carlo simulation.

5.2 Degree Characteristics

Table 1 reports the degree characteristics of the networks when viewed as networks over the active edges.[20] The table provides the average degree, the maximum degree, and the number of active edges (i.e., dyads). By construction, the maximum degree and the average degree increase monotonically as we increase the parameters in both specifications. The number of active edges increases with the sample size regardless of the specification. This reflects the fact that each node tends to have more far-away neighbors as the network becomes denser.

5.3 Results

In Table 2 we present the coverage probability for β and the average length of the confidence interval across simulations. To do so, we compute the t-statistic using the OLS estimator for β and different variance estimators under a Normal distribution approximation. We also follow Aronow et al. (2015) in calculating the empirical standard error[21] and plotting it in the boxplots of the variance

[19] In this simulation, we focus on cases of positive spillovers. Cases of negative spillovers can be analyzed in an analogous manner.

[20] The degree characteristics of the original networks are summarized in Table 4 in Appendix C.1.

[21] The empirical standard error is obtained as a standard deviation of the parameter estimates.

TABLE 1

The Average and Maximum Degrees of Networks among Edges in the Simulations

		Specification 1			Specification 2		
N		$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
500	d_{act}	401	788	1175	238	484	740
	d_{max}	32	45	70	4	9	14
	d_{ave}	3.6858	6.0063	9.1881	0.9580	2.0248	3.0027
1000	d_{act}	859	1699	2540	498	981	1501
	d_{max}	35	55	76	5	8	10
	d_{ave}	3.9581	6.7810	9.2047	1.0341	1.9888	2.9594
5000	d_{act}	4663	9305	13952	2476	5008	7557
	d_{max}	74	161	210	7	12	15
	d_{ave}	5.2989	9.5159	12.8521	1.0137	2.0228	3.0341

Note: Observation units in this table are active edges (dyads), which departs from the convention. Active edges are edges that are at work in the original network over the nodes. The number of active edges is denoted by d_{act}. The maximum degree, d_{max}, expresses the maximum number of edges that are adjacent to an edge, and the average degree, d_{ave}, is the average number of edges adjacent to each edge of the network.

estimates together with the variance estimates in Figure 2.[22]

The results for the empirical coverage probabilities depend on two dimensions: the sample size (N) and the denseness of the underlying network (parametrized by ν and λ). The coverage probability for each estimator improves with the sample size. However, when spillovers are high ($\gamma = 0.8$), only our proposed network-robust variance estimator has coverage close to 95%. Meanwhile, both the Eicker-Huber-White and the dyadic-robust variance estimators perform poorly as the underlying network becomes denser, no matter which specification of the network is involved. For example, in Specification 1 with $\nu = 3$ and the largest sample size ($N = 5000$), the confidence

[22]It is well known that estimates of a variance-covariance matrix may be negative semidefinite when the sample size is very small. This occurs in 4 out of 5000 simulations when $N = 500$. Rather than dropping such observations, we follow Cameron et al. (2011) and augment the eigenvalues of the matrix by adding a small constant, say 0.005, thereby obtaining a new variance estimate that is more conservative.

intervals based on the Eicker-Huber-White and the dyadic-robust variance estimators do not cover the true parameter 615 and 455 times out of 5000 simulations (12.3% and 9.1%), respectively. This is because as the network gets denser, each agent tends to have more direct or indirect links to other agents. This leads to larger higher-order correlations, which is neither accounted by the Eicker-Huber-White nor by the dyadic-robust variance estimator. On the other hand, the network-robust variance estimator is designed to capture such higher-order correlations and, thus, its coverage remains stable across network configurations. A similar conclusion is drawn from the average length of the confidence intervals. The confidence intervals for the Eicker-Huber-White and dyadic-robust variance estimators are typically shorter than those for our proposed estimator when γ is large and $S = 2$. This is because the former undercovers the true parameter (in the presence of positive spillovers), confirming our theoretical results. The difference is often around 10-20% of the length of the network-robust estimator.

However, as the magnitude of spillovers decreases (i.e. γ tends to 0), higher-order spillovers are less pronounced, and the biases from using the Eicker-Huber-White and dyadic-robust variance estimators disappear. This is shown in Table 6 of Appendix C when $S = 2$ and $\gamma = 0.2$. When $S = 1$, the dyadic-robust variance estimator coincides with our proposed estimator (i.e., there are no spillovers from non-adjacent links). This is shown in Table 7.

The finite-sample properties of the three variance estimators are further illustrated in Figure 2, where the horizontal axes represent the sample size and the vertical axes indicate the standard error of the regression coefficient. The boxplots show the 25th and 75th percentiles across simulations, as well as the median, with the whiskers indicating the bounds that are not considered as outliers. The whisker length is set to cover ± 2.7 times the standard deviation of the standard-error estimates. The light-, medium- and dark-grey boxplots describe the distribution of the Eicker-Huber-White, the dyadic-robust and our proposed network-robust variance estimates across simulations, respectively. The diamonds indicate the empirical standard errors of the estimates of the regression coefficients, what Aronow et al. (2015) call the true standard error. It is unsurprising that the empirical standard errors are the same across different variance estimators, as we use the same $\hat{\beta}$. The boxplots show that as the sample size increases, the variation of the network-robust variance estimator shrinks, reaching the empirical standard error (the diamonds). This is expected since this estimator is consistent for the true variance (Theorem 4.1). The estimates appear to vary little for moderate sample sizes (e.g., $N = 1000$). However, the other variance estimators (the light- and medium-grey boxplots) converge to lower values than the empirical standard errors (the diamonds),

TABLE 2

The empirical coverage probability and average length of confidence intervals for β at 95% nominal level: $S = 2$, $\gamma = 0.8$.

		Specification 1			Specification 2		
	N	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
		Coverage Probability					
Eicker-Huber-White	500	0.8774	0.8680	0.8712	0.8910	0.8700	0.8750
	1000	0.8798	0.8732	0.8728	0.8924	0.8810	0.8876
	5000	0.8792	0.8708	0.8770	0.8928	0.8890	0.8790
Dyadic-robust	500	0.9216	0.8976	0.8938	0.9320	0.9212	0.9172
	1000	0.9286	0.9134	0.9006	0.9366	0.9274	0.9242
	5000	0.9342	0.9118	0.9090	0.9388	0.9332	0.9224
Network-robust	500	0.9300	0.9166	0.9146	0.9370	0.9370	0.9410
	1000	0.9392	0.9338	0.9328	0.9456	0.9446	0.9476
	5000	0.9486	0.9440	0.9426	0.9468	0.9472	0.9480
		Average Length of the Confidence Intervals					
Eicker-Huber-White	500	0.3678	0.4093	0.4819	0.2874	0.2853	0.2962
	1000	0.2663	0.3024	0.3309	0.2048	0.2013	0.2068
	5000	0.1321	0.1585	0.1756	0.0916	0.0898	0.0938
Dyadic-robust	500	0.4260	0.4543	0.5204	0.3280	0.3285	0.3365
	1000	0.3124	0.3393	0.3611	0.2359	0.2323	0.2366
	5000	0.1580	0.1775	0.1920	0.1060	0.1043	0.1077
Network-robust	500	0.4413	0.4932	0.5681	0.3370	0.3487	0.3655
	1000	0.3263	0.3733	0.4079	0.2435	0.2482	0.2585
	5000	0.1666	0.1994	0.2215	0.1097	0.1117	0.1183

Note: The upper-half of the table displays the empirical coverage probability of the asymptotic confidence interval for β_1, and the lower-half showcases the average length of the estimated confidence intervals. As the sample size (N) increases, the empirical coverage probability for our estimator accounting for network spillovers approaches 0.95, the correct nominal level. However, that is not the case for alternative estimators.

verifying their inconsistency in this environment with network spillovers, as shown by Corollary 4.1. As we make such spillovers very small (e.g., $\gamma = 0.2$ in Appendix C.3), all estimators have similar performance. This highlights the role of condition (20): namely, the dyadic-robust variance estimator might perform satisfactorily well as long as higher-order correlations beyond immediate neighbors are negligible.

6 Empirical Illustration: Legislative Voting in the European Parliament

We now turn to an empirical application to demonstrate the performance of our variance estimator with real-world data. In doing so, we revisit the work of Harmon et al. (2019) on whether legislators who sit next to each other in Parliament tend to vote more alike on policy proposals.

In their original article, the authors study the existence of peer effects in legislative voting. They focus on the European Parliament, whose Members (MEPs) are voted in through elections in each European Union (EU) member country every five years. The Parliament convenes once or twice a month, in either Brussels or Strasbourg, to debate and vote on a series of proposals.

Once elected to the European Parliament (EP), these MEPs are organized into European Political Groups (EPGs), which aggregate similar ideological members/parties across countries. As Harmon et al. (2019) describe, these EPGs function as parties for many of the traditional party-functions in other legislatures, including coordination on policy and policy votes. Most importantly, MEPs sit within their EPG groups in the chamber. However, within each EPG group, non-party leaders traditionally sit in alphabetical order by last name. See Figure 3 for an example. Hence, to the extent that last names are exogenous (i.e., politicians are not selected or choose last names anticipating seating arrangements in the European Parliament), seating arrangements are used as a quasi-natural experiment: it is sometimes by chance that one legislator sits next to another. And across different periods, as politicians get replaced, they find themselves sitting with new neighbors.

6.1 Data

We adopt the same dataset used in Harmon et al. (2019), which collects the MEP-level data on votes cast in the EP. The dataset records what each MEP voted for (Yes or No), where she is seated, and a number of individual characteristics (e.g. country, age, education, gender, tenure, etc). Its sample period is the plenary sessions held in both Brussels and Strasbourg between October 2006 and November 2010. This corresponds to the sixth and seventh terms of the European Parliament.

For our empirical illustration, we restrict the sample to the policies voted in Strasbourg during the seventh term and we focus on the seating pattern between July 14th - July 16th, 2009 (which involved 116 different proposals being voted on). The resulting sample has 2,431,261 observations, which are split into 422 politicians forming 26,099 pairs (i.e., dyads) of MEPs[23] over 116 proposals. Further information on the construction of our sample is detailed in Appendix E. These restrictions keep the main set-up in Harmon et al. (2019), while allowing us to evaluate variance estimators with smaller sample sizes.

6.2 Empirical Setup

The empirical analysis in Harmon et al. (2019) studies whether MEPs who are seated together are more likely to vote similarly. To make it consistent with our notation above, let $m = d(i,j)$ denote the dyad between legislator i and j, with t being used to index the proposal. Let $Agree_{d(i,j),t}$ be an indicator that takes one if MEP i and j cast the same vote on proposal t, and zero otherwise. Likewise, $SeatNeighbors_{d(i,j),t}$ is a binary variable that equals one if MEP i and j are seated next to each other when the vote for proposal t is taken place, and zero otherwise. We view the seating arrangement as a network among the MEPs: i.e., two MEPs who are seated next to each other within the same political group are treated as adjacent. (See the note below Figure 3.)[24]

We follow Harmon et al. (2019) in assuming that such seating arrangements are exogenously determined, i.e., the adjacency relation is exogenously formed. Their main specification is a linear model:

$$Agree_{d(i,j),t} = \beta_0 + \beta_1 SeatNeighbors_{d(i,j),t} + \varepsilon_{d(i,j),t}.^{25} \tag{21}$$

The authors originally conducted inference using the estimator in Aronow et al. (2015), assuming that dyads $m = d(i,j)$ cannot be correlated with $m' = d(k,l)$ unless they share a common member (i.e., there is no correlation across errors if i, j and k, l do not include a common unit). We will now

[23]There are 334 pairs of adjacent dyads and 591 pairs of connected dyads.

[24]This accommodates the row-by-EP-by-EPG clustering implemented by the authors. This is already an extension to Harmon et al. (2019) which assume that only seating neighbors would have correlated ε. One could extend this further by allowing connections across parties.

[25]For the sake of notational simplicity, the triangular array structure is once again made implicit by suppressing the M subscript.

compare this approach to using the variance estimator introduced in Section 4, which allows the error terms to exhibit arbitrary correlations as long as they are connected on the network represented by the adjacency relation of seating arrangements in Parliament.

Inspired by Harmon et al. (2019), we consider three specifications: (I) a simple linear regression model as given in (21); (II) the model (21) augmented with a flexible set of other demographic variables;[26] and (III) the model (21) with both a flexible set of other demographic variables and day-specific fixed effects.[27] When fixed effects are present in their original estimation, we estimate a within-difference model via OLS.[28]

6.3 Results

The main results of our empirical analysis are summarized in Table 3. Panel A displays the parameter estimates with the standard errors obtained from our proposed variance estimator from equation (10). The full model with all covariates, Specification (III), shows that our point-estimates are consistent in magnitude with the original estimates of Harmon et al. (2019) (column 7 of Table 4), as they are close to 0.006 (their original results) and stable across specifications. Hence, changes to point-estimates are not due to sample selection.[29] The positive coefficient for *SeatNeighbors* indicates that the MEPs sitting together tend to vote more similarly than those sitting apart, providing evidence in favor of their original hypothesis. The coefficients on the covariates (Panel C) are also quantitatively and qualitatively similar to those in their original paper. For instance, our estimates for *SameCountry* are 0.056, while their estimates are around 0.051, suggesting politicians from the same country are more likely to vote similarly on policies.

Panel B shows the standard errors for the regression coefficient of *SeatNeighbors* using different variance estimators. Building on Section 5, the panel compares three different types of heteroskedasticity robust estimators: namely, the Eicker-Huber-White, the dyadic-robust estimator (used in their original work), and our proposed network-robust estimator (under the assumption of

[26]Following Harmon et al. (2019), we include indicators whether country of origins, quality of education, freshman status and gender, respectively, are the same, as well as differences in ages and tenures. See Table 3 for details.

[27]The day-specific fixed effects are meant to control for unmeasured factors specific to a single day, possibly correlated to the covariates.

[28]Since our focus is to study how the standard errors are different when the far-away neighbors are considered, we do not pursue instrumenting the possible endogeneity in sitting together by means of name adjacency.

[29]Note that our dependent variable is equal to one if two MEPs vote the same and zero otherwise, while Harmon et al. (2019) code it as one if MEPs vote differently. Hence, to compare our estimates with their paper, the signs on the estimates of $SeatNeighbors$ should be flipped around.

$\pi_{N,M} = 1$ for the reason entertained in Remark 4.3). We note that, due to the smaller sample size, the standard errors in our exercise are larger than the original authors'. Hence, we prefer not to compare our results to theirs but, rather, focusing on the different specifications within our chosen sample.

As foreshadowed in the Monte Carlo simulations, the Eicker-Huber-White estimates are the smallest, followed by the dyadic-robust estimates, which, in turn, are smaller than the network-robust estimates. In fact, for Specification (III), the Eicker-Huber-White estimate is roughly 73% smaller than using the estimator accounting for network spillovers across dyads, while the dyadic-robust one is 22% smaller. This fact entails two implications. First, our finding provides empirical evidence in support of the existence of *indirect* positive spillovers among the MEPs: even distant connections may indirectly generate correlated behavior among politicians i and j. Second, the use of alternative estimators not accounting for such spillovers undercovers the true parameter and may generate biased hypothesis testing about the regression coefficient of *SeatNeighbors*. The difference in estimates appears quantitatively meaningful in this empirical example.

7 Conclusion

Dyadic data have widespread applications in social sciences. Researchers typically assume that dyads are uncorrelated if they do not share a common unit, an assumption that is leveraged in inference. We showed that this is not the case in many models whereby strategic interactions occur on a network: while data may be dyadic, the cross-sectional dependence may be much more complex and spillover beyond pairwise interactions. For instance, trade between countries may depend on trade between auxiliary partners, even those beyond the dyads being considered. In political economy, whether a politician votes with a colleague may depend on all intermediate seating neighbors beyond one's own immediate ones. We verified this using both theoretical results and Monte Carlo simulations. A new consistent variance estimator for parameters in a linear model with dyadic data were derived, which has correct asymptotic coverage and good finite sample properties. This estimator can account for network spillovers while existing results (e.g., Aronow et al. (2015), Tabord-Meehan (2019)) cannot. It is also simple to implement and provided stable inference in an application to concordance in legislative voting across pairs of members using real-world data. We found that, when accounting for network spillovers across dyads, standard errors grew by at least 20% compared to the original specification.

TABLE 3

Spillovers in Legislative Voting – Main Analysis

	Specification (I)	Specification (II)	Specification (III)
Panel A: Parameter estimates for Seat neighbors			
Seat neighbors	0.0069	0.0060	0.0060
Panel B: Standard errors for Seat neighbors			
Eicker-Huber-White	0.0031	0.0030	0.0030
Dyadic-robust	0.0075	0.0082	0.0087
Network-robust	0.0095	0.0104	0.0112
Panel C: Parameter estimates for other covariates			
Same country		0.0561	0.0562
		(0.0008)	(0.0008)
Same quality education		0.0030	0.0028
		(0.0007)	(0.0007)
Same freshman status		-0.0070	-0.0070
		(0.0008)	(0.0008)
Same gender		0.0004	0.0004
		(0.0007)	(0.0006)
Age difference		0.0007	0.0004
		(0.0004)	(0.0004)
Tenure difference		-0.0149	-0.0149
		(0.0006)	(0.0006)
Day-level FE	No	No	Yes

Note: Adjacency of MEPs is defined at the level of a row-by-EP-by-EPG. (See the note below Figure 3.) Independent variables are as follows: *Seat neighbors* is an indicator variable denoting whether both MEPs sit together; *Same country* represents an indicator for whether both MEPs are from the same country; *Same quality education* is an indicator showing whether both MEPs have the same quality of education background, measured by if both have the degree from top 500 universities; *Same freshman status* encodes whether both MEPs are freshman or not; *Age difference* is the difference in the MEPs' ages; and *Tenure difference* measures the difference in the MEPs' tenures.

To conclude, we would like to clarify that our goal in this exercise is neither to criticize dyadic-robust variance estimators, which are a fundamental part of the empiricist's toolkit, nor to suggest our approach should always be used. Rather, we wish to draw attention that researchers should fully specify the cross-sectional dependence in their model. If it is a purely dyadic environment or one where such spillovers might be negligible, then previous approaches suffice and will be equal to the one proposed here. However, as we have discussed above, many existing applications seemingly apply the latter method even if it is seemingly inappropriate to their setting. Hence, we recommend researchers to continue to fully specify their model, including full specification of their covariance structure, thereby clarifying what type of inference procedure is most appropriate for their environment.

Acknowledgments

We thank Aimee Chin, Hugo Jales, Taisuke Otsu, Bent E. Sørensen, Vitor Possebom and seminar participants at the University of Houston and at the 2022 Texas Econometrics Camp for valuable comments and suggestions.

References

Acemoglu, D., García-Jimeno, C., & Robinson, J. A. (2015). State capacity and economic development: A network approach. *American Economic Review*, *105*(8), 2364-2409. doi: https://doi.org/10.1257/aer.20140044

Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, *93*(1), 170-192. doi: https://doi.org/10.1257/000282803321455214

Aronow, P. M., Samii, C., & Assenova, V. A. (2015). Cluster–robust variance estimation for dyadic data. *Political Analysis*, *23*(4), 564-577. doi: https://doi.org/10.1093/pan/mpv018

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509-512. doi: https://doi.org/10.1126/science.286.5439.509

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238-249. doi: https://doi.org/10.1198/jbes.2010.07136

Cameron, A. C., & Miller, D. L. (2014). Robust inference for dyadic data.. (Working Paper)

Canen, N., Schwartz, J., & Song, K. (2020). Estimating local interactions among many agents who observe their neighbors. *Quantitative Economics*, *11*(1), 917-956. doi: https://doi.org/10.3982/QE923

Chiang, H. D., Kato, K., Ma, Y., & Sasaki, Y. (2021). Multiway cluster robust double/debiased machine learning. *Journal of Business & Economic Statistics*, 1-11. doi: https://doi.org/10.1080/07350015.2021.1895815

Chiang, H. D., Matsushita, Y., & Otsu, T. (2021). *Multiway empirical likelihood.* (Working Paper)

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics*, *92*(1), 1-45. doi: https://doi.org/10.1016/S0304-4076(98)00084-0

Davezies, L., D'Haultfœuille, X., & Guyonvarch, Y. (2018). *Asymptotic results under multiway clustering.* (Working Paper)

de Paula, A., Richards-Shubik, S., & Tamer, E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, *86*(1), 263-288. doi: https://doi.org/10.3982/ECTA13564

Erdös, P., & Rényi, A. (1959). On random graphs i. *Publicationes Mathematicae*, *6*, 290-297.

Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, *5*, 17-61.

Fafchamps, M., & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, *83*(2), 326 - 350. doi: https://doi.org/10.1257/aer.97.2.75

Graham, B. S. (2020a). Dyadic regression. In B. Graham & Á. d. Paula (Eds.), *The econometric analysis of network data* (1st ed., p. 23-40). Academic Press. doi: https://doi.org/10.1016/B978-0-12-811771-2.00008-0

Graham, B. S. (2020b). Network data. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, & R. L. Matzkin (Eds.), *Handbook of econometrics, volume 7a* (Vol. 7, p. 111-218). Elsevier. doi: https://doi.org/10.1016/bs.hoe.2020.05.001

Harmon, N., Fisman, R., & Kamenica, E. (2019). Peer effects in legislative voting. *American Economic Journal: Applied Economics*, *11*(4), 156-80. doi: https://doi.org/10.1257/app.20180286

Ibragimov, R., & Müller, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics*, *28*(4), 453-468. doi: https://doi.org/10.1198/jbes.2009.08046

Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*(1), 53-67. doi: https://doi.org/10.1016/j.jeconom.2009.10.025

Kojevnikov, D., Marmer, V., & Song, K. (2021). Limit theorems for network dependent random variables. *Journal of Econometrics*, *222*(2), 882-908. doi: https://doi.org/10.1016/j.jeconom.2020.05.019

Leung, M. P. (2021). *Network cluster-robust inference.* (Working Paper)

Leung, M. P. (2022). Causal inference under approximate neighborhood interference. *Econometrica*, *90*(1), 267-293. doi: https://doi.org/10.3982/ECTA17841

Leung, M. P., & Moon, H. R. (2021). *Normal approximation in large network models.* (Working Paper)

Lustig, H., & Richmond, R. J. (2020). Gravity in the exchange rate factor structure. *The Review of Financial Studies*, *33*(8), 3492-3540. doi: https://doi.org/10.1093/rfs/hhz103

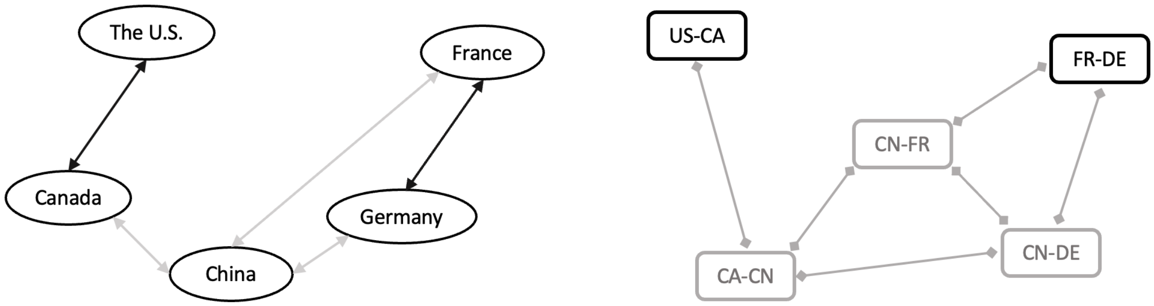
MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2022). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, *Forthcoming*. doi: https://doi.org/10.1016/j.jeconom.2022.04.001

Penrose, M. D., & Yukich, J. E. (2003). Weak laws of large numbers in geometric probability. *Annals of Applied Probability*, *13*(1), 277-303. doi: 10.1214/aoap/1042765669

Tabord-Meehan, M. (2019). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business & Economic Statistics*, *37*(4), 671-680. doi: https://doi.org/10.1080/07350015.2017.1409630

Vainora, J. (2020). *Network dependence.* (Working Paper)

FIGURE 1: Hypothetical Example of Trade Network

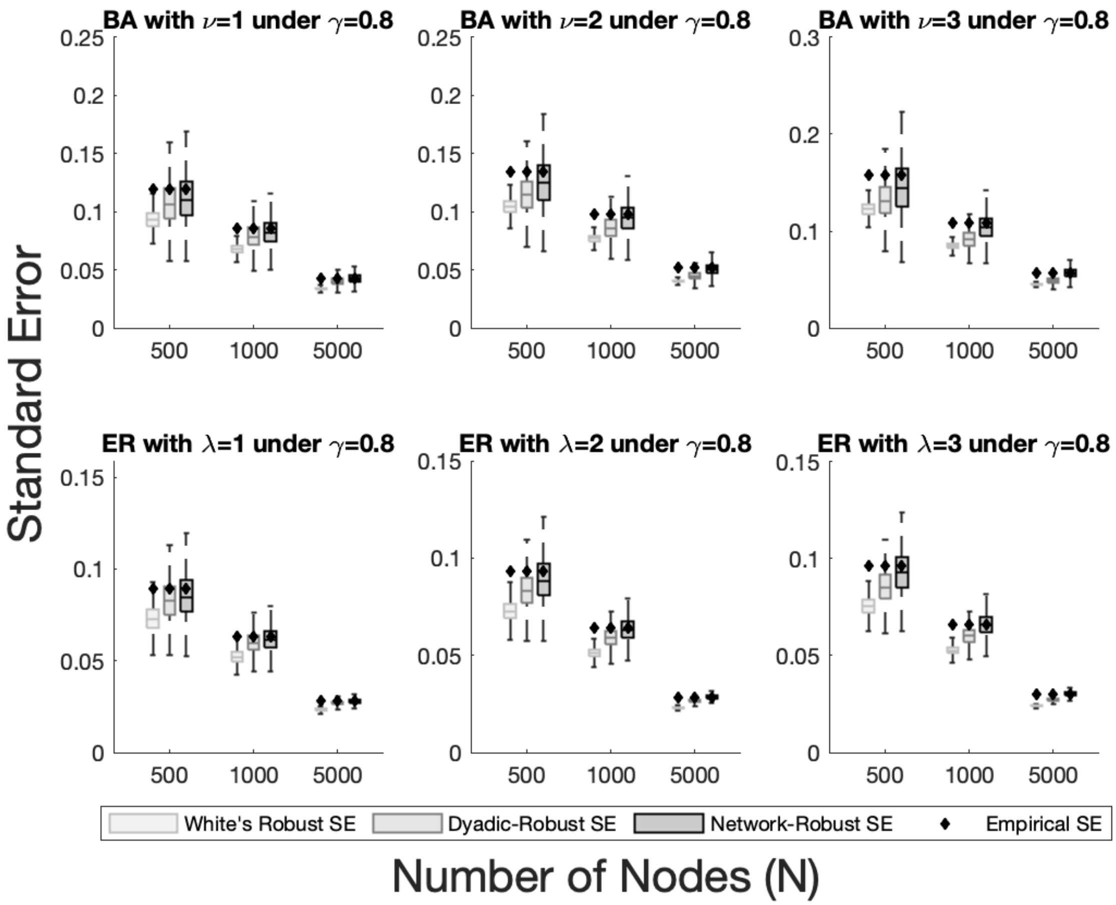


(A) Trade network across countries

(B) Trade network across pairs of countries

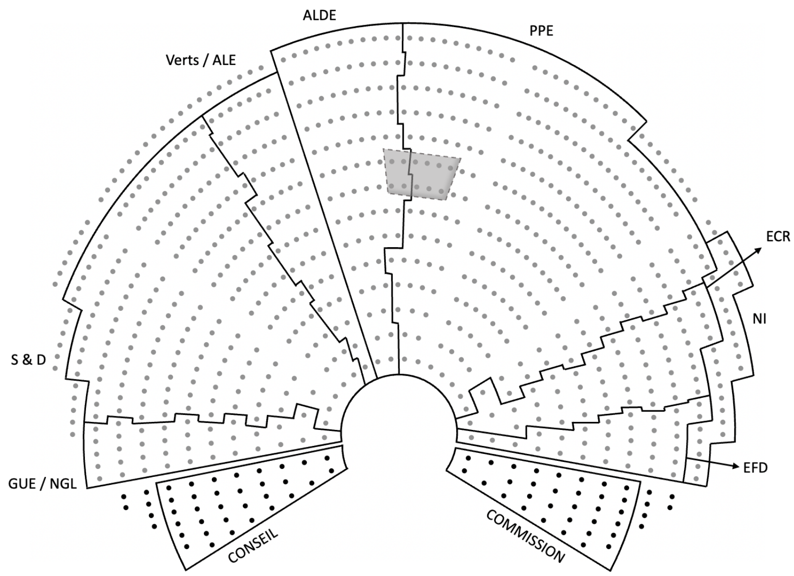
Note: Panel (A) shows a hypothetical example of trade networks across countries. Each double-headed arrow signifies the existence of a bilateral trade flow between countries. (The absence of a double-headed arrow is interpreted as the nonexistence of any type of trade flow between the countries.) In this example, though the Canada-US and France-Germany trades (denoted by black arrows) do not have a country in common, they might still be correlated through two indirect links (denoted by grey arrows): the US-Japan-South Korea-France trade, and the Germany-China-Canada trade. Panel (B) illustrates the network across pairs of countries (i.e., dyads). Round squares stand for bilateral trade flows between countries and the adjacency of these trade flows are represented by turned square arrows.

FIGURE 2: Boxplots of Standard Errors for Specifications 1 and 2 ($S = 2$, $\gamma = 0.8$)



Note: This figure shows boxplots describing the estimated standard errors and the empirical standard errors for various combinations of parameters under Specification 1 (Barabási-Albert networks) and Specification 2 (Erdös-Renyi networks). The horizontal axis shows the number of nodes and the vertical axis represents the the standard error of the coefficient. The shaded boxes represent the 25th, 50th and 75th percentiles of estimated standard errors with the whiskers indicating the most extreme values that are not considered as outliers. The light-grey box illustrates the Eicker-Huber-White standard error, the medium-grey one the dyadic-robust standard error and the dark-grey one the network-robust standard error. The diamonds stand for the empirical standard error, defined as the standard deviation of the estimates of the regression coefficient. The estimator is not covering the true standard error when the diamond is outside of the shaded area.

FIGURE 3: Seating Plan at the European Parliament: Strasbourg, September 14, 2009



(A) Zoomed-Out View

		ALDE			PPE			
Row 10	⋯	⋮	⋮	⋮	⋮	⋮	⋮	⋯
Row 9	⋯	Seat 36: ALFANO	Seat 37: ALVARO	Seat 38: CAVADA	Seat 39: COELHO	Seat 40: COLLINO	Seat 41: COMI	⋯
Row 8	⋯	Seat 32: OJULAND	Seat 33: OVIIR	Seat 34: BACH	Seat 35: BALDASSARRE	Seat 36: BALZ	⋯	
Row 7	⋯	⋮	⋮	⋮	⋮	⋮	⋮	⋯

(B) Zoomed-In View

Note: Panel (A) illustrates a zoomed-out view of a seating plan for the European parliament in Strasbourg on September 14, 2009. Grey circles are individual MEPs, while black circles embody members of conseil and commission. The associated party (EPG) is denoted at the top. Panel (B) provides a zoomed-in view elaborating on the part of Panel (A) marked by the dotted trapezoid shaded in grey. Alafano and Alvaro are treated as adjacent because they are sitting next to each other and belong to the same political party, i.e., ALE. Similarly, Ojuland and Oviir are considered to be adjacent. On the other hand, following the original authors, Alvaro and Cavada are not regarded as adjacent though they are seated together because they belong to different political parties, i.e., ALE and PPE, respectively. In terms of dyad-level adjacency, Cavada-Coelho and Coelho-Collino are adjacent dyads as they share Coelho, whereas Cavada-Coelho and Collino-Comi are not adjacent, but they are still connected as they have indirect paths to one another along the dyadic network.

A Mathematical Setup

This section lays out the mathematical setup of our model in more detail, heavily drawing from Kojevnikov et al. (2021).

We first define a collection of pairs of sets of dyads. For any positive integers a, b and s, define

$$\mathcal{P}_M(a, b; s) \coloneqq \{(A, B) : A, B \subset \mathcal{M}_N, |A| = a, |B| = b, \rho_M(A, B) \geq s\},$$

where

$$\rho_M(A, B) \coloneqq \min_{m \in A} \min_{m' \in B} \rho_M(m, m'), \tag{22}$$

with $\rho_M(m, m')$ denoting the geodesic distance between dyads m and m'. In words, the set $\mathcal{P}_M(a, b; s)$ collects all two distinct sets of active dyads whose sizes are a and b and that have no dyads in common.

Next we consider a collection of bounded Lipschitz functions. Define

$$\mathcal{L}_K \coloneqq \{\mathcal{L}_{K,c} : c \in \mathbb{N}\},$$

where

$$\mathcal{L}_{K,c} \coloneqq \left\{f : \mathbb{R}^{K \times c} \to \mathbb{R} : \|f\|_\infty < \infty, \text{Lip}(f) < \infty\right\},$$

with $\|\cdot\|_\infty$ representing the supremum norm and $\text{Lip}(f)$ being the Lipschitz constant.[30] In words, the set $\mathcal{L}_{K,c}$ collects all the bounded Lipschitz functions on $\mathbb{R}^{K \times c}$ and the set $\mathcal{L}_K$ moreover gathers such sets with respect to $c \in \mathbb{N}$.

Lastly, we write

$$Y_{M,A} \coloneqq (Y_{M,m})_{m \in A},$$

and $Y_{M,B}$ is analogously defined. Let $\{\mathcal{C}_M\}_{M \geq 1}$ denote a sequence of σ-algebras and be suppressed as $\{\mathcal{C}_M\}$.

The network dependent random variables are characterized by the upper bound of their covariances.

[30]It is immediate to see that $\mathbb{R}$ is a normed space with respect to the Euclidean norm, while the $\mathbb{R}^{K \times c}$ can be equipped with the norm $\rho_c(x, y) \coloneqq \sum_{\ell=1}^{c} \|x_\ell - y_\ell\|$ where $x, y \in \mathbb{R}^{K \times c}$ and $\|z\| \coloneqq (z'z)^{\frac{1}{2}}$, thereby the Lipschitz constant is defined as $\text{Lip}(f) \coloneqq \min\left\{w \in \mathbb{R} : |f(x) - f(y)| \leq w\rho_c(x, y) \; \forall x, y \in \mathbb{R}^{K \times c}\right\}$.

Definition A.1 (Conditional ψ-Dependence given $\{\mathcal{C}_M\}$: Kojevnikov et al. (2021), Definition 2.2)**.** A triangular array $\{Y_{M,m} \in \mathbb{R}^K : M \geq 1, m \in \{1, \dots, M\}\}$ is called conditionally ψ-dependent given $\{\mathcal{C}_M\}$, if for each $M \in \mathbb{N}$, there exist a $\mathcal{C}_M$-measurable sequence $\theta_M := \{\theta_{M,s}\}_{s\geq 0}$ with $\theta_{M,0} = 1$, and a collection of nonrandom function $(\psi_{a,b})_{a,b\in\mathbb{N}}$ where $\psi_{a,b} : \mathcal{L}_{K,a} \times \mathcal{L}_{K,b} \to [0.\infty)$, such that for all $(A, B) \in \mathcal{P}_M(a, b; s)$ with $s > 0$ and all $f \in \mathcal{L}_{K,a}$ and $g \in \mathcal{L}_{K,b}$,

$$|Cov\left(f(Y_{M,A}), g(Y_{M,B}) \mid \mathcal{C}_M\right)| \leq \psi_{a,b}(f, g)\theta_{M,s} \quad a.s.$$

Intuitively, this definition states that the upper bound must be decomposed into two components, one of which is deterministic and depends on nonlinear Lipschitz functions f and g, and the other of which is stochastic and depends only on the distance of the random variables on the underlying network. The former, nonrandom component reflects the scaling of the random variables as well as that of the Lipschitz transformations, while the latter random part stands for the covariability of the two random variables. We follow Kojevnikov et al. (2021) in assuming boundedness for these two components.

Assumption A.1 (Kojevnikov et al. (2021), Assumption 2.1)**.** The triangular array $\{Y_{M,m} \in \mathbb{R}^K : M \geq 1, m \in \{1, \dots, M\}\}$ is conditionally ψ-dependent given $\{\mathcal{C}\}$ with the dependence coefficients $\{\theta_{M,s}\}$ satisfying the following conditions:

(a) There exists a constant $C > 0$ such that

$$\psi_{a,b}(f, g) \leq C \times ab\left(\|f\|_\infty + \text{Lip}(f)\right)\left(\|g\|_\infty + \text{Lip}(g)\right);$$

(b) $\sup_{M\geq 1} \max_{s\geq 1} \theta_{M,s} < \infty$ a.s.

Assumption A.1 is maintained throughout the paper and employed to show asymptotic properties of our estimators such as the consistency and asymptotic normality and the consistency of the network-robust variance estimator for dyadic data.

To prove Theorem 3.2, we make use of the following assumptions which strengthen their counterparts in the proof of consistency.

Assumption A.2 (Conditional Finite Moment of ε_m)**.** There exists $p > 4$ such that

$$\sup_{N\geq 1} \max_{m\in\mathcal{M}_N} E\left[|\varepsilon_m|^p \mid \mathcal{C}_M\right] < \infty \quad a.s.$$

Assumption A.2 allows for the same interpretation as Assumption 3.1, i.e., the random error term ε_m cannot be too large, conditional on a common component. This assumption, however, is more stringent than the previous one because it now requires the finiteness of a higher moment of ε_m. When coupled with Assumption 2.3, Assumption A.2 implies that Assumption 3.3. of Kojevnikov et al. (2021) for each $Y^u_{M,m}$ with $u \in \{1, \dots, K\}$.

Assumption A.3 (Kojevnikov et al. (2021), Assumption 3.4)**.** There exists a positive sequence $r_M \to \infty$ such that for $k = 1, 2$,

$$\frac{M^2 \theta^{1-1/p}_{M,r_M}}{\sigma_M} \xrightarrow{a.s.} 0,$$
$$\frac{M}{\sigma_M^{2+k}} \sum_{s \geq 0} c_n(s, r_M; k) \theta_{M,s}^{1-\frac{2+k}{p}} \xrightarrow{a.s.} 0,$$

as $M \to \infty$, where $p > 4$ is the same as the one that appears in Assumption A.2.

Similar to Assumption 3.2, this assumption binds the covariance of the random variables, the dependence reflected in the dependence coefficients, and the underlying network. The first part of this assumption requires that σ_M grows at least as fast as M does. Kojevnikov et al. (2021) claim that it is satisfied if the long-run variance $\frac{\sigma_M^2}{M}$ is bounded away from c^2 for all $M \geq 1$ for some constant c independent of M, i.e., $\sigma_M \geq c\sqrt{M}$. The second part can be interpreted in an analogous manner to Assumption 3.2. That is, given σ_M growing at least at the same rate of M, the composite of the density of the network and the magnitude of the correlations of the random variables must decay fast enough.

B Related Literature

Remark B.1 (Vainora (2020))**.** Vainora (2020) develops a network version of a stationary concept, what he calls weakly $\mathcal{C}$-stationary. Weak $\mathcal{C}$-stationary, in our context, requires covariances between two random variables to be equal as long as the geodesic distance is the same. Definition 2.2, on the other hand, only refers to the upper bound of covariances with respect to the geodesic distance. Vainora (2020) further establishes a Law of Large Numbers and Central Limit Theorem (CLT) for random variables whose dependence structure is tied to the network, with applications to linear regression models. But the consistency of the standard errors, the central object of our interest, is left unexplored.

Remark B.2 (Leung and Moon (2021))**.** Leung and Moon (2021) derive a CLT of an estimator involving a complex form of dependence along the network. As they are primarily interested in network formation models, which could be rationalized by strategic games of network formation, their estimator is applicable only for a discrete-choice-type regression model. Our framework complements their work for the case of linear regression models with a continuous outcome but with exogenous network formation.

Remark B.3 (Leung (2021))**.** A notable literature which shares motivations with us is Leung (2021), who exploits the asymptotic theory of Kojevnikov et al. (2021) to perform robust inference under network dependence. His approach is different from ours in that he derives conditions under which the use of cluster-robust estimators for inference is justified. He finds a set of conditions that ensures the network is asymptotically segregated into disjoint subnetworks that are not connected with each other, and then makes use of the existing cluster-robust estimators. Our focus is on the role of network spillovers within the particular class of linear models with dyadic data, where the use of cluster-robust estimators still has the potential drawback raised in Example 2.1.

Remark B.4 (Leung (2022))**.** Leung (2022) considers robust causal inference under network interference by applying the results of Kojevnikov et al. (2021). Although he analyzes a linear regression model akin to (1), his work differs from ours in two important ways. First, the source of randomness in his model is the uncertainty in assignments (*design-based uncertainty*), thereby featuring no random shocks, from which the stochastic nature of our model stems. Second, his analysis presumes an environment where the population is finite and focuses on characterizing the bias of the variance estimator in a finite population model.

Remark B.5 (Multiway Clustering)**.** A separate line of research considers regression models for multiple index data and proposes variance estimators appropriate for multiway clustering: see, for example, Davezies, D'Haultfœuille, and Guyonvarch (2018), Chiang, Kato, et al. (2021) and Chiang, Matsushita, and Otsu (2021), to name but a few (see MacKinnon, Nielsen, and Webb (2022) for review). Multiway clustering methods typically assume that the data can be segmented into independent clusters. Networks, however, do not always allow for such divisible structures, which may cast doubts on the validity of any inference drawn upon multiway clustering.

C Additional Monte Carlo Simulation Results

C.1 Summary Statistics

Table 4 shows summary statistics (i.e., the average and maximum degrees) of the networks across nodes that are used in our simulation study. By construction, the maximum degree and the average degree monotonically increase in the parameters for both specifications.

TABLE 4

The Average and Maximum Degrees of Networks among Nodes in the Simulations

		Specification 1			Specification 2		
N		$\nu = 1$	$\nu = 2$	$\nu = 3$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
500	d_{max}	23	40	41	5	7	8
	d_{ave}	0.8020	1.5780	2.3540	0.4760	0.9680	1.4800
1000	d_{max}	26	36	47	4	7	8
	d_{ave}	0.8590	1.7000	2.5410	0.4980	0.9810	1.5010
5000	d_{max}	53	125	130	6	9	10
	d_{ave}	0.9326	1.8618	2.7910	0.4952	1.0016	1.5114

Notes: Observation units in this table are nodes (individuals) as usual in the literature. The maximum degree, d_{max}, means the maximum number of nodes that are adjacent to a node, and the average degree, d_{ave}, is the average number of nodes adjacent to each node of the network.

C.2 $S = 2$ and $\gamma = 0.8$

Table 5 describes the standard deviations of the estimated regression coefficients (what Aronow et al. (2015) calls the true standard errors) and the means of the estimated standard errors for each variance estimator. The round brackets indicate the biases of each estimate relative to the true standard error in percentage (%). For instance, the Eicker-Huber-White variance estimator and the dyadic-robust variance estimator, when applied to Specification 1 with $\nu = 3$, underestimate the true standard error by 21.45% and 14.14%, respectively.

TABLE 5

Means and Biases of the Standard Errors: $N = 5000$, $S = 2$, $\gamma = 0.8$.

	Specification 1			Specification 2		
	$\nu = 1$	$\nu = 2$	$\nu = 3$	$m = 1$	$m = 2$	$m = 3$
True	0.0430	0.0518	0.0570	0.0285	0.0283	0.0302
Eicker-Huber-White	0.0337	0.0404	0.0448	0.0234	0.0229	0.0239
(Bias %)	(-21.61)	(-21.92)	(-21.45)	(-17.91)	(-19.14)	(-20.68)
Dyadic-robust	0.0403	0.0453	0.0490	0.0270	0.0266	0.0275
(Bias %)	(-6.19)	(-12.54)	(-14.14)	(-5.01)	(-6.12)	(-8.93)
Network-robust	0.0425	0.0509	0.0565	0.0280	0.0285	0.0302
(Bias %)	(-1.09)	(-1.78)	(-0.92)	(-1.70)	(0.58)	(0.09)

Note: This table shows the standard deviations of the estimated regression coefficients (the true standard error) and the means of the estimated standard errors for each variance estimator with the round brackets indicating the biases relative to the true standard error in percentage (%). To facilitate the comparison, the biases are rounded off to the second decimal places.

C.3 $S = 2$ and $\gamma = 0.2$

Table 6 presents the empirical coverage probability and average length of confidence intervals for β at 5% nominal size when $S = 2$ and $\gamma = 0.2$. The associated boxplots are given in Figure 5. Since the magnitude of spillovers is now much smaller than the case of $\gamma = 0.8$, there are only minor differences in performance between the network-robust variance estimator and the other two existing methods (namely, the Eicker-Huber-White and dyadic-robust variance estimators). In terms of convergence, the comparable performance of the dyadic-robust-variance estimator is evident in Figure 5.

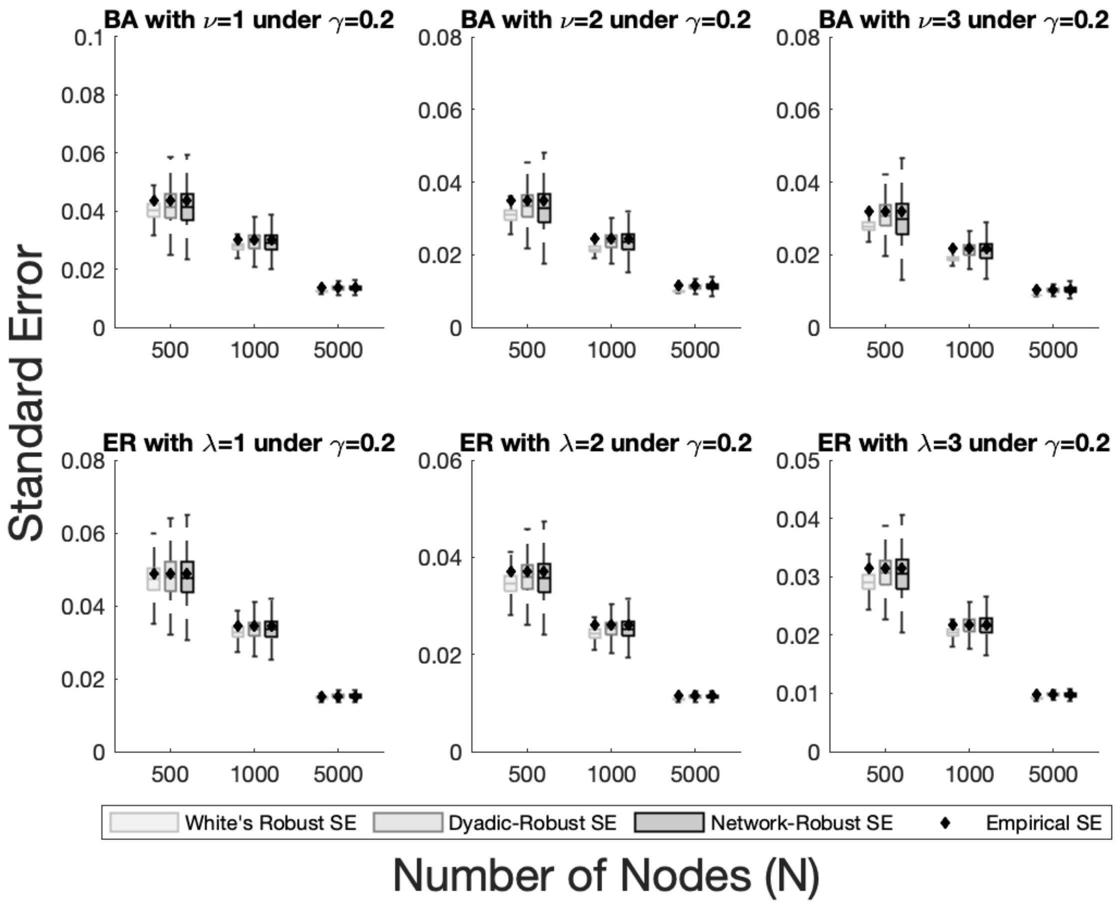
TABLE 6

The empirical coverage probability and average length of confidence intervals for β at 95% nominal level: $S = 2$, $\gamma = 0.2$.

	N	Specification 1 $\nu = 1$	$\nu = 2$	$\nu = 3$	Specification 2 $\lambda = 1$	$\lambda = 2$	$\lambda = 3$
		Coverage Probability					
Eicker-Huber-White	500	0.9286	0.9186	0.9108	0.9434	0.9292	0.9308
	1000	0.9320	0.9158	0.9148	0.9338	0.9322	0.9322
	5000	0.9200	0.9110	0.9124	0.9434	0.9382	0.9308
Dyadic-robust	500	0.9342	0.9350	0.9336	0.9454	0.9368	0.9422
	1000	0.9454	0.9376	0.9458	0.9398	0.9422	0.9486
	5000	0.9446	0.9448	0.9432	0.9486	0.9490	0.9472
Network-robust	500	0.9284	0.9246	0.9162	0.9428	0.9360	0.9384
	1000	0.9414	0.9294	0.9370	0.9392	0.9410	0.9456
	5000	0.9454	0.9476	0.9418	0.9494	0.9492	0.9470
		Average Length of the Confidence Intervals					
Eicker-Huber-White	500	0.1578	0.1214	0.1092	0.1860	0.1360	0.1141
	1000	0.1088	0.0846	0.0743	0.1290	0.0955	0.0799
	5000	0.0486	0.0388	0.0346	0.0579	0.0423	0.0357
Dyadic-robust	500	0.1648	0.1316	0.1213	0.1890	0.1410	0.1205
	1000	0.1158	0.0931	0.0833	0.1319	0.0994	0.0848
	5000	0.0532	0.0439	0.0398	0.0594	0.0443	0.0381
Network-robust	500	0.1637	0.1291	0.1174	0.1885	0.1404	0.1196
	1000	0.1154	0.0922	0.0825	0.1318	0.0993	0.0848
	5000	0.0533	0.0440	0.0401	0.0594	0.0444	0.0382

Note: The upper-half of the table displays the empirical coverage probability of the asymptotic confidence interval for β, and the lower-half showcases the average length of the estimated confidence intervals. One computational issue that plagues the Monte Carlo simulation is the potential lack of positive-semi-definiteness of the estimated variance-covariance matrix. In general, this problem prevails only when the sample size (N) is small. In our case, when $N = 500$, four variance estimates out of five thousands take negative values. We deal with this issue by first applying the eigenvalue decomposition to the estimated variance-covariance matrix and then augmenting the diagonal matrix of eigenvalues by a small constant, followed by pre- and post-multiplications by the matrix of eigenvectors to obtain the updated estimate for the variance-covariance matrix. As the sample size (N) increases, the empirical coverage probability approaches 0.95, the nominal level. This convergence is accompanied by the shrinking average length of confidence intervals.

FIGURE 5: Boxplots of Standard Errors for Specifications 1 and 2 ($S = 2$, $\gamma = 0.2$)



Note: This picture shows boxplots describing the estimated standard errors and the empirical standard errors for various combinations of parameters under Specification 1 (Barabási-Albert networks) and Specification 2 (Erdös-Renyi networks). The horizontal axis shows the number of nodes and the vertical axis represents the the standard error of the coefficient. The shaded boxes represent the 25th, 50th and 75th percentiles of estimated standard errors with the whiskers indicating the most extreme values that are not considered as outliers. The light-grey box illustrates the Eicker-Huber-White standard error, the medium-grey one the dyadic-robust standard error and the dark-grey one the network-robust standard error. The diamonds stand for the empirical standard error, defined as the standard deviation of the estimates of the regression coefficient. This figure showcases the boxplots for the case when $\gamma = 0.2$.

C.4 $S = 1$

For comparison purposes, this subsection explores the results for $S = 1$. If $S = 1$, there are no higher-order correlations beyond direct (adjacent) neighbors. Then, the network-robust variance estimator ought to coincide with the dyadic-robust variance estimator by definition, for any γ, as pointed out in Section 2.1 and Example 2.1. This is verified below for the case of $\gamma = 0.8$. Table 7 shows the simulation results.

TABLE 7

The empirical coverage probability and average length of confidence intervals for β at 95% nominal level: $S = 1, \gamma = 0.8$.

	N	Specification 1 $\nu = 1$	$\nu = 2$	$\nu = 3$	Specification 2 $\lambda = 1$	$\lambda = 2$	$\lambda = 3$
				Coverage Probability			
Eicker-Huber-White	500	0.8804	0.8676	0.8734	0.8906	0.8768	0.8692
	1000	0.8678	0.8810	0.8710	0.8984	0.8864	0.8856
	5000	0.8752	0.8652	0.8742	0.8996	0.8910	0.8778
Dyadic-robust	500	0.9292	0.9304	0.9384	0.9366	0.9416	0.9368
	1000	0.9364	0.9426	0.9432	0.9428	0.9454	0.9484
	5000	0.9474	0.9414	0.9498	0.9452	0.9518	0.9506
Network-robust	500	0.9292	0.9304	0.9384	0.9366	0.9416	0.9368
	1000	0.9364	0.9426	0.9432	0.9428	0.9454	0.9484
	5000	0.9474	0.9414	0.9498	0.9452	0.9518	0.9506
				Average Length of the Confidence Intervals			
Eicker-Huber-White	500	0.3282	0.2901	0.2881	0.2664	0.2377	0.2235
	1000	0.2321	0.2088	0.1964	0.1887	0.1665	0.1564
	5000	0.1131	0.1042	0.0980	0.0844	0.0742	0.0704
Dyadic-robust	500	0.3934	0.3591	0.3603	0.3104	0.2888	0.2776
	1000	0.2853	0.2625	0.2500	0.2227	0.2037	0.1950
	5000	0.1428	0.1330	0.1259	0.0998	0.0913	0.0882
Network-robust	500	0.3934	0.3591	0.3603	0.3104	0.2888	0.2776
	1000	0.2853	0.2625	0.2500	0.2227	0.2037	0.1950
	5000	0.1428	0.1330	0.1259	0.0998	0.0913	0.0882

Note: The upper-half of the table displays the empirical coverage probability of the asymptotic confidence interval for β, and the lower-half showcases the average length of the estimated confidence intervals. As the sample size (N) increases, the empirical coverage probability approaches 0.95, the nominal level. This convergence is accompanied by the shrinking average length of confidence intervals.

D Proof of Main Theorems

D.1 Identification of β

Proof of Proposition 2.1. For each $m \in \mathcal{M}_N$, premultiply the model (1) by $x_{M,m}$ to obtain

$$x_{M,m} y_{M,m} = x_{M,m} x'_{M,m} \beta + x_{M,m} \varepsilon_{M,m} \qquad \forall m \in \mathcal{M}_N.$$

Summing up the both sides over the index set $\mathcal{M}_N$ yields

$$\sum_{m \in \mathcal{M}_N} x_{M,m} y_{M,m} = \sum_{m \in \mathcal{M}_N} x_{M,m} x'_{M,m} \beta + \sum_{m \in \mathcal{M}_N} x_{M,m} \varepsilon_{M,m}.$$

Taking the expectation with respect to $\{(x_{M,m}, y_{M,m}, \varepsilon_{M,m})\}_{m \in \mathcal{M}_N}$,

$$E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} y_{M,m}\right] = E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} x'_{M,m}\right] \beta + E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} \varepsilon_{M,m}\right].$$

Since it holds by Assumption 2.2 (iv) that

$$\begin{aligned}
E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} \varepsilon_{M,m}\right] &= \sum_{m \in \mathcal{M}_N} E\left[E\left[x_{M,m} \varepsilon_{M,m} \mid X_M\right]\right] \\
&= \sum_{m \in \mathcal{M}_N} E\left[x_{M,m} \underbrace{E\left[\varepsilon_{M,m} \mid X_M\right]}_{0}\right] \\
&= 0,
\end{aligned}$$

then

$$E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} y_{M,m}\right] = E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} x'_{M,m}\right] \beta.$$

Next, Assumption 2.2 (iii) ensures existence of the inverse of the expectation term in the right hand side, so that we have

$$\beta = E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} x'_{M,m}\right]^{-1} E\left[\sum_{m \in \mathcal{M}_N} x_{M,m} y_{M,m}\right].$$

This proves the identification of the regression parameter β. $\square$

D.2 Consistency of $\hat{\beta}$

Proof of Theorem 3.1. From (8), (13) and (14), we can write

$$\hat{\beta} - \beta = \left(\sum_{j \in \mathcal{M}_N} x_{M,j} x'_{M,j}\right)^{-1} \sum_{m \in \mathcal{M}_N} x_{M,m} \varepsilon_{M,m}$$

$$= \frac{1}{M} \sum_{m \in \mathcal{M}_N} Y_{M,m},$$

and then

$$\hat{\beta}_u - \beta_u = \frac{1}{M} \sum_{m \in \mathcal{M}_N} Y^u_{M,m},$$

where $\hat{\beta}_u$ and β_u, respectively, denote the u-th entry of $\hat{\beta}$ and β, and thus are one-dimensional. In light of Assumption 2.2 (iv), it holds that for each $m \in \mathcal{M}_N$

$$\begin{aligned}
E\left[Y_{M,m} \mid \mathcal{C}_M\right] &= E\left[\left(\frac{1}{M} \sum_{j \in \mathcal{M}_N} x_{M,j} x'_{M,j}\right)^{-1} x_{M,m} \varepsilon_{M,m} \,\middle|\, \mathcal{C}_M\right] \\
&= \left(\frac{1}{M} \sum_{j \in \mathcal{M}_N} x_{M,j} x'_{M,j}\right)^{-1} x_{M,m} \underbrace{E\left[\varepsilon_{M,m} \mid \mathcal{C}_M\right]}_{0} \\
&= 0,
\end{aligned}$$

and thus

$$E\left[Y^u_{M,m} \mid \mathcal{C}_M\right] = 0.$$

By Theorem 3.1 of Kojevnikov et al. (2021),

$$\begin{aligned}
&\left\| \frac{1}{M} \sum_{m \in \mathcal{M}_N} \left(Y^u_{M,m} - \underbrace{E\left[Y^u_{M,m} \mid \mathcal{C}_M\right]}_{0} \right) \right\|_{\mathcal{C}_M, 1} \xrightarrow{a.s.} 0 \\
\therefore &\left\| \frac{1}{M} \sum_{m \in \mathcal{M}_N} Y^u_{M,m} \right\|_{\mathcal{C}_M, 1} \xrightarrow{a.s.} 0,
\end{aligned}$$

so that

$$\begin{aligned}
E\left[\left|\hat{\beta}_u - \beta_u\right|\right] &= E\left[E\left[\left|\hat{\beta}_u - \beta_u\right| \,\middle|\, \mathcal{C}_M\right]\right] \\
&= E\left[\|\hat{\beta}_u - \beta_u\|_{\mathcal{C}_M, 1}\right] \\
&= E\left[\left\| \frac{1}{M} \sum_{m \in \mathcal{M}_N} Y^u_{M,m} \right\|_{\mathcal{C}_M, 1}\right] \\
&\to 0,
\end{aligned}$$

where the last implication is a consequence of the Dominated Convergence Theorem.

Since it holds by the Markov inequality that for any $c > 0$

$$\Pr\left(\left|\hat{\beta}_u - \beta_u\right| > c\right) \leq \frac{E\left[\left|\hat{\beta}_u - \beta_u\right|\right]}{c},$$

it then follows that

$$\Pr\left(\left|\hat{\beta}_u - \beta_u\right| > c\right) \to 0,$$

as $N \to \infty$. Hence we have

$$\hat{\beta}_u \xrightarrow{p} \beta_u \qquad as \quad N \to \infty.$$

Finally, we can invoke the Cramér-Wold device to obtain

$$\hat{\beta} \xrightarrow{p} \beta \qquad as \quad N \to \infty,$$

as desired. □

D.3 Asymptotic Normality of $\hat{\beta}$

Proof of Proposition 3.2. First of all, we prove

$$\frac{S_M^u}{\sigma_M} \xrightarrow{d} \mathcal{N}(0,1),$$

as $N \to \infty$. Denote $\tilde{S}_M^u := \frac{S_M^u}{\sigma_M}$. Under A.2 and A.3, it holds by Theorem 3.2 of Kojevnikov et al. (2021) that for any $\epsilon > 0$, there exists $M_0 > 0$ such that for each $M > M_0$ and for each $x \in \mathbb{R}$,

$$\sup_{t \in \mathbb{R}} \left|\Pr(\tilde{S}_M^u \leq t \mid X = x) - \Phi(t)\right| < \epsilon. \tag{23}$$

Then, by the law of total probability, we have

$$\begin{aligned}
\left|\Pr(\tilde{S}_M^u \leq t) - \Phi(t)\right| &= \left|\int \Pr(\tilde{S}_M^u \leq t \mid X = x) dF_X(x) - \Phi(t)\right| \\
&= \left|\int \Pr(\tilde{S}_M^u \leq t \mid X = x) - \Phi(t) dF_X(x)\right| \\
&\leq \int \left|\Pr(\tilde{S}_M^u \leq t \mid X = x) - \Phi(t)\right| dF_X(x) \\
&\leq \int \sup_{t \in \mathbb{R}} \left|\Pr(\tilde{S}_M^u \leq t \mid X = x) - \Phi(t)\right| dF_X(x), \qquad (24)
\end{aligned}$$

where $F_X(\cdot)$ denotes the probability distribution function of X. Now pick arbitrarily $\epsilon > 0$. Then there exists $M_0 > 0$ such that for each $M > M_0$

$$\begin{aligned} \left|\Pr(\tilde{S}_M^u \leq t) - \Phi(t)\right| &\leq \int \underbrace{\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}_M^u \leq t \mid X = x) - \Phi(t)\right|}_{<\epsilon} dF_X(x) \\ &\leq \int \epsilon dF_X(x) \\ &\leq \epsilon, \end{aligned} \tag{25}$$

where the first and second inequalities come from (24) and (23), respectively. Since the right hand side of (25) does not depend on t, we then have that for each $M > M_0$,

$$\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}_M^u \leq t) - \Phi(t)\right| \leq \epsilon,$$

which implies

$$\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}_M^u \leq t) - \Phi(t)\right| \to 0 \qquad as \quad M \to \infty,$$

Here by Assumption 2.1, $N \to \infty$ implies $M \to \infty$, so that

$$\sup_{t\in\mathbb{R}} \left|\Pr(\tilde{S}_M^u \leq t) - \Phi(t)\right| \to 0 \qquad as \quad N \to \infty.$$

Then we have

$$\frac{S_M^u}{\sigma_M} \xrightarrow{d} \mathcal{N}(0,1) \qquad as \quad N \to \infty.$$

Next this can be combined with Assumption 3.3 by using the Slutsky's Theorem, yielding that

$$\frac{S_M^u}{\tau_M \sqrt{\pi_{N,M}}} \xrightarrow{d} \mathcal{N}(0,1) \qquad as \quad N \to \infty.$$

Moreover, applying the Cramér-Wold device gives

$$\frac{\tau_M^{-1}}{\sqrt{\pi_{N,M}}} S_M \xrightarrow{d} \mathcal{N}(0, I_K) \qquad as \quad N \to \infty,$$

where I_K is the $K \times K$ identity matrix and τ_M is understood as the variance-covariance matrix.[31] Now notice that by the definition of S_M we have

$$\hat{\beta} - \beta = \frac{1}{M} S_M,$$

[31] To economize on notation, we use the same notation τ_M to denote the case of one-dimensional parameter and the case of multiple-dimensional parameters as we believe this causes little confusion.

so that we obtain

$$\sqrt{N}\left(\hat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}(0,AVar(\hat{\beta}))\qquad as\quad N\to\infty,$$

where

$$AVar(\hat{\beta}):=\lim_{N\to\infty}\frac{N\pi_{N,M}}{M^2}E\left[\left(\sum_{k\in\mathcal{M}_N}x_{M,k}x'_{M,k}\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N}\varepsilon_{M,m}\varepsilon'_{M,m'}x_{M,m}x'_{M,m'}\right)\left(\sum_{k\in\mathcal{M}_N}x_{M,k}x'_{M,k}\right)^{-1}\right],$$

which is assumed to be well-defined.

□

D.4 Consistency of $\widehat{Var}(\hat{\beta})$

Proof of Theorem 4.1. Denote the variance of $\frac{S_M}{\sqrt{M}}$ as $V_{N,M}:=Var\left(\frac{S_M}{\sqrt{M}}\right)$. It can readily be shown that $V_{N,M}$ takes the form of

$$V_{N,M}=\sum_{s\geq 0}\Omega_{N,M}(s),$$

where

$$\Omega_{N,M}(s):=\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}E\left[Y_{M,m}Y'_{M,j}\right].$$

Following Kojevnikov et al. (2021), we define the kernel heteroskedasticity and autocorrelation consistent (HAC) estimator of $V_{N,M}$ as

$$\hat{V}_{N,M}:=\sum_{s\geq 0}\omega_M(s)\hat{\Omega}_{N,M}(s),$$

where

$$\omega_M(s):=\omega\left(\frac{s}{b_M}\right)$$
$$\hat{\Omega}_{N,M}(s):=\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\hat{Y}_{M,m}\hat{Y}'_{M,j}.$$

Moreover, we define an empirical analogy of $V_{N,M}$, though infeasible, by

$$\tilde{V}_{N,M}:=\sum_{s\geq 0}\omega_M(s)\tilde{\Omega}_{N,M}(s),$$

where

$$\tilde{\Omega}_{N,M}(s):=\frac{1}{M}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}Y_{M,m}Y'_{M,j}.$$

Additionally, we denote a conditional version of $V_{N,M}$ by $V_{N,M}^c \coloneqq Var(\frac{S_M}{\sqrt{M}} \mid \mathcal{C}_M)$, i.e.,

$$V_{N,M}^c = \sum_{s \geq 0} \omega_M(s) \Omega_{N,M}^c(s),$$

where

$$\Omega_{N,M}^c(s) \coloneqq \frac{1}{M} \sum_{m \in \mathcal{M}_N} \sum_{j \in \mathcal{M}_N^{\partial}(m;s)} E\left[Y_{M,m} Y_{M,j}^{'} \;\middle|\; \mathcal{C}_M \right].$$

Notice that since $E\left[Y_{M,m} \mid \mathcal{C}_M\right] = 0 \;\; a.s.$, it follows from the law of total variance that,

$$V_{N,M} = E\left[V_{N,M}^c\right].$$

Notice furthermore that it holds that

$$Var(\hat{\beta}) = \frac{N \pi_{N,M}}{M} V_{N,M},$$

and

$$\widehat{Var}(\hat{\beta}) = \frac{N \pi_{N,M}}{M} \hat{V}_{N,M}.$$

Since

$$\begin{aligned}
\|\widehat{Var}(\hat{\beta}) - Var(\hat{\beta})\|_F &= \left\| \frac{N \pi_{N,M}}{M} \hat{V}_{N,M} - \frac{N \pi_{N,M}}{M} V_{N,M} \right\|_F \\
&= \left\| \frac{N \pi_{N,M}}{M} \left(\hat{V}_{N,M} - V_{N,M} \right) \right\|_F \\
&= \frac{N \pi_{N,M}}{M} \left\| \hat{V}_{N,M} - V_{N,M} \right\|_F,
\end{aligned}$$

and $\frac{N \pi_{N,M}}{M}$ is bounded due to Assumption 4.4, it suffices to show that

$$\|\hat{V}_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0,$$

which, in light of the technique of “add and subtract,” is equivalent to

$$\|\hat{V}_{N,M} - \tilde{V}_{N,M} + \tilde{V}_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0.$$

Since by the triangular inequality,

$$\begin{aligned}
0 &\leq \|\hat{V}_{N,M} - \tilde{V}_{N,M} + \tilde{V}_{N,M} - V_{N,M}\|_F \\
&\leq \|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F + \|\tilde{V}_{N,M} - V_{N,M}\|_F,
\end{aligned}$$

we wish to prove that

(i) $\|\tilde{V}_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0$;

(ii) $\|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F \xrightarrow{p} 0$,

as $N \to \infty$.

(i) $\|\tilde{V}_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0$:

To economize on notation, we reference the (i, j) entry of $V_{N,M}$ by $V_{i,j}$. Analogous notations are applied to $\tilde{V}_{N,M}$ and $V^c_{N,M}$. The proof proceeds in multiple steps:

(1) $E\left[|\tilde{V}_{i,j} - V_{i,j}|\right] \to 0$;

(2) $|\tilde{V}_{i,j} - V_{i,j}| \xrightarrow{p} 0$,

as $N \to \infty$.

(1) $E\left[|\tilde{V}_{i,j} - V_{i,j}|\right] \to 0$:

We consider two cases:

(a) the case when $\tilde{V}_{i,j} - V_{i,j} \geq 0$;

(b) the case when $\tilde{V}_{i,j} - V_{i,j} < 0$,

separately.

(a) the case when $\tilde{V}_{i,j} - V_{i,j} \geq 0$:

If $\tilde{V}_{i,j} - V_{i,j} \geq 0$, it holds that

$$\begin{aligned}
E\left[|\tilde{V}_{i,j} - V_{i,j}|\right] &= E\left[\tilde{V}_{i,j} - V_{i,j}\right] \\
&= E\left[\tilde{V}_{i,j} - E\left[V^c_{i,j}\right]\right] \\
&= E\left[\tilde{V}_{i,j} - V^c_{i,j}\right] \\
&\leq \left|E\left[\tilde{V}_{i,j} - V^c_{i,j}\right]\right| \\
&\leq E\left[\left|\tilde{V}_{i,j} - V^c_{i,j}\right|\right] \\
&= E\left[\underbrace{E\left[\left|\tilde{V}_{i,j} - V^c_{i,j}\right| \,\middle|\, \mathcal{C}_M\right]}_{\xrightarrow{a.s.} 0}\right] \\
&\to 0,
\end{aligned}$$

where the last implication is the result of Proposition 4.1 of Kojevnikov et al. (2021)[32] and the Dominated Convergence Theorem.

(b) the case when $\tilde{V}_{i,j} - V_{i,j} < 0$:

If $\tilde{V}_{i,j} - V_{i,j} < 0$, it holds that

$$\begin{aligned}
E\left[|\tilde{V}_{i,j} - V_{i,j}|\right] &= E\left[-(\tilde{V}_{i,j} - V_{i,j})\right] \\
&= -E\left[\tilde{V}_{i,j} - V_{i,j}\right] \\
&= -E\left[\tilde{V}_{i,j} - E\left[V^c_{i,j}\right]\right] \\
&= -E\left[\tilde{V}_{i,j} - V^c_{i,j}\right] \\
&\leq \left|E\left[\tilde{V}_{i,j} - V^c_{i,j}\right]\right| \\
&\leq E\left[\left|\tilde{V}_{i,j} - V^c_{i,j}\right|\right] \\
&= E\left[\underbrace{E\left[\left|\tilde{V}_{i,j} - V^c_{i,j}\right| \,\middle|\, \mathcal{C}_M\right]}_{\xrightarrow{a.s.} 0}\right] \\
&\to 0,
\end{aligned}$$

where the last implication is a consequence of Proposition 4.1 of Kojevnikov et al. (2021) and the Dominated Convergence Theorem. Hence in either case, we have

$$E\left[|\tilde{V}_{i,j} - V_{i,j}|\right] \to 0.$$

(2) $|\tilde{V}_{i,j} - V_{i,j}| \xrightarrow{p} 0$:

By the Markov inequality, for any positive constant $c > 0$,

$$\begin{aligned}
Pr\left(|\tilde{V}_{i,j} - V_{i,j}| > c\right) &\leq \frac{1}{c}\underbrace{E\left[|\tilde{V}_{i,j} - V_{i,j}|\right]}_{\to 0} \\
&\to 0,
\end{aligned}$$

where the last implication is the result of part (1). Hence it holds that

$$|\tilde{V}_{i,j} - V_{i,j}| \xrightarrow{p} 0,$$

[32]Notice that the definitions of $V_{N,M}$, $\hat{V}_{N,M}$ and $\tilde{V}_{N,M}$ are slightly different from those used in Proposition 4.1 of Kojevnikov et al. (2021).

which in turn implies

$$\|\tilde{V}_{N,M} - V_{N,M}\|_F \xrightarrow{p} 0.$$

(ii) $\left\|\hat{V}_{N,M} - \tilde{V}_{N,M}\right\|_F \xrightarrow{p} 0$:

The proof mimics the standard proof of consistency of a heteroskedasticity-robust standard error. First, applying the properties of the Frobenius norm,[33] we have[34]

$$\begin{aligned}
\left\|\hat{V}_{N,M} - \tilde{V}_{N,M}\right\|_F &= \Bigg\| \sum_{s\geq 0} \omega_M(s) \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} x_m \hat{\varepsilon}_m \hat{\varepsilon}_j x_j' \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \\
&\quad - \sum_{s\geq 0} \omega_M(s) \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} x_m \varepsilon_m \varepsilon_j x_j' \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \Bigg\|_F \\
&= \left\| \sum_{s\geq 0} \omega_M(s) \frac{1}{M} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} x_m \left(\hat{\varepsilon}_m \hat{\varepsilon}_j - \varepsilon_m \varepsilon_j\right) x_j' \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \right\|_F \\
&\leq \left\| \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} x_m \left(\hat{\varepsilon}_m \hat{\varepsilon}_j - \varepsilon_m \varepsilon_j\right) x_j' \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \right\|_F \\
&= \left\| \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \left\{ \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\hat{\varepsilon}_m \hat{\varepsilon}_j - \varepsilon_m \varepsilon_j\right) x_m x_j' \right\} \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \right\|_F \\
&= \left\| \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \left\{ \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\hat{\varepsilon}_m \hat{\varepsilon}_j - \varepsilon_m \varepsilon_j\right) x_m x_j' \right\} \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \right\|_F \\
&\leq \left\| \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \right\|_F \left\| \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\hat{\varepsilon}_m \hat{\varepsilon}_j - \varepsilon_m \varepsilon_j\right) x_m x_j' \right\|_F \left\| \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \right\|_F .
\end{aligned}$$

In what follows we aim to show that the "bread" part ($\left\|\left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}\right\|_F$) remains finite, and that the "meat" part ($\left\|\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\left(\hat{\varepsilon}_m\hat{\varepsilon}_j - \varepsilon_m\varepsilon_j\right)x_m x_j'\right\|_F$) converges to zero in probability. We examine these by showing the convergence in mean-square, a strategy employed in Aronow et al. (2015) and Tabord-Meehan (2019). First, in light of the Continuous Mapping Theorem, the proof of the bread part amounts to showing the convergence in probability

[33] Recall that

$$\|AB\|_F \leq \|A\|_F \|B\|_F ,$$

where A and B are matrices whose product is well-defined. Also, It should be noted that by Assumption 4.2, there exists a nonnegative finite constant $C > 0$ such that

$$C = \sup_{N\geq 1} \left\| \left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \right\|_F .$$

[34] To lighten the notational burden, we drop the M subscript from $\{x_{M,m}\}_{m\in\mathcal{M}_N}$ and $\{\varepsilon_{M,m}\}_{m\in\mathcal{M}_N}$ in the rest of the proof.

of $\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'$. We prove this by considering an element-by-element convergence. Let $x_{k,i}$ denote the i-th element of x_k. Then the (i,j) entry of $\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'$ is given by $\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}$. Consider its expectation,

$$E\left[\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}\right] = \frac{1}{M}\sum_{k\in\mathcal{M}_N} E\left[x_{k,i}x_{k,j}\right].$$

From Assumption 4.2, there exists a nonnegative finite constant $C_{0,1}$ such that

$$C_{0,1} = \sup_{N\geq 1}\max_{m\in\mathcal{M}_N} E\left[x_{m,i}x_{m,j}\right].$$

Hence $E\left[\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}\right]$ exists and finite. Next, the variance can be expressed as a sum of covariance:

$$\begin{aligned}
Var\left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}\right) &= \frac{1}{M^2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} Cov\left(x_{m,i}x_{m,j}, x_{m',i}x_{m',j}\right) \\
&= \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N(m;s)} Cov\left(x_{m,i}x_{m,j}, x_{m',i}x_{m',j}\right).
\end{aligned}$$

Again from Assumption 4.2, there exists a nonnegative finite constant $C_{0,2}$ such that

$$C_{0,2} = \sup_{N\geq 1}\max_{m,m'\in\mathcal{M}_N} Cov\left(x_{m,i}x_{m,j}, x_{m',i}x_{m',j}\right).$$

Hence

$$\begin{aligned}
Var\left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}\right) &\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N(m;s)} C_{0,2} \\
&= \frac{C_{0,2}}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N(m;s)} 1 \\
&= \frac{C_{0,2}}{M^2}\sum_{s\geq 0}\delta_M^{\partial}(s;1) \\
&= C_{0,2}\underbrace{\frac{1}{M}}_{\to 0}\underbrace{\frac{1}{M}\sum_{s\geq 0}\delta_M^{\partial}(s;1)}_{\to 0} \\
&\to 0,
\end{aligned}$$

where the last implication is due to Assumption 4.3. Therefore, we obtain

$$\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j} \xrightarrow{p} E\left[\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_{k,i}x_{k,j}\right].$$

By repeating the argument above, we arrive at

$$\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k' \xrightarrow{p} E\left[\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right].$$

Furthermore, applying the Continuous Mapping Theorem yields

$$\left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1} \xrightarrow{p} \left(E\left[\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right]\right)^{-1},$$

obtaining the result.

Next the proof of the meat part consists of multiple steps. To begin with, observe that by definition, $\hat{\varepsilon}_m$ can be written as

$$\begin{aligned}
\hat{\varepsilon}_m &= \underbrace{y_m}_{x_m'\beta+\varepsilon_m} - x_m'\hat{\beta} \\
&= x_m'\beta + \varepsilon_m - x_m'\hat{\beta} \\
&= \varepsilon_m - x_m'(\hat{\beta}-\beta),
\end{aligned}$$

and then

$$\begin{aligned}
\hat{\varepsilon}_m\hat{\varepsilon}_j' &= \left\{\varepsilon_m - x_m'(\hat{\beta}-\beta)\right\}\left\{\varepsilon_j - x_j'(\hat{\beta}-\beta)\right\}' \\
&= \varepsilon_m\varepsilon_j' - \varepsilon_m(\hat{\beta}-\beta)'x_j - x_m'(\hat{\beta}-\beta)\varepsilon_j' + x_m'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_j.
\end{aligned}$$

Hence

$$\hat{\varepsilon}_m\hat{\varepsilon}_j' - \varepsilon_m\varepsilon_j' = -\varepsilon_m(\hat{\beta}-\beta)'x_j - x_m'(\hat{\beta}-\beta)\varepsilon_j' + x_m'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_j,$$

so that by the triangular inequality and the Cauchy-Schwartz inequality,

$$\begin{aligned}
\left|\hat{\varepsilon}_m\hat{\varepsilon}_j' - \varepsilon_m\varepsilon_j'\right| &= \left|-\varepsilon_m(\hat{\beta}-\beta)'x_j - x_m'(\hat{\beta}-\beta)\varepsilon_j' + x_m'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_j\right| \\
&\le \underbrace{\left|\varepsilon_m(\hat{\beta}-\beta)'x_j\right|}_{\le|\varepsilon_m|\left|(\hat{\beta}-\beta)'x_j\right|} + \underbrace{\left|x_m'(\hat{\beta}-\beta)\varepsilon_j'\right|}_{\le\left|x_m'(\hat{\beta}-\beta)\right||\varepsilon_j|} + \underbrace{\left|x_m'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_j\right|}_{\left|x_m'(\hat{\beta}-\beta)\right|\left|(\hat{\beta}-\beta)'x_j\right|} \\
&\le |\varepsilon_m|\underbrace{\left|(\hat{\beta}-\beta)'x_j\right|}_{\left\|\hat{\beta}-\beta\right\|_2\|x_j\|_2} + \underbrace{\left|x_m'(\hat{\beta}-\beta)\right|}_{\le\|x_m\|_2\left\|\hat{\beta}-\beta\right\|_2}|\varepsilon_j| + \underbrace{\left|x_m'(\hat{\beta}-\beta)\right|}_{\le\|x_m\|_2\left\|\hat{\beta}-\beta\right\|_2}\underbrace{\left|(\hat{\beta}-\beta)'x_j\right|}_{\le\left\|\hat{\beta}-\beta\right\|_2\|x_j\|_2} \\
&\le \left\|\hat{\beta}-\beta\right\|_2\|x_j\|_2|\varepsilon_m| + \left\|\hat{\beta}-\beta\right\|_2\|x_m\|_2|\varepsilon_j| + \left\|\hat{\beta}-\beta\right\|_2^2\|x_m\|_2\|x_j\|_2,
\end{aligned}$$

for each $m, j \in \mathcal{M}_N$. Hence the meat part can be bounded as

$$
\begin{aligned}
&\left\| \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\hat{\varepsilon}_m\hat{\varepsilon}_j - \varepsilon_m\varepsilon_j\right) x_m x_j' \right\|_F \\
&\leq \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left|\hat{\varepsilon}_m\hat{\varepsilon}_j - \varepsilon_m\varepsilon_j\right| \underbrace{\|x_m x_j'\|_F}_{\leq \|x_m\|_2\|x_j\|_2} \\
&\leq \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\left\|\hat{\beta}-\beta\right\|_2 \|x_j\|_2 |\varepsilon_m| + \left\|\hat{\beta}-\beta\right\|_2 \|x_m\|_2 |\varepsilon_j| + \left\|\hat{\beta}-\beta\right\|_2^2 \|x_m\|_2 \|x_j\|_2 \right) \|x_m\|_2\|x_j\|_2 \\
&= \left\|\hat{\beta}-\beta\right\|_2 \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2\|x_j\|_2^2|\varepsilon_m| + \left\|\hat{\beta}-\beta\right\|_2 \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2|\varepsilon_j| \\
&\quad + \left\|\hat{\beta}-\beta\right\|_2^2 \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2^2.
\end{aligned}
$$

Denote

$$
\begin{aligned}
R_{N,1} &:= \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2\|x_j\|_2^2|\varepsilon_m| \\
R_{N,2} &:= \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2|\varepsilon_j| \\
R_{N,3} &:= \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2^2.
\end{aligned}
$$

Now, since by Theorem 3.1,

$$\|\hat{\beta}-\beta\|_2 \xrightarrow{p} 0,$$

and the application of the Continuous Mapping Theorem yields

$$\|\hat{\beta}-\beta\|_2^2 \xrightarrow{p} 0,$$

it thus suffices to prove that each of $R_{N,1}$, $R_{N,2}$ and $R_{N,3}$ converges in probability to a finite number.

First let us study the expectation of $R_{N,1}$. By applying the Cauchy-Schwartz inequality repeatedly, it holds

$$
\begin{aligned}
E\left[R_{N,1}\right] &= E\left[\frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2\|x_j\|_2^2|\varepsilon_m| \right] \\
&= \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \underbrace{E\left[\|x_m\|_2\|x_j\|_2^2|\varepsilon_m|\right]}_{\leq \left(E\left[\|x_m\|_2^2\|x_j\|_2^4\right]\right)^{\frac{1}{2}} \left(E[|\varepsilon_m|^2]\right)^{\frac{1}{2}}} \\
&\leq \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\underbrace{E\left[\|x_m\|_2^2\|x_j\|_2^4\right]}_{\leq \left(E\left[\|x_m\|_2^4\right]\right)^{\frac{1}{2}} \left(E\left[\|x_j\|_2^8\right]\right)^{\frac{1}{2}}} \right)^{\frac{1}{2}} \left(\underbrace{E\left[|\varepsilon_m|^2\right]}_{=E[E[|\varepsilon_m|^2|\mathcal{C}_M]]} \right)^{\frac{1}{2}}
\end{aligned}
$$

$$\leq \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\left(\left(E\left[\|x_m\|_2^4\right]\right)^{\frac{1}{2}}\left(E\left[\|x_j\|_2^8\right]\right)^{\frac{1}{2}}\right)^{\frac{1}{2}}\left(E\left[E\left[|\varepsilon_m|^2 \mid \mathcal{C}_M\right]\right]\right)^{\frac{1}{2}}.$$

Here, in light of Assumption 4.1, there exists a nonnegative finite number $C_1 > 0$ such that

$$C_1 = \sup_{N\geq 1}\max_{m\in\mathcal{M}_N} E\left[|\varepsilon_m| \mid \mathcal{C}_M\right],$$

and moreover by Assumption 4.2, there exists a nonnegative finite number $C_2 > 0$ such that

$$C_2 = \sup_{N\geq 1}\max_{m\in\mathcal{M}_N} E\left[\|x_m\|_2^8\right].$$

With a slight abuse of notation, we have

$$\begin{aligned}
E\left[R_{N,1}\right] &\leq \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} C_1C_2 \\
&= \frac{C_1C_2}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} 1 \\
&= \frac{C_1C_2}{M}\sum_{s\geq 0}\delta_M^{\partial}(s;1) \\
&= C_1C_2\underbrace{\frac{1}{M}\sum_{s\geq 0}\delta_M^{\partial}(s;1)}_{\to 0} \\
&\to 0,
\end{aligned}$$

where the last implication is because of Assumption 4.3. Next let us study the variance of $R_{N,1}$. It suffices to show that

$$E\left[R_{N,1}^2\right] \to 0,$$

By the Cauchy-Schwartz inequality, it holds that

$$\begin{aligned}
E\left[R_{N,1}^2\right] &= E\left[\frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)}\|x_m\|_2\|x_j\|_2^2\|x_k\|_2\|x_l\|_2^2|\varepsilon_m||\varepsilon_k|\right] \\
&= \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} E\left[\|x_m\|_2\|x_j\|_2^2\|x_k\|_2\|x_l\|_2^2|\varepsilon_m||\varepsilon_k|\right] \\
&\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)}\left(E\left[\|x_m\|_2^2\|x_j\|_2^4\|x_k\|_2^2\|x_l\|_2^4\right]\right)^{\frac{1}{2}}\left(E\left[|\varepsilon_m|^2|\varepsilon_k|^2\right]\right)^{\frac{1}{2}}.
\end{aligned}$$

Here, by Assumption 4.2 and the Cauchy-Schwartz inequality, there exists a nonnegative finite constant $C_3 > 0$ such that

$$C_3 = \sup_{N\leq 1}\max_{m,j,k,l\in\mathcal{M}_N} E\left[\|x_m\|_2^2\|x_j\|_2^4\|x_k\|_2^2\|x_l\|_2^4\right].$$

Then, with a slight abuse of notation in writing $C_3^{\frac{1}{2}}$ as C_3, we have

$$
\begin{aligned}
E\left[R_{N,1}^2\right] &\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} \underbrace{\left(E\left[\|x_m\|_2^2\|x_j\|_2^4\|x_k\|_2^2\|x_l\|_2^4\right]\right)^{\frac{1}{2}}}_{\leq C_3^{\frac{1}{2}}}\left(E\left[|\varepsilon_m|^2|\varepsilon_k|^2\right]\right)^{\frac{1}{2}} \\
&\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} C_3^{\frac{1}{2}}\left(E\left[|\varepsilon_m|^2|\varepsilon_k|^2\right]\right)^{\frac{1}{2}} \\
&= \frac{C_3}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} \left(E\left[|\varepsilon_m|^2|\varepsilon_k|^2\right]\right)^{\frac{1}{2}} \\
&= \frac{C_3}{M^2}\sum_{s\geq 0}\sum_{(m,j,k,l)\in H_M(s;b_M)} \left(E\left[|\varepsilon_m|^2|\varepsilon_k|^2\right]\right)^{\frac{1}{2}} \\
&= \frac{C_3}{M^2}\sum_{s\geq 0}\sum_{(m,j,k,l)\in H_M(s;b_M)} \left(E\left[E\left[|\varepsilon_m|^2|\varepsilon_k|^2 \mid \mathcal{C}_M\right]\right]\right)^{\frac{1}{2}},
\end{aligned}
$$

Corollary A.2 of Kojevnikov et al. (2021) shows that there exists a nonnegative finite constant C_4 such that

$$
E\left[|\varepsilon_m|^2|\varepsilon_k|^2 \mid \mathcal{C}_M\right] \leq C_4\bar{\theta}\theta_{M,s}^{1-\frac{4}{p}},
$$

where

$$
\bar{\theta} \coloneqq \sup_{M\geq 1}\max_{s\geq 1}\theta_{M,s}.
$$

We now use the following lemma from Kojevnikov et al. (2021), p.903:

Lemma D.1. *Define*

$$
H_M(s,r) \coloneqq \left\{(m,j,k,l)\in\mathcal{M}_N^4 : j\in\mathcal{M}_N(m;r),\ l\in\mathcal{M}_N(k;r),\ x_m(\{m,j\},\{k,l\}) = s\right\}.
$$

Then

$$
|H_M(s,r)| \leq 4Mc_M(s,r;2).
$$

By Lemma D.1, we obtain

$$
\begin{aligned}
E\left[R_{N,1}^2\right] &= \frac{C_3}{M^2}\sum_{s\geq 0}\sum_{(m,j,k,l)\in H_M(s;b_M)}\left(E\left[\underbrace{E\left[|\varepsilon_m|^2|\varepsilon_k|^2 \mid \mathcal{C}_M\right]}_{\leq C_4\bar{\theta}\theta_{M,s}^{1-\frac{4}{p}}}\right]\right)^{\frac{1}{2}} \\
&\leq \frac{C_3}{M^2}\sum_{s\geq 0}\sum_{(m,j,k,l)\in H_M(s;b_M)}\left(E\left[C_4\bar{\theta}\theta_{M,s}^{1-\frac{4}{p}}\right]\right)^{\frac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{C_3 C_4'}{M^2} \sum_{s\geq 0} \sum_{(m,j,k,l)\in H_M(s;b_M)} \left(E\left[\theta_{M,s}^{1-\frac{4}{p}} \right] \right)^{\frac{1}{2}} \\
&= \frac{C_3 C_4'}{M^2} \sum_{s\geq 0} \left(E\left[\theta_{M,s}^{1-\frac{4}{p}} \right] \right)^{\frac{1}{2}} \sum_{(m,j,k,l)\in H_M(s;b_M)} 1 \\
&= \frac{C_3 C_4'}{M^2} \sum_{s\geq 0} \left(E\left[\theta_{M,s}^{1-\frac{4}{p}} \right] \right)^{\frac{1}{2}} \underbrace{|H_M(s;b_M)|}_{\leq 4Mc_M(s,b_M;2)} \\
&\leq \frac{C_3 C_4'}{M^2} \sum_{s\geq 0} \left(E\left[\theta_{M,s}^{1-\frac{4}{p}} \right] \right)^{\frac{1}{2}} 4Mc_M(s,b_M;2) \\
&= \frac{4C_3 C_4'}{M} \sum_{s\geq 0} \underbrace{\left(E\left[\theta_{M,s}^{1-\frac{4}{p}} \right] \right)^{\frac{1}{2}}}_{<\infty} c_M(s,b_M;2) \\
&= \frac{4C_3 C_4''}{M} \sum_{s\geq 0} c_M(s,b_M;2) \\
&= 4C_3 C_4'' \underbrace{\frac{1}{M} \sum_{s\geq 0} c_M(s,b_M;2)}_{\to 0} \\
&\to 0,
\end{aligned}
$$

where we apply Assumption 4.3 for the last implication, and C_4' and C_4'' are nonnegative finite constants defined appropriately. Hence we have shown that

$$
R_{N,1} \xrightarrow{p} 0.
$$

The proof of $R_{N,2}$ is analogous. It remains to show that $R_{N,3}$ converges to zero in probability. Let us first study the expectation of $R_{N,3}$.

$$
\begin{aligned}
E\left[R_{N,3}\right] &= E\left[\frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^\partial(m;s)} \|x_m\|_2^2 \|x_j\|_2^2 \right] \\
&= \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^\partial(m;s)} E\left[\|x_m\|_2^2 \|x_j\|_2^2 \right]
\end{aligned}
$$

In view of Assumption 4.2, there exists a nonnegative finite number $C_5 > 0$ such that

$$
C_5 = \sup_{N\geq 1} \max_{m\in\mathcal{M}_N} E\left[\|x_m\|_2^2 \|x_j\|_2^2 \right].
$$

Hence

$$
E\left[R_{N,3}\right] = \frac{1}{M} \sum_{s\geq 0} \sum_{m\in\mathcal{M}_N} \sum_{j\in\mathcal{M}_N^\partial(m;s)} \underbrace{E\left[\|x_m\|_2^2 \|x_j\|_2^2 \right]}_{\leq C_5}
$$

$$
\begin{aligned}
&\leq \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} C_5 \\
&= \frac{C_5}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} 1 \\
&= \frac{C_5}{M}\sum_{s\geq 0}\delta_M^{\partial}(s;1) \\
&= C_5 \underbrace{\frac{1}{M}\sum_{s\geq 0}\delta_M^{\partial}(s;1)}_{\to 0} \\
&\to 0,
\end{aligned}
$$

where we apply Assumption 4.3 in the last implication. Next let us consider the variance of $R_{M,3}$.

$$
\begin{aligned}
E\left[R_{N,3}^2\right] &= E\left[\frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2^2\|x_k\|_2^2\|x_l\|_2^2\right] \\
&= \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} E\left[\|x_m\|_2^2\|x_j\|_2^2\|x_k\|_2^2\|x_l\|_2^2\right].
\end{aligned}
$$

Once again, Assumption 4.2, in conjunction with the Cauchy-Schwartz inequality, implies that there exists a nonnegative finite number $C_6 > 0$ such that

$$
C_6 = \sup_{N\geq 1}\max_{m,j,k,l\in\mathcal{M}_N} E\left[\|x_m\|_2^2\|x_j\|_2^2\|x_k\|_2^2\|x_l\|_2^2\right].
$$

Then by Lemma D.1,

$$
\begin{aligned}
E\left[R_{N,3}^2\right] &= \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} \underbrace{E\left[\|x_m\|_2^2\|x_j\|_2^2\|x_k\|_2^2\|x_l\|_2^2\right]}_{\leq C_6} \\
&\leq \frac{1}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} C_6 \\
&= \frac{C_6}{M^2}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)}\sum_{t\geq 0}\sum_{k\in\mathcal{M}_N}\sum_{l\in\mathcal{M}_N^{\partial}(m;s)} 1 \\
&= \frac{C_6}{M^2}\sum_{s\geq 0}\sum_{(m,j,k,l)\in H_M(s;b_M)} 1 \\
&= \frac{C_6}{M^2}\sum_{s\geq 0}\underbrace{|H_M(s;b_M)|}_{\leq 4Mc_M(s,b_M;2)} \\
&= \frac{C_6}{M^2}\sum_{s\geq 0}4Mc_M(s,b_M;2) \\
&= \frac{4C_6}{M}\sum_{s\geq 0}c_M(s,b_M;2)
\end{aligned}
$$

$$
\begin{aligned}
&= 4C_6 \underbrace{\frac{1}{M}\sum_{s\geq 0} c_M(s, b_M; 2)}_{\to 0} \\
&\to 0,
\end{aligned}
$$

where the last implication is a consequence of Assumption 4.3 (ii). Therefore we have shown that

$$
\begin{aligned}
\left\| \frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\hat{\varepsilon}_m\hat{\varepsilon}_j - \varepsilon_m\varepsilon_j\right) x_m x_j' \right\|_F &= \left\|\hat{\beta}-\beta\right\|_2 \underbrace{\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2\|x_j\|_2^2|\varepsilon_m|}_{R_{M,1}} \\
&\quad + \left\|\hat{\beta}-\beta\right\|_2 \underbrace{\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2|\varepsilon_j|}_{R_{M,2}} \\
&\quad + \left\|\hat{\beta}-\beta\right\|_2^2 \underbrace{\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \|x_m\|_2^2\|x_j\|_2^2}_{R_{M,3}} \\
&= \underbrace{\left\|\hat{\beta}-\beta\right\|_2}_{\xrightarrow{p}0} \underbrace{R_{M,1}}_{\xrightarrow{p}0} + \underbrace{\left\|\hat{\beta}-\beta\right\|_2}_{\xrightarrow{p}0} \underbrace{R_{M,2}}_{\xrightarrow{p}0} + \underbrace{\left\|\hat{\beta}-\beta\right\|_2^2}_{\xrightarrow{p}0} \underbrace{R_{M,3}}_{\xrightarrow{p}0} \\
&\xrightarrow{p} 0,
\end{aligned}
$$

which proves the convergence of the meat part to zero.

Hence, we have established

$$
\begin{aligned}
\left\|\hat{V}_{N,M} - \tilde{V}_{N,M}\right\|_F &\leq \underbrace{\left\|\left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}\right\|_F}_{<\infty} \underbrace{\left\|\frac{1}{M}\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{j\in\mathcal{M}_N^{\partial}(m;s)} \left(\hat{\varepsilon}_m\hat{\varepsilon}_j - \varepsilon_m\varepsilon_j\right) x_m x_j'\right\|_F}_{\xrightarrow{p}0} \underbrace{\left\|\left(\frac{1}{M}\sum_{k\in\mathcal{M}_N} x_k x_k'\right)^{-1}\right\|_F}_{<\infty} \\
&\xrightarrow{p} 0,
\end{aligned}
$$

completing part (ii) of the proof. Combining part (i) and (ii), we arrive at

$$
\begin{aligned}
\|\hat{V}_{N,M} - V_{N,M}\|_F &= \|\hat{V}_{N,M} - \tilde{V}_{N,M} + \tilde{V}_{N,M} - V_{N,M}\|_F \\
&\leq \underbrace{\|\hat{V}_{N,M} - \tilde{V}_{N,M}\|_F}_{\xrightarrow{p}0} + \underbrace{\|\tilde{V}_{N,M} - V_{N,M}\|_F}_{\xrightarrow{p}0} \\
&\xrightarrow{p} 0,
\end{aligned}
$$

as desired. $\square$

Proof of Corollary 4.1. For simplicity we denote

$$
\hat{V}_{N,M}^{Dyad} := \left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1} \left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} \mathbb{1}_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'} x_m x_{m'}'\right) \left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1},
$$

and

$$\hat{V}_{N,M}^{Network} := \left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\right)\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}.$$ [35]

Define moreover $\widetilde{Var}(\hat{\beta})$ to be the same variance as (18), but is now applied to the network-regression model (1) and (2). By the triangular inequality,

$$\begin{aligned}\left\|N\pi_{N,M}\hat{V}_{N,M}^{Network} - \hat{V}_{N,M}^{Dyad}\right\|_F &= \left\|N\pi_{N,M}\hat{V}_{N,M}^{Network} - \widetilde{Var}(\hat{\beta}) + \widetilde{Var}(\hat{\beta}) - N\pi_{N,M}\hat{V}_{N,M}^{Dyad}\right\|_F \\ &\leq \left\|N\pi_{N,M}\hat{V}_{N,M}^{Network} - \widetilde{Var}(\hat{\beta})\right\|_F + \left\|\widetilde{Var}(\hat{\beta}) - N\pi_{N,M}\hat{V}_{N,M}^{Dyad}\right\|_F.\end{aligned}$$

Since Theorem 4.1 implies $\left\|N\pi_{N,M}\hat{V}_{N,M}^{Network} - \widetilde{Var}(\hat{\beta})\right\|_F \xrightarrow{p} 0$, then in the limit we are left with

$$\left\|\hat{V}_{N,M}^{Network} - \hat{V}_{N,M}^{Dyad}\right\|_F \leq \left\|\widetilde{Var}(\hat{\beta}) - N\pi_{N,M}\hat{V}_{N,M}^{Dyad}\right\|_F. \tag{26}$$

Now we prove the statement by way of contradiction. Assume for the sake of contradiction that the dyadic-robust variance estimator $\hat{V}_{N,M}^{Dyad}$ is consistent, i.e., $\left\|\widetilde{Var}(\hat{\beta}) - N\pi_{N,M}\hat{V}_{N,M}^{Dyad}\right\|_F \xrightarrow{p} 0$. This, combined with the inequality (26), implies $\left\|N\pi_{N,M}\hat{V}_{N,M}^{Network} - N\pi_{N,M}\hat{V}_{N,M}^{Dyad}\right\|_F \xrightarrow{p} 0$. Now, observe that

$$\begin{aligned}&\left\|N\pi_{N,M}\hat{V}_{N,M}^{Network} - N\pi_{N,M}\hat{V}_{N,M}^{Dyad}\right\|_F \\
&= \Bigg\|N\pi_{N,M}\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\right)\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1} \\
&\quad -N\pi_{N,M}\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\left(\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N} \mathbb{1}_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\right)\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\Bigg\|_F \\
&= \Bigg\|N\pi_{N,M}\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\left(\sum_{s\geq 0}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} h_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\right)\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1} \\
&\quad -N\pi_{N,M}\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\left(\sum_{s=\{0,1\}}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} \mathbb{1}_{m,m'}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\right)\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\Bigg\|_F \\
&= \Bigg\|N\pi_{N,M}\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\left(\sum_{s=\{0,1\}}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} \underbrace{(h_{m,m'} - \mathbb{1}_{m,m'})}_{0}\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\right)\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1} \\
&\quad +N\pi_{N,M}\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\left(\sum_{s\geq 2}\sum_{m\in\mathcal{M}_N}\sum_{m'\in\mathcal{M}_N^{\partial}(m;s)} \left(\underbrace{h_{m,m'}}_{1} - \underbrace{\mathbb{1}_{m,m'}}_{0}\right)\hat{\varepsilon}_m\hat{\varepsilon}_{m'}x_m x_{m'}'\right)\left(\sum_{m\in\mathcal{M}_N} x_m x_m'\right)^{-1}\Bigg\|_F\end{aligned}$$

[35] For the sake of brevity, we suppress the M from subscript throughout this proof.

$$= \left\| N\pi_{N,M} \left(\sum_{m \in \mathcal{M}_N} x_m x_m' \right)^{-1} \left(\sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \hat{\varepsilon}_m \hat{\varepsilon}_{m'} x_m x_{m'}' \right) \left(\sum_{m \in \mathcal{M}_N} x_m x_m' \right)^{-1} \right\|_F$$

$$= \frac{N\pi_{N,M}}{M} \left\| \left(\frac{1}{M} \sum_{m \in \mathcal{M}_N} x_m x_m' \right)^{-1} \left(\frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \hat{\varepsilon}_m \hat{\varepsilon}_{m'} x_m x_{m'}' \right) \left(\frac{1}{M} \sum_{m \in \mathcal{M}_N} x_m x_m' \right)^{-1} \right\|_F,$$

where we note that $\frac{N\pi_{N,M}}{M} < \infty$ by Assumption 4.4. We prove that the inside the Frobenius norm does not converge in probability to zero. First it can be shown that the "bread" part $\left(\frac{1}{M} \sum_{m \in \mathcal{M}_N} x_m x_m' \right)^{-1}$ converges to $\left(\frac{1}{M} \sum_{m \in \mathcal{M}_N} E\left[x_m x_m' \right] \right)^{-1}$ by the same argument as the one in the proof of Theorem 4.1. Next plugging the definition of $\hat{\varepsilon}$ into the middle part, we have

$$\begin{aligned}
&\frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \hat{\varepsilon}_m \hat{\varepsilon}_{m'} x_m x_{m'}' \\
&= \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \left\{ \varepsilon_m + x_m'(\beta - \hat{\beta}) \right\} \left\{ \varepsilon_{m'} + x_{m'}'(\beta - \hat{\beta}) \right\} x_m x_{m'}' \\
&= \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m \varepsilon_{m'} x_m x_{m'}' + \varepsilon_m (\beta - \hat{\beta}) x_{m'} x_m x_{m'}' + x_m'(\beta - \hat{\beta}) \varepsilon_{m'} x_m x_{m'}' + x_m'(\beta - \hat{\beta})(\beta - \hat{\beta})' x_{m'} x_m x_{m'}' \\
&= \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m \varepsilon_{m'} x_m x_{m'}' + \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m (\beta - \hat{\beta}) x_{m'} x_m x_{m'}' \\
&\quad + \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} x_m'(\beta - \hat{\beta}) \varepsilon_{m'} x_m x_{m'}' + \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} x_m'(\beta - \hat{\beta})(\beta - \hat{\beta})' x_{m'} x_m x_{m'}'.
\end{aligned}$$

Denote

$$\begin{aligned}
Q_{N,1} &\coloneqq \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m \varepsilon_{m'} x_m x_{m'}' \\
Q_{N,2} &\coloneqq \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m (\beta - \hat{\beta}) x_{m'} x_m x_{m'}' \\
Q_{N,3} &\coloneqq \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} x_m'(\beta - \hat{\beta}) \varepsilon_{m'} x_m x_{m'}' \\
Q_{N,4} &\coloneqq \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} x_m'(\beta - \hat{\beta})(\beta - \hat{\beta})' x_{m'} x_m x_{m'}'.
\end{aligned}$$

From Theorem 3.1, it can be seen that $Q_{M,2}$, $Q_{M,3}$ and $Q_{M,3}$ either converge to zero or diverge as N goes to infinity. When it comes to $Q_{M,1}$, observe that

$$\begin{aligned}
E\left[Q_{N,1} \right] &= E\left[\frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} \varepsilon_m \varepsilon_{m'} x_m x_{m'}' \right] \\
&= \frac{1}{M} \sum_{s \geq 2} \sum_{m \in \mathcal{M}_N} \sum_{m' \in \mathcal{M}_N^{\partial}(m;s)} E\left[\varepsilon_m \varepsilon_{m'} x_m x_{m'}' \right],
\end{aligned}$$

which never converges to zero due to the hypothesis (20) of this corollary. In either case, the middle part does not converge in probability to zero, meaning that $\left\| N\pi_{N,M}\hat{V}_{N,M}^{Network} - N\pi_{N,M}\hat{V}_{N,M}^{Dyad} \right\|_F \xrightarrow{p} 0$ is not true. This, however, contradicts the implication of the assumption that the dyadic-robust variance estimator is consistent. Hence, by means of contradiction, we conclude that the dyadic-robust variance estimator is not consistent, which completes the proof. $\square$

E Data Construction for the Empirical Illustration

Data construction for our empirical exercise in Section 6 proceeds in multiple steps:

Step 1 Our subsample consists of the location of interest, i.e., Strasbourg, for the period of interest, i.e., Term 7. We select a further subset of the extracted data by seating arrangement (i.e., we focus on Pattern 1 for the present analysis - see Table 8).

Step 2 Since our analysis is concerned with voting concordance, we follow the original authors in dropping entries with missing data or "abstain" in the variable "vote."[36]

Step 3 The resulting data still contains individuals belonging to "Identity, Tradition and Sovereignty (ITS)," one of the European Political Groups that dissolved in November 7, during the sixth term. We drop such MEPs from our analysis.

Step 4 The selected data is used to form the dyadic-data registering the pair-of-MEPs-specific information. When pairing two MEPs, we follow Harmon et al. (2019) in focusing on those pairs of MEPs, both of whom are

(i) in the same EPG;

(ii) from an alphabetically-seated EPG; and

(iii) non-leaders at the time of voting.

Our dyadic-data consists of two types of variables: binary variables and numerical variables. The dyad-level binary (i.e., indicator) variables are defined to be one if the individual-level binary variables are the same, and zero otherwise. The dyadic-specific numerical variables in our analysis are the differences between the individual-level numerical variables, such as age and tenure. When calculating the differences in ages and tenures, we take the absolute values as we do not consider directional dyads, and we then rescale them into ten-year units. See the note for Table 3 for details.

[36]This amounts to assuming that those observations are missing completely at random (MCAR).

TABLE 8

Patterns of Seating Arrangements: Strasbourg, Term 7

Pattern	Date			Number of Proposals
1	7/14/2009	∼	7/16/2009	116
2	8/18/2009	∼	8/21/2009	72
3	9/23/2009	∼	9/25/2009	114
4	10/13/2009	∼	10/16/2009	40
5	11/19/2009	∼	12/11/2009	94
6	1/5/2010	∼	1/8/2010	79
7	3/17/2010	∼	3/19/2010	45
8	4/14/2010	∼	4/16/2010	120
9	5/5/2010	∼	5/7/2010	79
10	7/7/2010	∼	7/9/2010	34
11	7/21/2010	∼	7/22/2010	50
12	8/18/2010	∼	8/20/2010	118

Note: This table presents patterns of seating arrangements with the corresponding dates and the number of total observations for each pattern. Since voting may be taken place for multiple proposals within the same day, the total number proposals tends to be higher than that of days in a single pattern. For example, the first line indicates that 116 proposals were discussed and votes were cast over the three days (from the 14th of July, 2009 to the 16th of July, 2009).