An empirical model of dyadic link formation in a network with unobserved heterogeneity

Andreas Dzemski*

April 19, 2018

I study a dyadic linking model in which agents form directed links that exhibit homophily and reciprocity. A fixed effect approach accounts for unobserved sources of degree heterogeneity. I consider inference with respect to homophily preferences and a reciprocity parameter, as well as a test of model specification. The specification test compares observed transitivity to the transitivity predicted by the dyadic linking model. My test statistics are robust to the incidental parameter problem (Neyman and Scott 1948). This is accomplished by using analytical formulas that approximate the effect of the incidental parameter on the bias and the variance of the test statistics. The approximation is justified under dense large network asymptotics. In an application to favor networks in Indian villages, the model specification test detects that the dyadic linking model underestimates the true transitivity of the network.

JEL codes: C33, C35

Keywords: Network formation, fixed effects, incidental parameter problem, transitive structure, favor networks

*Department of Economics and Statistics, University of Gothenburg. This paper is a substantially revised and extended version of Chapter 3 of my PhD thesis. I am grateful to the co-editor and three anonymous referees for comments that greatly improved the quality of the paper. Moreover, I benefited from discussions with Yann Bramoullé, Iván Fernández-Val, Markus Frölich, Bryan Graham, Geert Dhaene, Xavier d'Hautefeuille, Stephen Kastoryano, Enno Mammen, Jan Nimczik, Vladimir Pinheiro Ponczek, Andrea Weber, Martin Weidner and seminar participants at 2016 Cemmap Conference on Networks at Berkeley, 2015 Econometrics Journal Conference in Cambridge, ENSAI, Gothenburg, KU Leuven, 2014 ES Winter Meeting, Mannheim, Marseille and Paris School of Economics.

1. Introduction

A substantial amount of economic activity takes place outside of centralized markets, within networks of interpersonal relationships. The importance of interpersonal relationships has been documented, e.g., for information dissemination (Banerjee et al. 2013) and informal insurance (Fafchamps and Lund 2003). Social network data encodes interpersonal relationships as links in a network and makes them amendable to empirical investigation.

In models of dyadic link formation, linking decisions are a binary choice that depends only on the characteristics of the two agents connected by the link. Models of dyadic link formation are straight-forward to implement and often applied in practice (Mayer and Puller 2008; Fafchamps and Gubert 2007). Even though dyadic linking models ignore the strategic dimension of link formation, they can still replicate important stylized features of social networks (Jochmans 2017). Some of the agent characteristics entering the linking decisions may be unobserved to the Econometrician but can be accounted for using a fixed effects approach. Controlling for a high-dimensional vector of fixed effects complicates inference because of the incidental parameter problem (Neyman and Scott 1948). For dyadic linking models, the incidental parameter problem has been discussed in Charbonneau (2017), Graham (2017), and Jochmans (2017).

This paper studies inference in a dyadic linking model with fixed effects. I consider significance testing for the parameters that describe homophily preferences and the propensity of agents to reciprocate links, as well as a test of model specification based on the transitive structure of the network. Robustness to the incidental parameter problem is ensured by using new test statistics that are based on analytical formulas that approximate the effect of fixed effect estimation on the bias and variance of a naïve t-test. The approximation is theoretically justified for large networks.

In my linking model, agents form directed links if the link surplus exceeds a random threshold. The model is related to Holland and Leinhardt (1981) and accounts for all three drivers of linking behavior that they identify: homophily, degree heterogeneity and reciprocity. *Homophily* in linking decisions is the clustering of agents who share similar observed characteristics (McPherson, Smith-Lovin, and Cook 2001). *Degree heterogeneity* means that the number of linking partners varies a lot between agents. *Link reciprocity* refers to the fact that, conditional on agent characteristics, observing a link from one agent to another agent renders observing the link in the opposite direction more likely.

My linking network targets the linking behavior within dyads (groups of two). A test of model specification can be based on the predictive power of the linking model for network statistics that are not pinned down completely by pairwise interactions. My specification test looks at transitive relationships in triads (groups of three). A transitive relationship arises if two agents who are connected indirectly via a third agent form a link that connects them directly. The test statistic of the specification test compares the number of observed transitive relationships to the number of transitive relationships predicted by the dyadic linking model. The dyadic model is rejected if the test detects that it significantly under- or overestimates transitivity.

The idea of using network statistics to elicit the plausibility of dyadic linking was first suggested in Holland and Leinhardt (1978) and subsequently developed in Karlberg (1997) and Karlberg (1999). More recently, Chandrasekhar and Jackson (2016) use simulated network statistics to evaluate a dyadic linking model. They find that a dyadic model without fixed effects predicts too little transitivity.[1] Using a different approach, I replicate their finding. By using a dyadic model with fixed effects, I show that the conclusion in Chandrasekhar and Jackson (2016) is robust to allowing some determinants of the linking decisions to be unobserved.

[1]See also Davis (1970), Watts and Strogatz (1998), and Apicella et al. (2012).

My transitivity test can be interpreted as testing the dyadic model against models that target the formation of transitive relationships. This includes models of strategic network formation with agents who value transitive closure (Leung 2015; Mele 2016; Menzel 2017; Sheng 2016), as well as models in which transitivity is generated by an exogenous mechanism (Wasserman and Pattison 1996; Snijders et al. 2006; Chandrasekhar and Jackson 2016). Most of the models from this list are challenging to implement, computationally hard and make restrictive assumptions about unobserved heterogeneity.[2] My transitivity test can be used to detect networks in which the dyadic linking model, along with its ease of implementation and permissive assumptions about unobserved characteristics, provides a reasonable approximation of the true linking process. Even if the specification test rejects, the dyadic linking model can still serve as a tool to generate useful descriptive statistics such as a measure of link reciprocity that projects out homophily effects and degree heterogeneity.

Related literature Graham (2017) studies a directed version of the model discussed in the present paper. He focuses on inference about the homophily component and considers ML estimation with analytic bias correction as well as an alternative approach that conditions out the incidental parameter. The latter approach has the advantage of producing reliable estimates in sparse networks, i.e., in settings where agent degrees grow only slowly as the number of linking opportunities increases. I justify my approach under the assumption that the network is not sparse. The identification strategy for the conditioning approach in Graham (2017) relies on the assumption of logistic errors. Candelaria (2017), Toth (2017), and Gao (2017) study identification of homophily preferences under non-parametric distributional assumptions.

Shi and Chen (2016) study a linking model in which undirected links between two agents

[2]For example, Bhamidi, Bresler, and Sly (2011) show that the computational cost of fitting an exponential random graph model can be prohibitive.

are observed if the agents reciprocate links in a latent directed network. Similar to my analysis, they assume that the linking rule generates a network that is not too sparse.

T. Yan, Jiang, et al. (2018) study analytical bias correction for an estimator of homophily preferences in a directed dyadic linking model with logistic errors. They also characterize the joint asymptotic distribution of finite collections of estimated fixed effects. Such a result is useful, e.g., to test the hypothesis of no structural degree heterogeneity for subsets of agents.

Charbonneau (2017) identifies homophily preferences in the model with logistic errors using a conditioning approach. Jochmans (2017) demonstrates theoretically and in Monte Carlo simulations that the approach in Charbonneau (2017) is robust to sparsity of the network. His simulations also indicate that analytic bias correction of the type that is proposed in the present paper and in T. Yan, Jiang, et al. (2018), may not work well in sparse settings. The conditioning approach is specific to the homophily parameter in the model with logistic errors and does not extend readily to the inference problems that I consider.

The asymptotic analysis of my linking model benefits from arguments originally developed in the context of nonlinear large-T panel models with fixed effects (Hahn and Newey 2004; Hahn and Kuersteiner 2011; Dhaene and Jochmans 2015). For my proofs, I adapt arguments from Fernández-Val and Weidner (2016) who study nonlinear panel models in the context of a broad class of ML models with fixed effects. Their implicit key assumption is equivalent to assuming that certain derivatives of functionals of the fixed effects satisfy a sparsity assumption. For the dyadic linking model, I verify that this condition is satisfied for the functionals related to my parameters of interest.

Organization of paper Section 2 defines the dyadic linking model and discusses two-step maximum likelihood estimation. Section 3 introduces the asymptotic framework. Section 4 discusses t-tests for structural parameters and Section 5 discusses the specification test. Section 6 reports simulation evidence on the finite sample behavior of my procedures and Section 7 applies the specification test to Indian favor networks.

Notation for networks Let $V = V(N) = \{1, \ldots, N\}$ denote a set of agents (vertices). The set of all ordered tuples from V represents directed links (edges) between agents and is denoted by $E = E(N) = \{(i, j) : i, j \in V(N), i \neq j\}$. For a given link (i, j), i is the sender and j the receiver of the link. To conserve notation, I frequently shorten (i, j) to ij. For $A \subset V$ and $i \in V$, I write $V_{-A} = V \setminus A$ and $V_{-i} = V_{-\{i\}}$.

2. The dyadic linking model

2.1. Definition of model

We observe agents $V(N) = \{1, \ldots, N\}$ and their linking decisions. For every potential link $ij \in E(N)$, the dummy variable Y_{ij} takes the value one if agent i links to agent j and the value zero otherwise. Linking decisions are described by a version of the linking model in Holland and Leinhardt (1981) that models linking decisions as a binary choice. Other versions of this model have recently been studied in Jochmans (2017) and T. Yan, Jiang, et al. (2018). Each agent i is endowed with characteristics $(X_i', \gamma_i^{S,0}, \gamma_i^{R,0})'$, where X_i is an observed vector of agent characteristics and $\gamma_i^{S,0}$ and $\gamma_i^{R,0}$ are unobserved scalar parameters. The link ij is established if the latent surplus Z_{ij} exceeds a link-specific shock U_{ij},

$$Y_{ij} = \mathbf{1}\left(Z_{ij} \geq U_{ij}\right).$$

The link surplus is given by

$$Z_{ij} = X_{ij}'\theta^0 + \gamma_i^{S,0} + \gamma_j^{R,0},$$

where X_{ij} is a known transformation of $(X_i', X_j')'$ that takes values in $\mathbb{R}^{\dim(\theta)}$ and $\theta^0 \in \Theta \subset \mathbb{R}^{\dim(\theta)}$ is an unknown model parameter that parameterizes homophily preferences. We interpret $X_{ij}'\theta^0$ as a measure of social distance between agents i and j that drives homophily of linking decisions.[3] The *sender* or *productivity effect* $\gamma_i^{S,0}$ of sender i summarizes the effect of all characteristics of i that make her efficient at establishing outbound links. The *receiver* or *popularity effect* $\gamma_j^{R,0}$ of receiver j summarizes the effect of all characteristics of j that make her efficient at attracting inbound links. The vector of unobserved agent effects, denoted by $\boldsymbol{\gamma}^0 = (\gamma_i^{S,0}, \gamma_i^{R,0})_{i \in V(N)} \in \Gamma \subset \mathbb{R}^{2N}$, can be interpreted as a structural driver of degree heterogeneity (Graham 2017; Jochmans 2017). I take a fixed effect approach and treat $\boldsymbol{\gamma}^0$ as a parameter that has to be estimated. Identification of the agent effects is achieved by the normalization

$$\sum_{i \in V(N)} \left(\gamma_i^{S,0} - \gamma_i^{R,0}\right) = 0.$$

The shocks (U_{ij}, U_{ji}) are drawn independently across dyads $\{i, j\}$ from a bivariate normal distribution with covariance ρ^0 and marginal variances equal to one. If ρ^0 is positive then agents will tend to reciprocate links. This is why I refer to ρ^0 as the reciprocity parameter.[4]

The flavor of reciprocity modeled here can be interpreted as the effect of a shock at the

[3] For a discussion of homophily in dyadic linking models see Graham (2017) and Jochmans (2017). Toth (2017) discusses specification of X_{ij}.

[4] In models of dyadic link formation with random effects, reciprocity is modeled in a similar way (Hoff 2005; Hoff 2015).

dyad level. For positive ρ^0, suppose that U_{ij} and U_{ji} can be decomposed as

$$U_{ij} = \sqrt{\rho^0} U_{ij}^d + \sqrt{1-\rho^0} U_{ij}^l \quad \text{and} \quad U_{ji} = \sqrt{\rho^0} U_{ij}^d + \sqrt{1-\rho^0} U_{ji}^l,$$

where U_{ij}^d, U_{ij}^l and U_{ji}^l are independent draws from the standard normal distribution. Here, U_{ij}^d represents a shock that affects both linking decisions within the dyad and U_{ij}^l represents an idiosyncratic link-specific shock. Economically, the dyad-specific shock can be interpreted as modeling the effect of a meeting process that randomly introduces people to each other, reducing the cost of establishing links for both parties.[5] The parameter ρ^0 weighs the relative importance of the dyad-specific and link-specific components of U_{ij}.

In Holland and Leinhardt (1981), reciprocity is modeled in a different way, by letting the surplus Z_{ij} depend on the link indicator Y_{ji}. This can be interpreted as modeling agents that derive utility from reciprocated links. Such a specification renders the linking decision endogenous and introduces a strategic element to each linking decision with the possibility of multiple equilibria. Leung (2015), Mele (2016), and Ridder and Sheng (2017) study identifying assumptions for models in which agents make strategic decisions about whether to reciprocate links.

2.2. Two-stage estimation of model parameters

The model is fitted in two stages. The first stage is a pseudo-likelihood approach that ignores the within-dyad correlations and recovers estimates of the homophily parameter θ^0 and the incidental parameter $\boldsymbol{\gamma}^0$ from the marginal link distribution. In the second stage, ρ^0 is estimated by estimated maximum likelihood, substituting the first-stage estimates for unknown population parameters in the likelihood for the reciprocity parameter.

[5] An example in social media (e.g. Tinder) are recommender systems that encourage people to connect to each other.

An alternative to the two-stage procedure is to estimate all parameters simultaneously by maximizing the full information likelihood. This approach yields more efficient estimators but is computationally challenging. In contrast, the two-stage procedure is easy to implement in standard statistical software and numerically stable.[6]

The two-stage estimation proceeds as follows.

Stage 1 For parameter values $\theta \in \Theta$ and $\boldsymbol{\gamma} = \left(\gamma_i^S, \gamma_i^R\right)_{i \in V} \in \Gamma$, define the linking probability $p_{ij}(\theta, \boldsymbol{\gamma}) = \Phi\left(X_{ij}'\theta + \gamma_i^S + \gamma_j^R\right)$, where Φ is the cumulative distribution function of a standard normal random variable. The first-stage estimator $(\hat{\theta}', \hat{\boldsymbol{\gamma}}')'$ solves the constrained maximum likelihood program

$$(\hat{\theta}', \hat{\boldsymbol{\gamma}}')' = \arg\max_{\substack{\theta \in \Theta \\ \boldsymbol{\gamma} \in \Gamma}} \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \Big\{ Y_{ij} \log\left(p_{ij}(\theta, \boldsymbol{\gamma})\right) + (1 - Y_{ij}) \log\left(1 - p_{ij}(\theta, \boldsymbol{\gamma})\right) \Big\} \tag{2.1}$$

$$\text{subject to} \qquad \sum_{i \in V} \left(\gamma_i^S - \gamma_i^R\right) = 0.$$

In practice, the constraint can be eliminated by plugging it into the objective function. Elimination of the constraint yields a probit program with a $N + (N-1) + \dim(\theta)$ dimensional parameter. The unconstrained program can be solved by standard methods such as the `probit` command in Stata, the `glm` command in R, or the `glmfit` command in Matlab.[7]

[6] Each of the two stages solves a concave maximization problem.

[7] Algorithms that exploit the sparse structure of the design matrix, such as `speedglm` in the R package Enea (2013), can speed up the computation of the estimates.

Stage 2 Let $r(\cdot,\cdot,\rho)$ denote the distribution function of a bivariate normal random variable with marginal variances equal to one and covariance ρ, i.e.,

$$r(z_1, z_2, \rho) = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \phi_2(t_1, t_2, \rho)\, dt_2\, dt_1,$$

where $\phi_2(\cdot,\cdot,\rho)$ is the bivariate density

$$\phi_2(t_1, t_2, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{t_1^2 + t_2^2 - 2\rho t_1 t_2}{2(1-\rho^2)}\right].$$

For each dyad $\{i, j\}$, the indicator $Y_{ij}Y_{ji}$ takes the value one if both links within the dyad are observed and the value zero otherwise. For $ij \in E(N)$ define

$$r_{ij}(\theta, \boldsymbol{\gamma}, \rho) = r\big(X_{ij}'\theta + \gamma_i^S + \gamma_j^R, X_{ji}'\theta + \gamma_j^S + \gamma_i^R, \rho\big).$$

This function can be used to compute the conditional probability of observing a reciprocated link. Let $\bar{\mathbb{E}}$ denote the conditional expectation operator that integrates out the randomness in $(U_{ij})_{ij\in E(N)}$. Then,

$$\bar{\mathbb{E}}(Y_{ij}Y_{ji}) = \text{Prob}\big(U_{ij} \leq Z_{ij}, U_{ji} \leq Z_{ji} \mid X_i, X_j, \boldsymbol{\gamma}\big) = r_{ij}(\theta^0, \boldsymbol{\gamma}^0, \rho^0).$$

The second stage estimator $\hat{\rho}$ solves the maximization problem

$$\begin{aligned}\hat{\rho} = \arg\max_{\rho\in[-1+\kappa, 1-\kappa]} \frac{1}{N} \sum_{i\in V} \sum_{j\in V_{-i}} &\Big\{ Y_{ij}Y_{ji} \log\big(r_{ij}(\hat{\theta}, \hat{\boldsymbol{\gamma}}, \rho)\big) \\ &+ (1 - Y_{ij}Y_{ji}) \log\big(1 - r_{ij}(\hat{\theta}, \hat{\boldsymbol{\gamma}}, \rho)\big)\Big\},\end{aligned} \tag{2.2}$$

where $\kappa \in (0, 1)$ is a known constant.

3. Asymptotic framework

My approach is justified by an asymptotic approximation of the network that sends the number of agents to infinity ("large network asymptotics"). The proofs for the asymptotic results presented below can be found in Supplemental Appendix B.

For functions of θ and $\boldsymbol{\gamma}$, we adopt the convention that omitted function arguments indicate evaluation at the true parameter values θ^0 and $\boldsymbol{\gamma}^0$. For example, we write $p_{ij} = p_{ij}(\theta^0, \boldsymbol{\gamma}^0)$. We often consider functions $(z_1, z_2, \rho) \mapsto g(z_1, z_2, \rho)$ that are evaluated at $z_1 = Z_{ij}^*$ and $z_2 = Z_{ji}^*$. To indicate the point of evaluation, we write $g_{ij}(\rho) = g(Z_{ij}^*, Z_{ji}^*, \rho)$. We proceed similarly for partial derivatives and write, e.g., $\partial_\rho r_{ij}(\rho) = \partial_\rho r(z_1, z_2, \rho) \mid_{z_1 = Z_{ij}^*, z_2 = Z_{ji}^*, \rho = \rho^0}$. For functions $z \mapsto g(z)$, write $g_{ij} = g(Z_{ij}^*)$ and $\partial_{z^k} g_{ij} = \partial_{z^k} g(z) \mid_{z = Z_{ij}^*}$ for $k \in \mathbb{N} \cup \{0\}$. Moreover, write $p_{1,ij} = p_{ij}(1 - p_{ij})$ for the conditional variance of Y_{ij}; $r_{1,ij} = r_{ij}(1 - r_{ij})$ for the conditional variance of $Y_{ij}Y_{ji}$; and $\tilde{\rho}_{ij} = (r_{ij} - p_{ij}p_{ji})/\sqrt{p_{1,ij}p_{1,ji}}$ for the conditional correlation between Y_{ij} and Y_{ji}. Finally, let $H_{ij} = \partial_z p_{ij}/p_{1,ij}$ and $\omega_{ij} = H_{ij}(\partial_z p_{ij})$.[8]

The formulas presented below depend on appropriately projected link characteristics.[9] To define the projections, let $\mathcal{P}$ denote the projection operator that orthogonally projects vectors $v = (v_{ij})_{ij \in E(N)}$ onto the space spanned by the agent effects under an inner product weighted by the diagonal matrix with diagonal entries $(\omega_{ij})_{ij \in E(N)}$. In particular, $(\mathcal{P}v)_{ij} = \bar{\gamma}_i^S + \bar{\gamma}_j^R$, where

$$(\bar{\gamma}_i^S, \bar{\gamma}_i^R)_{i \in V} \in \arg\min_{\gamma_i^S, \gamma_i^R} \sum_{i \in V} \sum_{j \in V_{-i}} \omega_{ij} \big(v_{ij} - \gamma_i^S - \gamma_j^R \big)^2.$$

Let $\tilde{X}_k$ denote the residual of the projection of the k$^{\text{th}}$ component of the edge-specific covariate. Formally, let $X_k = (X_{ij,k})_{ij \in E(N)}$ and define $\tilde{X}_k = X_k - \mathcal{P}X_k$. Also, let $\tilde{X}_{ij}$ denote the column vector $(\tilde{X}_{ij,1}, \ldots, \tilde{X}_{ij,\dim(\theta)})'$.

[8] These quantities are linked to the score and the Hessian of the first stage maximum likelihood problem. In particular, writing $\ell_{ij} = Y_{ij} \log(p_{ij}) + (1 - Y_{ij}) \log(1 - p_{ij})$ for the likelihood contribution of link ij, we have $\partial_z \ell_{ij} = H_{ij}(Y_{ij} - p_{ij})$ and $\bar{\mathbb{E}}[-\partial_{z^2} \ell_{ij}] = \omega_{ij}$.

[9] See T. Yan, Jiang, et al. (2018) for an approach that does not rely on projection arguments.

The asymptotic results reported below hold under the following regularity assumptions:

Assumption 1 (Regularity assumptions). *(i)* $\rho^0 \in [-1+2\kappa, 1-2\kappa]$.

There is an event A_N with $P(A_N) \to 1$ such that on A_N:

(ii) Let $\lambda_1(M)$ denote the smallest eigenvalue of a matrix M. For $\bar{W}_{1,N}$ as defined in Theorem 1, $\liminf_{N\to\infty} \lambda_1(\bar{W}_{1,N}) > 0$.

(iii) For $k = 1, \dots, \dim(\theta)$ and $i \in V(N)$, $\limsup_{N\to\infty} \frac{1}{N} \sum_{j \in V_{-i}} \tilde{X}^4_{ij,k} < \infty$.

(iv) Let $\mathcal{L}$ as defined in (A.1) *and $\bar{\mathcal{H}}$ as defined in* (A.2) *in Supplemental Appendix A and let b denote the associated penalty parameter. There is $b > 0$ such that, for all N, $\mathcal{L}$ is concave on $\Theta \times \Gamma$ and $\bar{\mathcal{H}}$ is positive definite.*

(v) There are p_{min} and p_{max} such that $0 < p_{min} < p_{ij} < p_{max} < 1$ for all $ij \in E(N)$.

Assumption 1(i) rules out perfectly correlated within-dyad shocks. This implies that the errors U_{ij} cannot be fully explained by a dyad-level shock. Assumption 1(ii) ensures that, asymptotically, the variance of $\hat{\theta}$ is non-degenerate. The corresponding assumption in Fernández-Val and Weidner 2016 requires the limiting variance to be positive definite. Since I condition on covariates and fixed effects, this limit may not exist. The moment condition Assumption 1(iii) guarantees that the asymptotic bias and variance of $\hat{\theta}$ are finite. As in Fernández-Val and Weidner 2016, the theoretical analysis of the maximum likelihood program (2.1) imposes the normalization of the fixed effects using penalization. Assumption 1(iv) requires the sample and population versions of the penalized program to be concave. This can be interpreted as an assumption of sufficient "within variation".

Assumption 1(v) implies that the linking rule generates a dense network (i.e., a network that is not sparse). This assumption may be restrictive in some social networks (Graham 2017; Jochmans 2017). For a related dyadic linking model with logistic errors, T. Yan, Jiang, et al. (2018) show that analytic bias correction of the homophily parameter can be justified even with vanishing linking probabilities.[10] In my Monte Carlo simulations, I investigate the robustness of my procedures in sparse designs.

4. Inference with respect to the model parameters

4.1. A t-test for the homophily parameter

The dyadic linking model bears some similarity to panel models with individual and time fixed effects: In the dyadic model, agent i faces $(N-1)$ linking choices that each depend on i's own sender effect and the receiver effect of the potential linking partner. In a panel model, agent i makes choices in T time periods, each depending on her own individual effect and the time of the respective time period. Fernández-Val and Weidner (2016) study incidental parameter bias in the panel model with two-sided fixed effects. The following theorem establishes a companion result to Theorem 4.1 in Fernández-Val and Weidner (2016) for networks.[11]

Theorem 1 (Distribution of $\hat{\theta}$)**.** *Let* $B_N^{\theta} = B_N^{\theta,S} + B_N^{\theta,R}$, *where*

$$B_N^{\theta,S} = \left[\frac{1}{2N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}\omega_{ij}\tilde{X}_{ij}\tilde{X}_{ij}'}{\sum_{j\in V_{-i}}\omega_{ij}}\right]\theta^0, \qquad B_N^{\theta,R} = \left[\frac{1}{2N}\sum_{j\in V}\frac{\sum_{i\in V_{-j}}\omega_{ij}\tilde{X}_{ij}\tilde{X}_{ij}'}{\sum_{i\in V_{-j}}\omega_{ij}}\right]\theta^0,$$

[10] Their result requires that linking probabilities vanish sufficiently slowly to allow us to observe an infinite number of connections for all agents in the limit network. This is an intuitive requirement for a procedure that relies on point identification of all fixed effects.

[11] As noted in Y. Yan et al. (2016), the result for the panel model does not imply the corresponding result in the network setting. My proof builds on the results for general ML models with additive fixed effects in Fernández-Val and Weidner 2016. Checking that the linking model satisfies all assumptions of the general result is similar, but not completely congruent, to checking the assumptions for the panel setting. See also Candelaria (2017) for a discussion of how the incidental parameter problem in networks is different from the incidental parameter problem in panels.

and let

$$\begin{aligned}\bar{W}_{1,N} =& \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \omega_{ij} \tilde{X}_{ij} \tilde{X}_{ij}', \\ \bar{W}_{2,N} =& \bar{W}_{1,N} + \frac{1}{N(N-1)} \sum_{i \in V} \sum_{j \in V_{-i}} \tilde{\rho}_{ij} \sqrt{\omega_{ij}\omega_{ji}} \tilde{X}_{ij} \tilde{X}_{ji}'.\end{aligned}$$

Under Assumption 1

$$\bar{W}_{2,N}^{-1/2} \big(N \, \bar{W}_{1,N} (\hat{\theta} - \theta^0) - B_N^{\theta} \big) = \mathcal{N} \left(0, \mathbb{I}_{\dim(\theta)} \right) + o_p(1).$$

To converge to a normal distribution, the difference between the estimator $\hat{\theta}$ and true value θ^0 has to be inflated proportional to the number of agents N. In the dense network setting considered here, θ^0 is estimated based on the observed linking decisions about $N(N-1)$ potential links. Therefore, the rate of convergence N is the conventional parametric rate corresponding to the square root of the sample size (cf. Graham 2017). The expression for the asymptotic bias term B_N^{θ} corresponds to a "naïve" translation of the corresponding formula given in Fernández-Val and Weidner (2016) for panel models.

Let $\hat{B}_N^{\theta}$, $\widehat{W}_{1,N}$ and $\widehat{W}_{2,N}$ denote consistent estimators of B_N^{θ}, $\bar{W}_{1,N}$ and $\bar{W}_{2,N}$, respectively. Theorem 1 implies

$$\left(\widehat{W}_{1,N}^{-1} \widehat{W}_{2,N} \widehat{W}_{1,N}^{-1} / N^2 \right)^{-1/2} \left(\hat{\theta} - \theta^0 - \widehat{W}_{1,N}^{-1} \hat{B}_N^{\theta} / N \right) = \mathcal{N} \left(0, \mathbb{I}_{\dim(\theta)} \right) + o_p(1). \tag{4.1}$$

This result can be used to construct bias-corrected t-statistics to test, e.g., statistical significance of the estimated components of $\hat{\theta}$.

The estimators $\hat{B}_N^{\theta}$, $\widehat{W}_{1,N}$ and $\widehat{W}_{2,N}$ can be constructed by a plug-in approach, i.e., by replacing the population parameters in B_N^{θ}, $\bar{W}_{1,N}$ and $\bar{W}_{2,N}$ by the estimates obtained by

ML estimation of the model.[12] Preliminary estimation of ρ^0 can be avoided by estimating $\bar{W}_{2,N}$ by

$$\widehat{W}_{2,N} = \frac{1}{N^2} \sum_{\substack{i,j \in V \\ i<j}} \left(\hat{\tilde{X}}_{ij} \hat{H}_{ij} (Y_{ij} - \hat{p}_{ij}) + \hat{\tilde{X}}_{ji} \hat{H}_{ji} (Y_{ji} - \hat{p}_{ji}) \right)^2,$$

where $\hat{\tilde{X}}_{ij}$, $\hat{H}_{ij}$ and $\hat{p}_{ij}$ are the plug-in estimators of $\tilde{X}_{ij}$, H_{ij} and p_{ij}. This variance estimator clusters errors at the dyad level.[13]

4.2. A t-test for the reciprocity parameter

Let $m_{ij}(\theta, \boldsymbol{\gamma}, \rho) = Y_{ij} Y_{ji} \log\big(r_{ij}(\theta, \boldsymbol{\gamma}, \rho)\big) + (1 - Y_{ij} Y_{ji}) \log\big(1 - r_{ij}(\theta, \boldsymbol{\gamma}, \rho)\big)$ so that we can write the second-stage likelihood evaluated at the true structural parameters as

$$\mathcal{M}(\rho) = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} m_{ij}\big(\hat{\theta}, \hat{\boldsymbol{\gamma}}, \rho\big).$$

With $J_{ij} = \partial_\rho r_{ij} / r_{1,ij}$, the corresponding score is

$$\partial_\rho \mathcal{M} = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \partial_\rho m_{ij} = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} J_{ij} (Y_{ij} Y_{ji} - r_{ij}).$$

Let $\Omega = \mathcal{P} M$ for $M = (M_{ij})_{ij \in E(N)}$ and $M_{ij} = J_{ij} (\partial_{z_1} r_{ij}) / \omega_{ij}$. Let $\partial_z \ell_{ij} = H_{ij} (Y_{ij} - p_{ij})$ denote the contribution of link ij to the score of the first stage maximum likelihood problem.

[12] The plug-in approach is expected to yield consistent estimators (cf. Theorem 4.3 and Lemma S.1 in Fernández-Val and Weidner 2016).

[13] By default, most software packages for probit estimation estimate the variance matrix for $\hat{\theta}$ under the assumption that the information matrix equality holds, i.e., $\bar{W}_{1,N} = \bar{W}_{2,N}$. This assumption is justified if within-dyad shocks are uncorrelated in which case $\tilde{\rho}_{ij} = 0$.

The term

$$\overline{\text{corr}}_i = \frac{\sum_{j\in V_{-i}} \tilde{\rho}_{ij}\sqrt{\omega_{ij}\omega_{ji}}}{\left(\sum_{j\in V_{-i}} \omega_{ij}\right)^{1/2}\left(\sum_{j\in V_{-i}} \omega_{ji}\right)^{1/2}}.$$

measures the correlation of all $\partial_z \ell_{ij}$ in the neighborhood of agent i.[14] The following result characterizes the asymptotic behavior of $\hat{\rho}$.

Theorem 2 (Distribution of $\hat{\rho}$)**.** *Let*

$$Q_N = -\frac{1}{N^2}\sum_{i\in V}\sum_{j\in V_{-i}} J_{ij}(\partial_{z_1} r_{ij})\tilde{X}_{ij}$$

and

$$\begin{aligned} v^{\rho}_{1,N} =& \frac{1}{N(N-1)/2}\sum_{i\in V}\sum_{j\in V_{-i}} J_{ij}(\partial_{\rho} r_{ij}) \\ v^{\rho}_{2,N} =& v^{\rho}_{1,N} + \frac{1}{N(N-1)}\sum_{i\in V}\sum_{j\in V_{-i}} \Bigg\{ 4(q_{N,ij}-\Omega_{ij})J_{ij}(\partial_z p_{ij})\frac{r_{ij}}{p_{ij}} + 2(q_{N,ij}-\Omega_{ij})^2\omega_{ij} \\ & + 2(q_{N,ij}-\Omega_{ij})(q_{N,ji}-\Omega_{ji})\tilde{\rho}_{ij}\sqrt{\omega_{ij}\omega_{ji}} \Bigg\}, \end{aligned}$$

[14]Note that $\overline{\text{corr}}_i = \left(\sum_{j\in V_{-i}} \bar{\mathbb{E}}(\partial_z \ell_{ij}\partial_z \ell_{ji})\right) / \sqrt{\left(\sum_{j\in V_{-i}} \bar{\mathbb{E}}(\partial_z \ell_{ij})^2\right)\left(\sum_{j\in V_{-i}} \bar{\mathbb{E}}(\partial_z \ell_{ji})^2\right)}$.

where $q_{N,ij} = Q_N' \bar{W}_{1,N}^{-1} \tilde{X}_{ij}$. *Moreover, let* $B_N^\rho = B_N^{\rho,S} + B_N^{\rho,R} + B_N^{\rho,SR}$ *with*

$$
\begin{aligned}
B_N^{\rho,S} =& \frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} \left\{ \partial_{z_1} J_{ij} \left(\frac{r_{ij} \partial_z p_{ij}}{p_{ij}} - \partial_{z_1} r_{ij} \right) + \frac{1}{2} \Omega_{ij} H_{ij} (\partial_{z^2} p_{ij}) - \frac{1}{2} J_{ij} (\partial_{z_1^2} r_{ij}) \right\}}{\sum_{j \in V_{-i}} \omega_{ij}}, \\
B_N^{\rho,R} =& \frac{1}{N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} \left\{ \partial_{z_1} J_{ij} \left(\frac{r_{ij} \partial_z p_{ij}}{p_{ij}} - \partial_{z_1} r_{ij} \right) + \frac{1}{2} \Omega_{ij} H_{ij} (\partial_{z^2} p_{ij}) - \frac{1}{2} J_{ij} (\partial_{z_1^2} r_{ij}) \right\}}{\sum_{i \in V_{-j}} \omega_{ij}}, \\
B_N^{\rho,SR} =& - \frac{1}{N} \sum_{i \in V} \frac{\overline{\text{corr}}_i \sum_{j \in V_{-i}} \left\{ (\partial_{z_1} J_{ij})(\partial_{z_1} r_{ji}) + (\partial_{z_1} J_{ji})(\partial_{z_1} r_{ij}) + J_{ij} (\partial_{z_1 z_2} r_{ij}) \right\}}{\left(\sum_{j \in V_{-i}} \omega_{ij} \right)^{1/2} \left(\sum_{j \in V_{-i}} \omega_{ji} \right)^{1/2}}.
\end{aligned}
$$

Under Assumption 1,

$$
\frac{N(\hat{\rho} - \rho^0) - 2 \left(Q_N' \bar{W}_{1,N}^{-1} B_N^\theta + B_N^\rho \right) / v_{1,N}^\rho}{\sqrt{v_{2,N}^\rho} / v_{1,N}^\rho} = \mathcal{N}(0, 2) + o_p(1).
$$

Let $\hat{B}_N^\rho$, $\hat{v}_{1,N}^\rho$, $\hat{v}_{2,N}^\rho$ and $\hat{Q}_N$ denote consistent estimators of B_N^ρ, $v_{1,N}^\rho$, $v_{2,N}^\rho$ and Q_N, respectively.[15] Theorem 2 implies

$$
\frac{\hat{\rho} - \rho^0 - 2 \left(\hat{Q}_N' \widehat{W}_{1,N}^{-1} \hat{B}_N^\theta + \hat{B}_N^\rho \right) / (N \hat{v}_{1,N}^\rho)}{\sqrt{2 \hat{v}_{2,N}^\rho} / (N \hat{v}_{1,N}^\rho)} = \mathcal{N}(0, 1) + o_p(1). \tag{4.2}
$$

The term on the left-hand side of the equality is a bias-corrected t-statistic for $\hat{\rho}$. It can be used to test hypotheses about the true reciprocity parameter.

The proof of Theorem 2 exploits results in Fernández-Val and Weidner 2016 who study functionals of the incidental parameter in a class of ML models with additive fixed effects. They apply their results to panel models with individual and time fixed effects and derive an asymptotic bias that exhibits a factoring property: The bias in the model with both individual and time fixed effects can be recovered as the sum of the bias terms in the

[15] In Supplemental Appendix H, I discuss how to compute certain derivatives of bivariate normal probabilities that show up in the formulas in Theorem 2

two models with only individual or only time fixed effects. Because of the "cross term" $B_N^{\rho,SR}$, the asymptotic bias in Theorem 2 does not factor. This behavior is caused by the within-dyad correlation of linking decisions.[16]

However, as illustrated by Theorem 1, even with correlated within-dyad shocks, it is possible to derive an asymptotic bias that factors. The relevant difference between Theorem 1 and Theorem 2 is that they study functionals of the incidental parameter that exhibit differently structured Hessians. The appropriate Hessian for Theorem 1 has strong diagonal and weak off-diagonal elements.[17] In a Taylor expansion around the true incidental parameter, the interaction of $\partial_z \ell_{ij}$ and $\partial_z \ell_{ji}$ is weighed by a weak element and is not of first order. The corresponding Hessian for Theorem 2 has a two-by-two block structure where each block has strong diagonal and weak off-diagonal elements. In a Taylor expansion around the true incidental parameter, the interaction of $\partial_z \ell_{ij}$ and $\partial_z \ell_{ji}$ is weighed by a strong element and cannot be ignored in the limit.

5. Specification testing

5.1. Motivation of testing approach based on transitive relationships

The dyadic linking model induces a theoretical probability distribution of the random graph $\{Y_{ij}\}_{ij \in E(N)}$. We can construct tests of model specification by comparing the observed behavior of a particular network feature to the behavior that is expected under the dyadic model. The linking model targets the linking behavior within pairs of agents and will therefore always fit the network relationships within dyads (groups of two agents) fairly well. To test the model, we can check how well the dyadic linking model replicates the behavior

[16] If within-dyad shocks are not correlated, i.e., $\rho^0 = 0$, then $B_N^{\rho,SR} = 0$ and the bias term B_N^{ρ} factors.

[17] A strong element of the appropriately standardized Hessian is of asymptotic order $O(1)$, a weak element is of order $O(N^{-1})$.

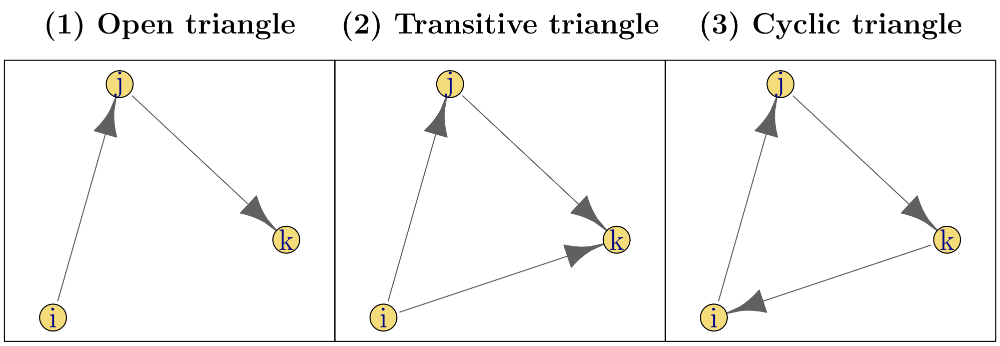

Figure 1: An open, a transitive and a cyclic triangle.

within groups of three or larger. In particular, I consider a test of model specification based on transitive relationships within triads (groups of three).

To introduce the notion of transitive relationships, consider a network where agent i has linked to agent j, and j has linked to agent k (see panel 1 in Figure 1). Agents i and k are already indirectly connected and can "close" the open triangle by adding a link that connects them directly. In a directed network, there are two ways of closing the triangle; i can link to k to form a *transitive triangle* (panel 2 in Figure 1), or k can link to i to form a *cyclic triangle* (panel 3 in Figure 1).[18] Whether it is more salient to test for closure in a transitive or cyclic sense, depends on the economics of the network. For ease of exposition, I focus on a test based on transitive triangles. In Supplemental Appendix F, I adapt my results to a test based on cyclic triangles.

For distinct $i, j, k \in V(N)$, the transitive triangle $\beta = \{ij, jk, ik\}$ is observed if $\beta \subset \{ij \in$

[18]The terms *transitive triangle* and *cyclic triangle* are adapted from the notion of transitive and cyclic triads in Davis and Leinhardt 1972.

$E(N) : Y_{ij} = 1\}$.[19] The set of all possible transitive triangles is given by

$$B = B(N) = \{\{(i,j),(j,k),(i,k)\} : \{i,j,k\} \subset V(N), |\{i,j,k\}| = 3\}.$$

For $\beta \in B$, the binary indicator $A_\beta = \prod_{e \in \beta} Y_e$ takes the value one if β is observed, and the value zero otherwise. The number of observed transitive triangles is given by

$$S_N = \sum_{\beta \in B(N)} A_\beta.$$

My test of model specification compares the observed transitivity S_N to the transitivity predicted by the dyadic linking model. For a given vector of agent characteristics $(X_i', \gamma_i^{S,0}, \gamma_i^{R,0})_{i \in V}$, the best prediction of the observed number of transitive triangles is given by $\bar{\mathbb{E}}\, S_N$. The discrepancy between the observed and the predicted level of transitivity can be summarized by a measure of excess transitivity defined as

$$T_N^{\text{oracle}} = \frac{S_N - \bar{\mathbb{E}}\, S_N}{N^3}, \tag{5.1}$$

where the denominator normalizes by the number of transitive triangles in the complete graph, $|B(N)| = N^3$.[20] Positive values of this statistic indicate that we observe *more* transitive relationships than expected, negative values of the statistic indicate that we observe *less* transitive relationships than expected. Under the dyadic linking model, the variance of T_N^{oracle} vanishes as the size of the network grows. Therefore, we can interpret "large" values of T_N^{oracle} as evidence against the validity of the dyadic model.

[19] There may be other interactions within the triad $\{i,j,k\}$, such as a link from k to j. These do not play a role in determining the presence of β. In contrast to *triadic configurations* (Davis and Leinhardt 1972), triangles are defined by the presence but not by the absence of links.

[20] This measure of excess transitivity translates a concept for undirected networks discussed in Karlberg (1997) to directed networks. An alternative is to standardize by the number of open triangles, yielding the *clustering coefficient* (Karlberg 1999; Jackson 2008, p. 37).

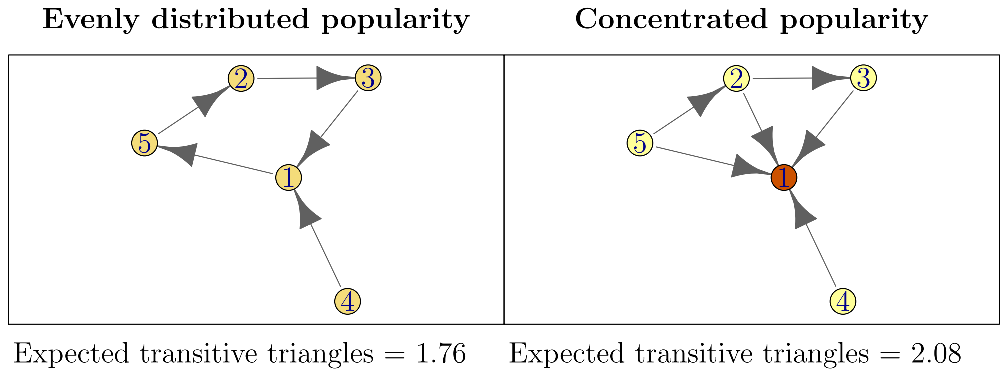

Figure 2: The effect of redistribution of agent popularity.

This kind of specification test can be interpreted in the tradition of transitivity tests in the sociometric literature (Holland and Leinhardt 1978; Karlberg 1997; Karlberg 1999). Transitivity tests assess the explanatory power of the transitive structure of a network. Holland and Leinhardt (1978) argue that it is important to compute the expected transitivity under a reference distribution that replicates key features of the dyadic interactions such as degree-heterogeneity and reciprocity.[21] Failure to account for dyadic sources of transitivity may lead a researcher to erroneously ascribe explanatory power to the transitive structure of the network ("spurious transitivity"). My reference distribution fulfills this requirement, since it is derived from a model of dyadic link formation that accounts for structural sources of reciprocity (correlation of within dyad shocks) and degree heterogeneity (productivity and popularity fixed effects).

Example 1. The effect of model specification on expected transitivity can be illustrated by a simple example. Consider networks on $N = 5$ agents with $\gamma_i^{S,0} = 0$ for $i \in V(5)$ and $\sum_{i \in V(5)} \gamma_i^{R,0} = 2.5$. The latent link surplus is given by

$$Z_{ij} = -1 + \gamma_j^{R,0}.$$

[21] Faust (2007) and Graham (2015) discuss the close relationship between the degree distribution and the triadic structure of a network.

First, consider distributing popularity evenly among agents by assigning $\gamma_i^{R,0} = 0.5$ for all $i \in V$. This scenario is depicted in the first panel of Figure 2. For an alternative scenario, set $\gamma_1^{R,0} = 2.5$ and $\gamma_i^{R,0} = 0$ for $i \in V_{-1}$. This scenario is depicted in the second panel of Figure 2. The redistribution of popularity increases the expected number of transitive triangles. Intuitively, concentrated popularity serves as a kind of coordination device that makes the occurrence of transitive relationships more likely.

Holland and Leinhardt (1978) and Karlberg (1999) do not explicitly model dyadic link formation. Instead, they condition on observed network characteristics that they assume to be driven by dyadic interactions. It is not clear how to compute critical values that appropriately account for the effect of conditioning on observed network features.[22] Karlberg (1999) computes critical values using a simulation approach, but does not justify this procedure theoretically. My approach is amenable to large sample arguments and I show that critical values can be computed from a normal approximation.

5.2. The test statistic for the transitivity test

Under the dyadic linking model, the conditional probability of observing a transitive triangle $\beta \in B(N)$ is given by $\bar{\mathbb{E}}\, A_\beta = \prod_{e \in \beta} p_e(\theta^0, \boldsymbol{\gamma}^0)$. In reality, θ^0 and $\boldsymbol{\gamma}^0$ are unknown and it is not feasible to compute $\bar{\mathbb{E}} S_N = \sum_{\beta \in B(N)} \bar{\mathbb{E}}\, A_\beta$ in T_N^{oracle}. A feasible test statistic is given by

$$T_N = \frac{S_N - \widehat{\mathbb{E}\, S_N}}{N^3},$$

[22]By conditioning on the observed degree sequence, Karlberg (1999) introduces a sample dependence that is reminiscent of the preliminary estimation of the structural linking model in my case.

where we replaced $\bar{\mathbb{E}}\, S_N$ by the naïve plug-in estimator

$$\widehat{\mathbb{E}\, S_N} = \sum_{\beta \in B(N)} \prod_{e \in \beta} p_e(\hat{\theta}, \hat{\boldsymbol{\gamma}})$$

with $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_i^S, \hat{\gamma}_i^R)_{i \in V}$. A theoretical analysis of T_N can be based on the decomposition

$$N\, T_N = N\, T_N^{\text{oracle}} - N^{-2} \sum_{\beta \in B(N)} \left(\prod_{e \in \beta} p_e\big(\hat{\theta}, \hat{\boldsymbol{\gamma}}\big) - \prod_{e \in \beta} p_e\big(\theta^0, \boldsymbol{\gamma}^0\big) \right). \tag{5.2}$$

Both terms on the right-hand side are of the same stochastic order and contribute to the asymptotic distribution. The first term is the appropriately scaled oracle statistic. Under the dyadic linking model it is centered at zero. The second term represents the effect of estimating linking probabilities. Because of the incidental parameter problem, this term is not centered at zero. Consequently, the sign of T_N cannot be interpreted in the same way as the sign of T_N^{oracle}. In particular, values of T_N that are close to zero do not indicate that the observed level of transitivity is consistent with the true dyadic linking model.

In preparation for a formal analysis of T_N, let

$$\boldsymbol{\beta}_{ij}^N = \frac{1}{H_{ij} N} \sum_{\substack{\beta \in B(N) \\ \beta \ni ij}} \bar{\mathbb{E}}[A_\beta \mid Y_{ij} = 1] = \frac{1}{H_{ij} N} \sum_{\substack{\beta \in B(N) \\ \beta \ni ij}} p_{-ij}^T(\beta),$$

where for $ij \in E(N)$ and $\beta \in B(N)$

$$p_{-ij}^T(\beta) = \bar{\mathbb{E}}[A_\beta \mid Y_{ij} = 1] = \prod_{e \in \beta \setminus \{ij\}} p_e$$

is the probability of observing the triangle β conditional on observing the edge ij. The sum in $\boldsymbol{\beta}_{ij}^N$ counts the expected number of observed triangles containing the link ij conditional on observing ij. Let $\boldsymbol{\beta}^N = (\boldsymbol{\beta}_{ij}^N)_{ij \in E(N)}$ and define the projected vector $\tilde{\boldsymbol{\beta}}^N = \boldsymbol{\beta}^N - \mathcal{P}\boldsymbol{\beta}^N$.

The following result establishes convergence of T_N to a normal limit and gives expressions for its asymptotic bias and variance.

Theorem 3 (Transitvity test)**.** *Let*

$$U_N = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \boldsymbol{\beta}_{ij}^N \omega_{ij} \tilde{X}_{ij}$$

and $\tilde{u}_{N,ij} = U_N' \bar{W}_{1,N}^{-1} \tilde{X}_{ij}$ *and suppose that Assumption 1 holds. Then*

$$\frac{N T_N + B_N^T + U_N' \bar{W}_{1,N}^{-1} B_N^{\theta}}{\sqrt{v_N^T}} = \mathcal{N}(0,1) + o_p(1),$$

where

$$v_N^T = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left\{ \left(\tilde{\boldsymbol{\beta}}_{ij}^N - \tilde{u}_{N,ij} \right)^2 \omega_{ij} + \left(\tilde{\boldsymbol{\beta}}_{ij}^N - \tilde{u}_{N,ij} \right) \left(\tilde{\boldsymbol{\beta}}_{ji}^N - \tilde{u}_{N,ji} \right) \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}} \right\}$$

and $B_N^T = B_N^{T,S} + B_N^{T,R} + B_N^{T,SR}$ *with*

$$\begin{aligned}
B_N^{T,S} =& \frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} H_{ij} (\partial_{z^2} p_{ij}) \tilde{\boldsymbol{\beta}}_{ij}^N}{\sum_{j \in V_{-i}} \omega_{ij}} \\
&+ \frac{1}{2N} \sum_{i \in V} \frac{N^{-1} \sum_{j \in V_{-i}} \sum_{k \in V_{-\{i,j\}}} (\partial_z p_{ij})(\partial_z p_{ik}) \left[p_{jk} + p_{kj} \right]}{\sum_{j \in V_{-i}} \omega_{ij}} \\
B^{T,R} =& \frac{1}{2N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} H_{ij} (\partial_{z^2} p_{ij}) \tilde{\boldsymbol{\beta}}_{ij}^N}{\sum_{j \in V_{-i}} \omega_{ij}} \\
&+ \frac{1}{2N} \sum_{j \in V} \frac{N^{-1} \sum_{i \in V_{-j}} \sum_{k \in V_{-\{i,j\}}} (\partial_z p_{ij})(\partial_z p_{kj}) \left[p_{ik} + p_{ki} \right]}{\sum_{i \in V_{-j}} \omega_{ij}} \\
B_N^{T,SR} =& \frac{1}{N} \sum_{i \in V} \frac{\overline{\text{corr}}_i \, N^{-1} \sum_{j \in V_{-i}} \sum_{k \in V_{-\{j,k\}}} (\partial_z p_{ij})(\partial_z p_{ki}) p_{kj}}{\left(\sum_{j \in V_{-i}} \omega_{ij} \right)^{1/2} \left(\sum_{j \in V_{-i}} \omega_{ji} \right)^{1/2}}.
\end{aligned}$$

If linking probabilities are sufficiently small, $p_{ij} \leq 1/2$ for all $ij \in E(N)$, and positively correlated within dyads, i.e., $\rho^0 \geq 0$, then the bias term B_N^T is positive. In particular, if the link surplus does not contain a homophily component, then T_N is guaranteed to be centered at a negative value if the dyadic linking model is the true model. In more general specifications, the sign of the bias depends on the numerical values of the structural parameters and can be positive or negative.

In the case of uncorrelated within-dyad shocks and no covariates, the asymptotic variance of NT_N is given by

$$v_N^T = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left(\tilde{\boldsymbol{\beta}}_{ij}^N\right)^2 \omega_{ij}$$

and, by Lemma A.4 in Supplemental Appendix A, the asymptotic variance of the oracle statistic NT_N^{oracle} is given by

$$v_N^{o,T} = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left(\boldsymbol{\beta}_{ij}^N\right)^2 \omega_{ij}.$$

When passing from known linking probabilities to estimated linking probabilities, we replace $\boldsymbol{\beta}_{ij}^N$ by its projection $\tilde{\boldsymbol{\beta}}_{ij}^N$ onto the space that is orthogonal to the space spanned by the fixed effects. By definition of the projection operator, we have $v_N^T < v_N^{o,T}$ so that the plug-in statistic T_N estimates the expected excess transitivity more precisely than the oracle statistic T_N^{oracle}.[23] Intuitively, T_N compares the observed transitivity against the transitivity predicted by the dyadic model that provides the best fit. Therefore, my test looks only at the variation in transitivity that cannot be explained by degree distributions that are spanned by the sender and receiver effects. For the oracle test, the sampling error of the

[23] Efficiency gains from replacing population quantities by estimated quantities have been observed elsewhere in the Econometric literature, e.g., in Hahn (1998) and Abadie and Imbens (2016).

degree distribution under the true dyadic linking model provides an additional source of uncertainty.

In the general setting with covariates and possibly correlated within-dyad shocks, it is not clear how v_N^T and $v_N^{o,T}$ are ranked. For the data in my empirical application, I estimate that $v_N^T < v_N^{o,T}$ by a substantial margin.

Let $\hat{B}_N^T$, $\hat{U}_N$ and $\hat{v}_N^T$ denote consistent estimators of B_N^T, U_N and v_N^T, respectively. Theorem 3 implies that the studentized statistic

$$\hat{T}_N^{\text{stud}} = \frac{T_N + \left(\hat{B}_N^T + \hat{U}_N' \left(\widehat{W}_{1,N}\right)^{-1} \hat{B}_N^{\theta}\right)/N}{\sqrt{\hat{v}_N^T/N}} \tag{5.3}$$

follows approximately a standard normal distribution. A feasible transitivity test can be based on the test statistic $\hat{T}_N^{\text{stud}}$. Its sign can be interpreted in the same way as the sign of the infeasible statistic T_N^{oracle}; large positive values indicate that the dyadic model underestimates transitivity and large negative values indicate that the dyadic model overestimates transitivity.

The estimators $\hat{B}_N^T$, $\hat{U}_N$ and $\hat{v}_N^T$ can be constructed by a plug-in approach, i.e., by replacing the population parameters in B_N^T, U_N and v_N^T by the estimates obtained by ML estimation of the linking model. To reduce the computational burden, the test statistic can be computed without a preliminary estimate of ρ^0, if the asymptotic variance is estimated by clustering at the dyad level,

$$\hat{v}_N^T = \frac{1}{N^2} \sum_{\substack{i,j \in V \\ i<j}} \left(\left(\hat{\tilde{\boldsymbol{\beta}}}_{ij}^N - \hat{\tilde{u}}_{N,ij} \right) \hat{H}_{ij}(Y_{ij} - \hat{p}_{ij}) + \left(\hat{\tilde{\boldsymbol{\beta}}}_{ji}^N - \hat{\tilde{u}}_{N,ji} \right) \hat{H}_{ji}(Y_{ji} - \hat{p}_{ji}) \right)^2,$$

and if $\overline{\text{corr}}_i$ is estimated by

$$\widehat{\text{corr}}_i = \frac{\sum_{j \in V_{-i}} \hat{H}_{ij}\hat{H}_{ji}(Y_{ij} - \hat{p}_{ij})(Y_{ji} - \hat{p}_{ji})}{\sqrt{\sum_{j \in V_{-i}} \hat{\omega}_{ij} \sum_{j \in V_{-i}} \hat{\omega}_{ji}}},$$

where $\hat{\bar{\boldsymbol{\beta}}}^N_{ij}$, $\hat{\bar{u}}_{N,ij}$, $\hat{p}_{ij}$ and $\hat{\omega}_{ij}$ are the obvious plug-in estimators.

6. Monte Carlo simulations

In this section, I investigate the finite sample performance of my procedures in Monte Carlo simulations.[24] Agent $i \in V(N)$ is characterized by an observed scalar covariate X_i,

$$X_i = 1 - 2\,\mathbf{1}\{i \text{ is odd}\}$$

and agent fixed effects

$$\gamma_i^{S,0} = \left(\frac{N-i}{N-1}\right) C_N$$

and $\gamma_i^{R,0} = \gamma_i^{S,0}$, where $C_N \in \{\log\log N, \log^{1/2} N, 2\log^{1/2} N, \log N\}$ is a sparsity parameter.

This parameterized family of fixed effect specifications has first been proposed in T. Yan, Leng, Zhu, et al. (2016) and has also been used in Jochmans (2017) and T. Yan, Jiang, et al. (2018). Let the density of a network be defined as the fraction of possible links that are observed, i.e., density $= \sum_{i \in V} \sum_{j \in V_{-i}} Y_{ij}/(N(N-1))$. The larger C_N, the denser the generated networks tend to be. For $C_N = \log N$, only about 3% of all possible links are realized in my simulation designs.

As in Graham (2017), the link-specific covariate is given by $X_{ij} = X_i X_j$. In this

[24] The simulations were carried out on computational resources at Chalmers Center for Computational Science and Engineering provided by the Swedish National Infrastructure for Computing.

				bias				rej prob	
N	ρ	C_N	density	$\hat{\theta}$	$\hat{\theta}$ bc	$\hat{\rho}$	$\hat{\rho}$ bc	$\hat{\theta}$	$\hat{\rho}$
50	0.0	$\log\log N$	0.19	1.22	0.13	-0.02	-0.05	0.07	0.09
50	0.0	$\log^{1/2} N$	0.12	1.10	0.22	-0.00	-0.06	0.08	0.08
50	0.0	$2\log^{1/2} N$	0.06	0.70	0.33	-0.13	-0.20	0.07	0.10
50	0.0	$\log N$	0.03	0.88	0.59	-0.10	-0.18	0.08	0.15
50	0.5	$\log\log N$	0.19	1.14	0.17	0.25	-0.17	0.13	0.11
50	0.5	$\log^{1/2} N$	0.12	1.00	0.24	0.50	-0.01	0.13	0.11
50	0.5	$2\log^{1/2} N$	0.06	0.71	0.34	0.62	0.03	0.11	0.12
50	0.5	$\log N$	0.03	-	-	-	-	-	-
70	0.0	$\log\log N$	0.18	1.11	0.03	0.04	0.01	0.10	0.07
70	0.0	$\log^{1/2} N$	0.11	1.12	0.16	-0.01	-0.07	0.11	0.12
70	0.0	$2\log^{1/2} N$	0.06	0.97	0.30	-0.02	-0.10	0.09	0.09
70	0.0	$\log N$	0.03	0.71	0.37	-0.18	-0.26	0.07	0.12
70	0.5	$\log\log N$	0.18	1.03	0.07	0.34	-0.07	0.14	0.10
70	0.5	$\log^{1/2} N$	0.11	1.12	0.23	0.46	-0.08	0.12	0.10
70	0.5	$2\log^{1/2} N$	0.06	1.05	0.27	0.65	0.06	0.13	0.11
70	0.5	$\log N$	0.03	0.69	0.37	0.58	-0.19	0.09	0.03

Table 1: Simulation results for $\hat{\theta}$ and $\hat{\rho}$. The bias is reported in terms of standard deviations. "$\hat{\theta}$ bc" and "$\hat{\rho}$ bc" give results for the bias-corrected estimators.[26] The empirical rejection probabilities ("rej prob") are for two-sided t-tests against the truth based on (4.1) and (4.2). Missing results ("-") are reported if simulation runs are aborted due to numerical instability.

specification, agents with an even index prefer links to agents with an even index over links to agents with an odd index, and vice versa for agents with an odd index. The homophily parameter is fixed at $\theta^0 = 1$. The reciprocity parameter is set to $\rho^0 = 0, 0.5$. I simulate networks with $N = 50, 70$ agents.[25] Network statistics for the simulated networks are given in Table E.1 in Supplemental Appendix E. Unless stated otherwise, the simulation results are based on 500 replications. All rejection probabilities are calculated based on a nominal level of $\alpha = 0.1$.

[25] Since the relevant sample size is the number of potential links $N(N-1)$, passing from $N = 50$ to $N = 70$ can be interpreted as doubling the sample size.

[26] The bias-corrected estimator of θ^0 is given by $\hat{\theta} - \widehat{W}_{1,N}^{-1}\hat{B}_N^{\theta}/N$, the bias-corrected estimator of ρ^0 is given

t-tests for θ^0 and ρ^0 Table 1 summarizes simulation results for the homophily and the reciprocity parameter.

The maximum likelihood estimator $\hat{\theta}$ exhibits a bias of up to more than than one standard deviation. The quality of the analytical bias correction decreases the sparser the design is. In the most sparse case, slightly less than half of the bias is eliminated. The empirical size of a t-test based on (4.1) that tests $\hat{\theta}$ against the truth concentrates around the nominal level. The observed size distortions slightly exceed those expected under the random Monte Carlo design.[27]

Without link reciprocity ($\rho^0 = 0$), the maximum likelihood estimator $\hat{\rho}$ of the reciprocity is approximately unbiased. In this case, analytical bias correction increases the bias slightly.[28] With link reciprocity ($\rho^0 = 0.5$), $\hat{\rho}$ exhibits a positive bias that is detected by the analytical bias correction. In all but the most sparse designs, the empirical size of a t-test based on (4.2) that tests $\hat{\rho}$ against the truth is close to the nominal level. In the designs with correlated within-dyad shocks and extreme sparsity ($C_N = \log N$), inference with respect to ρ^0 is unreliable. In the smaller sample, the maximum likelihood estimation becomes numerically unstable. In the larger sample, the analytical bias correction picks up only about two-thirds of the bias and the t-test testing against the truth is undersized.

Specification test The simulation results for the specification test suggest that the test statistic T_N^{stud} converges only slowly to its limit distribution. This can render the specification test based on analytical critical values oversized. As an alternative, I study bootstrap critical values based on a percentile bootstrap of the test statistic. The details of the bootstrap protocol are given in Supplemental Appendix D. The bootstrap procedure can

by $\hat{\rho} - 2\left(\hat{Q}_N' \widehat{W}_{1,N}^{-1} \hat{B}_N^{\theta} + \hat{B}_N^{\rho}\right) / (N \hat{v}_{1,N}^{\rho})$.

[27] The theoretical MC standard deviation for the rejection probabilities is ≈ 0.013.

[28] Analytical bias correction introduces estimation error that can cause a small bias at parameter values where the ML estimator is approximately unbiased.

be interpreted as a version of the double bootstrap in Kim and Sun (2016), where the inner loop is replaced by an analytical bias calculation.

As a benchmark, I consider a "naïve" implementation of a feasible test that ignores the effect of estimating the structural parameters. Lemma A.4 in Supplemental Appendix C suggests that the variance of the oracle excess transitivity T_N^{oracle} can be estimated by

$$\hat{v}_N^{T,o} = \frac{1}{N(N-1)} \sum_{\substack{i,j \in V \\ i<j}} \left(\hat{\boldsymbol{\beta}}_{ij}^N \hat{H}_{ij}(Y_{ij} - \hat{p}_{ij}) + \hat{\boldsymbol{\beta}}_{ji}^N \hat{H}_{ji}(Y_{ji} - \hat{p}_{ji}) \right)^2,$$

where $\hat{\boldsymbol{\beta}}_{ij}^N$ is the obvious plug-in estimator. The naïve test statistic is given by $\hat{T}_N^{\text{naive}} = T_N / \sqrt{\hat{v}_N^{T,o}/N}$. I also consider a bias-corrected version of $\hat{T}_N^{\text{naive}}$ that is defined by replacing $\hat{v}_N^T$ in (5.3) by $\hat{v}_N^{T,o}$.

The simulation results are summarized in Table 2. The estimated excess transitivity exhibits a negative bias of between -3.8 and -5.4 standard deviations. Even though the analytical correction removes a large portion of the bias, the magnitude of the remaining bias is still large, in particular in the sparser designs. As predicted by the asymptotic theory, increasing the sample size increases the quality of the bias correction. However, the rate at which the analytical correction improves is slow.

For the test using critical values calculated from the normal distribution, type-I error exceeds the nominal level by more than ten percentage points. The size distortion is caused by unaccounted bias and the fact that $\hat{v}_N^T$ underestimates the true variability of T_N. Again, increasing the sample size increases the quality of the asymptotic approximation, albeit not by a sufficient degree to appropriately control the size of the test. In contrast, the empirical size of the test with bootstrapped critical value is close to the nominal level, suggesting that the bootstrap distribution may replicate higher-order terms that are ignored by the analytical approximation.

				bias		rej prob				
N	ρ	C_N	density	T_N	T_N bc	analy	boot	naïve	naïve bc	ratio
50	0.0	$\log\log N$	0.19	-5.45	-0.27	0.28	0.10	0.00	0.00	0.15
50	0.0	$\log^{1/2} N$	0.12	-4.74	-0.47	0.29	0.11	0.00	0.00	0.22
50	0.0	$2\log^{1/2} N$	0.06	-4.09	-0.67	0.45	0.10	0.01	0.00	0.26
50	0.0	$\log N$	0.03	-4.03	-0.81	0.58	-	0.01	0.00	0.25
50	0.5	$\log\log N$	0.19	-4.82	-0.24	0.33	0.11	0.00	0.00	0.17
50	0.5	$\log^{1/2} N$	0.12	-4.44	-0.43	0.33	0.12	0.00	0.00	0.22
50	0.5	$2\log^{1/2} N$	0.06	-4.12	-0.65	0.45	0.11	0.01	0.00	0.26
50	0.5	$\log N$	0.03	-	-	-	-	-	-	-
70	0.0	$\log\log N$	0.18	-5.56	-0.24	0.23	0.10	0.00	0.00	0.16
70	0.0	$\log^{1/2} N$	0.11	-4.67	-0.36	0.23	0.08	0.00	0.00	0.22
70	0.0	$2\log^{1/2} N$	0.06	-4.64	-0.61	0.31	0.11	0.02	0.00	0.25
70	0.0	$\log N$	0.03	-4.33	-0.71	0.45	0.11	0.01	0.00	0.24
70	0.5	$\log\log N$	0.18	-5.16	-0.22	0.27	0.10	0.00	0.00	0.17
70	0.5	$\log^{1/2} N$	0.11	-4.56	-0.32	0.25	0.09	0.00	0.00	0.22
70	0.5	$2\log^{1/2} N$	0.06	-4.27	-0.49	0.34	0.09	0.01	0.00	0.25
70	0.5	$\log N$	0.03	-4.01	-0.63	0.51	0.07	0.01	0.00	0.28

Table 2: Simulation results for the specification test under the null hypothesis. Bias is reported in terms of standard deviations of estimated excess transitivity. "T_N bc" gives the empirical bias for the bias-corrected excess transitivity estimator $T_N + \left(\hat{B}_N^T + \hat{U}_N'(\widehat{W}_{1,N})^{-1}\hat{B}_N^{\theta}\right)/N$. For the empirical rejection probabilities ("rej prob"), "analy" and "boot" give results for the test based on (5.3) with analytical and bootstrap critical values, respectively, and "naïve" and "naïve bc" give results for the naïve test with and without bias correction. The bootstrap results are based on 250 simulations with 500 bootstrap replications each. The column "ratio" gives the ratio of the standard deviations of $\hat{T}_N^{\text{stud}}$ and T_N^{oracle}.

The naïve tests with and without bias correction are severely undersized. This is because $\hat{v}_N^{T,o}$ substantially overestimates the variance of T_N. The column "ratio" in Table 2 gives the standard deviation of T_N as a fraction of the standard deviation of T_N^{oracle}. The reported ratios indicate that estimating the structural parameters substantially decreases the variability of excess transitivity. A theoretical argument for why this happens is given in Section 5.2.

Specification test under a dynamic alternative To study the power properties of the specification test, I simulate an alternative model in which agents work endogenously towards transitive closure. The alternative model is a dynamic process with two stages. At stage $k = 1, 2$, the network is given by

$$Y_{ij}^{(k)} = \mathbf{1}\big(X_{ij} + \gamma_i^{S,0} + \gamma_j^{R,0} \geq U_{ij}^{(k)}\big).$$

The link covariate X_{ij} and the agent fixed effects are defined as above. In the network $\{Y_{ij}^{(1)}\}_{ij \in E(N)}$, the link ij is called an *unsupported link* if ij realizes, but none of the transitive triangles containing ij do. The link ij is called a *closing link* if, for some $k \in V_{-\{i,j\}}$, the links ik and kj realize, but ij does not. A closing link completes an open triangle and makes it transitive (see Figure 1). The first stage errors $(U_{ij}^{(1)})_{ij \in E(N)}$ are drawn as in the dyadic linking model. Let $(V_{ij})_{ij \in E(N)}$ denote a vector of $N(N-1)$ independent draws from the standard normal distribution. The second stage errors are given by

$$U_{ij}^{(2)} = \begin{cases} \min\{U_{ij}^{(1)}, V_{ij}\} & \text{if } ij \text{ is an unsupported link in } \{Y_{ij}^{(1)}\}_{ij \in E(N)} \\ \max\{U_{ij}^{(1)}, V_{ij}\} & \text{if } ij \text{ is a closing link in } \{Y_{ij}^{(1)}\}_{ij \in E(N)} \\ U_{ij}^{(1)} & \text{otherwise.} \end{cases}$$

N	ρ	C_N	density	rej prob					ratio
				analy	boot	oracle	naïve	naïve bc	
50	0.0	$\log\log N$	0.18	1.00	1.00	0.65	0.00	0.00	0.24
50	0.0	$\log^{1/2} N$	0.10	0.93	0.95	0.20	0.00	0.00	0.30
50	0.0	$2\log^{1/2} N$	0.05	-	-	-	-	-	-
50	0.0	$\log N$	0.02	-	-	-	-	-	-
50	0.5	$\log\log N$	0.18	0.99	1.00	0.67	0.00	0.00	0.25
50	0.5	$\log^{1/2} N$	0.10	0.90	0.93	0.21	0.00	0.00	0.29
50	0.5	$2\log^{1/2} N$	0.05	-	-	-	-	-	-
50	0.5	$\log N$	0.02	-	-	-	-	-	-
70	0.0	$\log\log N$	0.17	1.00	1.00	0.97	0.00	0.02	0.25
70	0.0	$\log^{1/2} N$	0.09	1.00	1.00	0.31	0.00	0.01	0.29
70	0.0	$2\log^{1/2} N$	0.04	0.86	0.93	0.08	0.00	0.00	0.31
70	0.0	$\log N$	0.02	-	-	-	-	-	-
70	0.5	$\log\log N$	0.17	1.00	1.00	0.95	0.00	0.02	0.27
70	0.5	$\log^{1/2} N$	0.09	1.00	1.00	0.30	0.00	0.01	0.30
70	0.5	$2\log^{1/2} N$	0.04	0.86	0.90	0.07	0.00	0.00	0.31
70	0.5	$\log N$	0.02	-	-	-	-	-	-

Table 3: Simulation results for the specification test under the dynamic alternative. "analy" and "boot" give results for the test based on (5.3) with analytical and bootstrap critical values, respectively, and "naïve" and "naïve bc" give results for the naïve test with and without bias correction. The bootstrap results are based on 250 simulations with 500 bootstrap replications each. The column "ratio" gives the ratio of the standard deviations of $\hat{T}_N^{\text{stud}}$ and T_N^{oracle}.

The second stage randomly removes some unsupported links and adds some closing links.[29] The final network is the observed network. Network statistics are given in Table E.2 in Supplemental Appendix E. Parameters are estimated by naïvely fitting the dyadic linking model.

The simulation results for the model specification test in the alternative model are summarized in Table 3.[30] The model specification test based on $\hat{T}_N^{\text{stud}}$ detects the alternative

[29] Manipulating the network by both adding and removing links, ensures that the second stage does not substantially change the network density.

[30] Under the dynamic alternative, the simulated parameters give rise to slightly sparser networks than under the simulated null model, rendering the maximum likelihood estimator non-existent for a wider range of designs.

reliably, with rejection probabilities ranging from .86 for the sparser designs to one for the denser designs. The difference between using analytical and bootstrap critical values is small, with bootstrap critical values yielding a slightly more powerful test.

As predicted by the theory, the test based on the infeasible statistic T_N^{oracle} is substantially *less* powerful than the test based on $\hat{T}_N^{\text{stud}}$. The naïve tests have barely any power. Only the naïve approach with bias correction leads to rejections, albeit with very small probability. The naïve test without bias correction is unable to detect any excess transitivity since the increase in the measured transitivity is not large enough to offset the negative bias in T_N.

7. Empirical application

I study excess transitivity in favor networks using the Indian village data from Banerjee et al. (2013) and Jackson, Rodriguez-Barraquer, and Tan (2012). This data set contains survey data from 75 Indian villages. In each village, about 30-40% of the adult population were handed out detailed questionnaires that elicit network relationships to other people in the same village as well as a wide range of socio-economic characteristics.

For each village, I define a directed network based on the survey questions "If you suddenly needed to borrow Rs. 50 for a day, whom would you ask?" and "If you needed to borrow kerosene or rice, to whom would you go?". To set up the network, I let every surveyed individual send directed links to each of the individuals nominated in one of the two questions, provided that the nominee was also included in the survey.[31]

Economists and Sociologists have long argued that transitive closure plays an important role in favor networks, where agents have to trust each other to repay favors in the future (see, e.g., Coleman 1988). Jackson, Rodriguez-Barraquer, and Tan (2012) study a game-

[31] The observed networks are defined to be the network of interest, sidestepping identification issues that arise when using a partial sample of the network (see Chandrasekhar and Lewis 2016).

theoretic model of favor exchange in which agents are punished by their network neighbors for reneging on reciprocating a favor. They show that networks with a large degree of transitive closure facilitate favor exchange, while satisfying certain optimality criteria.

The theoretical appeal of transitivity motivates the empirical study of excess transitivity in favor networks. Leung (2015) estimates a model in which agents endogenously form favor networks and finds that agents derive utility from being included in a transitive relationship. Chandrasekhar and Jackson (2016) use data on favor networks to test whether a dyadic linking model can explain the observed level of transitivity.[32] They find that the dyadic model generates an insufficient amount of transitivity. Using my model specification test, I replicate their finding.

My empirical finding complements the result in Chandrasekhar and Jackson (2016) by showing that it is robust against more sophisticated dyadic linking models. While the linking probabilities in Chandrasekhar and Jackson (2016) are a function of observables, my linking model can capture unobserved components of agent productivity and popularity using the fixed effect approach. As illustrated in Example 1, it is important to account for all dyadic sources of degree heterogeneity when testing transitivity. Moreover, my test does not rely on across-network variation and can be computed from one network observation. Therefore, it can be applied even if agents in different networks follow different linking rules.

For my transitivity test, I estimate dyadic linking models for each of the 75 village networks. The link-specific covariates for the homophily component are given in Table G.1 and test results are given in Table G.2 in Supplemental Appendix G. At level $\alpha = 0.1$, the test with analytical critical value detects excess transitivity in all networks, and the test with bootstrap critical value detects excess transitivity in all but one village.

Table G.2 also reports results for the naïve tests from Section 6. Even though these are not

[32] They refer to a model with dyadic linking as a block model and report a clustering coefficient that can be interpreted as measuring transitivity.

expected to work well, they are still instructive about the empirical relevance of accounting for the estimation of the dyadic model. The naïve approach with bias correction consistently detects excess transitivity, albeit with larger p-values than the preferred approach. This indicates that, for the data used in this application, my transitivity test is more powerful than the infeasible oracle test that uses the true dyadic linking probabilities. Without bias correction, the naïve test does not reject at level $\alpha = 0.1$ for eight of the 75 villages. As discussed in Section 6, failure to correct for a negative bias makes it harder to detect excess transitivity. Indeed, for all villages, the bias is estimated to be negative and large in absolute value. For the median village, the bias accounts for about half of the estimated excess transitivity after bias correction.

References

Abadie, Alberto and Guido Imbens (2016). "Matching on the Estimated Propensity Score". In: *Econometrica*.

Apicella, Coren et al. (2012). "Social networks and cooperation in hunter-gatherers". In: *Nature* 481.7382, pp. 497–501.

Banerjee, Abhijit et al. (2013). "The diffusion of microfinance". In: *Science* 341.6144.

Bhamidi, Shankar, Guy Bresler, and Allan Sly (2011). "Mixing time of exponential random graphs". In: *The Annals of Applied Probability* 21.6, pp. 2146–2170.

Candelaria, Luis (2017). "A semiparametric network formation model with multiple linear fixed effects". Working paper.

Chandrasekhar, Arun and Matthew Jackson (2016). "A network formation model based on subgraphs". Working paper.

Chandrasekhar, Arun and Randall Lewis (2016). "Econometrics of sampled networks". Working paper.

Charbonneau, Karyne (2017). "Multiple fixed effects in binary response panel data models". In: *The Econometrics Journal*.

Coleman, James S (1988). "Social Capital in the Creation of Human Capital". In: *American Journal of Sociology*, S95–S120.

Davis, James (1970). "Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices". In: *American Sociological Review*, pp. 843–851.

Davis, James and Samuel Leinhardt (1972). "The structure of positive interpersonal relations in small groups." In: *Sociological theories in progress*. Ed. by Joseph Berger. Vol. 2.

Dhaene, Geert and Koen Jochmans (2015). "Split-panel jackknife estimation of fixed-effect models". In: *The Review of Economic Studies* 82.3, pp. 991–1030.

Enea, Marco (2013). *speedglm: Fitting Linear and Generalized Linear Models to large data sets*. R package version 0.2. URL: `https://CRAN.R-project.org/package=speedglm`.

Fafchamps, Marcel and Flore Gubert (2007). "The formation of risk sharing networks". In: *Journal of Development Economics* 83.2, pp. 326–350.

Fafchamps, Marcel and Susan Lund (2003). "Risk-sharing networks in rural Philippines". In: *Journal of Development Economics* 71.2, pp. 261–287.

Faust, Katherine (2007). "Very local structure in social networks". In: *Sociological Methodology* 37.1, pp. 209–256.

Fernández-Val, Iván and Martin Weidner (2016). "Individual and time effects in nonlinear panel models with large N, T". In: *Journal of Econometrics*.

Gao, Wayne Yuan (2017). "Nonparametric Identification in Index Models of Link Formation". In: *arXiv preprint arXiv:1710.11230*.

Graham, Bryan S (2015). "Methods of identification in social networks". In: *Annu. Rev. Econ.* 7.1, pp. 465–485.

Graham, Bryan S (2017). “An econometric model of network formation with degree heterogeneity”. In: *Econometrica* 85.4, pp. 1033–1063.

Hahn, Jinyong (1998). “On the role of the propensity score in efficient semiparametric estimation of average treatment effects”. In: *Econometrica*, pp. 315–331.

Hahn, Jinyong and Guido Kuersteiner (2011). “Bias reduction for dynamic nonlinear panel models with fixed effects”. In: *Econometric Theory* 27.06, pp. 1152–1191.

Hahn, Jinyong and Whitney Newey (2004). “Jackknife and analytical bias reduction for nonlinear panel models”. In: *Econometrica* 72.4, pp. 1295–1319.

Hoff, Peter (2005). “Bilinear mixed-effects models for dyadic data”. In: *Journal of the American Statistical Association* 100.469, pp. 286–295.

— (2015). “Dyadic data analysis with amen”. Working paper.

Holland, Paul and Samuel Leinhardt (1978). “An omnibus test for social structure using triads”. In: *Sociological Methods & Research* 7.2, pp. 227–256.

— (1981). “An exponential family of probability distributions for directed graphs”. In: *Journal of the American Statistical Association* 76.373, pp. 33–50.

Horn, Roger A and Charles R Johnson (2012). *Matrix analysis*. Cambridge University Press.

Jackson, Matthew (2008). *Social and economic networks*. Princeton University Press.

Jackson, Matthew, Tomas Rodriguez-Barraquer, and Xu Tan (2012). “Social capital and social quilts: Network patterns of favor exchange”. In: *The American Economic Review* 102.5, pp. 1857–1897.

Jochmans, Koen (2017). “Semiparametric analysis of network formation”. In: *Journal of Business & Economic Statistics*, pp. 1–9.

Karlberg, Martin (1997). “Testing transitivity in graphs”. In: *Social Networks* 19.4, pp. 325–343.

Karlberg, Martin (1999). “Testing transitivity in digraphs”. In: *Sociological Methodology* 29.1, pp. 225–251.

Kim, Min Seong and Yixiao Sun (2016). “Bootstrap and k-step bootstrap bias correction for fixed effects estimators in nonlinear panel models”. In: *Econometric Theory*.

Leung, Michael P (2015). “Two-step estimation of network-formation models with incomplete information”. In: *Journal of Econometrics* 188.1, pp. 182–195.

Mayer, Adalbert and Steven Puller (2008). “The old boy (and girl) network: Social network formation on university campuses”. In: *Journal of Public Economics* 92.1, pp. 329–347.

McPherson, Miller, Lynn Smith-Lovin, and James Cook (2001). “Birds of a feather: Homophily in social networks”. In: *Annual Review of Sociology*, pp. 415–444.

Mele, Angelo (2016). “A structural model of dense network formation”. In: *Econometrica*. Forthcoming.

Menzel, Konrad (2017). “Strategic network formation with many agents”. Working paper.

Neyman, Jerzy and Elizabeth L Scott (1948). “Consistent estimates based on partially consistent observations”. In: *Econometrica*, pp. 1–32.

Ridder, Geert and Shuyang Sheng (2017). “Estimation of large network formation games”. Working paper.

Sheng, Shuyang (2016). “A Structural Econometric Analysis of Network Formation Games”. Working paper.

Shi, Zhentao and Xi Chen (2016). “A structural pairwise network model with individual heterogeneity”. Working paper.

Snijders, Tom et al. (2006). “New specifications for exponential random graph models”. In: *Sociological Methodology* 36.1, pp. 99–153.

Toth, Peter (2017). “Semiparametric estimation in network formation models with homophily and degree heterogeneity”. Working paper.

van der Vaart, Aad (2000). *Asymptotic Statistics*. Cambridge University Press.

Wasserman, Stanley and Philippa Pattison (1996). “Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p”. In: *Psychometrika* 61.3, pp. 401–425.

Watts, Duncan and Steven Strogatz (1998). “Collective dynamics of “small-world” networks”. In: *Nature* 393.6684, pp. 440–442.

Yan, Ting, Binyan Jiang, et al. (2018). “Statistical inference in a directed network model with covariates”. In: *Journal of the American Statistical Association*.

Yan, Ting, Chenlei Leng, Ji Zhu, et al. (2016). “Asymptotics in directed exponential random graph models with an increasing bi-degree sequence”. In: *The Annals of Statistics* 44.1, pp. 31–57.

Yan, Ying et al. (2016). “Statistical inference in a directed network model with covariates”. Working paper.

Supplemental Appendix
for
An empirical model of dyadic link formation in a network with unobserved heterogeneity

Andreas Dzemski

April 19, 2018

A. Applying results from Fernández-Val and Weidner (2016)

Fernández-Val and Weidner (2016) (henceforth FVW) study a panel model with time and individual fixed effects. Their results can be leveraged for the analysis of my network model. In particular, FVW derive a stochastic expansion for a broad class of maximum likelihood models with an incidental parameter (Theorem B.1 in FVW). It can be shown that this class contains the dyadic linking model. Below, I adapt some key results in FVW to the dyadic linking model.

Let 1_N denote an N-vector of ones and let $v_N = (1_N', -1_N')'$. For $b > 0$, the ML program (2.1) can be rewritten as

$$(\hat{\theta}, \hat{\boldsymbol{\gamma}}) = \arg\max_{\theta,\boldsymbol{\gamma}} \mathcal{L}(\theta, \boldsymbol{\gamma}),$$

where

$$\begin{aligned}\mathcal{L}(\theta, \boldsymbol{\gamma}) = \frac{1}{N} \sum_{i\in V} \sum_{j\in V_{-i}} \Big\{ & Y_{ij} \log\left(p_{ij}(\theta, \boldsymbol{\gamma})\right) \\ & + (1 - Y_{ij}) \log\left(1 - p_{ij}(\theta, \boldsymbol{\gamma})\right) \Big\} - \frac{b}{2N} (v_N' \boldsymbol{\gamma})^2.\end{aligned} \tag{A.1}$$

The penality imposes the normalization constraint in the ML program (2.1). Let

$$\mathcal{S}(\theta, \boldsymbol{\gamma}) = \partial_{\boldsymbol{\gamma}} \mathcal{L}(\theta, \boldsymbol{\gamma}) \qquad \mathcal{H}(\theta, \boldsymbol{\gamma}) = -\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'} \mathcal{L}(\theta, \boldsymbol{\gamma}).$$

We adapt the convention that omitting function argument indicates that the function is eliminated at the true paramters, e.g., $\mathcal{S} = \mathcal{S}(\theta^0, \boldsymbol{\gamma}^0)$ and $\mathcal{H} = \mathcal{H}(\theta^0, \boldsymbol{\gamma}^0)$. For a random variable W we set $\bar{W} = \bar{\mathbb{E}}\, W$ and $\tilde{W} = W - \bar{W}$. In particular, let

$$\bar{\mathcal{H}} = \begin{bmatrix} \bar{\mathcal{H}}^*_{SS} & \bar{\mathcal{H}}^*_{SR} \\ (\bar{\mathcal{H}}^{SR})' & \bar{\mathcal{H}}^*_{RR} \end{bmatrix} + b\left(v_N v_N'\right)/N,$$

where

$$\bar{\mathcal{H}}^*_{SS} = \operatorname{diag}\left(\left(\frac{1}{N} \sum_{j\in V_{-i}} \omega_{ij}\right)_{i\in V}\right) \quad \text{and} \quad \bar{\mathcal{H}}^*_{RR} = \operatorname{diag}\left(\left(\frac{1}{N} \sum_{i\in V_{-j}} \omega_{ij}\right)_{j\in V}\right), \tag{A.2}$$

and $\bar{\mathcal{H}}^*_{SR}$ is the $N \times N$ matrix with diagonal entries equal to zero and off-diagonal entries ω_{ij}/N for $i, j \in V$ and $i \neq j$.

Lemma A.1 (Network version of Lemma D.1 in FVW)**.** *Under Assumption 1*

$$\left\| \bar{\mathcal{H}}^{-1} - \left(\operatorname{diag}(\bar{\mathcal{H}}^*_{SS}, \bar{\mathcal{H}}^*_{RR}) \right)^{-1} \right\|_{\max} = O_p\left(N^{-1}\right).$$

Proof. For the purposes of this proof define $\omega_{ii} = 0$. Then we can write

$$\bar{\mathcal{H}}^*_{SS} = \operatorname{diag}\left(\left(\frac{1}{N} \sum_{j=1}^{N} \omega_{ij}\right)_{i\in V}\right), \quad \text{and} \quad \bar{\mathcal{H}}^*_{RR} = \operatorname{diag}\left(\left(\frac{1}{N} \sum_{i=1}^{N} \omega_{ij}\right)_{j\in V}\right)$$

and $\bar{\mathcal{H}}^*_{SR}$ for the matrix with entries $\bar{\mathcal{H}}^*_{SR,ij} = \omega_{ij}/N$ for $i, j \in V$. Assumption 1(v) implies that there are constants $b_{\min}$ and $b_{\max}$ that are independent of N and satisfy $0 < b_{\min} < \omega_{ij} < b_{\max}$ for $i \neq j$. If ω_{ii} could be assumed to satisfy the same inequality then the Hessian would exhibit the structure that is exploited by the proof of Lemma D.1 in FVW. Define $\omega^{\dagger}_{ij} = \max\{\omega_{ij}, b_{\min}\}$. Define $\bar{\mathcal{H}}^{\dagger}_{SS}$, $\bar{\mathcal{H}}^{\dagger}_{RR}$ and $\bar{\mathcal{H}}^{\dagger}_{SR}$ similar to $\bar{\mathcal{H}}_{SS}$, $\bar{\mathcal{H}}_{RR}$ and $\bar{\mathcal{H}}_{SR}$ with ω_{ij} replaced by $\omega^{\dagger}_{ij}$. Let $D = \operatorname{diag}(\bar{\mathcal{H}}^{*}_{SS}, \bar{\mathcal{H}}^{*}_{RR})$ and $D^{\dagger} = \operatorname{diag}(\bar{\mathcal{H}}^{\dagger}_{SS}, \bar{\mathcal{H}}^{\dagger}_{RR})$. Lemma D.1 in FVW implies that $\left\|(\bar{\mathcal{H}}^{\dagger})^{-1} - (D^{\dagger})^{-1}\right\|_{\max} = O_p\left(N^{-1}\right)$. By the inequality on p 351 in Horn and Johnson 2012

$$\begin{aligned}\left\|(\bar{\mathcal{H}}^{\dagger})^{-1}\right\|_{\max} \leq& \left\|(D^{\dagger})^{-1}\right\|_{\max} \left\|\left(\mathbb{I}_{2N} - (D^{\dagger} - \bar{\mathcal{H}}^{\dagger})\right)^{-1}\right\|_{\max} \\ \leq& b_{\max}\left(1 - \left\|D^{\dagger} - \bar{\mathcal{H}}^{\dagger}\right\|_{\max}\right)^{-1} \leq 2b_{\max}.\end{aligned}$$

By construction $\left\|\bar{\mathcal{H}}^{\dagger} - \bar{\mathcal{H}}\right\|_{\max} \leq b_{\min}/N$ and $\left\|D^{\dagger} - D\right\|_{\max} \leq b_{\min}/N$. Therefore

$$\begin{aligned}\left\|\bar{\mathcal{H}}^{-1} - (\bar{\mathcal{H}}^{\dagger})^{-1}\right\|_{\max} \leq& \left\|(\bar{\mathcal{H}}^{\dagger})^{-1}(\bar{\mathcal{H}}^{\dagger} - \bar{\mathcal{H}})(\bar{\mathcal{H}}^{\dagger})^{-1}\left(\mathbb{I}_{2N} - (\bar{\mathcal{H}}^{\dagger} - \bar{\mathcal{H}})\right)^{-1}\right\|_{\max} \\ \leq& \left\|(\bar{\mathcal{H}}^{\dagger})^{-1}\right\|^{2}_{\max} \left\|\bar{\mathcal{H}}^{\dagger} - \bar{\mathcal{H}}\right\|_{\max} \left(1 - \left\|\bar{\mathcal{H}}^{\dagger} - \bar{\mathcal{H}}\right\|_{\max}\right)^{-1} \\ \leq& 4b^{2}_{\max}\frac{b_{\min}}{N}\left(1 - \frac{b_{\min}}{N}\right)^{-1} = O_p\left(N^{-1}\right).\end{aligned}$$

Then, by the triangle inequality

$$\begin{aligned}\left\|\bar{\mathcal{H}}^{-1} - D^{-1}\right\|_{\max} \leq& \left\|\bar{\mathcal{H}}^{-1} - (\bar{\mathcal{H}}^{\dagger})^{-1}\right\|_{\max} \\ &+ \left\|(\bar{\mathcal{H}}^{\dagger})^{-1} - (D^{\dagger})^{-1}\right\|_{\max} + \left\|(D^{\dagger})^{-1} - D^{-1}\right\|_{\max} = O_p\left(N^{-1}\right).\end{aligned}$$

□

Lemma A.2 (Network version of Theorem B.1 in FVW)**.** *Let*

$$\hat{\boldsymbol{\gamma}}(\theta) = \arg\max\nolimits_{\boldsymbol{\gamma}} \mathcal{L}(\theta, \boldsymbol{\gamma}).$$

denote the concentrated likelihood and suppose that Assumption 1 holds. Then

$$\hat{\boldsymbol{\gamma}}(\theta) - \boldsymbol{\gamma}^{0} = \mathcal{H}^{-1}\mathcal{S} + \mathcal{H}^{-1}[\partial_{\boldsymbol{\gamma}\theta'}\mathcal{L}](\theta - \theta^{0}) + \frac{1}{2}\mathcal{H}^{-1}\sum_{g=1}^{2N}[\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\gamma}_g}\mathcal{L}]\mathcal{H}^{-1}\mathcal{S}[\mathcal{H}^{-1}\mathcal{S}]_g + R^{\boldsymbol{\gamma}}(\theta)$$

and

$$\partial_{\theta}\mathcal{L}(\theta, \boldsymbol{\gamma}) = U - \bar{W}N(\theta - \theta^{0}) + R(\theta),$$

where $U = U^{(0)} + U^{(1)}$, *and*

$$\begin{aligned}
\bar{W} =& -\frac{1}{N}\left(\partial_{\theta,\theta'}\bar{\mathcal{L}} + [\partial_{\theta\gamma}\bar{\mathcal{L}}]\bar{H}^{-1}[\partial_{\gamma\theta}\bar{\mathcal{L}}]\right), \\
U^{(0)} =& \partial_{\theta}\mathcal{L} + [\partial_{\theta\gamma'}\bar{\mathcal{L}}]\bar{\mathcal{H}}^{-1}\mathcal{S}, \\
U^{(1)} =& [\partial_{\theta\gamma}\tilde{\mathcal{L}}]\bar{\mathcal{H}}^{-1}\mathcal{S} - [\partial_{\theta\gamma'}\mathcal{L}]\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} \\
&+ \frac{1}{2}\sum_{g}^{2N}\left(\partial_{\theta\gamma'\gamma_g}\bar{\mathcal{L}} + [\partial_{\theta\gamma'}\bar{\mathcal{L}}]\bar{H}^{-1}[\partial_{\gamma\gamma'\gamma_g}\bar{\mathcal{L}}]\right)[\bar{H}^{-1}\mathcal{S}]_g\bar{\mathcal{H}}^{-1}\mathcal{S}.
\end{aligned}$$

The remainder terms R^{γ} *and* R *satisfy*

$$\sup_{\theta\in\Theta_N}\frac{N^{\frac{7}{8}}\left\|R^{\gamma}(\theta)\right\|_8}{1+N\left\|\theta-\theta^0\right\|} = o_p(1), \qquad \sup_{\theta\in\Theta_N}\frac{R(\theta)}{1+N\left\|\theta-\theta^0\right\|} = o_p(1),$$

where $\Theta_N \subset \mathbb{R}^{\dim(\theta)}$ *satisfies* $\mathrm{Prob}(\hat{\theta}\in\Theta_N) \to 1$.

Proof. Let C denote a generic constant. To apply Theorem B.1 in FVW we have to verify their Assumption B.1. It is without loss of generality to operate conditional on the event A_N. We set $T = N$, $q = 8$, $r_{\beta} = \log(N)N^{-1/4}$ and $r_{\phi} = N^{-1/8}$. B.1(i) holds trivially. B.1(ii) holds since p_{ij} is bounded away from zero and one. B.1(iv) holds by Assumption (iv) and Lemma A.1 (see the proof of Theorem 4.1 in FVW). Condition B.1(v) can be checked by following the arguments for the panel case. The arguments are very similar to the ones employed in the respective proofs of Lemma C.7 and Lemma C.8. For condition B.1(vi) the arguments for the panel model carry over almost ad verbatim. For example, to prove that $\|\tilde{\mathcal{H}}\| = o_p(N^{1/4})$ let $\bar{\mathcal{H}}_{SS}$ and $\bar{\mathcal{H}}_{RR}$ and $\bar{\mathcal{H}}_{SR}$ as defined in the proof of Lemma A.1. By the triangle inequality

$$\left\|\tilde{\mathcal{H}}\right\| \leq \left\|-\partial_{\gamma_S\gamma_S}\mathcal{L} - \bar{\mathcal{H}}^*_{SS}\right\| + \left\|-\partial_{\gamma_R\gamma_R}\mathcal{L} - \bar{\mathcal{H}}^*_{RR}\right\| + 2\left\|-\partial_{\gamma_S\gamma_R}\mathcal{L} - \bar{\mathcal{H}}^*_{SR}\right\|.$$

Let $\xi_{ij} = (Y_{ij} - p_{ij})\partial_{\gamma_i^S}\left[\frac{\phi(Y^*_{ij})}{p_{ij}(1-p_{ij})}\right]$ for $i \neq j$ and $\xi_{ii} = 0$ for $i \in V$. Then $-\partial_{\gamma_S\gamma_S}\mathcal{L} - \bar{\mathcal{H}}^*_{SS} = -\operatorname{diag}\left((\frac{1}{N}\sum_{j\in V_{-i}}\xi_{ij})_{i\in V}\right)$. The $\|\cdot\|$ matrix norm is given by the largest eigenvalue of a matrix. Therefore

$$\begin{aligned}
\bar{\mathbb{E}}\left\|-\partial_{\gamma_S\gamma_S}\mathcal{L} - \bar{\mathcal{H}}^*_{SS}\right\|^8 =& \bar{\mathbb{E}}\left[\max_{i\in V}\left(\frac{1}{N}\sum_{j\in V_{-i}}\xi_{ij}\right)^8\right] \\
\leq& \sum_{i\in V}\bar{\mathbb{E}}\left[\left(\frac{1}{N}\sum_{j\in V_{-i}}\xi_{ij}\right)^8\right] = O_p\left(N^{-3}\right).
\end{aligned}$$

A similar argument applies to $\left\|-\partial_{\gamma_R\gamma_R}\mathcal{L} - \bar{\mathcal{H}}^*_{RR}\right\|$. This shows that

$$\left\|-\partial_{\gamma_S\gamma_S}\mathcal{L} - \bar{\mathcal{H}}^*_{SS}\right\| = O_p\left(N^{-\frac{3}{8}}\right),$$
$$\left\|-\partial_{\gamma_R\gamma_R}\mathcal{L} - \bar{\mathcal{H}}^*_{RR}\right\| = O_p\left(N^{-\frac{3}{8}}\right).$$

Now apply Lemma S.6 in FVW with $T = N$ and $e_{it} = \xi_{ij}$. For $i, j \in V$ we have $\bar{\mathbb{E}}[\xi^2_{ij}] = 0$ if $i = j$ and $\bar{\mathbb{E}}[\xi^2_{ij}] \leq \left(\partial_{\gamma_i^S}\left[\frac{\phi(Y^*_{ij})}{p_{ij}(1-p_{ij})}\right]\right)^2$ if $i \neq j$. Assumption 1(v) ensures that $\bar{\mathbb{E}}[\xi^2_{ij}] < C$ so that $\bar{\sigma}^2_i = \frac{1}{N}\sum_{j\in V}\bar{\mathbb{E}}(\xi^2_{ij}) < C$. The matrix Ω with elements is a $N \times N$ matrix with elements given by $\Omega_{j\ell} = \frac{1}{N}\sum_{i\in V}\bar{\mathbb{E}}(e_{ij}e_{i\ell})$. It is easy to see that Ω is a diagonal matrix whose diagonal elements are bounded by C. Therefore, $\frac{1}{N}\operatorname{Tr}(\Omega) = O_p(1)$. Let $\eta_{ij} = \frac{1}{\sqrt{N}}\sum_{\ell\in V}\left(\xi_{i\ell}\xi_{j\ell} - \bar{\mathbb{E}}[\xi_{i\ell}\xi_{j\ell}]\right)$. We have

$$\bar{\mathbb{E}}(\eta_{ii})^4 \leq N^{-2}(N + N^2)(2C)^4$$

and therefore $\frac{1}{N}\sum_{i=1}^N \bar{\mathbb{E}}(\eta_{ii})^4 = O_p(1)$. For $i \neq j$ we have $\eta_{ij} = \frac{1}{\sqrt{N}}\sum_{\ell\in V}\xi_{i\ell}\xi_{j\ell}$. Taking the 4th power of η_{ij} gives a long sum where each term has the form

$$\bar{\mathbb{E}}[\xi_{ik_1}\xi_{ik_2}\xi_{ik_3}\xi_{ik_4}\xi_{jk_1}\xi_{jk_2}\xi_{jk_3}\xi_{jk_4}].$$

Clearly, the term is equal to 0 if there is a $k_\ell \in \{i, j\}$. Therefore we may assume that the $k_\ell \in V \setminus \{i, j\}$. The term is also equal to zero if a k_ℓ gets picked only once. Therefore we can bound the term by (relabelling the k's if necessary)

$$\bar{\mathbb{E}}[\xi_{ik_{\ell_1}}\xi_{ik_{\ell_2}}\xi_{jk_{\ell_1}}\xi_{jk_{\ell_2}}] \leq \left(\bar{\mathbb{E}}[(\xi_{ik_{\ell_1}})^8]\right)^{\frac{1}{4}}\left(\bar{\mathbb{E}}[(\xi_{ik_{\ell_2}})^8]\right)^{\frac{1}{4}}\left(\bar{\mathbb{E}}[(\xi_{jk_{\ell_1}})^8]\right)^{\frac{1}{4}}\left(\bar{\mathbb{E}}[(\xi_{jk_{\ell_2}})^8]\right)^{\frac{1}{4}} < C.$$

There are $(N-2)^2$ ways of picking the k_ℓ's so that $\frac{1}{N^2}\sum_{i,j\in V}\bar{\mathbb{E}}(\eta^4_{ij}) < C$. Thus, all the conditions of Lemma S.6 in FVW are satisfied and the matrix $\xi = (\xi_{ij})_{i\neq j}$ has $\|\xi\| = O_p(N^{5/8})$. Since $N\xi = -\partial_{\gamma_R\gamma_R}\mathcal{L} - \bar{\mathcal{H}}^*_{RR}$ this implies

$$\left\|-\partial_{\gamma_R\gamma_R}\mathcal{L} - \bar{\mathcal{H}}^*_{RR}\right\| = O_p\left(N^{-\frac{3}{8}}\right).$$

The other arguments in FVW can be adapted similarly. To verify assumption B.1(iii) in FVW, we apply Theorem B.3 in FVW. A condition of the theorem is that $\bar{W}$ has a positive definite limit. Inspection of the proof in FVW shows that this condition can be replaced by assuming that the eigenvalues of $\bar{W}$ are positive and bounded away from zero. It can be shown that $\bar{W} - \bar{W}_{N,1} = o_p(1)$. Therefore $\bar{W}$ satisifies this eigenvalue condition by Assumption 1(ii). It remains to check that $U = O_p(1)$. It can be shown that $U = B^\theta_N + \frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} H_{ij}\tilde{X}_{ij}(Y_{ij} - p_{ij}) + o_p(1)$. We first show $B^\theta_N = O_p(1)$. We only show $B^{\theta,S}_N = O_p(1)$. The proof for $B^{\theta,R}_N$ is similar. Since $\sum_{j\in V_{-i}}\omega_{ij}/N$ is bounded away from zero uniformly over $i \in V$ by Assumption 1(v), it suffices to show that $\|\sum_{j\in V_{-i}} \omega_{ij}\tilde{X}_{ij}\tilde{X}'_{ij}/N\|$ is bounded uniformly over $i \in V$. It suffices to bound each of the elements of the matrix. For $i \in V$ and $k_1, k_2 = 1, \ldots, \dim(\theta)$, applying the the Cauchy-Schwarz inequality twice gives the bound

$$\Big|\frac{1}{N}\sum_{j\in V_{-i}} \omega_{ij}\tilde{X}_{ij,k_1}\tilde{X}_{ij,k_2}\Big| \leq \left(\frac{1}{N}\sum_{j\in V_{-i}} \omega_{ij}^2\right)^{1/2} \left(\frac{1}{N}\sum_{j\in V_{-i}} \tilde{X}_{ij,k_1}^4 + \frac{1}{N}\sum_{j\in V_{-i}} \tilde{X}_{ij,k_2}^4\right)^{1/2}.$$

The first term in the product is bounded by Assumption 1(v), the second term is bounded by Assumption 1(iii). This verifies that B_N^θ is. Boundedness of

$$\bar{\mathbb{E}}\Big[\frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} H_{ij}\tilde{X}_{ij}(Y_{ij}-p_{ij})\Big]^2 = \bar{W}_{2,N}$$

can be shown using similar arguments. Therefore $U = O_p(1)$ and Theorem B.3 in FVW can be applied. It follows that Assumption B.1(iii) in FVW is met and

$$\|\hat{\theta}-\theta^0\| = O_p\left(N^{-\frac{1}{2}}\right).$$

If we define $\Theta_N = \{\theta \in \Theta : \|\theta-\theta^0\| < N^{-\frac{1}{4}}\}$ then $\text{Prob}(\hat{\theta} \in \Theta_N) \to 1$ and also $\Theta_N \subset \{\theta \in \Theta : \|\theta-\theta^0\| < r_\beta\}$. The conclusion now follows from Theorem B.1 in FVW. □

For convenience I restate some bounds originally derived in FVW.

Lemma A.3. *Suppose Assumption 1 holds. Then*

$$\begin{aligned}\big\|\mathcal{H}^{-1}-\bar{\mathcal{H}}^{-1}\big\| &= o_p\big(N^{-1/4}\big),\\ \big\|\mathcal{H}^{-1}-\big(\bar{\mathcal{H}}^{-1}-\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\big)\big\| &= o_p\big(N^{-1/2}\big),\\ \Big\|\sum_{g,h=1}^{\dim(\gamma)}\partial_{\gamma\gamma_g\gamma_h}\tilde{\mathcal{L}}[\bar{\mathcal{H}}^{-1}\mathcal{S}]_g[\bar{\mathcal{H}}^{-1}\mathcal{S}]_h\Big\| &= o_p\big(N^{-1/2}\big).\end{aligned}$$

Moreover, $\big\|\mathcal{H}^{-1}\big\| = O_p(1)$, $\big\|\tilde{\mathcal{H}}\big\| = o_p\big(N^{-1/4}\big)$, $\big\|\partial_{\gamma\theta'}\mathcal{L}\big\| = O_p\big(N^{1/2}\big)$, $\big\|\partial_{\gamma\theta'}\tilde{\mathcal{L}}\big\| = O_p(1)$.

Proof. Fernández-Val and Weidner 2016. See also proof of Lemma A.2. □

Lemma A.4 (Stochastic expansion of $\hat{\theta}$)**.** *Under Assumption 1*

$$N\,\bar{W}_{1,N}(\hat{\theta}-\theta^0) = B_N^\theta + \frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} H_{ij}\tilde{X}_{ij}(Y_{ij}-p_{ij}) + o_p(1)$$

and

$$\bar{W}_{2,N}^{-1/2}\left(N\,\bar{W}_{1,N}(\hat{\theta}-\theta^0)-B_N^{\theta}\right)=\mathcal{N}(0,1)+o_p(1)$$

where B_N^{θ}, $\bar{W}_{1,N}$ *and* $\bar{W}_{2,N}$ *are defined in Theorem 1.*

Proof. This follows from Lemma A.2 employing similar arguments as in the proof of Theorem 4.1 in FVW. □

B. Proofs of main results

Proof of Theorem 1. Apply Lemma A.2 and employ similar arguments as in the proof of Theorem 4.1 in FVW to derive the linear asymptotic expansion. For the distributional result write

$$\begin{aligned}
&\frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}}H_{ij}\tilde{X}_{ij}(Y_{ij}-p_{ij})\\
=&\frac{1}{\sqrt{2}}\frac{1}{\sqrt{N(N-1)/2}}\sum_{\substack{i,j\in V\\ i<j}}\left\{H_{ij}\tilde{X}_{ij}(Y_{ij}-p_{ij})+H_{ji}\tilde{X}_{ji}(Y_{ji}-p_{ji})\right\}+o_p(1)
\end{aligned}$$

and apply a multi-variate CLT to the normalized sum on the right-hand side. The (conditional) variance of the normalized sum is given by

$$\begin{aligned}
2\bar{W}_{2,N}=&\frac{1}{N(N-1)/2}\sum_{\substack{i,j\in V\\ i<j}}\left\{H_{ij}(\partial_\pi p_{ij})\tilde{X}_{ij}\tilde{X}_{ij}'+H_{ji}(\partial_\pi p_{ji})\tilde{X}_{ji}\tilde{X}_{ji}'\right.\\
&\left.\qquad+2H_{ij}H_{ji}\tilde{X}_{ij}\tilde{X}_{ji}'\tilde{\rho}_{ij}\sqrt{p_{1,ij}p_{1,ji}}\right\}\\
=&2\bar{W}_{1,N}+\frac{4}{N(N-1)}\sum_{\substack{i,j\in V\\ i<j}}H_{ij}H_{ji}\tilde{X}_{ij}\tilde{X}_{ji}'\tilde{\rho}_{ij}\sqrt{p_{1,ij}p_{1,ji}}+o_p(1).
\end{aligned}$$

□

Proof of Theorem 2. This theorem follows from the stochastic expansion in Lemma C.1.

By Lemma C.3,

$$
\begin{aligned}
&\partial_\rho \mathcal{M} + \left(\partial_{\rho\theta'}\mathcal{M} + \partial_{\rho\gamma'}\bar{\mathcal{M}}\bar{\mathcal{H}}^{-1}\left[\partial_{\gamma\theta'}\bar{\mathcal{L}}\right]\right)(\hat{\theta}-\theta^0) + \left(\partial_{\rho\gamma'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\mathcal{S} \\
=&\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}} J_{ij}(Z_{ij}-r_{ij}) + T_N' N(\hat{\theta}-\theta^0) - \frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} \Omega_{ij}H_{ij}(Y_{ij}-p_{ij}) \\
=&\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}} J_{ij}(Z_{ij}-r_{ij}) + \frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} (\tilde{t}_{ij}-\Omega_{ij})H_{ij}(Y_{ij}-p_{ij}) \\
=&\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}} U_{ij} + T_N' W_{1,N}^{-1} B_N^{\theta},
\end{aligned}
$$

where

$$
U_{ij} = J_{ij}(Z_{ij}-r_{ij}) + (\tilde{t}_{ij}-\Omega_{ij})H_{ij}(Y_{ij}-p_{ij}) + (\tilde{t}_{ji}-\Omega_{ji})H_{ji}(Y_{ji}-p_{ji}).
$$

The sum on the right-hand side is over $\binom{N}{2} = \frac{N(N-1)}{2}$ independent observations. To verify that the (conditional) variance of the normalized sum is given by $v_{2,N}$ note that

$$
\begin{aligned}
\bar{\mathbb{E}}[(Z_{ij}-r_{ij})(Y_{ij}-p_{ij})] &= r_{ij}(1-p_{ij}) = p_{1,ij}\frac{r_{ij}}{p_{ij}}, \\
\bar{\mathbb{E}}[(Z_{ij}-r_{ij})(Y_{ji}-p_{ji})] &= r_{ij}(1-p_{ji}) = p_{1,ji}\frac{r_{ij}}{p_{ji}}, \\
\bar{\mathbb{E}}[(Y_{ij}-p_{ij})(Y_{ji}-p_{ji})] &= r_{ij} - p_{ij}p_{ji} = \tilde{\rho}_{ij}\sqrt{p_{1,ij}p_{1,ji}},
\end{aligned}
$$

and $\bar{\mathbb{E}}[(Y_{ij}-p_{ij})^2] = p_{1,ij}$, $\bar{\mathbb{E}}[(Y_{ji}-p_{ji})^2] = p_{1,ji}$, and $\bar{\mathbb{E}}[(Z_{ij}-r_{ij})^2] = r_{1,ij}$. Now straightforward calculations yield

$$
\begin{aligned}
&\bar{\text{var}}\left(\frac{1}{N(N-1)/2}\sum_{\substack{i,j\in V\\ i<j}} U_{ij}\right) \\
=&\frac{1}{N(N-1)/2}\sum_{\substack{i,j\in V\\ i<j}} \Big\{ J_{ij}^2 r_{1,ij} + [(\tilde{t}_{ij}-\Omega_{ij})H_{ij}]^2 p_{1,ij} + [(\tilde{t}_{ji}-\Omega_{ji})H_{ji}]^2 p_{1,ji} \\
&\qquad + 2J_{ij}(\tilde{t}_{ij}-\Omega_{ij})H_{ij}p_{1,ij}\frac{r_{ij}}{p_{ij}} + 2J_{ij}(\tilde{t}_{ji}-\Omega_{ji})H_{ji}p_{1,ji}\frac{r_{ij}}{p_{ji}} \\
&\qquad + 2(\tilde{t}_{ij}-\Omega_{ij})(\tilde{t}_{ji}-\Omega_{ji})H_{ij}H_{ji}\tilde{\rho}_{ij}\sqrt{p_{1,ij}p_{1,ji}} \Big\} = v_{2,N}.
\end{aligned}
$$

Setting $\partial_\rho \mathcal{M}(\hat{\boldsymbol{\gamma}}, \hat{\theta}, \hat{\rho}) = 0$ and rearranging from Lemma C.1 now gives

$$\begin{aligned}\big(- \partial_{\rho^2} \mathcal{M}\big)[\hat{\rho} - \rho^0] \ =& \frac{1}{N} \sum_{\substack{i,j \in V \\ i<j}} U_{ij} + T_N' \bar{W}_{1,N}^{-1} B_N^\theta + B_N^{\rho,*} \\ & + O_p\Big(\sqrt{N}\left|\hat{\rho} - \rho^0\right| + N\left|\hat{\rho} - \rho^0\right|^2\Big) + o_p(1).\end{aligned}$$

Plugging in for $\partial_{\rho^2}\mathcal{M}$ from Lemma C.3 and for $B_N^{\rho,*}$ from Lemma C.2 now gives

$$\begin{aligned}v_{1,N} N[\hat{\rho} - \rho^0] \ =& \frac{\sqrt{2}}{\sqrt{N(N-1)/2}} \sum_{\substack{i,j \in V \\ i<j}} U_{ij} + 2T_N' \bar{W}_{1,N}^{-1} B_N^\theta + 2B_N^{\rho} \\ & + O_p\Big(\sqrt{N}\left|\hat{\rho} - \rho^0\right| + N\left|\hat{\rho} - \rho^0\right|^2\Big) + o_p(1).\end{aligned}$$

By an appropriate CLT

$$\frac{1}{\sqrt{v_{2,N} N(N-1)/2}} \sum_{\substack{i,j \in V \\ i<j}} U_{ij} \to \mathcal{N}(0,1).$$

□

Proof of Theorem 3. Write

$$N^{-2}\Big(S_N - \widehat{\mathbb{E}\,S_N}\Big) = N^{-2}\Big(S_N - \bar{\mathbb{E}} S_N\Big) - N^{-2}\Big(\widehat{\mathbb{E}\,S_N} - \bar{\mathbb{E}} S_N\Big).$$

We first analyze the second term. By definition

$$N^{-2}\Big(\widehat{\mathbb{E}\,S_N} - \bar{\mathbb{E}} S_N\Big) = s_N(\hat{\boldsymbol{\gamma}}, \hat{\theta}) - s_N(\boldsymbol{\gamma}^0, \theta^0)$$

Therefore, by Lemma C.4 and Lemma C.5

$$\begin{aligned}N^{-2}\Big(\widehat{\mathbb{E}\,S_N} - \bar{\mathbb{E}} S_N\Big) =& \left\{ (\partial_{\theta'} s_N) + (\partial_{\boldsymbol{\gamma}'} s_N)\, \bar{\mathcal{H}}^{-1} \left[\partial_{\boldsymbol{\gamma}\theta'} \bar{\mathcal{L}}\right] \right\} (\hat{\theta} - \theta^0) \\ & + (\partial_{\boldsymbol{\gamma}'} s_N)\, \bar{\mathcal{H}}^{-1} \mathcal{S} + B_N^S + o_p\big(1\big).\end{aligned}$$

Straightforward calculations give

$$\partial_\theta s_N = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \boldsymbol{\beta}_{ij}^N \omega_{ij} X_{ij}.$$

As in the proof of Lemma C.3, for $k = 1, \ldots, \dim(\theta)$ let

$$\Xi_{ij,k} = -\frac{1}{N} \sum_{k_1 \in V} \sum_{k_2 \in V_{-k_1}} \left(\bar{\mathcal{H}}^{-1}_{SS,ik_1} + \bar{\mathcal{H}}^{-1}_{RS,jk_1} + \bar{\mathcal{H}}^{-1}_{SR,ik_2} + \bar{\mathcal{H}}^{-1}_{RR,jk_2} \right) \bar{\mathbb{E}}(\partial_{\theta_k \pi} \ell_{k_1 k_2}).$$

and let $\Xi_{ij} = (\Xi_{ij,1}, \ldots, \Xi_{ij,\dim(\theta)})'$. By Lemma S.8(i) in FVW and the matrix representation of $\partial_{\boldsymbol{\gamma}} s_N$ from the proof of Lemma C.5

$$(\partial_{\boldsymbol{\gamma}'} s_N) \bar{\mathcal{H}}^{-1} (\partial_{\boldsymbol{\gamma}\theta'} \bar{\mathcal{L}}) = -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \boldsymbol{\beta}^N_{ij} \omega_{ij} \Xi'_{ij},$$

$$(\partial_{\boldsymbol{\gamma}'} s_N) \bar{\mathcal{H}}^{-1} \mathcal{S} = -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} (\mathcal{P}\boldsymbol{\beta}_N)_{ij} H_{ij} (Y_{ij} - p_{ij}).$$

Straightforward calculations give $X_{ij} - \Xi_{ij} = \tilde{X}_{ij}$ so that

$$\partial_\theta s_N + (\partial_{\boldsymbol{\gamma}'} s_N) \bar{\mathcal{H}}^{-1} (\partial_{\boldsymbol{\gamma}\theta'} \bar{\mathcal{L}}) = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \boldsymbol{\beta}^N_{ij} \omega_{ij} \tilde{X}'_{ij}.$$

Plugging in the linear representation of $\hat{\theta}$ from Lemma A.4 gives

$$\begin{aligned}
&N^{-2} \left(\widehat{\mathbb{E}\, S_N} - \bar{\mathbb{E}} S_N \right) \\
=& B^S_N + U'_N \bar{W}^{-1}_{1,N} B^\theta_N + \left(U'_N \bar{W}^{-1}_{1,N} \right)' \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \tilde{X}_{ij} H_{ij} (Y_{ij} - p_{ij}) \\
&+ \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} (\mathcal{P}\boldsymbol{\beta}_N)_{ij} H_{ij} (Y_{ij} - p_{ij}) + o_p(1).
\end{aligned}$$

Then, by Lemma C.6

$$\begin{aligned}
&N^{-2} \left(S_N - \widehat{\mathbb{E}\, S_N} \right) \\
=& - B^S_N - U'_N W^{-1}_{1,N} B^\theta_N + \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \left(\boldsymbol{\beta}^N_{ij} - (\mathcal{P}\boldsymbol{\beta}_N)_{ij} - \tilde{u}_{N,ij} \right) H_{ij} (Y_{ij} - p_{ij}) + o_p(1).
\end{aligned}$$

The sum on the right-hand side has conditional variance v^S_N. The conclusion now follows by applying a CLT. □

C. Supporting lemmas

Lemma C.1 (Expansion of score of second stage likelihood)**.** *Under Assumption 1,*

$$\begin{aligned}\partial_\rho \mathcal{M}\big(\hat{\boldsymbol{\gamma}},\hat{\theta},\rho\big) =&\partial_\rho\mathcal{M} + (\partial_{\rho^2}\mathcal{M})(\rho-\rho^0) + \big(\partial_{\rho\theta'}\mathcal{M} + \partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\bar{\mathcal{H}}^{-1}\big[\partial_{\boldsymbol{\gamma}\theta'}\bar{\mathcal{L}}\big]\big)(\hat{\theta}-\theta^0)\\ &+\big(\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\big)\bar{\mathcal{H}}^{-1}\mathcal{S} + B_N^{\rho,*} + O_p\Big(\sqrt{N}\left|\rho-\rho^0\right| + N\left|\rho-\rho^0\right|^2\Big) + o_p(1)\end{aligned}$$

where the order of the higher-order terms is uniform in $\rho\in[-1+\kappa,1-\kappa]$ *and*

$$\begin{aligned}B_N^{\rho,*} =&\big(\partial_{\rho\boldsymbol{\gamma}'}\tilde{\mathcal{M}}\big)\bar{\mathcal{H}}^{-1}\mathcal{S} - \big(\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\big)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}\\ &+\frac{1}{2}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)'\left\{\partial_{\rho\boldsymbol{\gamma}\boldsymbol{\gamma}'}\bar{\mathcal{M}} + \sum_{g=1}^{\dim(\boldsymbol{\gamma})}\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'\gamma_g}\bar{\mathcal{L}}\big[\bar{\mathcal{H}}^{-1}\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\big]_g\right\}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big).\end{aligned}$$

Proof. Throughout the proof $q=8$. By a Taylor expansion

$$\begin{aligned}\partial_\rho\mathcal{M}\big(\hat{\boldsymbol{\gamma}},\hat{\theta},\rho\big) - \partial_\rho\mathcal{M}\big(\hat{\boldsymbol{\gamma}},\theta^0,\rho^0\big) =&\partial_{\rho^2}\bar{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)(\rho-\rho^0)\\ &+\partial_{\rho\theta'}\bar{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)(\hat{\theta}-\theta^0) + R_{N,1}(\rho) + R_{N,2}\end{aligned}$$

where

$$\begin{aligned}R_{N,1}(\rho) =&\partial_{\rho^2}\tilde{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)(\rho-\rho^0) + (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0)'\left[\partial_{\rho^2\boldsymbol{\gamma}}\mathcal{M}\big(\bar{\boldsymbol{\gamma}},\bar{\theta},\rho^0\big)\right](\rho-\rho^0)\\ &+(\hat{\theta}-\theta^0)'\left[\partial_{\rho^2\theta}\mathcal{M}\big(\bar{\boldsymbol{\gamma}},\bar{\theta},\rho^0\big)\right](\rho-\rho^0) + \partial_{\rho^3}\mathcal{M}\big(\hat{\boldsymbol{\gamma}},\hat{\theta},\bar{\rho}\big)(\rho-\rho^0)^2\\ R_{N,2} =&\partial_{\rho\theta'}\tilde{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)(\hat{\theta}-\theta^0) + (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0)'\left[\partial_{\rho\boldsymbol{\gamma}\theta'}\mathcal{M}\big(\tilde{\boldsymbol{\gamma}},\theta^0,\rho^0\big)\right](\hat{\theta}-\theta^0)\\ &+(\hat{\theta}-\theta^0)'\Big[\partial_{\rho\theta\theta'}\mathcal{M}\big(\hat{\boldsymbol{\gamma}},\tilde{\theta},\rho^0\big)\Big](\hat{\theta}-\theta^0)\end{aligned}$$

and $\bar{\boldsymbol{\gamma}},\tilde{\boldsymbol{\gamma}},\bar{\theta},\tilde{\theta},\bar{\rho}$ are intermediate values. It is easy to see that

$$\begin{aligned}\sup_{\boldsymbol{\gamma}\in\Gamma,\theta\in\Theta}\left\|\partial_{\rho\theta\theta'}\mathcal{M}\big(\boldsymbol{\gamma},\theta,\rho^0\big)\right\| &= O_p(N),\\ \sup_{\boldsymbol{\gamma}\in\Gamma,\theta\in\Theta}\left\|\partial_{\rho^2\theta}\mathcal{M}\big(\boldsymbol{\gamma},\theta,\rho^0\big)\right\| &= O_p(N),\\ \sup_{\boldsymbol{\gamma}\in\Gamma,\theta\in\Theta,\rho\in[-\kappa,\kappa]}\left\|\partial_{\rho^3}\mathcal{M}\big(\boldsymbol{\gamma},\theta,\rho\big)\right\| &= O_p(N),\end{aligned}$$

and

$$\begin{aligned}\left\|\partial_{\rho^2}\tilde{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)\right\| &= O_p(1),\\ \left\|\partial_{\rho\theta'}\tilde{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)\right\| &= O_p(1).\end{aligned}$$

Moreover, applying Lemma C.7(i) gives

$$\sup_{\gamma\in\Gamma,\theta\in\Theta}\left\|\partial_{\rho^2\gamma}\mathcal{M}\big(\boldsymbol{\gamma},\theta,\rho^0\big)\right\|_q = O_p\Big(N^{\frac{1}{q}}\Big),$$
$$\sup_{\gamma\in\Gamma,\theta\in\Theta}\left\|\partial_{\rho\gamma\theta'}\mathcal{M}\big(\boldsymbol{\gamma},\theta,\rho^0\big)\right\|_q = O_p\Big(N^{\frac{1}{q}}\Big).$$

Noting that $\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\|_q = O_p(N^{-1/2+1/q})$ and $\|\hat{\theta}-\theta\| = O_p(N^{-1})$, we have

$$\begin{aligned}
|R_{N,1}(\rho)| \leq & \left\|\partial_{\rho^2}\tilde{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)\right\| \left|\hat{\rho}-\rho^0\right| \\
& + N^{1-2/q}\left\|\partial_{\rho^2\gamma}\mathcal{M}\big(\bar{\boldsymbol{\gamma}},\bar{\theta},\rho^0\big)\right\|_q \left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\right\|_q \left|\rho-\rho^0\right| \\
& + \left\|\hat{\theta}-\theta^0\right\| \left|\rho-\rho^0\right| \left\|\partial_{\rho^2\theta}\mathcal{M}\big(\bar{\boldsymbol{\gamma}},\bar{\theta},\rho^0\big)\right\| + \left|\rho-\rho^0\right|^2 \left\|\partial_{\rho^3}\mathcal{M}\big(\hat{\boldsymbol{\gamma}},\hat{\theta},\rho^0\big)\right\| \\
= & O_p\left(\sqrt{N}\left|\rho-\rho^0\right| + N\left|\rho-\rho^0\right|^2\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
|R_{N,2}| \leq & \left\|\partial_{\rho\theta'}\tilde{\mathcal{M}}\big(\boldsymbol{\gamma}^0,\theta^0,\rho^0\big)\right\| \left\|\hat{\theta}-\theta^0\right\| \\
& + N^{1-2/q}\left\|\partial_{\rho\gamma\theta'}\mathcal{M}\big(\tilde{\boldsymbol{\gamma}},\theta^0,\rho^0\big)\right\|_q \left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\right\|_q \left\|\hat{\theta}-\theta^0\right\| \\
& + \left\|\theta-\theta^0\right\|^2 \left\|\partial_{\rho\theta\theta'}\mathcal{M}\big(\hat{\boldsymbol{\gamma}},\tilde{\theta},\rho^0\big)\right\| = o_p(1).
\end{aligned}$$

Next, Taylor-expanding $\partial_\rho\mathcal{M}\big(\hat{\boldsymbol{\gamma}},\theta^0,\rho^0\big)$ and plugging in the expansion for $\hat{\boldsymbol{\gamma}}$ from Lemma A.2

$$\begin{aligned}
& \partial_\rho\mathcal{M}(\hat{\boldsymbol{\gamma}},\theta^0,\rho^0) - \partial_\rho\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0) \\
= & \partial_{\rho\gamma'}\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0)\Bigg\{\mathcal{H}^{-1}\mathcal{S} + \mathcal{H}^{-1}\big[\partial_{\gamma\theta'}\mathcal{L}\big](\hat{\theta}-\theta^0) \\
& \qquad + \frac{1}{2}\mathcal{H}^{-1}\sum_{g=1}^{\dim(\boldsymbol{\gamma})}\big[\partial_{\gamma\gamma'\gamma_g}\mathcal{L}\big]\mathcal{H}^{-1}S\big[\mathcal{H}^{-1}S\big]_g\Bigg\} \\
& + \frac{1}{2}\big(\mathcal{H}^{-1}S\big)'\Big[\partial_{\rho\gamma\gamma'}\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0)\Big]\big(\mathcal{H}^{-1}S\big) + R_{N,3}
\end{aligned}$$

with

$$\begin{aligned}
R_{N,3} = & \big[\partial_{\rho\gamma'}\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0)\big]R_{\boldsymbol{\gamma}}(\hat{\theta}) \\
& + \frac{1}{2}\big(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0-\mathcal{H}^{-1}\mathcal{S}\big)\big[\partial_{\rho\gamma\gamma'}\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0)\big]\big(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0+\mathcal{H}^{-1}\mathcal{S}\big) \\
& + \frac{1}{6}\sum_{g=1}^{\dim(\boldsymbol{\gamma})}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0)'\Big[\partial_{\rho\gamma\gamma'\gamma_g}\mathcal{M}(\bar{\boldsymbol{\gamma}},\theta^0,\rho^0)\Big](\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0)[\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0]_g,
\end{aligned}$$

where $\bar{\boldsymbol{\gamma}}$ is an intermediate value By Lemma C.7,

$$\begin{aligned}\left\|\partial_{\rho\boldsymbol{\gamma}'}\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0)\right\|_q &= O_p\left(N^{\frac{1}{q}}\right)\\ \sup_{\boldsymbol{\gamma}\in\Gamma}\left\|\partial_{\rho\boldsymbol{\gamma}\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathcal{M}(\boldsymbol{\gamma},\theta^0,\rho^0)\right\|_q &= O_p(1).\end{aligned}$$

Noting that $\left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0-\mathcal{H}^{-1}\mathcal{S}\right\|_q = O_p\left(N^{-1+2/q}\right)$,

$$\begin{aligned}|R_{N,3}| \leq& N^{1-2/q}\left\|\partial_{\rho\boldsymbol{\gamma}'}\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0)\right\|_q\left\|R_{\boldsymbol{\gamma}}(\hat{\theta})\right\|_q\\ &+\frac{1}{2}N^{1-2/q}\left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0-\mathcal{H}^{-1}\mathcal{S}\right\|_q\left(\left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\right\|_q+\left\|\mathcal{H}^{-1}\mathcal{S}\right\|_q\right)\left\|\partial_{\rho\boldsymbol{\gamma}\boldsymbol{\gamma}'}\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0)\right\|_q\\ &+\frac{1}{6}N^{1-2/q}\left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\right\|_q^3\left\|\partial_{\rho\boldsymbol{\gamma}\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathcal{M}(\bar{\boldsymbol{\gamma}},\theta^0,\rho^0)\right\|_q\\ \leq&(1+N\|\hat{\theta}-\theta^0\|)N^{1-1/q}\frac{\left\|R_{\boldsymbol{\gamma}}(\hat{\theta})\right\|_q}{1+N\|\hat{\theta}-\theta^0\|}+O_p\left(N^{-1/2+1/q}\right)=o_p(1).\end{aligned}$$

From now on, drop the arguments of $\mathcal{M}$ and its derivatives whenever they are evaulated at their true values. Then

$$\left(\partial_{\rho\boldsymbol{\gamma}'}\mathcal{M}\right)\mathcal{H}^{-1}\mathcal{S}=\left(\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\mathcal{S}+\left(\partial_{\rho\boldsymbol{\gamma}'}\tilde{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\mathcal{S}-\left(\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}+R_{N,4}$$

with

$$R_{N,4}=-\left(\partial_{\rho\boldsymbol{\gamma}'}\tilde{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}+\left(\partial_{\rho\boldsymbol{\gamma}'}\mathcal{M}\right)\left(\mathcal{H}^{-1}-\left(\bar{\mathcal{H}}^{-1}-\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\right)\right)\mathcal{S}.$$

Lemma C.7 applied with $q=2$ yields $\left\|\partial_{\rho\boldsymbol{\gamma}'}\tilde{\mathcal{M}}\right\|=O_p(1)$ and $\left\|\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\right\|=O_p\left(N^{1/2}\right)$. Then, by Lemma A.3

$$|R_{N,4}|\leq\left\|\partial_{\rho\boldsymbol{\gamma}'}\tilde{\mathcal{M}}\right\|\left\|\bar{\mathcal{H}}^{-1}\right\|^2\left\|\tilde{\mathcal{H}}\right\|\left\|\mathcal{S}\right\|+\left\|\partial_{\rho\boldsymbol{\gamma}'}\mathcal{M}\right\|\left\|\mathcal{H}^{-1}-\left(\bar{\mathcal{H}}^{-1}-\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\right)\right\|\|\mathcal{S}\|=o_p(1).$$

Next,

$$\left(\partial_{\rho\boldsymbol{\gamma}'}\mathcal{M}\right)\mathcal{H}^{-1}\left[\partial_{\boldsymbol{\gamma}\theta'}\mathcal{L}\right](\hat{\theta}-\theta^0)=\left(\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\left[\partial_{\boldsymbol{\gamma}\theta'}\bar{\mathcal{L}}\right](\hat{\theta}-\theta^0)+R_{N,5}$$

with

$$\begin{aligned}R_{N,5}=&\left(\partial_{\rho\boldsymbol{\gamma}'}\tilde{\mathcal{M}}\right)\mathcal{H}^{-1}\left[\partial_{\boldsymbol{\gamma}\theta'}\mathcal{L}\right](\hat{\theta}-\theta^0)+\left(\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\right)\left(\mathcal{H}^{-1}-\bar{\mathcal{H}}^{-1}\right)\left[\partial_{\boldsymbol{\gamma}\theta'}\mathcal{L}\right](\hat{\theta}-\theta^0)\\ &+\left(\partial_{\rho\boldsymbol{\gamma}'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\left[\partial_{\boldsymbol{\gamma}\theta'}\tilde{\mathcal{L}}\right](\hat{\theta}-\theta^0)\end{aligned}$$

and

$$|R_{N,5}| \leq \left\|\partial_{\rho\gamma'}\tilde{\mathcal{M}}\right\|\left\|\mathcal{H}^{-1}\right\|\left\|\left[\partial_{\gamma\theta'}\mathcal{L}\right]\right\|\left\|\hat{\theta}-\theta^0\right\| + \left\|\partial_{\rho\gamma'}\bar{\mathcal{M}}\right\|\left\|\hat{\theta}-\theta^0\right\|\Big\{\left\|\mathcal{H}^{-1}-\bar{\mathcal{H}}^{-1}\right\|\left\|\partial_{\gamma\theta'}\mathcal{L}\right\| + \left\|\bar{\mathcal{H}}^{-1}\right\|\left\|\partial_{\gamma\theta'}\tilde{\mathcal{L}}\right\|\Big\} = o_p(1).$$

Repeating the last argument in the proof of Theorem B.1, Part 2 in FVW almost ad verbum gives

$$\begin{aligned}&\left(\partial_{\rho\gamma'}\mathcal{M}\right)\mathcal{H}^{-1}\sum_{g=1}^{\dim(\boldsymbol{\gamma})}\left[\partial_{\gamma\gamma'\gamma_g}\mathcal{L}\right]\mathcal{H}^{-1}S\left[\mathcal{H}^{-1}S\right]_g \\ =&\left(\partial_{\rho\gamma'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\sum_{g=1}^{\dim(\boldsymbol{\gamma})}\left[\partial_{\gamma\gamma'\gamma_g}\bar{\mathcal{L}}\right]\bar{\mathcal{H}}^{-1}S\left[\bar{\mathcal{H}}^{-1}S\right]_g + o_p(1).\end{aligned}$$

Now write

$$\left(\mathcal{H}^{-1}S\right)'\Big[\partial_{\rho\gamma\gamma'}\mathcal{M}\Big]\left(\mathcal{H}^{-1}S\right) = \left(\bar{\mathcal{H}}^{-1}S\right)'\Big[\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}}\Big]\left(\bar{\mathcal{H}}^{-1}S\right) + R_{N,6}$$

with

$$\begin{aligned}|R_{N,6}| \leq&\left\|\mathcal{S}\right\|^2\left\|\mathcal{H}^{-1}-\bar{\mathcal{H}}^{-1}\right\|\left(\left\|\mathcal{H}^{-1}\right\|+\left\|\bar{\mathcal{H}}^{-1}\right\|\right)\left\|\partial_{\rho\gamma\gamma'}\mathcal{M}\right\| \\ &+\left\|\mathcal{S}\right\|^2\left\|\bar{\mathcal{H}}\right\|^2\left\|\partial_{\rho\gamma\gamma'}\tilde{\mathcal{M}}\right\| = o_p(1).\end{aligned}$$

The last inequality uses that by Lemma C.7(iii)

$$\left\|\partial_{\rho\gamma\gamma'}\tilde{\mathcal{M}}\right\| = O_p\left(N^{-3/8}\right).$$

We may now conclude that

$$\begin{aligned}&\partial_\rho\mathcal{M}(\hat{\gamma},\theta^0,\rho^0) - \partial_\rho\mathcal{M}(\boldsymbol{\gamma}^0,\theta^0,\rho^0) \\ =&\left(\partial_{\rho\gamma'}\tilde{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\mathcal{S} + \bar{\mathcal{H}}^{-1}\left[\partial_{\gamma\theta'}\bar{\mathcal{L}}\right](\hat{\theta}-\theta^0) + \left(\partial_{\rho\gamma'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\mathcal{S} - \left(\partial_{\rho\gamma'}\bar{\mathcal{M}}\right)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} \\ &+\frac{1}{2}\left(\bar{\mathcal{H}}^{-1}\mathcal{S}\right)'\left\{\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}} + \sum_{g=1}^{\dim(\boldsymbol{\gamma})}\left[\partial_{\gamma\gamma'\gamma_g}\bar{\mathcal{L}}\right]\bar{\mathcal{H}}^{-1}S\left[\bar{\mathcal{H}}^{-1}\partial_{\rho\gamma'}\bar{\mathcal{M}}\right]_g\right\}\left(\bar{\mathcal{H}}^{-1}\mathcal{S}\right) + o_p(1).\end{aligned}$$

□

Lemma C.2 (Behavior of deterministic term of $\hat{\rho}$)**.** *Under Assumption 1*

$$B_N^{\rho;*} = B_N^{\rho} + O_p\left(N^{-1/2}\right)$$

with B_N^{ρ} as defined in Theorem 2.

Proof.
Step 1: behavior of $(\partial_{\rho\gamma}\tilde{\mathcal{M}})\bar{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\rho\gamma}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}$.
Let $D_{ij}^{(m)} = \partial_{\rho z_1} m_{ij} - \Omega_{ij}(\partial_{z^2}\ell_{ij})$ and

$$\Lambda_{ij} = -\frac{1}{N}\sum_{k\in V}\sum_{l\in V_{-k}}\left(\bar{\mathcal{H}}^{-1}_{SS,ik} + \bar{\mathcal{H}}^{-1}_{RS,jk} + \bar{\mathcal{H}}^{-1}_{SR,il} + \bar{\mathcal{H}}^{-1}_{RR,jl}\right)(\partial_z \ell_{kl}).$$

By Lemma S.8(i) and (iii) in FVW

$$\begin{aligned}
&(\partial_{\rho\gamma}\tilde{\mathcal{M}})\bar{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\rho\gamma}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}\\
=& -\frac{1}{N}\sum_{i\neq j}\Lambda_{ij}(\partial_{\rho z_1}\tilde{m}_{ij}) + \frac{1}{N}\sum_{i\neq j}\Lambda_{ij}\Omega_{ij}\big(\partial_{z^2}\ell_{ij} - \bar{\mathbb{E}}[\partial_{z^2}\ell_{ij}]\big)\\
=& -\frac{1}{N}\sum_{i\neq j}\Lambda_{ij}\left(D_{ij}^{(m)} - \bar{\mathbb{E}}D_{ij}^{(m)}\right) = U_1 + U_2 + U_3 + U_4
\end{aligned}$$

with

$$\begin{aligned}
U_1 &= \frac{1}{N^2}\sum_{i\in V}\left\{\Big(\sum_{\substack{k\in V\\ l\in V_{-k}}}\bar{\mathcal{H}}^{-1}_{SS,ik}(\partial_z\ell_{k,l})\Big)\sum_{j\in V_{-i}}\left(D_{ij}^{(m)} - \bar{\mathbb{E}}D_{ij}^{(m)}\right)\right\},\\
U_2 &= \frac{1}{N^2}\sum_{j\in V}\left\{\Big(\sum_{\substack{k\in V\\ l\in V_{-k}}}\bar{\mathcal{H}}^{-1}_{RS,jk}(\partial_z\ell_{k,l})\Big)\sum_{i\in V_{-j}}\left(D_{ij}^{(m)} - \bar{\mathbb{E}}D_{ij}^{(m)}\right)\right\},\\
U_3 &= \frac{1}{N^2}\sum_{i\in V}\left\{\Big(\sum_{\substack{k\in V\\ l\in V_{-k}}}\bar{\mathcal{H}}^{-1}_{SR,il}(\partial_z\ell_{k,l})\Big)\sum_{j\in V_{-i}}\left(D_{ij}^{(m)} - \bar{\mathbb{E}}D_{ij}^{(m)}\right)\right\},\\
U_4 &= \frac{1}{N^2}\sum_{j\in V}\left\{\Big(\sum_{\substack{k\in V\\ l\in V_{-k}}}\bar{\mathcal{H}}^{-1}_{RR,jl}(\partial_z\ell_{k,l})\Big)\sum_{i\in V_{-j}}\left(D_{ij}^{(m)} - \bar{\mathbb{E}}D_{ij}^{(m)}\right)\right\}.
\end{aligned}$$

Next,

$$\begin{aligned}
U_1 =& \frac{1}{N^2}\sum_{i\in V}\left\{\Big(\sum_{k\in V}\sum_{l\in V_{-k}}\big[(\bar{\mathcal{H}}^*_{SS})^{-1}\big]_{ik}(\partial_z\ell_{k,l})\Big)\sum_{j\in V_{-i}}\left(D_{ij}^{(m)} - \bar{\mathbb{E}}D_{ij}^{(m)}\right)\right\}\\
&+ \frac{1}{N^2}\sum_{i\in V}\left\{\Big(\sum_{k\in V}\sum_{l\in V_{-k}}\big[\bar{\mathcal{H}}^{-1}_{SS} - (\bar{\mathcal{H}}^*_{SS})^{-1}\big]_{ik}(\partial_z\ell_{k,l})\Big)\sum_{j\in V_{-i}}\left(D_{ij}^{(m)} - \bar{\mathbb{E}}D_{ij}^{(m)}\right)\right\}\\
=& U_{1a} + U_{1b}.
\end{aligned}$$

A straightforward application of Cauchy-Schwarz yields

$$
\begin{aligned}
(U_{1b})^2 \le & N^{-1}\frac{1}{N}\sum_{i\in V}\left(\frac{1}{N}\sum_{k\in V}\sum_{l\in V_{-k}} N\big[\bar{\mathcal{H}}_{SS}^{-1}-(\bar{\mathcal{H}}_{SS}^{*})^{-1}\big]_{ik}(\partial_z\ell_{k,l})\right)^2 \\
& \times \frac{1}{N}\sum_{i\in V}\left(\frac{1}{\sqrt{N}}\sum_{j\in V_{-i}}\left(D_{ij}^{(m)}-\bar{\mathbb{E}}D_{ij}^{(m)}\right)\right)^2.
\end{aligned}
$$

Now observe that for $k_1,k_2,l_1,l_2\in V$, $\bar{\mathbb{E}}[\partial_z\ell_{k_1,l_1}\partial_z\ell_{k_2,l_2}]$ is bounded if $\{k_1,l_1\}=\{k_2,l_2\}$ and 0 otherwise. The cardinality of the set $\{(k_1,k_2,l_1,l_2): k_1,k_2,l_1,l_2\in V,\{k_1,l_1\}=\{k_2,l_2\}\}$ is $O(N^2)$. Moreover, by Lemma A.1

$$
N\big\|\bar{\mathcal{H}}_{SS}^{-1}-(\bar{\mathcal{H}}_{SS}^{*})^{-1}\big\|_{\max}=O_p(1).
$$

Therefore,

$$
\begin{aligned}
& \sup_{i\in V}\bar{\mathbb{E}}\left\{\frac{1}{N}\sum_{k\in V}\sum_{l\in V_{-k}} N\big[\bar{\mathcal{H}}_{SS}^{-1}-(\bar{\mathcal{H}}_{SS}^{*})^{-1}\big]_{ik}(\partial_z\ell_{k,l})\right\}^2 \\
\le & \sup_{i\in V}\frac{1}{N^2}\sum_{k_1,k_2,l_1,l_2\in V}\Bigg\{N\big[\bar{\mathcal{H}}_{SS}^{-1}-(\bar{\mathcal{H}}_{SS}^{*})^{-1}\big]_{ik_1}N\big[\bar{\mathcal{H}}_{SS}^{-1}-(\bar{\mathcal{H}}_{SS}^{*})^{-1}\big]_{ik_2} \\
& \qquad \times \bar{\mathbb{E}}[\partial_z\ell_{k_1l_1}\partial_z\ell_{k_2l_2}]\Bigg\}^2 = O_p(1).
\end{aligned}
$$

Furthermore,

$$
\sup_{i\in V}\bar{\mathbb{E}}\left\{\frac{1}{\sqrt{N}}\sum_{j\in V_{-i}}\left(D_{ij}^{(m)}-\bar{\mathbb{E}}D_{ij}^{(m)}\right)\right\}^2=\sup_{i\in V}\frac{1}{N}\sum_{j\in V_{-i}}\bar{\mathbb{E}}\left(D_{ij}^{(m)}-\bar{\mathbb{E}}D_{ij}^{(m)}\right)^2=O_p(1).
$$

This implies that $\bar{\mathbb{E}}(U_{1b})^2=O_p(N^{-1})$ and therefore $U_{1b}=O_p(N^{-1/2})$. Moving on to the analysis of the term U_{1a}, we can write

$$
U_{1a}=\frac{1}{N}\sum_{i\in V}(\bar{\mathcal{H}}_{SS}^{*})_{ii}^{-1}\left(\frac{1}{\sqrt{N}}\sum_{l\in V_{-i}}\partial_z\ell_{il}\right)\left(\frac{1}{\sqrt{N}}\sum_{j\in V_{-i}}\left(D_{ij}^{(m)}-\bar{\mathbb{E}}D_{ij}^{(m)}\right)\right).
$$

Let

$$
\begin{aligned}
s_{i_1} =& \left(\sum_{l_1 \in V_{-i_1}} \partial_z \ell_{i_1 l_1}\right)\left(\sum_{j_1 \in V_{-i_1}} \tilde{D}_{i_1 j_1}\right) \\
=& \left(\partial_z \ell_{i_1 i_2} + \sum_{l_1 \in V_{-\{i_1,i_2\}}} \partial_z \ell_{i_1 l_1}\right)\left(\tilde{D}_{i_1 i_2} + \sum_{j_1 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_1 j_1}\right).
\end{aligned}
$$

Under our assumptions, for $i_1 \neq i_2$

$$
\begin{aligned}
& \bar{\mathbb{E}}(s_{i_1} s_{i_2}) \\
=& \bar{\mathbb{E}}\left(\sum_{l_1 \in V_{-\{i_1,i_2\}}} \partial_z \ell_{i_1 l_1} \sum_{j_1 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_1 j_1}\right) \bar{\mathbb{E}}\left(\sum_{l_2 \in V_{-\{i_1,i_2\}}} \partial_z \ell_{i_2 l_2} \sum_{j_2 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_2 j_2}\right) \\
& + \bar{\mathbb{E}}\left[\left(\tilde{D}_{i_1 i_2} \sum_{l_1 \in V_{-\{i_1,i_2\}}} \partial_z \ell_{i_1 l_1}\right)\left(\tilde{D}_{i_2 i_1} \sum_{l_2 \in V_{-\{i_1,i_2\}}} \partial_z \ell_{i_2 l_2}\right)\right] \\
& + \bar{\mathbb{E}}\left[\left(\partial_z \ell_{i_1 i_2} \sum_{l_1 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_1 l_1}\right)\left(\partial_z \ell_{i_2 i_1} \sum_{l_2 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_2 l_2}\right)\right] + O_p(N) \\
=& \bar{\mathbb{E}}\left(\sum_{l_1 \in V_{-\{i_1,i_2\}}} \partial_z \ell_{i_1 l_1} \sum_{j_1 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_1 j_1}\right) \bar{\mathbb{E}}\left(\sum_{l_2 \in V_{-\{i_1,i_2\}}} \partial_z \ell_{i_2 l_2} \sum_{j_2 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_2 j_2}\right) + O_p(N)
\end{aligned}
$$

where the $O_p(N)$ term is uniform in i_1, i_2. Similarly,

$$
\bar{\mathbb{E}}(s_{i_1}) = \bar{\mathbb{E}}\left(\sum_{l_1 \in V_{-i_1}} \partial_z \ell_{i_1 l_1} \sum_{j_1 \in V_{-\{i_1,i_2\}}} \tilde{D}_{i_1 j_1}\right) + O_p(1).
$$

Then,

$$
\begin{aligned}
\bar{\text{var}}(U_{1a}) =& N^{-4}\sum_{i\in V}(\bar{\mathcal{H}}^*_{SS})^{-2}_{ii}\bar{\mathbb{E}}\bigg(\sum_{l\in V_{-i}}(\partial_z\ell_{il})^2+\sum_{l\in V_{-i}}\sum_{k\in V_{-\{i,l\}}}\partial_z\ell_{il}\partial_z\ell_{ik}\bigg)^2\\
&+N^{-4}\sum_{i_1\in V}\sum_{i_2\in V_{-i_1}}(\bar{\mathcal{H}}^*_{SS})^{-1}_{i_1i_1}(\bar{\mathcal{H}}^*_{SS})^{-1}_{i_2i_2}\bar{\mathbb{E}}(s_{i_1}s_{i_2})-\bigg(\sum_{i_1\in V}(\bar{\mathcal{H}}^*_{SS})^{-1}_{i_1i_1}\bar{\mathbb{E}}(s_{i_1})\bigg)^2\\
=& N^{-4}\sum_{i\in V}(\bar{\mathcal{H}}^*_{SS})^{-2}_{ii}\bigg(\sum_{l\in V_{-i}}\bar{\mathbb{E}}(\partial_z\ell_{il})^4+\sum_{l\in V_{-i}}\sum_{k\in V_{-\{i,l\}}}\bar{\mathbb{E}}(\partial_z\ell_{il})^2\bar{\mathbb{E}}(\partial_z\ell_{ik})^2\bigg)\\
&+N^{-4}\sum_{i_1\in V}\sum_{i_2\in V_{-i_1}}(\bar{\mathcal{H}}^*_{SS})^{-1}_{i_1i_1}(\bar{\mathcal{H}}^*_{SS})^{-2}_{i_2i_2}\bar{\mathbb{E}}(s_{i_1}s_{i_2})\\
&-N^{-4}\sum_{i_1\in V}\sum_{i_2\in V_{-i_1}}(\bar{\mathcal{H}}^*_{SS})^{-2}_{i_1i_1}(\bar{\mathcal{H}}^*_{SS})^{-2}_{i_2i_2}\bar{\mathbb{E}}(s_{i_1})\bar{\mathbb{E}}(s_{i_2})+O_p\big(N^{-3}\big)=O_p\big(N^{-1}\big).
\end{aligned}
$$

Therefore, $U_{1a}=\bar{\mathbb{E}}(U_{1a})+O_p\big(N^{-1/2}\big)$ or

$$
\begin{aligned}
U_{1a} =& \frac{1}{N}\sum_{i\in V}(\bar{\mathcal{H}}^*_{SS})^{-1}_{ii}\bigg(\frac{1}{N}\sum_{j\in V_{-i}}\bar{\mathbb{E}}\big(\partial_z\ell_{ij}\tilde{D}^{(m)}_{ij}\big)\bigg)+O_p\big(N^{-1/2}\big)\\
=& \frac{1}{N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}\bar{\mathbb{E}}\big(\partial_z\ell_{ij}D^{(m)}_{ij}\big)}{\sum_{j\in V_{-i}}\bar{\mathbb{E}}\big(-\partial_{z^2}\ell_{ij}\big)}+O_p\big(N^{-1/2}\big).
\end{aligned}
$$

Similar arguments can be used to show that

$$
\begin{aligned}
U_2 =& O_p\big(N^{-1/2}\big),\\
U_3 =& O_p\big(N^{-1/2}\big),\\
U_4 =& \frac{1}{N}\sum_{j\in V}\frac{\sum_{i\in V_{-j}}\bar{\mathbb{E}}\big(\partial_z\ell_{ij}D^{(m)}_{ij}\big)}{\sum_{i\in V_{-j}}\bar{\mathbb{E}}\big(-\partial_{z^2}\ell_{ij}\big)}+O_p\big(N^{-1/2}\big).
\end{aligned}
$$

In summary,

$$
\begin{aligned}
&(\partial_{\rho\gamma}\tilde{\mathcal{M}})\bar{\mathcal{H}}^{-1}\mathcal{S}-(\partial_{\rho\gamma}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}\\
=&\frac{1}{N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}\bar{\mathbb{E}}\big(\partial_z\ell_{ij}D_{ij}^{(m)}\big)}{\sum_{j\in V_{-i}}\bar{\mathbb{E}}\big(-\partial_{z^2}\ell_{ij}\big)}+\frac{1}{N}\sum_{j\in V}\frac{\sum_{i\in V_{-j}}\bar{\mathbb{E}}\big(\partial_z\ell_{ij}D_{ij}^{(m)}\big)}{\sum_{i\in V_{-j}}\bar{\mathbb{E}}\big(-\partial_{z^2}\ell_{ij}\big)}+O_p\big(N^{-1/2}\big)\\
=&\frac{1}{N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}(\partial_z p_{ij})\big\{(\partial_{z_1}J_{ij})\frac{r_{ij}}{p_{ij}}-\Omega_{ij}(\partial_z H_{ij})\big\}}{\sum_{j\in V_{-i}}(H_{ij}\partial_z p_{ij})}\\
&+\frac{1}{N}\sum_{j\in V}\frac{\sum_{i\in V_{-j}}(\partial_z p_{ij})\big\{(\partial_{z_1}J_{ij})\frac{r_{ij}}{p_{ij}}-\Omega_{ij}(\partial_z H_{ij})\big\}}{\sum_{i\in V_{-j}}(H_{ij}\partial_z p_{ij})}+O_p\big(N^{-1/2}\big).
\end{aligned}
$$

The second equality follows by noting that

$$
\begin{aligned}
\bar{\mathbb{E}}[(\partial_z\ell_{ij})(\partial_{z^2}\ell_{ij})]=&H_{ij}(\partial_z H_{ij})\bar{\mathbb{E}}[(Y_{ij}-p_{ij})^2]+0=(\partial_z H_{ij})(\partial_z p_{ij}),\\
\bar{\mathbb{E}}[(\partial_z\ell_{ij})(\partial_{\rho z_1}m_{ij})]=&H_{ij}(\partial_{z_1}J_{ij})\bar{\mathbb{E}}[(Y_{ij}-p_{ij})(Z_{ij}-r_{ij})]+0\\
=&H_{ij}(\partial_{z_1}J_{ij})(r_{ij}/p_{ij})p_{1,ij}=(\partial_z p_{ij})(\partial_{z_1}J_{ij})(r_{ij}/p_{ij}),\\
\bar{\mathbb{E}}[-(\partial_{z^2}\ell_{ij})]=&H_{ij}(\partial_z p_{ij}).
\end{aligned}
$$

Step 2: behavior of $\frac{1}{2}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)'\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)$.
Inspection of the proof of Lemma C.7 shows that $\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}}$ can be written as

$$
\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}}=\begin{bmatrix}D_{SS,1}+D_{SS,2} & D_{SR,1}+D_{SR,2}\\ D'_{SR,1}+D'_{SR,2} & D_{RR,1}+D_{RR,2}\end{bmatrix}
$$

where $D_{SS,1},D_{SR,1},D_{RR,1}$ are $N\times N$ diagonal matrices with entries

$$
\begin{aligned}
(D_{SS,1})_{ii}=&\partial_{\gamma_i^S\gamma_i^S}\partial_\rho\bar{\mathcal{M}}\\
(D_{SR,1})_{ii}=&\partial_{\gamma_i^S\gamma_i^R}\partial_\rho\bar{\mathcal{M}}\\
(D_{RR,1})_{ii}=&\partial_{\gamma_i^R\gamma_i^R}\partial_\rho\bar{\mathcal{M}}
\end{aligned}
$$

and $D_{SS,2},D_{SR,2},D_{RR,2}$ are $O_p(N^{-1})$ in the $\|\cdot\|_{\max}$-norm. Let Υ denote the $N\times N$ matrix with entries $\Upsilon_{ij}=\partial_z\ell_{ij}$. By Lemma A.1, $\bar{\mathcal{H}}^{-1}$ can be written as

$$
\bar{\mathcal{H}}^{-1}=\begin{bmatrix}(\bar{\mathcal{H}}^*_{SS})^{-1} & 0\\ 0 & (\bar{\mathcal{H}}^*_{RR})^{-1}\end{bmatrix}+R_N,
$$

where $\|R_N\|_{\max} = O_p(N^{-1})$. By the first assertion of Lemma C.12,

$$
\begin{aligned}
&N^2\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)'\Big(\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}}\Big)\bar{\mathcal{H}}^{-1}\mathcal{S} \\
=&(\iota_N'\Upsilon', \iota_N'\Upsilon)\,\bar{\mathcal{H}}^{-1}(\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\begin{pmatrix}\Upsilon\iota_N \\ \Upsilon'\iota_N\end{pmatrix} \\
=&\iota_N'\Upsilon'(\bar{\mathcal{H}}_{SS}^*)^{-1}D_{SS,1}(\bar{\mathcal{H}}_{SS}^*)^{-1}\Upsilon\iota_N + \iota_N'\Upsilon(\bar{\mathcal{H}}_{RR}^*)^{-1}D_{RR,1}(\bar{\mathcal{H}}_{RR}^*)^{-1}\Upsilon'\iota_N \\
&+ 2\iota_N'\Upsilon'(\bar{\mathcal{H}}_{SS}^*)^{-1}D_{SR,1}(\bar{\mathcal{H}}_{RR}^*)^{-1}\Upsilon'\iota_N + O_p(N).
\end{aligned}
$$

By the second assertion of Lemma C.12 and a Bartlett equality

$$
\begin{aligned}
\iota_N'\Upsilon'(\bar{\mathcal{H}}_{SS}^*)^{-1}D_{SS,1}(\bar{\mathcal{H}}_{SS}^*)^{-1}\Upsilon\iota_N =&\sum_{i\in V}\sum_{j\in V_{-i}}\big[(\bar{\mathcal{H}}_{SS}^*)^{-1}\big]_{ii}^2[D_{SS,1}]_{ii}\bar{\mathbb{E}}[(\partial_z\ell_{ij})^2] + O_p\big(N^{3/2}\big) \\
=&\sum_{i\in V}\frac{[D_{SS,1}]_{ii}\sum_{j\in V_{-i}}\bar{\mathbb{E}}[(\partial_{z^2}\ell_{ij})]}{\Big(N^{-1}\sum_{j\in V_{-i}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]\Big)^2} + O_p\big(N^{3/2}\big) \\
=&N\sum_{i\in V}\frac{[D_{SS,1}]_{ii}}{\Big(N^{-1}\sum_{j\in V_{-i}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]\Big)} + O_p\big(N^{3/2}\big).
\end{aligned}
$$

Similarly,

$$
\iota_N'\Upsilon(\bar{\mathcal{H}}_{RR}^*)^{-1}D_{RR,1}(\bar{\mathcal{H}}_{RR}^*)^{-1}\Upsilon'\iota_N = N\sum_{j\in V}\frac{[D_{RR,1}]_{jj}}{\Big(N^{-1}\sum_{i\in V_{-j}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]\Big)} + O_p\big(N^{3/2}\big)
$$

and

$$
\begin{aligned}
&\iota_N'\Upsilon'(\bar{\mathcal{H}}_{SS}^*)^{-1}D_{SR,1}(\bar{\mathcal{H}}_{RR}^*)^{-1}\Upsilon'\iota_N \\
=&N\sum_{i\in V}\frac{(D_{SR,1})_{ii}\,\overline{\mathrm{corr}}_i}{\Big(N^{-1}\sum_{j=1}^N\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]\Big)^{1/2}\Big(N^{-1}\sum_{j=1}^N\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ji}]\Big)^{1/2}},
\end{aligned}
$$

where

$$
\begin{aligned}
\overline{\text{corr}}_i =& \frac{\sum_{j \in V_{-i}} \bar{\mathbb{E}}[\partial_z \ell_{ij} \partial_z \ell_{ji}]}{\Big(\sum_{j \in V_{-i}} \bar{\mathbb{E}}[(\partial_z \ell_{ij})^2] \Big)^{1/2} \Big(\sum_{j \in V_{-i}} \bar{\mathbb{E}}[(\partial_z \ell_{ji})^2] \Big)^{1/2}} \\
=& \frac{\sum_{j \in V_{-i}} H_{ij} H_{ji} (r_{ij} - p_{ij} p_{ji})}{\Big(\sum_{j \in V_{-i}} H_{ij} (\partial_z p_{ij}) \Big)^{1/2} \Big(\sum_{j \in V_{-i}} H_{ji} (\partial_z p_{ji}) \Big)^{1/2}} \\
=& \frac{\sum_{j \in V_{-i}} \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}}}{\Big(\sum_{j \in V_{-i}} \omega_{ij} \Big)^{1/2} \Big(\sum_{j \in V_{-i}} \omega_{ji} \Big)^{1/2}}.
\end{aligned}
$$

Closed-form expressions for the elements of $D_{SS,1}, D_{SR,1}, D_{RR,1}$ are given in the proof of Lemma C.7. Re-writing them using Lemma C.9 yields

$$
\begin{aligned}
\partial_{(\gamma_i^S)^2} \partial_\rho \mathcal{M} &= \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\rho z_1^2} m_{ij} + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\rho z_2^2} m_{ji} = \frac{1}{N} \sum_{j \in V_{-i}} \partial_{\rho z_1^2} m_{ij}, \\
\partial_{(\gamma_i^R)^2} \partial_\rho \mathcal{M} &= \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\rho z_2^2} m_{ij} + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\rho z_1^2} m_{ji} = \frac{1}{N} \sum_{j \in V_{-i}} \partial_{\rho z_1^2} m_{ji}, \\
\partial_{\gamma_i^S \gamma_i^R} \partial_\rho \mathcal{M} &= \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\rho z_1 z_2} m_{ij} + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\rho z_2 z_1} m_{ji} = \frac{1}{N} \sum_{j \in V_{-i}} \partial_{\rho z_1 z_2} m_{ij}.
\end{aligned}
$$

Computing the derivatives and re-writing using Lemma C.9 gives

$$
\begin{aligned}
\partial_{\rho z_1^2} m_{ij} =& \big(\partial_{z_1^2} J_{ij}\big)(Z_{ij} - r_{ij}) - 2\big(\partial_{z_1} J_{ij}\big)\big(\partial_{z_1} r_{ij}\big) - J_{ij}\big(\partial_{z_1^2} r_{ij}\big) \\
\partial_{\rho z_1 z_2} m_{ij} =& \big(\partial_{z_1 z_2} J_{ij}\big)(Z_{ij} - r_{ij}) - \big(\partial_{z_1} J_{ij}\big)\big(\partial_{z_2} r_{ij}\big) - \big(\partial_{z_2} J_{ij}\big)\big(\partial_{z_1} r_{ij}\big) - J_{ij}\big(\partial_{z_1 z_2} r_{ij}\big) \\
=& \big(\partial_{z_1 z_2} J_{ij}\big)(Z_{ij} - r_{ij}) - \big(\partial_{z_1} J_{ij}\big)\big(\partial_{z_1} r_{ji}\big) - \big(\partial_{z_1} J_{ji}\big)\big(\partial_{z_1} r_{ij}\big) - J_{ij}\big(\partial_{z_1 z_2} r_{ij}\big)
\end{aligned}
$$

and therefore

$$
\begin{aligned}
(D_{SS})_{ii} =& -\frac{1}{N} \sum_{j \in V_{-i}} \Big\{ 2\bar{\mathbb{E}}\big[\big(\partial_{z_1} J_{ij}\big)\big(\partial_{z_1} r_{ij}\big)\big] + \bar{\mathbb{E}}\big[J_{ij}\big(\partial_{z_1^2} r_{ij}\big)\big] \Big\} \\
(D_{RR})_{ii} =& -\frac{1}{N} \sum_{j \in V_{-i}} \Big\{ 2\bar{\mathbb{E}}\big[\big(\partial_{z_1} J_{ji}\big)\big(\partial_{z_1} r_{ji}\big)\big] + \bar{\mathbb{E}}\big[J_{ji}\big(\partial_{z_1^2} r_{ji}\big)\big] \Big\} \\
(D_{SR})_{ii} =& -\frac{1}{N} \sum_{j \in V_{-i}} \Big\{ \bar{\mathbb{E}}\big[\big(\partial_{z_1} J_{ij}\big)\big(\partial_{z_1} r_{ji}\big)\big] + \bar{\mathbb{E}}\big[\big(\partial_{z_1} J_{ji}\big)\big(\partial_{z_1} r_{ij}\big)\big] + \bar{\mathbb{E}}\big[J_{ij}\big(\partial_{z_1 z_2} r_{ij}\big)\big] \Big\}.
\end{aligned}
$$

Step 3: behavior of $\frac{1}{2}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)'\big\{ \sum_{g=1}^{\dim(\gamma)} \partial_{\gamma\gamma'\gamma_g} \bar{\mathcal{L}}\big[\bar{\mathcal{H}}^{-1} \partial_{\rho\gamma'} \bar{\mathcal{M}}\big]_g \big\}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)$.

Following the argument in the proof of Theorem C.1 part(ii) in FVW and letting C denote the $N \times N$ matrix with elements $(C)_{ij} = \Omega_{ij}\bar{\mathbb{E}}\big(\partial_{z^3}\ell_{ij}\big)$ and

$$\mathcal{C} = \frac{1}{N}\begin{bmatrix}\operatorname{diag}(C\iota_N) & C \\ C' & \operatorname{diag}(C'\iota_N)\end{bmatrix}$$

gives

$$\frac{1}{2}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)'\left\{\sum_{g=1}^{\dim(\gamma)}\partial_{\gamma\gamma'\gamma_g}\bar{\mathcal{L}}\big[\bar{\mathcal{H}}^{-1}\partial_{\rho\gamma'}\bar{\mathcal{M}}\big]_g\right\}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big) = -\frac{1}{2N}\sum_{i\in V}\sum_{j\in V_{-i}}\Lambda_{ij}^2\Omega_{ij}\bar{\mathbb{E}}\big(\partial_{z^3}\ell_{ij}\big).$$

Lemma S.8(iii) in FVW yields

$$-\frac{1}{2N}\sum_{i\in V}\sum_{j\in V_{-i}}\Lambda_{ij}^2\Omega_{ij}\bar{\mathbb{E}}\big(\partial_{z^3}\ell_{ij}\big) = -\frac{1}{2}(\iota_N'\Upsilon', \iota_N'\Upsilon)\bar{\mathcal{H}}^{-1}\mathcal{C}\bar{\mathcal{H}}^{-1}\begin{pmatrix}\Upsilon\iota_N \\ \Upsilon'\iota_N\end{pmatrix}.$$

By Lemma C.12 the right-hand side of the preceding equation is equivalent to

$$\begin{aligned}
&-\frac{1}{2N}\iota_N'\Upsilon'(\bar{\mathcal{H}}^*_{SS})^{-1}\big(N^{-1}\operatorname{diag}(C\iota_N)\big)(\bar{\mathcal{H}}^*_{SS})^{-1}\Upsilon\iota_N \\
&-\frac{1}{2}\iota_N'\Upsilon(\bar{\mathcal{H}}^*_{RR})^{-1}\big(N^{-1}\operatorname{diag}(C'\iota_N)\big)(\bar{\mathcal{H}}^*_{RR})^{-1}\Upsilon'\iota_N + O_p\big(N^{-1}\big) \\
=&-\frac{1}{2N^2}\sum_{i\in V}\left\{[(\bar{\mathcal{H}}^*_{SS})^{-1}]_{ii}^2\left(N^{-1}\sum_{j\in V_{-i}}\Omega_{ij}\bar{\mathbb{E}}[\partial_{z^3}\ell_{ij}]\right)\sum_{j\in V_{-i}}\bar{\mathbb{E}}\big[(\partial_z\ell_{ij})^2\big]\right\} \\
&-\frac{1}{2N^2}\sum_{j\in V}\left\{[(\bar{\mathcal{H}}^*_{RR})^{-1}]_{jj}^2\left(N^{-1}\sum_{i\in V_{-j}}\Omega_{ij}\bar{\mathbb{E}}[\partial_{z^3}\ell_{ij}]\right)\sum_{i\in V_{-j}}\bar{\mathbb{E}}\big[(\partial_z\ell_{ij})^2\big]\right\} + O_p\big(N^{-1/2}\big).
\end{aligned}$$

By the definition of $\bar{\mathcal{H}}^*_{SS}$ and a Bartlett equality

$$\begin{aligned}
&-\frac{1}{2N^2}\sum_{i\in V}\left\{[(\bar{\mathcal{H}}^*_{SS})^{-1}]_{ii}^2\left(N^{-1}\sum_{j\in V_{-i}}\Omega_{ij}\bar{\mathbb{E}}[\partial_{z^3}\ell_{ij}]\right)\sum_{j\in V_{-i}}\bar{\mathbb{E}}\big[(\partial_z\ell_{ij})^2\big]\right\} \\
=&-\frac{1}{2N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}\Omega_{ij}\bar{\mathbb{E}}[\partial_{z^3}\ell_{ij}]}{\sum_{j\in V_{-i}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]} \\
=&\frac{1}{2N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}\Omega_{ij}\big\{2(\partial_z p_{ij})(\partial_z H_{ij}) + H_{ij}(\partial_{z^2}p_{ij})\big\}}{\sum_{j\in V_{-i}}H_{ij}(\partial_z p_{ij})},
\end{aligned}$$

where we use

$$\partial_{z^3}\ell_{ij} = \partial_{z^2}H_{ij}(Y_{ij} - p_{ij}) - H_{ij}(\partial_{z^2}p_{ij}) - 2(\partial_z H_{ij})(\partial_z p_{ij}).$$

Similarly,

$$
\begin{aligned}
&-\frac{1}{2N^2}\sum_{j\in V}\left\{[(\bar{\mathcal{H}}^*_{RR})^{-1}]^2_{jj}\left(N^{-1}\sum_{i\in V_{-j}}\Omega_{ij}\bar{\mathbb{E}}[\partial_{z^3}\ell_{ij}]\right)\sum_{i\in V_{-j}}\bar{\mathbb{E}}\big[(\partial_z\ell_{ij})^2\big]\right\}\\
=&\frac{1}{2N}\sum_{j\in V}\frac{\sum_{i\in V_{-j}}\Omega_{ij}\big\{2(\partial_z p_{ij})(\partial_z H_{ij})+H_{ij}(\partial_{z^2}p_{ij})\big\}}{\sum_{i\in V_{-j}}H_{ij}(\partial_z p_{ij})}.
\end{aligned}
$$

□

Lemma C.3 (Behavior of stochastic part of $\hat{\rho}$)**.** *Under Assumption 1*

$$
\begin{aligned}
\partial_{\rho^2}\mathcal{M}=&-\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}}J_{ij}(\partial_\rho r_{ij})+O_p(1),\\
\partial_{\rho\theta'}\mathcal{M}+(\partial_{\rho\gamma'}\mathcal{M})\bar{\mathcal{H}}^{-1}(\partial_{\gamma\theta'}\bar{\mathcal{L}})=&-\frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}}J_{ij}(\partial_{z_1}r_{ij})\tilde{X}'_{ij}+O_p(1),\\
(\partial_{\rho\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\mathcal{S}=&-\frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}}\Omega_{ij}H_{ij}(Y_{ij}-p_{ij}).
\end{aligned}
$$

Proof. We have

$$
\partial_{\rho^2}\mathcal{M}=\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}}\Big\{\partial_\rho J_{ij}(Z_{ij}-r_{ij})-J_{ij}(\partial_\rho r_{ij})\Big\}.
$$

It is easy to see that

$$
\bar{\mathbb{E}}\left[\left(\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}}\partial_\rho J_{ij}(Z_{ij}-r_{ij})\right)^2\right]=O_p(1)
$$

and therefore

$$
\partial_{\rho^2}\mathcal{M}=-\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}}J_{ij}(\partial_\rho r_{ij})+O_p(1).
$$

Arguing similarly we get

$$\begin{aligned}\partial_{\rho\theta'}\mathcal{M} &= -\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}} J_{ij}(\partial_{\theta'} r_{ij}) + O_p(1)\\ &= -\frac{1}{N}\sum_{\substack{i,j\in V\\ i<j}} J_{ij}\big\{(\partial_{z_1} r_{ij})X'_{ij} + (\partial_{z_2} r_{ij})X'_{ji}\big\} + O_p(1)\\ &= -\frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} J_{ij}(\partial_{z_1} r_{ij})X'_{ij} + O_p(1),\end{aligned}$$

where the first equality is by the chain rule for derivatives and the second equality follows from Lemma C.9 and symmetry of J_{ij}. For $k = 1, \ldots, \dim(\theta)$ let

$$\Xi_{ij,k} = -\frac{1}{N}\sum_{k_1\in V}\sum_{k_2\in V_{-k_1}}\left(\bar{\mathcal{H}}^{-1}_{SS,ik_1} + \bar{\mathcal{H}}^{-1}_{RS,jk_1} + \bar{\mathcal{H}}^{-1}_{SR,ik_2} + \bar{\mathcal{H}}^{-1}_{RR,jk_2}\right)\bar{\mathbb{E}}(\partial_{\theta_k z}\ell_{k_1k_2}).$$

and let $\Xi_{ij} = (\Xi_{ij,1}, \ldots, \Xi_{ij,\dim(\theta)})'$. By Lemma S.8(i) in FVW and Lemma C.11

$$(\partial_{\rho\gamma'}\mathcal{M})\bar{\mathcal{H}}^{-1}(\partial_{\gamma\theta'}\bar{\mathcal{L}}) = \frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} J_{ij}(\partial_{z_1} r_{ij})\Xi'_{ij} + O_p(1).$$

Straightforward calculations give $X_{ij} - \Xi_{ij} = \tilde{X}_{ij}$ so that

$$\partial_{\rho\theta}\mathcal{M} + (\partial_{\rho\gamma'}\mathcal{M})\bar{\mathcal{H}}^{-1}(\partial_{\gamma\theta'}\bar{\mathcal{L}}) = -\frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} J_{ij}(\partial_{z_1} r_{ij})\tilde{X}'_{ij} + O_p(1).$$

Lemma S.8(i) in FVW in conjunction with Lemma C.11 gives

$$(\partial_{\rho\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\mathcal{S} = -\frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}} \Omega_{ij}H_{ij}(Y_{ij} - p_{ij}).$$

□

Lemma C.4 (Stochastic expansion of $s_N(\hat{\boldsymbol{\gamma}}, \hat{\theta})$)**.** *Under Assumption 1*

$$\begin{aligned}&s_N(\hat{\boldsymbol{\gamma}}, \hat{\theta}) - s_N(\boldsymbol{\gamma}^0, \theta^0)\\ &= \left\{(\partial_{\theta'} s_N) + (\partial_{\boldsymbol{\gamma}'} s_N)\,\bar{\mathcal{H}}^{-1}\left[\partial_{\boldsymbol{\gamma}\theta'}\bar{\mathcal{L}}\right]\right\}(\hat{\theta} - \theta^0) + (\partial_{\boldsymbol{\gamma}'} s_N)\,\bar{\mathcal{H}}^{-1}\mathcal{S} + B_N^{S,*} + o_p\big(1\big),\end{aligned}$$

where

$$
\begin{aligned}
B_N^{S,*} =& - (\partial_{\boldsymbol{\gamma}'} s_N)\, \bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} + \frac{1}{2}(\partial_{\boldsymbol{\gamma}'} s_N) \sum_{g=1}^{\dim(\boldsymbol{\gamma})} \left[\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\gamma}_g}\bar{\mathcal{L}}\right] \bar{\mathcal{H}}^{-1}\mathcal{S}\left[\bar{\mathcal{H}}^{-1}\mathcal{S}\right]_g \\
&+ \frac{1}{2}\left(\bar{\mathcal{H}}^{-1}\mathcal{S}\right)' \left[\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'} s_N\right]\left(\bar{\mathcal{H}}^{-1}\mathcal{S}\right).
\end{aligned}
$$

Proof. In the following set $q = 8$. By straightforward Taylor expansions

$$
\begin{aligned}
s_N(\hat{\boldsymbol{\gamma}}, \hat{\theta}) - s_N(\boldsymbol{\gamma}^0, \theta^0) =& \partial_{\theta'} s_N(\boldsymbol{\gamma}^0, \theta^0)(\hat{\theta} - \theta^0) + \partial_{\boldsymbol{\gamma}'} s_N(\boldsymbol{\gamma}^0, \theta^0)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) \\
&+ \frac{1}{2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)' \left(\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'} s_N(\boldsymbol{\gamma}^0, \theta^0)\right)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) + R_{1,N},
\end{aligned}
$$

where

$$
\begin{aligned}
R_{1,N} =& (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)' \left(\partial_{\theta\boldsymbol{\gamma}'} s_N(\bar{\boldsymbol{\gamma}}, \theta^0)\right)(\hat{\theta} - \theta^0) + \frac{1}{2}(\hat{\theta} - \theta^0)' \left(\partial_{\theta\theta'} s_N(\hat{\boldsymbol{\gamma}}, \bar{\theta})\right)(\hat{\theta} - \theta^0) \\
&+ \frac{1}{6}\sum_{g=1}^{\dim(\boldsymbol{\gamma})} (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)' [\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\gamma}_g} s_N(\boldsymbol{\gamma}^*, \theta^0)](\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)[\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0]_g.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\sup_{\theta \in \Theta, \boldsymbol{\gamma} \in \Gamma} \left\|\partial_{\theta\theta'} s_N(\boldsymbol{\gamma}, \theta)\right\|_2 &= O_p(N), \\
\left\|\partial_{\theta\boldsymbol{\gamma}'} s_N(\boldsymbol{\gamma}^0, \theta^0)\right\|_q &= O_p\left(N^{\frac{1}{q}}\right), \\
\sup_{\boldsymbol{\gamma} \in \Gamma} \left\|\partial_{\boldsymbol{\gamma}^3} s_N(\boldsymbol{\gamma}, \theta^0)\right\|_q &= O_p(1), \\
\left\|\partial_{\boldsymbol{\gamma}} s_N(\boldsymbol{\gamma}^0, \theta^0)\right\|_2 &= O_p\left(N^{\frac{1}{2}}\right),
\end{aligned}
$$

where $\bar{\theta}$ and $\boldsymbol{\gamma}^*$ are intermediate values. The first equality follows by inspection and the other equalities follow from Lemma C.8. Therefore, since $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\|_q = O_p\left(N^{-\frac{1}{2}+\frac{1}{q}}\right)$ we have

$$
\begin{aligned}
|R_{1,N}| \leq& N^{1-\frac{2}{q}} \left\|\partial_{\theta\boldsymbol{\gamma}'} s_N(\bar{\boldsymbol{\gamma}}, \theta^0)\right\|_q \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\|_q \left\|\hat{\theta} - \theta^0\right\|_2 + \frac{1}{2}\left\|\partial_{\theta\theta'} s_N(\hat{\boldsymbol{\gamma}}, \bar{\theta})\right\|_2 \left\|\hat{\theta} - \theta^0\right\|_q^2 \\
&+ \frac{1}{6} N^{1-\frac{2}{q}} \left\|\partial_{\boldsymbol{\gamma}^3} s_N(\boldsymbol{\gamma}^*, \theta^0)\right\|_q \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\|_q^3 = O_p\left(N^{-\frac{1}{2}+\frac{1}{q}}\right) = o_p(1).
\end{aligned}
$$

From now on, drop the arguments of s_N and its derivatives whenever they are evaulated at their true values. Then,

$$\begin{aligned}
s_N(\hat{\boldsymbol{\gamma}},\hat{\theta}) - s_N(\boldsymbol{\gamma}^0,\theta^0) = &\left(\partial_{\theta'} s_N\right)(\hat{\theta}-\theta^0) + \left(\partial_{\boldsymbol{\gamma}'} s_N\right)\mathcal{H}^{-1}\mathcal{S} \\
&+ \left(\partial_{\boldsymbol{\gamma}'} s_N\right)\mathcal{H}^{-1}\left[\partial_{\boldsymbol{\gamma}\theta'}\mathcal{L}\right](\hat{\theta}-\theta^0) \\
&+ \frac{1}{2}\left(\partial_{\boldsymbol{\gamma}'} s_N\right)\sum_{g=1}^{\dim(\boldsymbol{\gamma})}\left[\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'\gamma_g}\mathcal{L}\right]\mathcal{H}^{-1}\mathcal{S}\left[\mathcal{H}^{-1}\mathcal{S}\right]_g \\
&+ \frac{1}{2}\left(\mathcal{H}^{-1}\mathcal{S}\right)'\left[\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'} s_N\right]\left(\mathcal{H}^{-1}\mathcal{S}\right) + R_{2,N},
\end{aligned}$$

with

$$R_{2,N} = \left(\partial_{\boldsymbol{\gamma}'} s_N\right) R_{\boldsymbol{\gamma}}(\hat{\theta}) + \frac{1}{2}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0 - \mathcal{H}^{-1}\mathcal{S}\right)'\left[\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'} s_N\right]\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0 + \mathcal{H}^{-1}\mathcal{S}\right),$$

where $R_{\boldsymbol{\gamma}}(\hat{\theta})$ is the remainder term from Theorem B.1 in FVW (compare also proof of Lemma C.7). By Lemma C.8,

$$\left\|\partial_{\boldsymbol{\gamma}'} s_N\right\|_q = O_p\left(N^{\frac{1}{q}}\right) \quad \text{and} \quad \left\|\partial_{\boldsymbol{\gamma}^3} s_N\right\|_q = O_p(1).$$

Noting that $\left\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0 - \mathcal{H}^{-1}\mathcal{S}\right\|_q = O_p\left(N^{-1+2/q}\right)$,

$$\begin{aligned}
|R_{N,2}| \leq &N^{1-2/q}\left\|\partial_{\boldsymbol{\gamma}'} s_N\right\|_q\left\|R_{\boldsymbol{\gamma}}(\hat{\theta})\right\|_q \\
&+ \frac{1}{2}N^{1-2/q}\left\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0 - \mathcal{H}^{-1}\mathcal{S}\right\|_q\left(\left\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right\|_q + \left\|\mathcal{H}^{-1}\mathcal{S}\right\|_q\right)\left\|\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'} s_N\right\|_q \\
\leq &(1 + N\|\hat{\theta}-\theta^0\|)N^{1-1/q}\frac{\left\|R_{\boldsymbol{\gamma}}(\hat{\theta})\right\|_q}{1 + N\|\hat{\theta}-\theta^0\|} + O_p\left(N^{-1/2+1/q}\right) = o_p(1).
\end{aligned}$$

Following closely the proof of Lemma C.1 it is now easy to prove the assertion of the lemma. □

Lemma C.5 (Behavior of deterministic part of T_N)**.** *Suppose that Assumption 1 holds. For $B_N^{S,*}$ in the statement of Lemma C.4 we have*

$$B_N^{S,*} = B_N^S + O_p\left(N^{-1/2}\right),$$

where B_N^S is given in Theorem 3.

Proof. Tedious calculations yield

$$\partial_{\gamma_i^S}\Big\{s_N(\boldsymbol{\gamma}^0,\theta^0)\Big\} = \frac{1}{N}\sum_{j\in V_{-i}}(\partial_z p_{ij})H_{ij}\boldsymbol{\beta}_{ij}^N,$$
$$\partial_{\gamma_i^R}\Big\{s_N(\boldsymbol{\gamma}^0,\theta^0)\Big\} = \frac{1}{N}\sum_{j\in V_{-i}}(\partial_z p_{ji})H_{ji}\boldsymbol{\beta}_{ji}^N.$$

This implies that

$$\partial_{\boldsymbol{\gamma}} s_N(\boldsymbol{\gamma}^0,\theta^0) = \frac{1}{N}\begin{bmatrix} A\iota_N \\ A'\iota_N \end{bmatrix}$$

for a $N\times N$ matrix A with entries

$$(A)_{ij} = \begin{cases} \omega_{ij}\boldsymbol{\beta}_{ij}^N & \text{for } i\neq j \\ 0 & \text{for } i=j \end{cases}.$$

We will exploit this representation in our projection arguments below.
Step 1: behavior of $-(\partial_{\boldsymbol{\gamma}} s_N)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}$.
Let

$$\Lambda_{ij} = -\frac{1}{N}\sum_{k\in V}\sum_{l\in V_{-k}}\Big(\bar{\mathcal{H}}^{-1}_{SS,ik}+\bar{\mathcal{H}}^{-1}_{RS,jk}+\bar{\mathcal{H}}^{-1}_{SR,il}+\bar{\mathcal{H}}^{-1}_{RR,jl}\Big)(\partial_z\ell_{kl}).$$

By Lemma S.8(iii) of FVW

$$-(\partial_{\boldsymbol{\gamma}} s_N)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} = \frac{1}{N}\sum_{i\in V}\sum_{j\in V_{-i}}\Lambda_{ij}(\mathcal{P}\boldsymbol{\beta}^N)_{ij}\Big\{\partial_{z^2}\ell_{ij}-\bar{\mathbb{E}}\,\partial_{z^2}\ell_{ij}\Big\}.$$

Following similar arguments as in the proof of Lemma C.2 it can then be shown that

$$\begin{aligned}
-(\partial_{\boldsymbol{\gamma}} s_N)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} &= -\frac{1}{N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}(\mathcal{P}\boldsymbol{\beta}^N)_{ij}\bar{\mathbb{E}}\big[\partial_z\ell_{ij}(\partial_{z^2}\ell_{ij})\big]}{\sum_{j\in V_{-i}}\bar{\mathbb{E}}(-\partial_{z^2}\ell_{ij})} + O_p\big(N^{-1/2}\big) \\
&= -\frac{1}{N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}(\mathcal{P}\boldsymbol{\beta}^N)_{ij}(\partial_z H_{ij})(\partial_z p_{ij})}{\sum_{j\in V_{-i}}\omega_{ij}} + O_p\big(N^{-1/2}\big).
\end{aligned}$$

Step 2: behavior of $\frac{1}{2}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)'(\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'}s_N)\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)$.

We have

$$\begin{aligned}
\partial_{(\gamma_i^S)^2}\Big\{s_N(\boldsymbol{\gamma}^0,\theta^0)\Big\} =& \frac{1}{N}\sum_{j\in V_{-i}}(\partial_{z^2}p_{ij})H_{ij}\boldsymbol{\beta}_{ij}^N, \\
&+\frac{1}{N^2}\sum_{j\in V_{-i}}\sum_{k\in V_{-\{i,j\}}}(\partial_z p_{ij})(\partial_z p_{ik})\left[p_{jk}+p_{kj}\right], \\
\partial_{(\gamma_i^R)^2}\Big\{s_N(\boldsymbol{\gamma}^0,\theta^0)\Big\} =& \frac{1}{N}\sum_{j\in V_{-i}}(\partial_{z^2}p_{ji})H_{ji}\boldsymbol{\beta}_{ji}^N \\
&+\frac{1}{N^2}\sum_{j\in V_{-i}}\sum_{k\in V_{-\{i,j\}}}(\partial_z p_{ji})(\partial_z p_{ki})\left[p_{jk}+p_{kj}\right], \\
\partial_{\gamma_i^S\gamma_i^R}\Big\{s_N(\boldsymbol{\gamma}^0,\theta^0)\Big\} =& \frac{1}{N^2}\sum_{j\in V_{-i}}\sum_{k\in V_{-\{i,k\}}}(\partial_z p_{ij})(\partial_z p_{ki})p_{kj}.
\end{aligned}$$

Moreover, for all $i\neq j$ the "cross derivatives" $\partial_{\gamma_i^S\gamma_j^S}(s_N)$, $\partial_{\gamma_i^R\gamma_j^R}(s_N)$ and $\partial_{\gamma_i^S\gamma_j^R}(s_N)$ are bounded by N times a universal constant. This implies that

$$\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'}s_N(\boldsymbol{\gamma}^0,\theta^0) = \begin{bmatrix} D_{SS}+M_{SS} & D_{SR}+M_{SR} \\ D_{SR}'+M_{SR}' & D_{RR}+M_{RR}\end{bmatrix},$$

where D_{SS} is a diagonal matrix with entries $\big(\partial_{(\gamma_i^S)^2}s_N\big)_{i\in V}$, D_{RR} is a diagonal matrix with entries $\big(\partial_{(\gamma_i^R)^2}s_N\big)_{i\in V}$, and D_{SR} is a diagonal matrix with entries $\big(\partial_{\gamma_i^S\gamma_i^R}s_N\big)_{i\in V}$. The matrices M_{SS}, M_{RR} and M_{SR} are off-diagonal matrices that are bounded in terms of the $\|\cdot\|_{\max}$-norm. Arguing similarly as in Lemma C.2 it can now be shown that

$$\begin{aligned}
&\frac{1}{2}(\bar{\mathcal{H}}^{-1}\mathcal{S})'\Big(\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'}s_N\Big)\bar{\mathcal{H}}^{-1}\mathcal{S} \\
=&\frac{1}{2N}\sum_{i\in V}\frac{[D_{SS}]_{ii}}{\Big(N^{-1}\sum_{j\in V_{-i}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]\Big)} \\
&+\frac{1}{2N}\sum_{j\in V}\frac{[D_{RR}]_{jj}}{\Big(N^{-1}\sum_{i\in V_{-j}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]\Big)} \\
&+\frac{1}{N}\sum_{i\in V}\frac{(D_{SR})_{ii}\,\overline{\mathrm{corr}}_i}{\Big(N^{-1}\sum_{j=1}^N\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]\Big)^{1/2}\Big(N^{-1}\sum_{j=1}^N\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ji}]\Big)^{1/2}}+O_p\big(N^{3/2}\big).
\end{aligned}$$

Step 3: behavior of $\frac{1}{2}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)'\big\{\sum_{g=1}^{\dim(\boldsymbol{\gamma})}\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'\gamma_g}\bar{\mathcal{L}}\big[\bar{\mathcal{H}}^{-1}\partial_{\gamma}s_N\big]_g\big\}\big(\bar{\mathcal{H}}^{-1}\mathcal{S}\big)$.

Following the arguments in the proof of Lemma C.2 yields

$$\begin{aligned}
&\frac{1}{2}(\bar{\mathcal{H}}^{-1}\mathcal{S})'\{\sum_{g=1}^{\dim(\boldsymbol{\gamma})}\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\gamma}_g}\bar{\mathcal{L}}\big[\bar{\mathcal{H}}^{-1}\partial_{\gamma}s_N\big]_g\}(\bar{\mathcal{H}}^{-1}\mathcal{S})\\
=&-\frac{1}{2N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}(\mathcal{P}\boldsymbol{\beta}^N)_{ij}\bar{\mathbb{E}}[\partial_{z^3}\ell_{ij}]}{\sum_{j\in V_{-i}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]}-\frac{1}{2N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}(\mathcal{P}\boldsymbol{\beta}^N)_{ij}\bar{\mathbb{E}}[\partial_{z^3}\ell_{ij}]}{\sum_{j\in V_{-i}}\bar{\mathbb{E}}[-\partial_{z^2}\ell_{ij}]}\\
=&\frac{1}{2N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}}(\mathcal{P}\boldsymbol{\beta}^N)_{ij}\big\{2(\partial_z p_{ij})(\partial_z H_{ij})+H_{ij}(\partial_{z^2}p_{ij})\big\}}{\sum_{j\in V_{-i}}H_{ij}(\partial_z p_{ij})}\\
&+\frac{1}{2N}\sum_{j\in V}\frac{\sum_{i\in V_{-j}}(\mathcal{P}\boldsymbol{\beta}^N)_{ij}\big\{2(\partial_z p_{ij})(\partial_z H_{ij})+H_{ij}(\partial_{z^2}p_{ij})\big\}}{\sum_{i\in V_{-j}}H_{ij}(\partial_z p_{ij})}.
\end{aligned}$$

□

Lemma C.6 (Linear representation of T_N^{oracle})**.** *Under Assumption 1*

$$S_N-\bar{\mathbb{E}}\,S_N=N\sum_{ij\in E(N)}\boldsymbol{\beta}_{ij}^N H_{ij}(Y_{ij}-p_{ij})+o_p\Big(\sqrt{\bar{\mathrm{var}}(S_N)}\Big)$$

and

$$\bar{\mathrm{var}}(S_N)=N^2\sum_{ij\in E(N)}\Big\{p_{1,ij}\Big(H_{ij}\boldsymbol{\beta}_{ij}^N\Big)^2+\tilde{\rho}_{ij}\sqrt{p_{1,ij}p_{1,ji}}\Big(H_{ij}H_{ji}\boldsymbol{\beta}_{ij}^N\boldsymbol{\beta}_{ji}^N\Big)\Big\}+O_p\big(N^3\big).$$

Proof. We start by computing the conditional variance of S_N. Since triangles β and β' are conditionally independent provided that $V(\beta)\cap V(\beta')=\emptyset$ we have

$$\bar{\mathbb{E}}\left[(A_\beta-\bar{\mathbb{E}}A_\beta)(A_{\beta'}-\bar{\mathbb{E}}A_{\beta'})\right]=0$$

for such triangles. Now,

$$\begin{aligned}
\bar{\text{var}}(S_N) =& \bar{\mathbb{E}}\Big(\sum_{\beta \in B(N)} \big(A_\beta - \bar{\mathbb{E}} A_\beta\big)\Big)^2 \\
=& \sum_{\substack{\beta,\beta' \in B^2(N) \\ |V(\beta)\cap V(\beta')|=2}} \big(A_\beta - \bar{\mathbb{E}} A_\beta\big)\big(A_{\beta'} - \bar{\mathbb{E}} A_{\beta'}\big) + H_N \\
=& \sum_{ij \in E(N)} \Bigg(\sum_{\substack{\beta,\beta' \ni ij \\ |V(\beta)\cap V(\beta')|=2}} p^T_{-ij}(\beta) p^T_{-ij}(\beta') \bar{\mathbb{E}}\left[(Y_{ij} - p_{ij})^2\right] \\
& \quad + \sum_{\substack{\beta \in ij, \beta' \ni ji \\ |V(\beta)\cap V(\beta')|=2}} p^T_{-ij}(\beta) p^T_{-ji}(\beta') \bar{\mathbb{E}}\left[(Y_{ij} - p_{ij})(Y_{ji} - p_{ji})\right] \Bigg) + H_N \\
=& \sum_{ij \in E(N)} \Bigg(p_{1,ij} \Big(\sum_{\beta \ni ij} p^T_{-ij}(\beta) \Big)^2 \\
& \quad + \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \Big(\sum_{\beta \ni ij} p^T_{-ij}(\beta) \Big) \Big(\sum_{\beta \ni ji} p^T_{-ji}(\beta) \Big) \Bigg) + H^*_N
\end{aligned}$$

where H_N is the contribution of triangle pairs that share the same vertex set and

$$H^*_N = H_N + \sum_{ij \in E(N)} \Delta_{N,ij}$$

with

$$\begin{aligned}
\Delta_{N,ij} =& p_{1,ij} \sum_{\substack{\beta,\beta' \ni ij \\ |V(\beta)\cap V(\beta')|=2}} p^T_{-ij}(\beta) p^T_{-ij}(\beta') + \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \sum_{\substack{\beta,\beta' \ni ij \\ |V(\beta)\cap V(\beta')|=2}} p^T_{-ij}(\beta) p^T_{-ji}(\beta') \\
& - p_{1,ij} \Big(\sum_{\beta \ni ij} p^T_{-ij}(\beta) \Big)^2 - \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \Big(\sum_{\beta \ni ij} p^T_{-ij}(\beta) \Big) \Big(\sum_{\beta \ni ji} p^T_{-ji}(\beta) \Big)
\end{aligned}$$

Clearly, H_N is of order $O_p(N^3)$. Each $\Delta_{N,ij}$ can be bounded by N times a universal constant. Therefore, H^*_N is $O_p(N^3)$ as well. The assumption of non-vanishing linking probabilities ensures that $\bar{\text{var}}(S_N) \asymp N^4$. We now consider the Hajek projection $\hat{S}_N$ of $S_N - \bar{\mathbb{E}} S_N$ onto dyads, i.e.

$$\hat{S}_N = \sum_{ij \in V^2(N)} \bar{\mathbb{E}}\big[(S_N - \bar{\mathbb{E}}\, S_N) \mid Y_{ij}, Y_{ji}\big]$$
$$= \sum_{\beta \in B(N)} \sum_{ij \in E(N)} \bar{\mathbb{E}}\big[(A_\beta - \bar{\mathbb{E}}\, A_\beta) \mid Y_{ij}\big]$$
$$= \sum_{ij \in E(N)} \Big\{(Y_{ij} - p_{ij}) \sum_{\beta \ni ij} p^T_{-ij}(\beta)\Big\}.$$

Here, the second equality uses that every transitive triangle β may contain the link ij or ji but not both. We now compute the conditional variance of $\hat{S}_N$:

$$\bar{\text{var}}(\hat{S}_N) = \sum_{ij \in V^2(N)} \bar{\mathbb{E}}\Big\{(Y_{ij} - p_{ij}) \sum_{\beta \ni ij} p^T_{-ij}(\beta) + (Y_{ji} - p_{ji}) \sum_{\beta \ni ji} p^T_{-ji}(\beta)\Big\}^2$$
$$= \sum_{ij \in E(N)} \Big\{p_{1,ij}\Big(\sum_{\beta \ni ij} p^T_{-ij}(\beta)\Big)^2 + \tilde{\rho}_{ij}\sqrt{p_{1,ij}p_{1,ji}}\Big(\sum_{\beta \ni ij} p^T_{-ij}(\beta)\Big)\Big(\sum_{\beta \ni ji} p^T_{-ji}(\beta)\Big)\Big\}.$$

From the previous results it is easy to see that

$$\frac{\bar{\text{var}}(S_N - \bar{\mathbb{E}} S_N)}{\bar{\text{var}}(\hat{S}_N)} = \frac{\bar{\text{var}}(S_N)}{\bar{\text{var}}(\hat{S}_N)} \to 1.$$

We now apply a conditional version of Theorem 11.2 in van der Vaart (2000). To prove the conditional version of the theorem simply replace the convergence in squared mean argument in the proof given in van der Vaart (2000) by an analogous squared condtional mean argument. It follows that

$$S_N - \bar{\mathbb{E}} S_N = \hat{S}_N + o_p\left(\sqrt{\bar{\text{var}}(S_N)}\right) = \hat{S}_N + o_p(N^2).$$

□

Lemma C.7 (Sparse bounded functionals of the incidental parameter I)**.** *Let K denote a finite constant and let $(\boldsymbol{\pi}_k)_{k=1}^K$ denote a collection of N-dimensional parameters. Define $\boldsymbol{\pi} = (\boldsymbol{\pi}'_1, \ldots, \boldsymbol{\pi}'_K)'$. Let $\{g_{ij}\}_{i<j}$ denote an an array of functions such that*

$$g_{ij}(\boldsymbol{\pi}) = g_{ij}(\pi_{1,i}, \ldots, \pi_{K,i}, \pi_{1,j}, \ldots, \pi_{K,j})$$

with

$$\|\partial_{\boldsymbol{\pi}^\ell} g_{ij}(\boldsymbol{\pi})\|_{\max} \le C \quad \textit{for } \ell = 0, 1, 2, 3$$

for a universal constant C. Let

$$g(\boldsymbol{\pi}) = \frac{1}{N}\sum_{i<j} g_{ij}(\boldsymbol{\pi}).$$

Then

(i) $\left\|\partial_{\boldsymbol{\pi}} g(\boldsymbol{\pi})\right\|_q = O_p\left(N^{\frac{1}{q}}\right)$*, moreover, if*

$$\frac{1}{N}\sum_i \left|\frac{1}{\sqrt{N}}\sum_{\substack{j\in V\\ j>i}} \partial_{\pi_{k,1}} g_{ij}(\boldsymbol{\pi})\right|^q + \frac{1}{N}\sum_i \left|\frac{1}{\sqrt{N}}\sum_{\substack{j\in V\\ j<i}} \partial_{\pi_{k,2}} g_{ji}(\boldsymbol{\pi})\right|^q = O_p(1)$$

then $\left\|\partial_{\boldsymbol{\pi}} g(\boldsymbol{\pi})\right\|_q = O_p\left(N^{-\frac{1}{2}+\frac{1}{q}}\right)$.

(ii) Moreover, $\left\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\right\|_q = O_p(1)$ *and* $\left\|\partial_{\boldsymbol{\pi}^3} g(\boldsymbol{\pi})\right\|_q = O_p(1)$.

(iii) For a σ-field $\mathcal{A}$ let $\bar{\mathbb{E}} = E[\cdot \mid \mathcal{A}]$. Suppose that conditional on $\mathcal{A}$ the elements of the arrays $(\partial_{\pi_{i,1}\pi_{j,2}} g_{ij})_{i,j=1}^N$ and $(\partial_{\pi_{i,2}\pi_{j,1}} g_{ji})_{i,j=1}^N$ are independent. Then

$$\left\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}\big[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\big]\right\| = O_p\left(N^{-3/8}\right).$$

Proof. First proof (i). Note that

$$\partial_{\pi_{k,i}} g(\boldsymbol{\pi}) = \frac{1}{N}\sum_{\substack{j\in V\\ j>i}} \partial_{\pi_{k,1}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N}\sum_{\substack{j\in V\\ j<i}} \partial_{\pi_{k,2}} g_{ji}(\boldsymbol{\pi}).$$

By definition of the $\|\cdot\|_q$-norm and the Minkowski inequality,

$$\begin{aligned}
\left\|\partial_{\boldsymbol{\pi}} g(\boldsymbol{\pi})\right\|_q &\leq \left(\sum_{\ell=1}^{KN} \left|\partial_{\boldsymbol{\pi}_\ell} g(\boldsymbol{\pi})\right|^q\right)^{\frac{1}{q}} \\
&\leq \left(\sum_{k=1}^{K}\sum_{i=1}^{N} \left|\frac{1}{N}\sum_{\substack{j\in V\\ j>i}} \partial_{\pi_{k,1}} g_{ij}(\boldsymbol{\pi})\right|^q\right)^{\frac{1}{q}} + \left(\sum_{k=1}^{K}\sum_{i=1}^{N} \left|\frac{1}{N}\sum_{\substack{j\in V\\ j<i}} \partial_{\pi_{k,2}} g_{ji}(\boldsymbol{\pi})\right|^q\right)^{\frac{1}{q}}.
\end{aligned}$$

To prove the first claim of (ii) note that

$$\partial_{\pi_{k,i}\pi_{\ell,i}} g(\boldsymbol{\pi}) = \frac{1}{N}\sum_{\substack{j\in V\\ j>i}} \partial_{\pi_{k,1}\pi_{\ell,1}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N}\sum_{\substack{j\in V\\ j<i}} \partial_{\pi_{k,2}\pi_{\ell,2}} g_{ji}(\boldsymbol{\pi})$$

and for $j \neq i$

$$\partial_{\pi_{k,i}\pi_{\ell,j}} g(\boldsymbol{\pi}) = \begin{cases} \frac{1}{N}\partial_{\pi_{k,1}\pi_{\ell,2}} g_{ij}(\boldsymbol{\pi}) & \text{for } j > i \\ \frac{1}{N}\partial_{\pi_{k,2}\pi_{\ell,1}} g_{ji}(\boldsymbol{\pi}) & \text{for } j < i \end{cases}.$$

Every element $\boldsymbol{\pi}_s$ of the parameter vector $\boldsymbol{\pi}$ corresponds to a unique $\pi_{k,i}$. Use the notation $\boldsymbol{\pi}_s = \pi_{k(s),i(s)}$. For every $1 \leq s \leq KN$

$$\begin{aligned} \sum_{t=1}^{KN} |[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})]_{s,t}| =& \frac{1}{N}\sum_{\ell=1}^{K} \Bigg| \sum_{\substack{j\in V \\ j>i(s)}} \partial_{\pi_{k(s),1}\pi_{\ell,1}} g_{i(s)j}(\boldsymbol{\pi}) + \sum_{\substack{j\in V \\ j<i(s)}} \partial_{\pi_{k(s),2}\pi_{\ell,2}} g_{ji(s)}(\boldsymbol{\pi}) \Bigg| \\ &+ \frac{1}{N}\sum_{\ell=1}^{K} \Bigg\{ \sum_{\substack{j\in V \\ j>i(s)}} \left|\partial_{\pi_{k(s),1}\pi_{\ell,2}} g_{i(s)j}(\boldsymbol{\pi})\right| + \sum_{\substack{j\in V \\ j<i(s)}} \left|\partial_{\pi_{k(s),2}\pi_{\ell,1}} g_{ji(s)}(\boldsymbol{\pi})\right| \Bigg\} \\ \leq& 2KC. \end{aligned}$$

By the symmetry of partial derivatives

$$\sum_{s=1}^{KN} |[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})]_{s,t}| = \sum_{s=1}^{KN} |[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})]_{t,s}| \leq 2KC.$$

It follows that

$$\begin{aligned} \|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\|_\infty &= \max_{1\leq s\leq KN} \sum_{t=1}^{KN} |[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})]_{s,t}| \leq 2KC \\ \|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\|_1 = \|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})'\|_\infty &= \max_{1\leq t\leq KN} \sum_{s=1}^{KN} |[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})]_{s,t}| \leq 2KC. \end{aligned}$$

By Lemma S.4 in FVW

$$\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\|_q \leq \|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\|_1^{\frac{1}{q}} \|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\|_\infty^{1-\frac{1}{q}} \leq 2KC.$$

Turning to the second claim of (ii) note that for $\{k, \ell, m\} \subset \{1, \ldots, K\}$

$$\partial_{\pi_{k,i}\pi_{\ell,i}\pi_{m,i}} g(\boldsymbol{\pi}) = \frac{1}{N}\sum_{\substack{j\in V \\ j>i}} \partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,1}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N}\sum_{\substack{j\in V \\ j<i}} \partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,2}} g_{ji}(\boldsymbol{\pi})$$

and for $i \neq j$

$$\partial_{\pi_{k,i}\pi_{\ell,i}\pi_{m,j}} g(\boldsymbol{\pi}) = \begin{cases} \frac{1}{N}\partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,2}} g_{ij}(\boldsymbol{\pi}) & \text{for } j > i \\ \frac{1}{N}\partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,1}} g_{ji}(\boldsymbol{\pi}) & \text{for } j < i \end{cases}.$$

For $i_1, i_2, i_3 \in V$ such that $\{i_1\} \cap \{i_2\} \cap \{i_3\} = \emptyset$ we have

$$\partial_{\pi_{k,i_1}\pi_{\ell,i_2}\pi_{m,i_3}} g(\boldsymbol{\pi}) = 0.$$

For convenience of notation, define the tensor D with

$$D = \left(\partial_{\pi_{k(s_1),i(s_1)}\pi_{k(s_2),i(s_2)}\pi_{k(s_3),i(s_3)}} g(\boldsymbol{\pi})\right)_{s_1,s_2,s_3 \in \{1,\ldots,KN\}}.$$

Also, let $\mathcal{P}(e_1, \ldots, e_n)$ denote the set of all permutations of the finite tupel $(e_1, \ldots, e_n)$ and let $\mathcal{C}_k(e_1, \ldots, e_n)$ denote all k-combinations from the finite set $\{e_1, \ldots, e_n\}$. Use

$$\sum_{\substack{s_1,s_2,s_3 \\ i=(i_1,i_2,i_3) \\ k=(\ell_1,\ell_2,\ell_3)}} \quad \text{as a shorthand for} \quad \sum_{\substack{s_1,s_2,s_3 \\ i(s_1)=i_1, i(s_2)=i_2, i(s_3)=i_3 \\ k(s_1)=\ell_1, k(s_2)=\ell_2, k(s_3)=\ell_3}}.$$

As in the proof of Lemma S.5 in FVW exploit that the $\|\cdot\|_q$ vector norm is dual to the $\|\cdot\|_{\frac{q}{q-1}}$ vector norm

$$\begin{aligned} \|D\|_q &= \max_{\|u^{(1)}\|_{\frac{q}{q-1}}=1} \max_{\|u^{(2)}\|_q=1} \max_{\|u^{(3)}\|_q=1} \left| \sum_{s_1,s_2,s_3=1}^{KN} u^{(1)}_{s_1} u^{(2)}_{s_2} u^{(3)}_{s_3} D_{s_1,s_2,s_3} \right| \\ &\leq \sum_{(\ell_1,\ell_2,\ell_3) \in \mathcal{C}_3(1,\ldots,K)} D^{(\ell_1,\ell_2,\ell_3)}_{s_1,s_2,s_3}, \end{aligned}$$

with

$$
\begin{aligned}
D^{(\ell_1,\ell_2,\ell_3)}_{s_1,s_2,s_3} = & \max_{\left\|u^{(1)}\right\|_{\frac{q}{q-1}}=1} \max_{\left\|u^{(2)}\right\|_{q}=1} \max_{\left\|u^{(3)}\right\|_{q}=1} \left| \sum_{i=1}^{N} \sum_{\substack{s_1,s_2,s_3 \\ i=(i,i,i) \\ k=(\ell_1,\ell_2,\ell_3)}} u^{(1)}_{s_1} u^{(2)}_{s_2} u^{(3)}_{s_3} D_{s_1,s_2,s_3} \right| \\
& + \max_{\left\|u^{(1)}\right\|_{\frac{q}{q-1}}=1} \max_{\left\|u^{(2)}\right\|_{q}=1} \max_{\left\|u^{(3)}\right\|_{q}=1} \left| \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \sum_{\substack{s_1,s_2,s_3 \\ i=(i,i,j) \\ k=(\ell_1,\ell_2,\ell_3)}} u^{(1)}_{s_1} u^{(2)}_{s_2} u^{(3)}_{s_3} D_{s_1,s_2,s_3} \right| \\
& + \max_{\left\|u^{(1)}\right\|_{\frac{q}{q-1}}=1} \max_{\left\|u^{(2)}\right\|_{q}=1} \max_{\left\|u^{(3)}\right\|_{q}=1} \left| \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \sum_{\substack{s_1,s_2,s_3 \\ i=(i,j,i) \\ k=(\ell_1,\ell_2,\ell_3)}} u^{(1)}_{s_1} u^{(2)}_{s_2} u^{(3)}_{s_3} D_{s_1,s_2,s_3} \right| \\
& + \max_{\left\|u^{(1)}\right\|_{\frac{q}{q-1}}=1} \max_{\left\|u^{(2)}\right\|_{q}=1} \max_{\left\|u^{(3)}\right\|_{q}=1} \left| \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \sum_{\substack{s_1,s_2,s_3 \\ i=(j,i,i) \\ k=(\ell_1,\ell_2,\ell_3)}} u^{(1)}_{s_1} u^{(2)}_{s_2} u^{(3)}_{s_3} D_{s_1,s_2,s_3} \right| \\
= & E_1 + E_2 + E_3 + E_4.
\end{aligned}
$$

Let

$$
d^{(\ell_1,\ell_2,\ell_3)} = \left(\partial_{\pi_{\ell_1,i}\pi_{\ell_2,i}\pi_{\ell_3,i}} g(\boldsymbol{\pi}) \right)_{i=1,\dots,N}.
$$

Then

$$
\begin{aligned}
E_1 \leq & \max_{\tilde{u}^{(1)}\in\mathbb{R}^N, \left\|\tilde{u}^{(1)}\right\|_{\frac{q}{q-1}}\leq 1} \max_{\tilde{u}^{2}\in\mathbb{R}^N, \left\|\tilde{u}^{(2)}\right\|_{q}\leq 1} \left| \sum_{i,j=1}^{N} \tilde{u}^{(1)}_i \tilde{u}^{(2)}_j \left[\operatorname{diag}\left(d^{(\ell_1,\ell_2,\ell_3)}\right) \right]_{i,j} \right| \\
= & \left\| \operatorname{diag}\left(d^{(\ell_1,\ell_2,\ell_3)}\right) \right\|_q \\
\leq & \left\| \operatorname{diag}\left(d^{(\ell_1,\ell_2,\ell_3)}\right) \right\|_1^{\frac{1}{q}} \left\| \operatorname{diag}\left(d^{(\ell_1,\ell_2,\ell_3)}\right) \right\|_\infty^{1-\frac{1}{q}} \\
= & \left\| \operatorname{diag}\left(d^{(\ell_1,\ell_2,\ell_3)}\right) \right\|_\infty = \max_{i\in V} \left| \left[\operatorname{diag}\left(d^{(\ell_1,\ell_2,\ell_3)}\right) \right]_{i,i} \right|
\end{aligned}
$$

To see why the first inequality holds construct feasible values of $\tilde{u}^{(1)}$ and $\tilde{u}^{(2)}$ from feasible values of $u^{(1)}, u^{(2)}, u^{(3)}$ in the following way. To determine the i's element of $\tilde{u}^{(1)}$ find the *unique* elements $u^{(1)}_{s_1}$ and $u^{(2)}_{s_2}$ such that $k(s_1) = \ell_1$, $k(s_2) = \ell_2$ and $i(s_1) = i(s_2) = i$. Then let $\tilde{u}^{(1)}_i = u^{(1)}_{s_1} u^{(2)}_{s_2}$. Note that $\left\|u^{(2)}\right\|_q = 1$ implies $\left\|u^{(2)}\right\|_{\max} \leq 1$ and therefore $\left\|\tilde{u}^{(1)}\right\|_{\frac{q}{1-q}} \leq 1$. Also, to determine the j's element of $\tilde{u}^{(2)}$ find the *unique* element $u^{(3)}_{s_3}$ such that $k(s_3) = \ell_3$ and $i(s_3) = j$. Note that $\left\|u^{(3)}\right\|_q = 1$ implies $\left\|\tilde{u}^{(2)}\right\|_q \leq 1$. The second inequality follows by Lemma S.4 in FVW and the last two equalities follows from the diagonal structure of the matrix whose norm we are considering. Therefore,

$$
\begin{aligned}
E_1 &\le \max_{i\in V}\left|\left[\operatorname{diag}\left(d^{(\ell_1,\ell_2,\ell_3)}\right)\right]_{i,i}\right| \\
&= \max_{i\in V}\left|\partial_{\pi_{k,i}\pi_{\ell,i}\pi_{m,i}} g(\boldsymbol{\pi})\right| \\
&= \max_{i\in V}\left|\frac{1}{N}\sum_{\substack{j\in V\\ j>i}}\partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,1}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N}\sum_{\substack{j\in V\\ j<i}}\partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,2}} g_{ji}(\boldsymbol{\pi})\right| \le C.
\end{aligned}
$$

Let

$$
e^{(\ell_1,\ell_2,\ell_3)} = \left(\partial_{\pi_{\ell_1,i}\pi_{\ell_2,i}\pi_{\ell_3,j}} g(\boldsymbol{\pi})\right)_{i,j=1,\dots,N}.
$$

Then

$$
\begin{aligned}
E_2 &\le \max_{\tilde u^{(1)}\in\mathbb{R}^N,\left\|\tilde u^{(1)}\right\|_{\frac{q}{q-1}}\le 1}\ \max_{\tilde u^{2}\in\mathbb{R}^N,\left\|\tilde u^{(2)}\right\|_{q}\le 1}\left|\sum_{i,j=1}^{N}\tilde u_i^{(1)}\tilde u_j^{(2)} e_{i,j}^{(\ell_1,\ell_2,\ell_3)}\right| \\
&= \left\|e^{(\ell_1,\ell_2,\ell_3)}\right\|_q \\
&\le \left\|e^{(\ell_1,\ell_2,\ell_3)}\right\|_1^{\frac{1}{q}}\left\|e^{(\ell_1,\ell_2,\ell_3)}\right\|_\infty^{1-\frac{1}{q}} \\
&= \left\{\max_{j\in V}\left|\sum_{i\in V} e_{i,j}^{(\ell_1,\ell_2,\ell_3)}\right|\right\}^{\frac{1}{q}}\left\{\max_{i\in V}\left|\sum_{j\in V} e_{i,j}^{(\ell_1,\ell_2,\ell_3)}\right|\right\}^{1-\frac{1}{q}}. \\
&\le \left\{\max_{j\in V}\left|\frac{1}{N}\sum_{\substack{i\in V\\ i<j}}\partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,2}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N}\sum_{\substack{i\in V\\ i>j}}\partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,1}} g_{ji}(\boldsymbol{\pi})\right|\right\}^{\frac{1}{q}} \\
&\quad\times\left\{\max_{i\in V}\left|\frac{1}{N}\sum_{\substack{j\in V\\ j>i}}\partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,2}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N}\sum_{\substack{j\in V\\ j<i}}\partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,1}} g_{ji}(\boldsymbol{\pi})\right|\right\}^{1-\frac{1}{q}} \le C.
\end{aligned}
$$

The first inequality can be argued similarly to the argument for the bound on E_1. The second inequality follows, again, from Lemma S.4 in FVW. The same bound can be derived for E_3 and E_4 in a similar way. In summary,

$$
\|D\|_q \le \sum_{(\ell_1,\ell_2,\ell_3)\in\mathcal{C}_3(1,\dots,K)} D_{s_1,s_2,s_3}^{(\ell_1,\ell_2,\ell_3)} \le 4K^3 C,
$$

concluding the proof of (ii). For (iii), write $G_{\boldsymbol{\pi}^2}^{11,k\ell}$ for the diagonal matrix with entries

$$
\left(\tilde G_{\boldsymbol{\pi}^2}^{11,k\ell}\right)_{i,i} = \partial_{\pi_{k,i}\pi_{\ell,i}} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}\left[\partial_{\pi_{k,i}\pi_{\ell,i}} g(\boldsymbol{\pi})\right]
$$

and $G^{12,k\ell}_{\boldsymbol{\pi}^2}$ for the matrix with entries

$$\left(\tilde{G}^{12,k\ell}_{\boldsymbol{\pi}^2}\right)_{i,j} = \begin{cases} \partial_{\pi_{k,i}\pi_{\ell,j}} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}\big[\partial_{\pi_{k,i}\pi_{\ell,j}} g(\boldsymbol{\pi})\big] & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}.$$

Now, we can write for a constant $C_{K,8}$ depending only on K

$$\begin{aligned} \big\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}\big[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\big]\big\|^8 \leq & \left(\sum_{k,\ell=1}^{K} \left\{\left\|\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}\right\| + 2\left\|\tilde{G}^{12,k\ell}_{\boldsymbol{\pi}^2}\right\|\right\}\right)^8 \\ \leq & C_{K,8}\left(\left\|\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}\right\|^8 + \left\|\tilde{G}^{12,k\ell}_{\boldsymbol{\pi}^2}\right\|^8\right). \end{aligned}$$

Fix $k, \ell \in \{1, \dots, K\}$. Let

$$\begin{aligned} \tilde{g}_{ij} &= \partial_{\pi_{k,1}\pi_{\ell,1}} g_{ij} - \bar{\mathbb{E}}\left[\partial_{\pi_{k,1}\pi_{\ell,1}} g_{ij}\right], \\ \tilde{g}_{ji} &= \partial_{\pi_{k,2}\pi_{\ell,2}} g_{ji} - \bar{\mathbb{E}}\left[\partial_{\pi_{k,2}\pi_{\ell,2}} g_{ji}\right]. \end{aligned}$$

As $\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}$ is a diagonal matrix

$$\bar{\mathbb{E}}\left\|\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}\right\|^8 = \bar{\mathbb{E}}\left(\max_{i\in V}\left|\left[\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}\right]_{i,i}\right|\right)^8 \leq \sum_{i\in V}\bar{\mathbb{E}}\left|\left[\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}\right]_{i,i}\right|^8.$$

Then,

$$\bar{\mathbb{E}}\left[\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}\right]^8_{i,i} \leq 2^7\bar{\mathbb{E}}\left(\frac{1}{N}\sum_{j\in V, j>i}\tilde{g}_{ij}\right)^8 + 2^7\bar{\mathbb{E}}\left(\frac{1}{N}\sum_{j\in V, j<i}\tilde{g}_{ji}\right)^8 = O_p\left(N^{-4}\right).$$

To prove the claim about the stochastic order of the right-hand side consider expanding $\bar{\mathbb{E}}\left(\sum_{j\in V, j>i}\tilde{g}_{ij}\right)^8$ (the argument for the second term is similar). A typically term in the expansion will look like $\bar{\mathbb{E}}\tilde{g}_{ij_1}\tilde{g}_{ij_2}\tilde{g}_{ij_3}\tilde{g}_{ij_4}$. The boundedness assumption gives us a universal upper bound on this term. By conditional independence and $\bar{\mathbb{E}}\tilde{g}_{ik} = 0$, whenever there is a $m = 1, \dots, 4$ such that $j_m \cap \{j_n : n = 1, \dots, 4; n \neq m\} = \emptyset$ we will have $\bar{\mathbb{E}}\tilde{g}_{ij_1}\tilde{g}_{ij_2}\tilde{g}_{ij_3}\tilde{g}_{ij_4} = 0$. The set of permissable $j_1, \dots, j_4$ that do not have one distinct index has cardinality less than $\binom{N}{3}$. Now, we can conclude that

$$\bar{\mathbb{E}}\left\|\tilde{G}^{11,k\ell}_{\boldsymbol{\pi}^2}\right\|^8 = O_p\left(N^{-3}\right).$$

Next, let's turn to bounding $\bar{\mathbb{E}}\left\|\tilde{G}^{12,k\ell}_{\boldsymbol{\pi}^2}\right\|^8$. Let

$$\hat{g}_{ij} = N\left(\partial_{\pi_{k,1}\pi_{\ell,2}} g_{ij} - \bar{\mathbb{E}}\left[\partial_{\pi_{k,1}\pi_{\ell,2}} g_{ij}\right]\right),$$
$$\hat{g}_{ji} = N\left(\partial_{\pi_{k,2}\pi_{\ell,1}} g_{ji} - \bar{\mathbb{E}}\left[\partial_{\pi_{k,2}\pi_{\ell,1}} g_{ji}\right]\right).$$

We will apply Lemma S.6 in FVW. Note that the assertion of this lemma remains true if E_ϕ is replaced by $\bar{\mathbb{E}}$ and independence conditional on ϕ is replaced by independence conditional on $\mathcal{A}$. This can easily be seen by inspection of their proof. Let e denote the matrix with entries

$$(e)_{i,j} = \left[N\tilde{G}^{12,k\ell}_{\boldsymbol{\pi}^2}\right]_{i,j}$$

Let $\bar{\sigma}_i^2 = \frac{1}{N}\sum_{j=1}^N \bar{\mathbb{E}} e_{i,j}^2$. Since there is a bound on the second derivative of g_{ij} there is a universal constant C such that

$$\bar{\sigma}_i^2 \leq \frac{2}{N}\Bigg\{\sum_{\substack{j\in V\\ i<j}} \hat{g}_{ij}^2 + \sum_{\substack{j\in V\\ i>j}} \hat{g}_{ji}^2\Bigg\} \leq N.$$

Therefore, $\frac{1}{N}\sum_{i=1}^N \left(\bar{\sigma}_i^2\right)^4 = O_p(1)$. Let Ω denote the matrix with entries $(\Omega)_{j_1,j_2} = \frac{1}{N}\sum_{i=1}^N \bar{\mathbb{E}}(e_{i,j_1}e_{i,j_2})$. Under our boundedness assumptions $\|\Omega\|_{\max} \leq C$ and therefore[33]

$$\frac{1}{N}\operatorname{Tr}(\Omega^4) \leq \|\Omega\|^4 \leq \|\Omega\|_{\max}^4.$$

Let $\eta_{i_1,i_2} = \frac{1}{\sqrt{N}}\sum_{j=1}^N \left[e_{i_1,j}e_{i_2,j} - \bar{\mathbb{E}}(e_{i_1,j}e_{i_2,j})\right]$. By conditional independence

$$\begin{aligned}\bar{\mathbb{E}}(\eta_{i_1,i_2}^4) =& \frac{1}{N^2}\sum_{j_1,j_2=1}^N \bar{\mathbb{E}}\left[(e_{i_1,j_1}e_{i_2,j_1} - \bar{\mathbb{E}}(e_{i_1,j_1}e_{i_1,j_1}))(e_{i_1,j_2}e_{i_2,j_2} - \bar{\mathbb{E}}(e_{i_1,j_2}e_{i_1,j_2}))\right]\\ \leq& \left(\frac{2}{N}\sum_{j=1}^N \bar{\mathbb{E}}(e_{i_1,j})^2\bar{\mathbb{E}}(e_{i_2,j})^2\right)^2 \leq C\end{aligned}$$

and therefore $\frac{1}{N}\sum_{i=1}^N \bar{\mathbb{E}}(\eta_{i,i}^4) = O_p(1)$ and $\frac{1}{N^2}\sum_{i_1,i_2=1}^N \bar{\mathbb{E}}\left(\eta_{i_1,i_2}^4\right) = O_p(1)$. Thus, Lemma S.6 is applicable and we can conclude that $\bar{\mathbb{E}}\,\|e\| = O_p\left(N^{5/8}\right)$ or, equivalently, $\bar{\mathbb{E}}\left\|\tilde{G}^{12,k\ell}_{\boldsymbol{\pi}^2}\right\| =$

[33] For every symmetric $N\times N$ matrix M we have $\frac{1}{N}\operatorname{Tr}(M^2) \leq \|M\|^2$. To prove this, consider a slightly more general case and let A, B denote symmetric $N\times N$ matrices with eigenvalues $\alpha_1 \leq \cdots \leq \alpha_N$ and $\beta_1 \leq \cdots \leq \beta_N$, respectively. By the von-Neumann trace inequality, $\operatorname{Tr}(AB) \leq \sum_{i=1}^N \alpha_i\beta_i$. For symmetric square matrices it is well-known that $\|A\| = \alpha_N$ and $\|B\| = \beta_N$. Therefore, $\operatorname{Tr}(AB) \leq N\,\|A\|\,\|B\|$. For any square matrix Ω, $M = \Omega'\Omega$ is symmetric. Therefore, $\frac{1}{N}\operatorname{Tr}(\Omega'\Omega\Omega'\Omega) \leq \|\Omega'\Omega\|^2 \leq \|\Omega\|^2\,\|\Omega'\|^2$. The first inequality now follows from noting that Ω as defined above is symmetric.

$O_p\left(N^{-3/8}\right)$. In summary, we have shown that

$$\bar{\mathbb{E}}\left\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}\big[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\big]\right\| = O_p\left(N^{-3/8}\right).$$

This implies

$$\left\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}\big[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\big]\right\| = O_p\left(N^{-3/8}\right),$$

concluding the proof of (iii). □

Lemma C.8 (Sparse bounded functionals of the incidental parameter II)**.** *Let K denote a finite constant and let $(\boldsymbol{\pi}_k)_{k=1}^K$ denote a collection of N-dimensional parameters. Define $\boldsymbol{\pi} = (\boldsymbol{\pi}_1', \ldots, \boldsymbol{\pi}_K')'$. Let $\{g_{i_1,\ldots,i_L}\}_{i_1<\cdots<i_L}$ denote an array of functions such that*

$$g_{i_1,\ldots,i_L}(\boldsymbol{\pi}) = g_{i_1,\ldots,i_L}(\pi_{1,i_1}, \ldots, \pi_{K,i_1}, \ldots, \pi_{1,i_L}, \ldots, \pi_{K,i_L})$$

with

$$\|\partial_{\boldsymbol{\pi}^\ell} g_{i_1,\ldots,i_L}(\boldsymbol{\pi})\|_{\max} \leq C \quad \text{for } \ell = 0,1,2,3$$

for a universal constant C. Let

$$g(\boldsymbol{\pi}) = \frac{1}{N^{L-1}} \sum_{i_1<\cdots<i_L} g_{i_1,\ldots,i_L}(\boldsymbol{\pi}).$$

Then $\|\partial_{\boldsymbol{\pi}} g(\boldsymbol{\pi})\|_q = O_p\left(N^{-\frac{1}{q}}\right)$, $\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})\|_q = O_p(1)$ and $\|\partial_{\boldsymbol{\pi}^3} g(\boldsymbol{\pi})\|_q = O_p(1)$.

Proof. The proof is very similar to that of Lemma C.7. For example,

$$\begin{aligned}\partial_{\boldsymbol{\pi}_{k,i}} g(\boldsymbol{\pi}) = &\sum_{i<i_1<\cdots<i_L} \partial_{\boldsymbol{\pi}_{k,i}} g_{i_1,\ldots,i_L}(\boldsymbol{\pi}) \\ &+ \sum_{i_1<i<i_3<\cdots<i_L} \partial_{\boldsymbol{\pi}_{k,i}} g_{i_1,\ldots,i_L}(\boldsymbol{\pi}) + \cdots + \sum_{i_1<\cdots<i_{L-1}<i} \partial_{\boldsymbol{\pi}_{k,i}} g_{i_1,\ldots,i_L}(\boldsymbol{\pi}).\end{aligned}$$

Therefore, there is a constant C such that

$$\begin{aligned}\|\partial_{\boldsymbol{\pi}} g(\boldsymbol{\pi})\|_q \leq \left(\sum_{\ell=1}^{KN} |\partial_{\boldsymbol{\pi}_\ell} g(\boldsymbol{\pi})^q|\right)^{\frac{1}{q}} &\leq \left(\frac{1}{N^{L-1}} \sum_{k=1}^K \sum_{i_1<\cdots<i_L} \sum_{\ell=1}^L \left|\partial_{\boldsymbol{\pi}_{k,i_\ell}} g_{i_1,\ldots,i_L}(\boldsymbol{\pi})\right|^q\right)^{\frac{1}{q}} \\ &< C(KLN)^{\frac{1}{q}}.\end{aligned}$$

□

Lemma C.9 (Symmetric functions)**.** *Define a class of symmetric functions,*

$$\mathcal{G} = \left\{g : \mathbb{R}^2 \to \mathbb{R} \mid g(z_1, z_2) = g(z_2, z_1)\right\}$$

The function class $\mathcal{G}$ is closed under multiplication and addition , i.e.,

$$\begin{aligned} g, h \in \mathcal{G} \Rightarrow & gh \in \mathcal{G} \\ g, h \in \mathcal{G} \Rightarrow & g + h \in \mathcal{G}. \end{aligned}$$

If $g \in \mathcal{G}$ is (partially) differentiable in the first component, then g is also differentiable in the second component and

$$\partial_1 g(z_1, z_2) = \partial_2 g(z_2, z_1).$$

Moreover, if $g \in \mathcal{G}$ is twice (partially) differentiable in the first component, then g is also twice differentiable in the second component and

$$\begin{aligned} \partial_{22} g(z_1, z_2) &= \partial_{11} g(z_2, z_1), \\ \partial_{21} g(z_1, z_2) &= \partial_{12} g(z_2, z_1). \end{aligned}$$

Let ϕ denote a scalar parameter and let $\mathcal{B}_\epsilon$ denote an open ball on the real line. Suppose that $g(z_1, z_2, \phi) \in \mathcal{G}$ for all $\phi \in \mathcal{B}_\epsilon$ and that g is differentiable in ϕ on $\mathcal{B}_\epsilon$. Then,

$$\partial_\phi g(z_1, z_2, \phi) \in \mathcal{G} \quad \textit{for } \phi \in \mathcal{B}_\epsilon.$$

Proof. Suppose that $g \in \mathcal{G}$ is differentiable in the first component. Then

$$\partial_1 g(z_1, z_2) = \lim_{\delta \to 0} \frac{g(z_1 + \delta, z_2) - g(z_1, z_2)}{\delta} = \lim_{\delta \to 0} \frac{g(z_2, z_1 + \delta) - g(z_2, z_1)}{\delta} = \partial_2 g(z_2, z_1).$$

Existence of the limit on the right-hand side follows from existence of the limit on the left-hand side. Furthermore,

$$\partial_{22} g(z_1, z_2) = \frac{d}{dz_2} \Big(\partial_2 g(z_1, z_2) \Big) = \frac{d}{dz_2} \Big(\partial_1 g(z_2, z_1) \Big) = \partial_{11} g(z_2, z_1).$$

The claim about the cross-derivative follows in a similar way. The last claim follows by noting that

$$\begin{aligned} \partial_\phi g(z_1, z_2, \phi) =& \lim_{\delta \to 0} \frac{g(z_1, z_2, \phi + \delta) - g(z_1, z_2, \phi)}{\delta} \\ =& \lim_{\delta \to 0} \frac{g(z_2, z_1, \phi + \delta) - g(z_2, z_1, \phi)}{\delta} = \partial_\phi g(z_2, z_1, \phi). \end{aligned}$$

□

Lemma C.10. *For a function $g : \mathbb{R}^3 \to \mathbb{R}$ write $g_{ij} = g(Z_{ij}, Z_{ji}, Y_{ij} Y_{ji})$. Let*

$$G(\boldsymbol{\gamma}) = \sum_{i<j} g_{ij} = \sum_{i<j} g(Z_{ij}, Z_{ji}, Y_{ij} Y_{ji}).$$

Define the matrix $A = (A_{ij})_{i,j\in V}$ *where*

$$A_{ij} = \begin{cases} \partial_1 g(Z_{ij}, Z_{ji}, Y_{ij}Y_{ji}) & \text{if } i < j \\ \partial_2 g(Z_{ji}, Z_{ij}, Y_{ij}Y_{ji}) & \text{if } i > j \\ 0 & \text{if } i = j \end{cases}.$$

Then

$$\partial_{\boldsymbol{\gamma}} G(\boldsymbol{\gamma}) = \begin{bmatrix} A\,\iota_N \\ A'\,\iota_N \end{bmatrix}.$$

Proof. This follows from a straightforward inspection. In particular,

$$\begin{aligned} \partial_{\gamma_i^S} G &= \sum_{\substack{j\in V \\ j>i}} \partial_1 g(Z_{ij}, Z_{ji}, Y_{ij}Y_{ji}) + \sum_{\substack{j\in V \\ j<i}} \partial_2 g(Z_{ji}, Z_{ij}, Y_{ij}Y_{ji}) \\ &= \sum_{j\in V_{-i}} \left\{ 1_{\{i<j\}} \partial_1 g(Z_{ij}, Z_{ji}, Y_{ij}Y_{ji}) + 1_{\{i>j\}} \partial_2 g(Z_{ji}, Z_{ij}, Y_{ij}Y_{ji}) \right\} = \sum_{j\in V_{-i}} A_{ij} \end{aligned}$$

and

$$\begin{aligned} \partial_{\gamma_j^R} G &= \sum_{\substack{i\in V \\ i<j}} \partial_1 g(Z_{ij}, Z_{ji}, Y_{ij}Y_{ji}) + \sum_{\substack{i\in V \\ i>j}} \partial_2 g(Z_{ji}, Z_{ij}, Y_{ij}Y_{ji}) \\ &= \sum_{i\in V_{-j}} \left\{ 1_{\{i<j\}} \partial_1 g(Z_{ij}, Z_{ji}, Y_{ij}Y_{ji}) + 1_{\{i>j\}} \partial_2 g(Z_{ji}, Z_{ij}, Y_{ij}Y_{ji}) \right\} = \sum_{i\in V_{-j}} A_{ij}. \end{aligned}$$

□

Lemma C.11. *Under Assumption 1(i) and Assumption 1(v)*

$$\partial_{\rho\boldsymbol{\gamma}} \mathcal{M} = \begin{bmatrix} A\iota_N \\ A'\iota_N \end{bmatrix}$$

where A *is a* $N \times N$ *matrix with entries*

$$A_{i,j} = \begin{cases} \partial_{\rho y_1} m_{ij} & \text{for } i \neq j \\ 0 & \text{for } i = j. \end{cases}$$

Moreover,

$$\partial_{\rho\boldsymbol{\gamma}} \mathcal{M} = \begin{bmatrix} \bar{A}\iota_N \\ \bar{A}'\iota_N \end{bmatrix}$$

where $\bar{A}$ is a $N \times N$ matrix with entries

$$\bar{A}_{i,j} = \begin{cases} \partial_{\rho y_1} \bar{m}_{ij} = -J_{ij}(\partial_{z_1} r_{ij}) & \text{for } i \neq j \\ 0 & \text{for } i = j. \end{cases}$$

Proof. We will apply Lemma C.10 with $g_{ij} = \partial_\rho m_{ij}$. Lemma C.10 gives that

$$\partial_{\rho\gamma}\mathcal{M} = \begin{bmatrix} A\iota_N \\ A'\iota_N \end{bmatrix},$$

where A is a $N \times N$ matrix with entries

$$A_{i,j} = \begin{cases} \partial_{\rho z_1} m_{ij} & \text{for } i < j \\ \partial_{\rho z_2} m_{ji} & \text{for } i > j \\ 0 & \text{for } i = j. \end{cases}$$

It remains to show that $\partial_{\rho z_2} m_{ji} = \partial_{\rho z_1} m_{ij}$. By construction r_{ij} is a symmetric function in the sense of Lemma C.9. Repeated application of Lemma C.9 shows that J_{ij} is also a symmetric function and therefore $\partial_{z_2} J_{ji} = \partial_{z_1} J_{ij}$ by Lemma C.9. Similary, one can show that $\partial_{z_2}(J_{ji} r_{ji}) = \partial_{z_1}(J_{ij} r_{ij})$. The second assertion is proved similarly. □

Lemma C.12. *Let $\mathcal{A}$ denote a σ-field and let $\mathbb{E}_\mathcal{A}$ denote the expectation operator conditional on $\mathcal{A}$. Let $(Y_{i,j}, Y_{j,i})_{i,j=1,\ldots,n}$ denote an array of random tuples that are mutually independent conditional on $\mathcal{A}$ and satisfy $\mathbb{E}_\mathcal{A} |Y_{i,j}|^4 \leq C$ for a constant C. Suppose that $\mathbb{E}_\mathcal{A} Y_{i,j} = 0$ for $i, j = 1, \ldots, n$ and let Υ denote the matrix random entries $(\Upsilon)_{i,j} = Y_{i,j}$. Let M denote a matrix with $\mathcal{A}$-measurable random entries such that $\|M\|_{\max} = O_p(n^{-1})$ and let D denote a diagonal matrix with with $\mathcal{A}$-measurable random entries such that $\|D\|_{\max} = O_p(1)$. Then for $A, B \in \{\Upsilon, \Upsilon'\}$*

$$\begin{aligned} 1_n' A' M B 1_n &= O_p(n), \\ 1_n' A' D B 1_n &= \mathbb{E}_\mathcal{A}[1_n' A' D B 1_n] + O_p\big(n^{3/2}\big) \\ &= \sum_{i,j=1}^n (D)_{i,i} \mathbb{E}_\mathcal{A}[a_{i,j} b_{i,j}] + O_p\big(n^{3/2}\big). \end{aligned}$$

Proof. To prove the first statement note that

$$n\, 1_n' A' M B 1_n = \sum_{i,j,k,\ell=1}^n a_{ik} b_{\ell j} [n\, m_{k,\ell}]$$

For $\kappa = 3, 4$ let

$$\mathcal{P}_\kappa(i_1, \ldots, i_\kappa) = \{(j_1 j_2, \ldots, j_{2\kappa-1} j_{2\kappa}) : j_1, \ldots, j_{2\kappa} \in \{i_1, \ldots, i_\kappa\}\}$$

the set of all 4-tuples of index pairs that can be generated from a given set of four (not necessarily distinct) indices $i_1, \ldots, i_\kappa$. Conditional independence and the zero mean property of the $Y_{i,j}$ yields

$$\mathbb{E}_{\mathcal{A}}[n\,1_n' A' M B 1_n]^2 = \sum_{i,j,k,\ell=1}^{n} \sum_{(p_1,\ldots,p_4)\in\mathcal{P}_4(i,j,k,\ell)} [n\,m]_{p_{12}p_{21}}[n\,m]_{p_{32}p_{41}} \mathbb{E}_{\mathcal{A}}[a_{p_1} b_{p_2} a_{p_3} b_{p_4}]$$
$$\leq C \sum_{i,j,k,\ell=1}^{n} \sum_{(p_1,\ldots,p_4)\in\mathcal{P}_4(i,j,k,\ell)} [n\,m]_{p_{12}p_{21}}[n\,m]_{p_{32}p_{41}} = O_p(n^4),$$

where the inequality follows from Cauchy-Schwarz. The first claim follows now immediately. To prove the second claim note that

$$1_n' A' D B 1_n = \sum_{i,j,k=1}^{n} a_{ij} b_{ik} d_{ii}.$$

Taking the squared expectation gives

$$\mathbb{E}_{\mathcal{A}}[1_n' A' D B 1_n]^2 = \sum_{i,j,k=1}^{n} \sum_{(p_1,\ldots,p_3)\in\mathcal{P}_3(i,j,k)} d_{p_{11}p_{11}} d_{p_{31}p_{31}} \mathbb{E}_{\mathcal{A}}[a_{p_1} b_{p_2} a_{p_3} b_{p_4}] = O_p(n^3).$$

It follows that $1_n' A' M B 1_n = \mathbb{E}_{\mathcal{A}}[1_n' A' M B 1_n] + O_p(n^{3/2})$. Finally, it is easy to see that

$$\mathbb{E}_{\mathcal{A}}[1_n' A' M B 1_n] = \sum_{i,j,k=1}^{n} d_{ii} \mathbb{E}_{\mathcal{A}}[a_{ij} b_{ik}] = \sum_{i,j=1}^{n} d_{ii} \mathbb{E}_{\mathcal{A}}[a_{ij} b_{ij}].$$

□

D. Bootstrap protocol for percentile bootstrap of $\hat{T}_N^{\text{stud}}$

The following bootstrap protocol is a variation of the double bootstrap procedure in Kim and Sun (2016). Draw B times from the bootstrap distribution as follows:

1. Draw $\binom{N}{2}$ independent pairwise bootstrap shocks $\{(U_{ij}^*, U_{ji}^*)\}_{\substack{i,j\in V\\ i<j}}$ from the bivariate normal distribution with marginal variance equal to one and covariance $\hat{\rho}$.

2. The bootstrapped network is given by $\{Y_{ij}^*\}_{ij\in E(N)}$ where

$$Y_{ij}^* = \mathbf{1}\left(X_{ij}'\hat{\theta} + \hat{\gamma}_i^S + \hat{\gamma}_j^R \geq U_{ij}^*\right)$$

3. Estimate θ^* and $\boldsymbol{\gamma}^* = (\gamma_i^{S,0,*}, \gamma_i^{S,0,*})_{i \in V}$ on the bootstrapped network using the first stage of the ML procedure from Section 2.2.

4. Compute $\big(\hat{T}_N^{\text{stud}}\big)^*$ on the bootstrapped network, using θ^* and $\boldsymbol{\gamma}^*$ as inputs for the plug-in estimators.

This gives a vector of draws $\{\big(\hat{T}_N^{\text{stud}}\big)^*\}_{j=1,\ldots,B}$. Let $t_{\alpha/2}$ and $t_{1-\alpha/N}$ denote the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of $\{\big(\hat{T}_N^{\text{stud}}\big)^*\}_{j=1,\ldots,B}$. The transitivity test with bootstrap critical values rejects if $\hat{T}_N^{\text{stud}} \notin [t^*_{\alpha/N}, t^*_{1-\alpha/2}]$.

E. Network statistics for simulated designs

				in-degree		out-degree				
N	ρ	C_N	density	mean	median	mean	median	comp	min cut	clust
50	0.0	$\log\log N$	0.19	9.50	9.16	9.50	9.14	1.00	1.51	0.57
50	0.0	$\log^{1/2} N$	0.12	5.65	4.97	5.65	4.92	0.90	0.00	0.47
50	0.0	$2\log^{1/2} N$	0.06	3.05	2.06	3.05	2.04	0.62	0.00	0.42
50	0.0	$\log N$	0.03	1.57	0.45	1.57	0.50	0.35	0.00	0.40
50	0.5	$\log\log N$	0.19	9.50	9.18	9.50	9.16	1.00	1.52	0.51
50	0.5	$\log^{1/2} N$	0.12	5.67	4.96	5.67	4.94	0.91	0.00	0.42
50	0.5	$2\log^{1/2} N$	0.06	3.05	2.03	3.05	2.04	0.63	0.00	0.38
50	0.5	$\log N$	0.03	1.58	0.52	1.58	0.51	0.37	0.00	0.36
70	0.0	$\log\log N$	0.18	12.56	12.03	12.56	12.00	1.00	1.95	0.56
70	0.0	$\log^{1/2} N$	0.11	7.50	6.43	7.50	6.42	0.93	0.00	0.47
70	0.0	$2\log^{1/2} N$	0.06	3.99	2.52	3.99	2.56	0.67	0.00	0.43
70	0.0	$\log N$	0.03	1.90	0.41	1.90	0.40	0.39	0.00	0.41
70	0.5	$\log\log N$	0.18	12.54	12.01	12.54	12.05	1.00	1.90	0.50
70	0.5	$\log^{1/2} N$	0.11	7.49	6.40	7.49	6.40	0.94	0.01	0.42
70	0.5	$2\log^{1/2} N$	0.06	3.99	2.53	3.99	2.51	0.68	0.00	0.38
70	0.5	$\log N$	0.03	1.89	0.39	1.89	0.41	0.40	0.00	0.37

Table E.1: Statistics for simulated networks under the null hypothesis (averaged over 500 replications): "comp" = share of agents belonging to the largest connected component, "min cut" = minimum cut of the network, "clust" = clustering coefficient.

				in-degree		out-degree				
N	ρ	C_N	density	mean	median	mean	median	comp	min cut	clust
50	0.0	$\log\log N$	0.18	8.74	8.21	8.74	8.21	0.98	0.53	0.57
50	0.0	$\log^{1/2} N$	0.10	4.75	3.50	4.75	3.49	0.70	0.00	0.50
50	0.0	$2\log^{1/2} N$	0.05	2.24	1.00	2.24	1.00	0.33	0.00	0.47
50	0.0	$\log N$	0.02	1.13	0.00	1.13	0.00	0.18	0.00	0.48
50	0.5	$\log\log N$	0.18	8.74	8.21	8.74	8.20	0.98	0.50	0.56
50	0.5	$\log^{1/2} N$	0.10	4.77	3.54	4.77	3.56	0.70	0.00	0.49
50	0.5	$2\log^{1/2} N$	0.05	2.26	0.98	2.26	1.01	0.34	0.00	0.46
50	0.5	$\log N$	0.02	1.15	0.00	1.15	0.00	0.18	0.00	0.48
70	0.0	$\log\log N$	0.17	11.70	10.66	11.70	10.69	0.99	0.82	0.56
70	0.0	$\log^{1/2} N$	0.09	6.38	4.47	6.38	4.50	0.82	0.00	0.49
70	0.0	$2\log^{1/2} N$	0.04	3.08	1.31	3.08	1.32	0.46	0.00	0.47
70	0.0	$\log N$	0.02	1.43	0.00	1.43	0.00	0.21	0.00	0.49
70	0.5	$\log\log N$	0.17	11.69	10.67	11.69	10.65	0.99	0.82	0.55
70	0.5	$\log^{1/2} N$	0.09	6.40	4.55	6.40	4.55	0.82	0.00	0.48
70	0.5	$2\log^{1/2} N$	0.04	3.09	1.30	3.09	1.34	0.46	0.00	0.46
70	0.5	$\log N$	0.02	1.44	0.00	1.44	0.00	0.21	0.00	0.48

Table E.2: Statistics for simulated networks under the dynamic alternative (averaged over 500 replications): "comp" = share of agents belonging to the largest connected component, "min cut" = minimum cut of the network, "clust" = clustering coefficient.

F. Model specification test based on cyclic triangles

A model specification test based on cyclic triangles can be implemented similarly to the specification test based on transitive triangles. The theory carries over in a straightforward way. It is convenient to use notation that is similar to the notation used for the test based on transitive triangles. To indicate that symbols refer to the test based on cyclic triangles, I add a "$\circ$" superscript. The set of all cyclic triangles is given by

$$B^\circ = B^\circ(N) = \big\{\{(i,j),(j,k),(k,i)\} : \{i,j,k\} \subset V(N), \{i,j,k\} = 3\big\}.$$

On N vertices, there are $\binom{N}{3}$ cyclic triangles. Let

$$S_N^\circ = \sum_{\beta \in B^\circ} \left(\prod_{ij \in \beta} Y_{ij} \right)$$
$$\widehat{\mathbb{E}\, S_N^\circ} = \sum_{\beta \in B^\circ} \left(\prod_{ij \in \beta} \hat{p}_{ij} \right),$$

denote the observed and the (estimated) predicted number of cyclic triangles, respectively. Excess cyclic closure can be measured by

$$T_N^\circ = 6 \left(\frac{S_N^\circ - \widehat{\mathbb{E}\, S_N^\circ}}{N^3} \right).$$

Let

$$\boldsymbol{\beta}_{ij}^{\circ;N} = \frac{1}{H_{ij} N} \sum_{\substack{\beta \in B^\circ(N) \\ \beta \ni ij}} \prod_{\substack{e \in \beta \\ e \neq ij}} p_e.$$

The following result characterizes the asymptotic distribution of T_N°. Similar to the the corresponding result for transitive triangles (Theorem 3) it can be used to construct a specification test based on cyclic closure (see Section 5.2).

Theorem 4 (Model specification test based on excess cyclic closure)**.** *Let*

$$U_N^\circ = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \boldsymbol{\beta}_{ij}^{\circ;N} \omega_{ij} \tilde{X}_{ij}$$

and $\tilde{u}_{N,ij}^\circ = (U_N^\circ)' \bar{W}_{1,N}^{-1} \tilde{X}_{ij}$ *and suppose that Assumption 1 holds. Then*

$$\frac{N T_N^\circ + 6\big(B_N^{T^\circ} + (U_N^\circ)' \bar{W}_{1,N}^{-1} B_N^\theta\big)}{\sqrt{v_N^{T^\circ}}} = \mathcal{N}(0, 36) + o_p(1),$$

where

$$v_N^{T^\circ} = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left\{ \big(\tilde{\boldsymbol{\beta}}_{ij}^{\circ,N} - \tilde{u}_{N,ij}^\circ\big)^2 \omega_{ij} + \big(\tilde{\boldsymbol{\beta}}_{ij}^{\circ,N} - \tilde{u}_{N,ij}^\circ\big)\big(\tilde{\boldsymbol{\beta}}_{ji}^{\circ,N} - \tilde{u}_{N,ji}^\circ\big) \tilde{\rho}_{ij} \sqrt{\omega_{ij}\omega_{ji}} \right\}$$

and $B_N^{T^\circ} = B_N^{T^\circ,S} + B_N^{T^\circ,R} + B_N^{T^\circ,SR}$ *with*

$$B_N^{T^\circ,S} = \frac{1}{2N}\sum_{i\in V}\frac{\sum_{j\in V_{-i}} H_{ij}(\partial_{z^2}p_{ij})\tilde{\boldsymbol{\beta}}_{ij}^{\circ,N}}{\sum_{j\in V_{-i}}\omega_{ij}}$$

$$B^{T^\circ,R} = \frac{1}{2N}\sum_{j\in V}\frac{\sum_{i\in V_{-j}} H_{ij}(\partial_{z^2}p_{ij})\tilde{\boldsymbol{\beta}}_{ij}^{\circ,N}}{\sum_{j\in V_{-i}}\omega_{ij}}$$

$$B_N^{T^\circ,SR} = \frac{1}{N}\sum_{i\in V}\frac{\overline{\text{corr}}_i\, N^{-1}\sum_{j\in V_{-i}}\sum_{k\in V_{-\{j,k\}}}(\partial_z p_{ij})(\partial_z p_{ji})p_{kj}}{\left(\sum_{j\in V_{-i}}\omega_{ij}\right)^{1/2}\left(\sum_{j\in V_{-i}}\omega_{ji}\right)^{1/2}}.$$

Proof. The proof is analogous to the proof of Theorem 3 and is omitted. □

G. Tables for application to favor networks

Variable	Description
same caste	i and j belong to the same caste
age diff	absolute value of age difference between i and j
same family	i and j belong to the same family
same latrine	i and j both (don't) live in a house with an own latrine
same gender	i and j have the same gender
both hh heads	both i and j are household heads
same village native	both i and j were born in the village
educ None-Some	one of i and j has no education, the other has finished primary education
educ None-SSLC	one of i and j has no education, the other has a obtained a SSL certificate
educ Some-SSLC	one of i and j has finished primary education, the other has obtained a SSL certificate

Table G.1: Description of variables measuring homophily (X_{ij}).

Village	N	T_N	T_N bc	rel bias	pval anal	pval BS	pval naïve	pval naïve bc
1	203	4.2e-06	6.3e-06	-0.34	< 0.01	< 0.1	< 0.01	< 0.01
2	203	1.5e-06	4.8e-06	-0.69	< 0.01	< 0.1	0.025	< 0.01
3	345	8.5e-07	1.3e-06	-0.35	< 0.01	< 0.1	< 0.01	< 0.01
4	256	2e-06	3.4e-06	-0.41	< 0.01	< 0.1	< 0.01	< 0.01
5	164	1.5e-06	4.8e-06	-0.68	< 0.01	< 0.1	0.035	< 0.01
6	110	3.6e-06	1.3e-05	-0.73	< 0.01	< 0.1	0.23	< 0.01

Continued on next page

Table G.2 – continued from previous page

Village	N	T_N	T_N bc	rel bias	pval anal	pval BS	pval naïve	pval naïve bc
7	172	1.1e-05	1.7e-05	-0.34	< 0.01	< 0.1	< 0.01	< 0.01
8	109	1.1e-05	3.6e-05	-0.69	< 0.01	< 0.1	0.099	< 0.01
9	247	2.9e-06	4.9e-06	-0.40	< 0.01	< 0.1	< 0.01	< 0.01
10	95	3.2e-06	1.5e-05	-0.79	< 0.01	< 0.1	0.5	< 0.01
11	142	5.4e-06	1.5e-05	-0.65	< 0.01	< 0.1	0.075	< 0.01
12	195	7.2e-06	1.2e-05	-0.40	< 0.01	< 0.1	< 0.01	< 0.01
14	150	1.2e-05	2.7e-05	-0.56	< 0.01	< 0.1	< 0.01	< 0.01
15	212	8.6e-07	4.3e-06	-0.80	< 0.01	< 0.1	0.18	< 0.01
16	178	6.6e-06	1.4e-05	-0.52	< 0.01	< 0.1	< 0.01	< 0.01
17	200	1.5e-06	4.7e-06	-0.69	< 0.01	< 0.1	0.086	< 0.01
18	284	4.8e-07	1.6e-06	-0.69	< 0.01	< 0.1	0.015	< 0.01
19	243	2.5e-06	5.2e-06	-0.52	< 0.01	< 0.1	< 0.01	< 0.01
20	159	6.7e-06	1.4e-05	-0.51	< 0.01	< 0.1	0.01	< 0.01
21	210	2.1e-06	6.1e-06	-0.65	< 0.01	< 0.1	< 0.01	< 0.01
23	280	2.6e-06	4e-06	-0.35	< 0.01	< 0.1	< 0.01	< 0.01
24	211	4.2e-06	1e-05	-0.60	< 0.01	< 0.1	< 0.01	< 0.01
25	304	1.3e-06	2.4e-06	-0.47	< 0.01	< 0.1	< 0.01	< 0.01
26	149	1.1e-05	2.1e-05	-0.50	< 0.01	< 0.1	< 0.01	< 0.01
27	174	1.6e-06	1.1e-05	-0.85	< 0.01	≥ 0.1	0.13	< 0.01
28	395	6.6e-07	1.1e-06	-0.42	< 0.01	< 0.1	< 0.01	< 0.01
29	303	2.7e-06	4.6e-06	-0.42	< 0.01	< 0.1	< 0.01	< 0.01
30	170	1.2e-05	2.4e-05	-0.49	< 0.01	< 0.1	< 0.01	< 0.01
31	200	4.7e-06	9.7e-06	-0.51	< 0.01	< 0.1	< 0.01	< 0.01
32	301	1.2e-06	2.8e-06	-0.56	< 0.01	< 0.1	< 0.01	< 0.01
33	219	4.4e-06	7.4e-06	-0.40	< 0.01	< 0.1	< 0.01	< 0.01
34	181	1e-05	2e-05	-0.50	< 0.01	< 0.1	< 0.01	< 0.01
35	216	8.8e-06	1.5e-05	-0.43	< 0.01	< 0.1	< 0.01	< 0.01
36	293	6.1e-06	7.8e-06	-0.23	< 0.01	< 0.1	< 0.01	< 0.01
37	132	2.8e-05	4.1e-05	-0.31	< 0.01	< 0.1	< 0.01	< 0.01
38	182	1.5e-06	5.6e-06	-0.73	< 0.01	< 0.1	0.15	< 0.01
39	370	1.4e-06	2.8e-06	-0.49	< 0.01	< 0.1	< 0.01	< 0.01
40	266	1.4e-05	2e-05	-0.30	< 0.01	< 0.1	< 0.01	< 0.01
41	181	3.3e-05	4e-05	-0.16	< 0.01	< 0.1	< 0.01	< 0.01
42	206	8.9e-06	1.5e-05	-0.39	< 0.01	< 0.1	< 0.01	< 0.01
43	227	1.5e-05	1.7e-05	-0.14	< 0.01	< 0.1	< 0.01	< 0.01
44	258	1e-05	1.3e-05	-0.21	< 0.01	< 0.1	< 0.01	< 0.01
45	263	2.3e-06	4.5e-06	-0.49	< 0.01	< 0.1	< 0.01	< 0.01
46	279	1.2e-06	2.3e-06	-0.49	< 0.01	< 0.1	< 0.01	< 0.01
47	160	2e-06	7.3e-06	-0.73	< 0.01	< 0.1	0.16	< 0.01
48	217	4.9e-06	1.1e-05	-0.53	< 0.01	< 0.1	< 0.01	< 0.01
49	184	5.6e-06	1.2e-05	-0.54	< 0.01	< 0.1	< 0.01	< 0.01
50	261	1e-05	1.5e-05	-0.33	< 0.01	< 0.1	< 0.01	< 0.01
51	309	6.5e-06	1.1e-05	-0.43	< 0.01	< 0.1	< 0.01	< 0.01
52	395	3.6e-06	5.7e-06	-0.37	< 0.01	< 0.1	< 0.01	< 0.01
53	170	2.4e-05	4.4e-05	-0.46	< 0.01	< 0.1	< 0.01	< 0.01
54	124	7.9e-06	3.4e-05	-0.77	< 0.01	< 0.1	0.14	< 0.01

Continued on next page

Table G.2 – continued from previous page

Village	N	T_N	T_N bc	rel bias	pval anal	pval BS	pval naïve	pval naïve bc
55	279	6e-06	1e-05	-0.42	< 0.01	< 0.1	< 0.01	< 0.01
56	148	2.7e-05	3.2e-05	-0.16	< 0.01	< 0.1	< 0.01	< 0.01
57	234	1.5e-06	2.3e-06	-0.37	< 0.01	< 0.1	< 0.01	< 0.01
58	210	2.1e-06	4.9e-06	-0.58	< 0.01	< 0.1	< 0.01	< 0.01
59	384	2.1e-06	3e-06	-0.32	< 0.01	< 0.1	< 0.01	< 0.01
60	413	1.1e-06	2e-06	-0.44	< 0.01	< 0.1	< 0.01	< 0.01
61	155	2.4e-05	3.8e-05	-0.37	< 0.01	< 0.1	< 0.01	< 0.01
62	242	7.7e-06	1.1e-05	-0.32	< 0.01	< 0.1	< 0.01	< 0.01
63	190	4.6e-06	9.2e-06	-0.50	< 0.01	< 0.1	< 0.01	< 0.01
64	294	4.7e-06	6.6e-06	-0.29	< 0.01	< 0.1	< 0.01	< 0.01
65	341	5.8e-06	1e-05	-0.43	< 0.01	< 0.1	< 0.01	< 0.01
66	189	3.2e-06	6.5e-06	-0.50	< 0.01	< 0.1	< 0.01	< 0.01
67	231	1.1e-06	3.6e-06	-0.68	< 0.01	< 0.1	< 0.01	< 0.01
68	164	-8.3e-07	6.9e-06	-1.12	< 0.01	< 0.1	0.47	< 0.01
69	220	1.4e-05	2.7e-05	-0.48	< 0.01	< 0.1	< 0.01	< 0.01
70	233	5.5e-06	9.4e-06	-0.42	< 0.01	< 0.1	< 0.01	< 0.01
71	298	4.3e-06	7.6e-06	-0.43	< 0.01	< 0.1	< 0.01	< 0.01
72	238	1.9e-06	4.2e-06	-0.56	< 0.01	< 0.1	< 0.01	< 0.01
73	217	5e-06	9.8e-06	-0.49	< 0.01	< 0.1	< 0.01	< 0.01
74	193	9e-06	1.6e-05	-0.42	< 0.01	< 0.1	< 0.01	< 0.01
75	210	8.3e-06	1.3e-05	-0.35	< 0.01	< 0.1	< 0.01	< 0.01
76	269	4.9e-06	7.8e-06	-0.37	< 0.01	< 0.1	< 0.01	< 0.01
77	172	9e-06	2e-05	-0.55	< 0.01	< 0.1	< 0.01	< 0.01

Table G.2: Transitivity tests for all village. T_N is the estimated excess transitivity, "T_N bc" is the bias-corrected estimated excess transitivity, "rel bias" = (T_N - T_N bc) / (T_N bc), "pval anal" gives analytical p-value based on Theorem 1, "pval BS" gives rejection based on BS critical value at nominal level $\alpha = 0.1$, "pval naïve" and "pval naïve bc" are p-values for the naïve tests (with and without bias correction) introduced in Section 5.2.

H. Derivatives of bivariate normal probabilities

Let $U = (U_1, U_2)'$ denote a bivariate random vector with zero-mean and covariance matrix

$$V = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

where ρ is a parameter giving the correlation between the marginal normals. Let

$$r(y_1, y_2, \rho) = P(U_1 \leq y_1, U_2 \leq y_2)$$

The formula for conditional distributions of a joint normal gives

$$U_2 \mid U_1 \sim \mathcal{N}(\rho U_1, 1-\rho^2).$$

By a conditioning argument

$$\begin{aligned}
r(y_1, y_2, \rho) =& P(U_1 \le y_1) P(U_2 \le y_2 \mid U_1 \le y_1) \\
=& P(U_1 \le y_1) \int_{-\infty}^{y_1} P(U_2 \le y_2 \mid U_1 = t) \frac{\phi(t)}{\Phi(t)} \, dt \\
=& \int_{-\infty}^{y_1} \Phi\left(\frac{y_2 - \rho t}{\sqrt{1-\rho^2}} \right) \phi(t) \, dt.
\end{aligned}$$

Then,

$$\partial_{y_1} r(y_1, y_2, \rho) = \Phi\left(\frac{y_2 - \rho y_1}{\sqrt{1-\rho^2}} \right) \phi(y_1)$$

and

$$\begin{aligned}
\partial_{y_1 \rho} r(y_1, y_2, \rho) =& -\left(\frac{y_1 - \rho y_2}{(1-\rho^2)^{3/2}} \right) \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1-\rho^2}} \right) \phi(y_1), \\
\partial_{(y_1)^2} r(y_1, y_2, \rho) =& -\frac{\rho}{\sqrt{1-\rho^2}} \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1-\rho^2}} \right) \phi(y_1) + \Phi\left(\frac{y_2 - \rho y_1}{\sqrt{1-\rho^2}} \right) \phi'(y_1) \\
=& -\frac{\rho}{\sqrt{1-\rho^2}} \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1-\rho^2}} \right) \phi(y_1) - y_1 \Phi\left(\frac{y_2 - \rho y_1}{\sqrt{1-\rho^2}} \right) \phi(y_1), \\
\partial_{y_1 y_2} r(y_1, y_2, \rho) =& \frac{1}{\sqrt{1-\rho^2}} \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1-\rho^2}} \right) \phi(y_1).
\end{aligned}$$

Moreover,

$$\partial_\rho r(y_1, y_2, \rho) = \int_{-\infty}^{y_1} \left(\frac{\rho y_2 - t}{(1-\rho^2)^{3/2}} \right) \phi\left(\frac{y_2 - \rho t}{\sqrt{1-\rho^2}} \right) \phi(t) \, dt.$$

The integral on the right-hand side can be solved numerically using the R function `integrate`.[34] For the case $\rho = 0$ no numerical integration is needed since

$$\partial_\rho r(y_1, y_2, 0) = -\phi(y_2) \int_{-\infty}^{y_1} t \, \phi(t) \, dt = \phi(y_1)\phi(y_2).$$

[34]I am grateful to Harry Joe for sharing his thoughts on how to compute this derivative in modern R.