Estimating Nonlinear Network Data Models with Fixed Effects*

David W. Hughes
MIT

November 8, 2021

Abstract

This paper considers estimation of a directed network model in which outcomes are driven by dyad-specific variables (such as measures of homophily) as well as unobserved agent-specific parameters that capture degree heterogeneity. I develop a jackknife bias correction to deal with the incidental parameters problem that arises from fixed effect estimation of the model. In contrast to previous proposals, the jackknife approach is easily adaptable to different models and allows for non-binary outcome variables. Additionally, since the jackknife estimates all parameters in the model, including fixed effects, it allows researchers to construct estimates of average effects and counterfactual outcomes. I also show how the jackknife can be used to bias-correct fixed effect averages over functions that depend on multiple nodes, e.g. triads or tetrads in the network. As an example, I implement specification tests for dependence across dyads, such as reciprocity or transitivity. Finally, I demonstrate the usefulness of the estimator in an application to a gravity model for import/export relationships across countries.

*MIT Economics job market paper. Contact at: dwhughes@mit.edu. I am extremely grateful for the support and advice of Whitney Newey, Anna Mikusheva, and Alberto Abadie. I would also like to thank Isaiah Andrews, Victor Chernozhukov, Claire Lazar Reich, Ben Deaner, and Sylvia Klosin for helpful feedback. All errors are my own.

1 Introduction

Networks are common in both economic and social contexts, and it is important to understand the factors that play a role in both the formation and strength of the links between agents. The econometric analysis of networks faces a number of challenges that have received much attention in recent literature (see de Paula (2020) and Graham (2020) for reviews of this literature). One common modeling approach is to assume a dyadic network structure (one in which decisions are made bilaterally between agents), but allow for linking decisions to depend on unobserved agent-specific heterogeneity. These models are common in practice since they are straightforward to implement while still being able to capture important aspects of observed networks. Controlling for agent-specific heterogeneity is important since in many real world networks agents vary significantly in the number and strength of connections made. Ignoring this heterogeneity can lead to large biases in estimated effects.

In this paper, we consider the estimation of dyadic models, where the presence of unobserved heterogeneity is accounted for by two sets of agent-specific fixed effects – a sender and a receiver effect. The fixed effects approach is appealing as it does not require strong assumptions about the unobserved component as in random effects models. In addition, the network setting does not suffer from the 'fixed T' issues of panel data, since we observe every agent interacting with $N-1$ other agents in the network, so that fixed effect estimates are consistent. However, the large number of fixed effects (proportional to the square root of the sample size) does create an incidental parameter problem (Neyman and Scott, 1948). This paper proposes a jackknife approach to bias correction, which has a number of benefits over existing methods. Importantly, the jackknife is easily adaptable to a range of settings, including models for non-binary outcome variables. Additionally, since the jackknife estimates all of the parameters in the model, including the fixed effects, we are able to construct estimates of average effects and counterfactual outcomes. We also show how the jackknife can be used to bias correct averages over functions of multiple observations (e.g. dyads or triads in the network), which we show is useful for constructing specification test statistics, such as tests for the presence of certain strategic interactions like reciprocity or transitivity.

We demonstrate the consistency and asymptotic normality of the jackknife estimator under asymptotic sequences in which a single network grows in size, while the network remains 'dense'. The network model we consider is one in which agents make bilateral decisions about link-specific outcomes, independently of other relationships. This type of dyadic model (a *dyad* is a pair of agents) has received much attention in the literature, because of its tractability and its ability to replicate some key features of observed networks. In particular, it allows for: *homophily*, the tendency of agents to form stronger ties with other agents that are similar to them; and *degree heterogeneity*, where the number/strength of links in the networks can vary substantially across nodes.

In the case of a binary outcome variable, the model we consider is one of link formation, and is an extension of the model by Holland and Leinhardt (1981). There are several alternative approaches to address the incidental parameters problem in this setting. Graham (2017), Charbonneau (2017), and Jochmans (2018) all consider versions of this model in which the latent disturbances follow a logistic distribution, and use conditioning arguments to remove dependence on the fixed effects. The conditioning approach has the advantage of being applicable under certain *sparse network* asymptotic sequences, but is limited to models in which sufficient statistics for the fixed effects exist, and is not able to recover counterfactuals or average effects. Yan et al. (2019) also studies the logistic model and provides asymptotic results for the incidental parameters. Graham (2017) considers an analytical correction for the logistic model, while Dzemski (2019) derives the analytical correction for a probit model. The analytical bias correction approach is limited to *dense network* sequences, as in this paper, and similarly to this paper can recover average effects. The advantage of the jackknife correction relative to an analytical approach is that it provides an off-the-shelf approach that researchers may apply to new settings, without the need to first derive bias expressions. Candelaria (2020), (Toth, 2017), and Gao (2020) study identification of the common parameters without a known parametric form for the disturbance term, while Zeleneev (2020) allows for nonparametric structure in the unobserved heterogeneity term.

Although the focus of the literature on dyadic network models has been on the binary outcome case, researchers often have access to outcome variables that are non-binary. Examples of these settings include the value of exports between countries, the value of loans between banks, or the number of workers migrating between states. The results in this paper are derived for a general M-estimator satisfying basic regularity conditions and so cover a range of models for both binary and non-binary outcome variables, as well as a range of estimation approaches, including MLE, quasi-MLE and nonlinear least squares estimation.

As a demonstration of the technique in an empirical setting, we estimate a model of international trade relationships. Gravity models have been a workhorse model in the trade literature for many years, and the importance of including country-specific fixed effects is well known ((Anderson and Van Wincoop, 2003)). We estimate the zero-inflated negative binomial model of Burger et al. (2009), which combines both a model for the decision of countries to engage in trade, as well as a model for the *value* of exports conditional on some trade occurring. The model addresses to key issues in the gravity model literature: it allows for a large proportion of zero trade flows in the network, and it captures the observed high dispersion of export values across countries. We obtain bias-corrected estimates of both the model parameters, as well as average effects.

The jackknife bias correction also allows for the construction of various specification tests. Many models of network formation include strategic aspects in which agents' decisions are influenced by the state of the network. For instance, agent i may find it more beneficial to link with j if they already share many other links in common. Graham and Pelican (2020) derives the locally best similar test for a class of alternatives in a logit model, using conditioning arguments. Dzemski (2019) tests for the presence of transitive links with triads (groups of three agents) in a probit model, and derives an analytical bias correction for the statistic. We demonstrate that a range of test statistics, including that of Dzemski (2019), can be bias-corrected using the jackknife. This extends the set of tests available to researchers, as well as the range of models they can be applied to. As an example, we test for reciprocity and transitivity in trade links between countries and find evidence that the decisions of countries to engage in trade are reciprocal (if country i exports to country j then it is likely that j also exports to i), but do not find evidence of transitive relationships.

The network jackknife extends previous results on jackknife bias correction in panel data. Hahn and Newey (2004) introduced a jackknife correction for panel estimators with individual fixed effects, based on re-estimating the parameters on data sets that exclude a single time period. Dhaene and Jochmans (2015) present a split-sample version of this idea based on estimating the model in the first and second halves of time periods separately. Fernández-Val and Weidner (2016) develop a general framework that allows for both time and individual fixed effects. The analysis in this paper builds heavily off of the asymptotic expansions in Fernández-Val and Weidner (2016).

Analogously to the panel data setting, the network jackknife is constructed by forming 'leave-out' estimates that exclude certain subsets of the data. Cruz-Gonzalez et al. (2017) and Chen et al. (2021) have suggested jackknife approaches for network data, although without formal proof, based on either a split-sample approach or a leave-one-out approach that drops a single agent at a time. We propose a different approach to jackknifing network data that is based on a novel partitioning of the data set that constructs leave-out estimates that remove bias from both sender and receiver of fixed effects in one step. We extend the asymptotic expansions of Fernández-Val and Weidner (2016) to allow for formal analysis of the jackknife estimator. The jackknife proposed here drops a single observation per fixed effect at each step, so that our approach is likely to have better finite sample variance properties than the split-sample approach (see Hughes and Hahn (2020) for a formal argument in the panel setting). In contrast to a jackknife that drops all observations from a single agent in the network, our jackknife retains all agents in each leave-out estimation, so that the distribution of unobserved effects is held constant. We demonstrate the small-sample effectiveness of our approach in comparison to previous suggestions in simulations that show that our jackknife is more robust to settings with meaningful levels of unobserved heterogeneity and in networks with lower density.

In addition, we introduce a weighted jackknife, that differs from standard implementations of the jackknife approach by taking a weighted-average of the leave-out estimates. This version puts less weight on noisier leave-out estimates, which improves the finite-sample properties of the jackknife in sparser settings. The weighted jackknife idea may be useful elsewhere, for example in binary-outcome panel data models with few successes for some individuals (so called 'rare events'). Finally, we also introduce a 'leave-l-out' version of the jackknife. This version requires only $(N-1)/l$ additional estimations of the model, and may allow researchers to reduce the computational burden in settings where model estimation is difficult.

The rest of the paper is organized as follows. Section 2 introduces the network model and discusses implementation of the jackknife procedure for the estimation of model parameters, while Section 3 discusses estimation of average effects, and the construction of specification tests. Section 4 provides asymptotic results for the estimators, and discusses the main assumptions under which they hold. In Section 5 we demonstrate the method by estimating a model of international trade flows, while Section 6 reports simulation results that are consistent with the jackknife theory.

2 Dyadic linking model and jackknife correction

2.1 Model

The researcher observes a network of N agents; these agents may, for example, be individuals, firms, or countries. For each potential *directed* connection, $i \to j$, we observe an associated link-specific outcome variable Y_{ij}. The variable Y_{ij} may capture the presence (or absence) of a link between two agents, in which case Y_{ij} is binary, or may represent a measure of the strength of the link between agents. For example, Y_{ij} may be the value of exports from country i to country j in a particular year, or the number of times agents i and j interacted in some period. Links are directed, meaning that $Y_{ij} \neq Y_{ji}$ in general, and so, following the literature, we term i the 'sender' and j the 'receiver' in link Y_{ij}.

The researcher also observes a set of link-specific covariates X_{ij}. The covariates capture characteristics of the relationship between agents that may impact the linking outcome. Often these will be interpreted as measures of *homophily*, that is, the tendency for agents to link with other agents that are similar to themselves. For example, countries may engage in greater levels of trade if they share a common language, or are geographically close.

Agents are endowed with two fixed effects, α_i and γ_i, which capture unobserved *degree heterogeneity*, that is, the tendency of some agents to form more (or stronger) links than others. The 'sender' fixed effect α_i accounts for heterogeneity in out-degree (the number or strength of links from agent i to other agents), while the 'receiver' fixed effect accounts for in-degree heterogeneity. Degree heterogeneity is an important feature of many networks, for example, we would expect countries with larger GDPs to engage in more trade than smaller countries *ceteris paribus* (see Anderson and Van Wincoop (2003) for an example of such a model). Since the network considered here is a directed one, we allow for the sender and receiver fixed effects to differ; some countries may have structural tastes for importing goods over exporting, that is, they run trade deficits (or vice versa), so that $\alpha_i < \gamma_i$. Failure to account for degree heterogeneity in a network can lead to incorrect conclusions about the strength of homophily in a network. For example, observing that the United States imports more from China than from Canada may lead to the conclusion that distance between countries is unimportant for trade if we do not account for a China export effect. Graham (2017) provides some further intuition for why failing to account for degree heterogeneity can bias conclusions about homophily in a network.

We make the assumption that linking decisions are bilateral in nature, so that

$$Y_{ij} \perp\!\!\!\perp Y_{kl} | X, \beta, \alpha, \gamma \qquad \forall (k,l) \notin \{(i,j),(j,i)\}, \tag{1}$$

where $\perp\!\!\!\perp$ denotes independence of the outcomes conditional on observed covariates and fixed effects. This assumption does allow for dependence between the two links within a pair of agents (a dyad), but implies the decision between i and j is independent of that between

i and k for instance. Importantly, this independence is conditional on the covariates and agent-specific fixed effects. Unconditionally, country i's exports to country j are correlated with their exports to country k, since both are determined by the exporter effect α_i. In many settings, the inclusion of fixed effects will be important in establishing the plausibility of (1). Assumption (1) may not be appropriate in situations where linking decisions are strategic. Estimation of models with strategic interactions is substantially more challenging, and is likely to require multiple observations of the network over time. Nonetheless, the dyadic model presented here still represents an important baseline model, and can be used to construct tests for the presence of strategic interactions against the null hypothesis of (1). We discuss examples of such tests in Section 3.2.

We leave the specific model for the network outcomes unspecified, and assume only that the model parameters $(\beta_0, \alpha_0, \gamma_0)$ are solutions to the population maximization problem

$$\max_{(\beta,\alpha,\gamma)\in\mathbb{R}^{\dim\beta+2N}} \bar{E}\big[\mathcal{L}_N(\beta,\alpha,\gamma)\big],$$

$$\mathcal{L}_N(\beta,\alpha,\gamma) = \frac{1}{N-1}\sum_i\sum_{j\neq i}\ell(Y_{ij}, X_{ij}, \beta, \alpha_i+\gamma_j) - \frac{b}{2N}\big(\sum_i \alpha_i - \sum_i \gamma_i\big)^2, \tag{2}$$

where $\bar{E}$ represents expectation conditional on the exogenous covariates and fixed effects, and ℓ is a known function that is maximized in expectation at the true parameters. Many models can be estimated by maximizing objective functions of the form in (2), including MLE, quasi-MLE, and nonlinear least squares estimators. The researcher need only specify the objective function $\mathcal{L}$ and does not need to specify the distribution of the fixed effects, or how they relate to the covariates.

We assume that the unobserved effects enter in an additively separable manner, i.e. as $\alpha_i + \gamma_j$, identification of the two sets of fixed effects parameters requires a normalization, for which we choose

$$\sum_i \alpha_i = \sum_i \gamma_i.$$

The term $\frac{b}{2N}\big(\sum_i \alpha_i - \sum_i \gamma_i\big)^2$ is a penalty term intended to impose this normalization on the fixed effect parameters, where $b > 0$ as an arbitrary constant.[1] Note that the vectors of fixed effects α and γ are dependent on the network size N, although we leave this

[1] In practice the constraint could simply be eliminated by substituting it into the objective. We follow Fernández-Val and Weidner (2016) in choosing this normalization as it is convenient in the proofs.

dependence implicit in the notation. When functions are evaluated at the true parameters $(\beta_0, \alpha_0, \gamma_0)$ we will typically drop them from notation. We will also use the shorthand $\ell_{ij} = \ell(Y_{ij}, X_{ij}, \beta_0, \alpha_{0,i} + \gamma_{0,j})$ when convenient.

Example. ***Maximum likelihood***

If we specify the full conditional distribution of the outcome variable as

$$Y_{ij}|X, \beta, \alpha, \gamma \sim f(Y_{ij}|X_{ij}, \beta, \alpha_i + \gamma_j),$$

then ℓ will be the log-likelihood function

$$\ell(Y_{ij}, X_{ij}, \beta, \alpha_i + \gamma_j) = \log f(Y_{ij}|X_{ij}, \beta, \alpha_i + \gamma_j).$$

Note that $\mathcal{L}$ need not be a true log-likelihood, since it may be that the observations Y_{ij} and Y_{ji} are conditionally dependent, in which case $\mathcal{L}$ is a pseudo log-likelihood.

As an example, consider a model of directed link formation according to

$$Y_{ij} = \mathbb{1}\{X_{ij}'\beta + \alpha_i + \gamma_j - \varepsilon_{ij} \geq 0\},$$

where ε_{ij} follows a known distribution F. This linking rule is compatible with a model in which utility is transferable across linked agents as in Bloch and Jackson (2007). Given the distributional assumption for ε_{ij}, the probability of a link forming, conditional on covariates and fixed effects is $p_{ij} = F(X_{ij}'\beta + \alpha_i + \gamma_j)$, and the log-likelihood is $\ell_{ij} = Y_{ij} \log p_{ij} + (1 - Y_{ij}) \log(1 - p_{ij})$. This is an extension of the linking model of Holland and Leinhardt (1981) and has been used extensively in empirical literature.

Example. ***Nonlinear least squares***

The researcher may choose to specify only the conditional mean function for the outcome, rather than its full distribution, e.g $E\big[Y_{ij}|X, \beta, \alpha, \gamma\big] = h(X_{ij}, \beta, \alpha_i + \gamma_j)$. In this case, we may estimate the parameters of the model by setting

$$\ell(Y_{ij}, X_{ij}, \beta, \alpha_i + \gamma_j) = -\big(Y_{ij} - h(X_{ij}, \beta, \alpha_i + \gamma_j)\big)^2.$$

2.2 Empirical setting - gravity equation

Although much of the focus of the incidental parameters bias-correction literature has been on the binary outcome case, researchers often have access to non-binary outcome variables and will be interested in modeling networks in which the links are weighted. As a working example throughout the paper, we will consider a model of country-level trade relationships using a data set consisting of a directed network of export volumes between 136 countries (136×135 country pair observations) in 1990. The data are taken from Santos Silva and Tenreyro (2006), and additional details on their construction can be found in their paper. The outcome variable is the value of exports from country i to country j. We also use several covariates to capture homophily in trade relationships, which include: *log distance*, the log of the distance between the capitals of the countries; *border*, an indicator of whether the countries share a common border; *language*, an indicator for whether the countries share a language; *colonial*, and indicator for whether either country had colonized the other at some point in history; and *trade agreement*, an indicator for the presence of a joint preferential trade agreement between the two countries.

The Anderson and Van Wincoop (2003) gravity equation expresses the trade volume from country i to country j as

$$Y_{ij} = \alpha_0 G_i G_j d_{ij}^{\beta} e^{\alpha_i + \gamma_j} \tag{3}$$

where Y_{ij} is the trade flow from country i to country j, G_i is GDP of country i, and d_{ij} is inversely proportional to the distance between the two countries (which is generally taken to include all factors that create resistance to trade). The inclusion of exporter and importer fixed effects (α_i and γ_i) is intended to control for multilateral resistance terms, which may bias results if excluded.

A simple method for estimating the parameters in (3) is to first log-linearize the model. Unfortunately, this raises the issue of how to deal with the presence of zero outcomes that are common in trade data. In the country-level trade data introduced above, just under half of all country pairs engage in no trade. A number of solutions to this problem have been suggested. Several authors use Tobit models or two-step Heckman style models, which combine a binary selection equation (predicting whether or not any trade occurs) and a separate equation for the value of trade (conditional on selection); see for example Helpman et al. (2008), Rose (2004), Linders and de Groot (2006).

Another popular approach, suggested by Santos Silva and Tenreyro (2006), is to use a Poisson pseudo-maximum-likelihood estimator, which provides a natural way to incorporate zero-valued outcomes, as well as being robust to heteroskedasticity issues that can arise when log-linearizing multiplicative equations. However, there are two key drawbacks to modeling trade flows with a Poisson distribution. Firstly, the proportion of zeroes observed in typical trade data is much larger than that predicted by a Poisson model. Secondly, since the variance of a Poisson is restricted to be equal to its mean, outcomes are typically much more dispersed than would be expected under the Poisson model. In order to address these two issues, Burger et al. (2009) propose a zero-inflated negative binomial model, in which the value of trade between i and j is given by the product of two variables, $Y_{ij} = z_{ij} Y_{ij}^*$, where $z_{ij} \in \{0, 1\}$ is a binary decision to enter into a trading relationship, while Y_{ij}^* is the value of exports that will be realized, conditional on $z_{ij} = 1$. The binary decision is modeled using as a probit function, while the latent outcome Y_{ij}^* is modeled as a negative binomial variable, which allows for overdispersion in the model for Y_{ij}^*, that is, it allows the variance to differ from the mean. Since the distribution of Y_{ij} is parametrically specified, we may estimate the model using maximum likelihood, so that this model is captured by the framework in (2).

2.3 Incidental parameters problem

In total, the model contains $\dim(\beta) + 2N$ parameters to be estimated, from the $N(N-1)$ observations (Y_{ij}, X_{ij}') — we will typically refer to these observations using the shorthand (i, j). As is well known, nonlinear estimators with fixed effects suffer from an incidental parameter problem (Neyman and Scott, 1948). To describe the problem, consider the maximization problem (2) after first concentrating out the fixed effect parameters

$$\widehat{\alpha}(\beta), \widehat{\gamma}(\beta) = \arg\max_{\alpha,\gamma} \mathcal{L}_N\big(\beta, \alpha, \gamma\big),$$

$$\widehat{\beta} = \arg\max_{\beta} \mathcal{L}_N\big(\beta, \widehat{\alpha}(\beta), \widehat{\gamma}(\beta)\big). \tag{4}$$

Replacing the population functions $\alpha(\beta), \gamma(\beta) = \arg\max_{\alpha,\gamma} \bar{E}\big[\mathcal{L}_N(\beta, \alpha, \gamma)\big]$ with their sample values, results in an objective function that is biased, in the sense that

$$\beta_0 \neq \beta_N = \arg\max_{\beta} \bar{E}\big[\mathcal{L}_N\big(\beta, \widehat{\alpha}(\beta), \widehat{\gamma}(\beta)\big)\big]. \tag{5}$$

To see why, observe that the first-order condition for $\widehat{\alpha}_i(\beta)$ depends only on the $N-1$ observations (i, j) for $j \neq i$. Similarly, the first-order condition for $\widehat{\gamma}_i(\beta)$ depends on the

$N-1$ observations (j,i) for $j \neq i$. Expanding $\widehat{\alpha}_i(\beta)$ around $\alpha_i(\beta)$ (and similarly for γ) will therefore result in a bias of order $O(N^{-1})$. Under regularity conditions discussed in Section 4, we show that the bias of the maximizer of the profile objective function (5) is approximately given by

$$\beta_N - \beta_0 \approx \frac{B_N}{N-1} \tag{6}$$

for some bias term B_N. Analogously to the panel literature, the bias is inversely proportional to the number of observations used to estimate each of the fixed effect parameters; the exporter effect for country i is estimated using the data on the $N-1$ other countries in the network that i may export to. As the size of the network grows, $N \to \infty$, we will have that $\beta_N \to \beta_0$ so that parameter estimates are consistent. Considering $\widehat{\beta}$ as an estimator for β_N, we can show that $N(\widehat{\beta} - \beta_N) \Rightarrow \mathcal{N}(0, V)$. However, since the bias $\beta_N - \beta_0$ is of the same order as the estimation error, $O(N^{-1})$, $\widehat{\beta}$ will be *asymptotically biased*, that is

$$N(\widehat{\beta} - \beta_0) \Rightarrow \mathcal{N}\big(B, V\big).$$

The incidental parameters generate an *asymptotic bias* in the network model analogous to the panel setting with both N and T growing to infinity at the same rate. Similar asymptotic expansion arguments have been used in the panel data literature on nonlinear models with fixed effects. Hahn and Newey (2004) derive expansions for models with individual fixed effects, while Fernández-Val and Weidner (2016) derive expansions that apply to general models with additively separable unobserved effects, and Chen et al. (2021) consider the setting with interactive effects. The expansions used in this paper rely heavily on these prior results. Dzemski (2019) also applies the Fernández-Val and Weidner (2016) expansions to the network model structure to derive bias expressions for a probit model. In this paper, we extend the asymptotic expansions to higher order so that they may be used to justify jackknife bias correction procedures. We demonstrate that, under mild additional regularity conditions, the jackknife estimator introduced below is asymptotically normal and mean zero, so that valid inference can be performed on model parameters.

2.4 Jackknife bias correction

The jackknife bias-corrected estimator is constructed as a linear combination of the full-sample parameter estimates and an average of 'leave-out' estimators that exclude certain observations in the data set. The particular linear combination chosen can be motivated by asymptotic expansions of the estimator, in particular the form of the bias in (6).

Suppose we were to drop observations from our data set in such a way that for every country i we exclude one observation in which i is the exporter, and one observation in which i is the importer (recall that a single observation is one export-import relationship, of which we observe $N(N-1)$ in total). We show that the new estimator using this 'leave-out' sample, $\tilde{\beta}$, has a bias that is approximately

$$\bar{E}[\tilde{\beta}] - \beta \approx \frac{B_N}{N-2}.$$

The form of the bias for $\tilde{\beta}$ can be explained by two important factors: (i) $\tilde{\beta}$ is estimated using only $N-2$ observations per fixed effect, since we excluded one observation related to each fixed effect parameter, and (ii) $\tilde{\beta}$ is estimated on a random sample generated from the same set of fixed effects, so that the bias expression B_N is the same as that in (6).

Taking advantage of the fact that the estimate $\tilde{\beta}$ has a larger bias than the full-sample estimate $\widehat{\beta}$ by the factor $\frac{N-1}{N-2}$, we can construct a new estimator $\widehat{\beta}_{jack} = (N-1)\widehat{\beta} - (N-2)\tilde{\beta}$ which has no asymptotic bias

$$\bar{E}\big[(N-1)\widehat{\beta} - (N-2)\tilde{\beta}\big] - \beta \approx 0.$$

To describe the construction of the leave-out estimators, we first define a partition of the $N(N-1)$ observations of directed pairs (i,j) into $N-1$ sets of the form[2]

$$\mathcal{I}_k = \big\{(i,j) : j = (i+k) \pmod N\big\}, \quad \text{for } k = 1, \dots, N-1$$

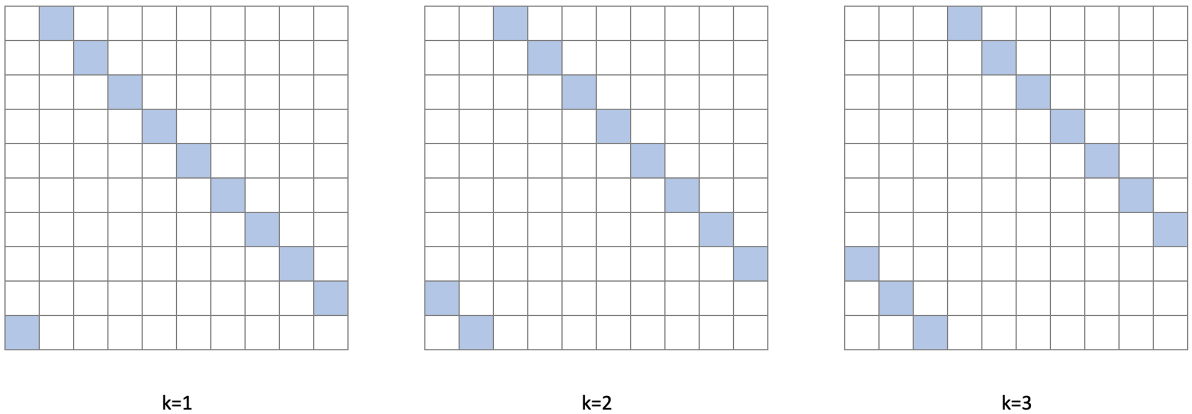
that is, the set of directed pairs $\{(1, 1+k), \dots, (N-k, N), (N-k+1, 1), \dots, (N, k)\}$.[3]

Figure 1 represents the structure of the first three sets $\mathcal{I}_k$ for a network of $N = 10$ agents. Observations are ordered in an $N \times N$ matrix so that the (i,j) cell represents the corresponding observation in the network (the diagonal elements are empty since there are no (i,i) observations). The leave-out sets take diagonal sections from the data matrix. Importantly, constructing the sets this way ensures that each contains exactly one observation related to each sender and receiver fixed effect; i.e. there is one observation taken from every row and every column.

[2]In the modulo notation we consider agent N equivalent to agent 0.

[3]The construction of the leave-out sets assumes that the labelling of the nodes is arbitrary. This will generally be true, but the researcher may ensure this by randomizing the ordering of nodes prior to estimation.

Figure 1: Diagram of leave-out sets $\mathcal{I}_k$ for $k = 1, 2, 3$



Observation (i, j) in the network is represented by the corresponding position in each matrix. The blue squares are the observations contained in the leave-out sets $\mathcal{I}_k$.

Let $1_{ij}^k = \mathbb{1}\{(i, j) \notin \mathcal{I}_k\}$, be an indicator variable that is equal to one whenever the observation (i, j) is *not* in the k-th leave-out set. The k-th leave-out estimates are

$$(\widehat{\beta}_{(k)}, \widehat{\phi}_{(k)}) = \arg \max_{(\beta,\phi)\in\Theta} \frac{1}{N-2} \sum_i \sum_{j\neq i} \ell_{ij}(\beta, \phi) \times \mathbb{1}_{ij}^k, \tag{7}$$
$$\text{subject to } \sum_i \alpha_i = \sum_i \gamma_i,$$

that is, the estimates obtained by excluding the observations in $\mathcal{I}_k$ from the data. We can then construct the jackknife bias-corrected estimator

$$\widehat{\beta}_J = (N-1)\widehat{\beta}_N - (N-2)\frac{1}{N-1}\sum_{k=1}^{N-1} \widehat{\beta}_{(k)}. \tag{8}$$

The construction of the leave-out estimators is analogous to jackknife bias correction in the panel data setting; however, the structure of the jackknife proposed here is new. The procedure relies on dropping sets of observations that contain a single observation related to every sender fixed effect α_i as well as every receiver fixed effect γ_i. In this way, the bias from both types of fixed effects can be addressed simultaneously, while holding the distribution of fixed effects constant across the leave-out samples. This is in contrast to an approach which drops all observations from a single agent, which removes that agents' fixed effects from the leave-out sample and alters the distribution of unobserved heterogeneity. We show in simulations that the method proposed here is more robust to networks that have more unobserved heterogeneity or are less dense.

We prove in Section 4 that the jackknife bias correction is consistent, and asymptotically normal, with mean zero and variance equal to that of the full-sample estimate

$$N(\widehat{\beta}_J - \beta_0) \Rightarrow \mathcal{N}\big(0, V\big).$$

The fact that the jackknife is able to remove bias without affecting the asymptotic variance of the estimator may seem surprising. This important feature is achieved by averaging across the $N-1$ different leave-out estimators $\widehat{\beta}_{(k)}$. Since the sets $\mathcal{I}_1, \ldots, \mathcal{I}_{N-1}$ form a partition of the $N(N-1)$ observations in the network, each observation is excluded from exactly one of the leave-out estimates. This balanced treatment of observations ensures that the jackknife procedure does not affect the first-order asymptotic approximation of the estimator.[4]

Remark 1. The construction of the leave-out sets depends on the labelling of nodes, and so the final estimator will be dependent on this labelling (since the make-up of the leave-out sets will change). While the researcher could re-randomize the node labels, construct $\widehat{\beta}_J$ for each randomization, and then average to remove some of the arbitrariness of the node labels, this is not necessary. The estimators with different labelings should be very similar so that the additional computations will have little effect.[5]

The jackknife estimator requires N estimations of the model, and so may be computationally intensive for large networks, although speed may be improved by computing the leave-out estimates in parallel, and using good starting values such as the full sample estimates. As an alternative, we present a 'leave-l-out' version of the jackknife, which reduces the number of additional estimations of the model by dropping l observations per fixed effect, as opposed to just one. To describe the estimator, let $N_l = \frac{N-1}{l}$ (we assume here that $N-1$ is divisible by l for simplicity). We can construct the k-th leave-l-out set by combining l of the leave-one-out sets as follows $\mathcal{I}_k^l = \cup_{j=0}^{l-1} \mathcal{I}_{k+jN_l}$, for $k = 1, \ldots, N_l$. This results in $N_l = \frac{N-1}{l}$ non-overlapping leave-l-out sets, with corresponding estimates $\hat{\beta}_{l,(k)}$, which are the estimates from using all

[4] In the panel data setting, Dhaene and Jochmans (2015) note that forming a jackknife using overlapping subpanels (across time) results in an inflation of the asymptotic variance, since some time periods are used more that others. That the $\mathcal{I}_k$ form a partition ensures that each observation is used an equal number of times in the average $\frac{1}{N-1}\sum_{k=1}^{N-1} \widehat{\beta}_{(k)}$.

[5] In simulations (see Section 6 for the design) the estimates based on different labelings were close to identical and so the additional calculations had almost no effect on the estimation.

observations except those in the k-th leave-l-out set $\mathcal{I}^l_k$. A jackknife bias-corrected estimate can then be constructed as

$$\hat{\beta}_{J,l} = \frac{N-1}{l}\hat{\beta}_N - \frac{N-1-l}{l}\frac{1}{N_l}\sum_{k=1}^{N_l}\hat{\beta}_{l,(k)}, \tag{9}$$

where $N_l = \frac{N-1}{l}$.[6]

Remark 2. The leave-l-out jackknife bias correction has the same asymptotic variance as the standard leave-one-out jackknife and the full-sample estimator. However, there may be some finite-sample efficiency loss, particularly when l is large or when the network is not sufficiently dense. Hughes and Hahn (2020) show in the panel case that the leave-one-out jackknife is higher-order more efficient than the split-sample jackknife (i.e. its variance to $O(N^{-1})$ is smaller), and it is likely that the same result applies here, although this is beyond the scope of the present paper.

2.5 A weighted jackknife

The jackknife relies on large dense network asymptotics, but in finite samples it is possible for some leave-out estimates to drop a number of important observations all at once. This is more likely to occur when N is small, there are few links for some nodes, or when we are using the leave-l-out jackknife with large l.

The performance of the jackknife can be improved in these settings by taking a weighted average of the estimates $\widehat{\beta}_{(k)}$. Define the weights

$$\widehat{W}_{(k)} = -\frac{1}{N}\Big(\partial_{\beta\beta'}\widehat{\mathcal{L}}_{(k)} - \frac{1}{N}(\partial_{\phi\beta'}\widehat{\mathcal{L}}_{(k)})(\partial_{\phi\phi'}\widehat{\mathcal{L}}_{(k)})^{-1}(\partial_{\beta\phi}\widehat{\mathcal{L}}_{(k)})\Big).$$

[6] In the case that $N-1$ is not divisible by l, let $N_l = \lfloor \frac{N-1}{l} \rfloor$. Then we may partition the data into $r = N-1-lN_l$ leave-$(l+1)$-out sets and $N_l - r$ leave-l-out sets. Denote $\hat{\beta}_{(k)}$ as the leave-$(l+1)$-out estimates for $k = 1, \ldots, r$ and the leave-l-out estimates for $k = r+1, \ldots, N_l$. Then, let

$$\bar{\beta} = \frac{1}{(N-1)(N_l-1)}\Big(\sum_{k=1}^{r}(N-l-2)\hat{\beta}_{(k)} + \sum_{k=r+1}^{N_l}(N-l-1)\hat{\beta}_{(k)}\Big)$$

The jackknife bias-corrected estimator is then given by

$$\hat{\beta}_J = N_l\hat{\beta}_N - (N_l - 1)\bar{\beta}$$

The weighted-jackknife estimator is given by

$$\widehat{\beta}_{wJ} = (N-1)\widehat{\beta}_N - (N-2)\bar{W}_J^{-1}\Big(\frac{1}{N-1}\sum_{k=1}^{N-1}\widehat{W}_{(k)}\widehat{\beta}_{(k)}\Big), \tag{10}$$

where $\bar{W}_J = \frac{1}{N-1}\sum_{k=1}^{N-1}\widehat{W}_{(k)}$.

The weights $\widehat{W}_{(k)}$ are the Hessian for β, after concentrating out the fixed effects. In the special case that $\mathcal{L}$ is a log-likelihood function, W_N is the Fisher information for β, and so is equal to the inverse of the asymptotic variance. In this case, we are using an inverse variance weighting scheme, which down-weights leave-out samples that produce particularly noisy estimates of the common parameters. The weighting scheme is equally applicable to non-likelihood settings, although it no longer carries the inverse variance interpretation. In simulations, this weighted version of the jackknife significantly improves the performance of the estimator in sparser networks (see Section 6 for more details).

Remark 3. Asymptotically the weights have no effect on the estimator, since all $\widehat{W}_{(k)}$ converge to the same quantity. This implies that the asymptotic variance of the weighted jackknife is the same as that of the standard jackknife. In finite samples, variation in the weights depends on the number of nodes N, as well as the density of the network (i.e. the variation in outcomes for each node). The weighting scheme is likely to have a large impact for small or less dense networks, but in denser (or larger) networks we will have $\bar{W}_J^{-1}\widehat{W}_{(k)} \approx I_{\dim\beta}$, so that the weighted and unweighted jackknife estimates are very similar.

Remark 4. The motivation behind the particular choice of weights comes from the first-order asymptotic expansion of the estimator. A Taylor expansion of the first-order conditions of the objective function gives an expression of the form

$$W_N(\widehat{\beta} - \beta) \approx A + B,$$

where A is mean zero and asymptotically normal with variance $\bar{\Omega}$ (as in Theorem 1), while B is an additional term responsible for the asymptotic bias of the estimator.

As demonstrated in Lemmas 3 and 4 in the Appendix, the jackknife procedure applied to

$A + B$ successfully demeans B and leaves A unchanged, i.e.

$$(N-1)A - (N-2)\frac{1}{N-1}\sum_{k=1}^{N-1} A_{(k)} = A,$$
$$\bar{E}\Big[(N-1)B - (N-2)\frac{1}{N-1}\sum_{k=1}^{N-1} B_{(k)}\Big] = 0.$$

This results in an estimator with no asymptotic bias and unchanged asymptotic variance. In practice however, the jackknife is applied to $\widehat{\beta}$, so we must consider the effect of the jackknife procedure on $W_N^{-1}(A+B)$.

The validity of the jackknife relies on the fact that $W_{(k)} \approx W_N$, which is guaranteed in large samples under our assumptions, in particular, dense network asymptotics. However, in finite samples, $W_{(k)}$ could vary substantially in some leave-out samples when N is small, there are few links for some nodes, or we are using the leave-l-out jackknife with large l. This motivates instead averaging over $\widehat{W}_{(k)}\widehat{\beta}_{(k)}$, so that variation in $W_{(k)}^{-1}$ has less impact on the quality of the asymptotic approximation. Intuitively, we are jackknifing the first-order condition for $\widehat{\beta}$, rather than $\widehat{\beta}$ itself.[7]

3 Estimating average effects

In addition to estimation of the common parameter β, researchers may also be interested in estimating certain averages over the distribution of exogenous regressors and fixed effects. An important advantage of the jackknife bias correction, over methods based on conditioning on sufficient statistics (e.g. Graham (2017), Jochmans (2018)), is that by estimating the fixed effect parameters we are able to construct estimators for these averages. Common examples include average and marginal effects, as well as counterfactual outcomes. In the network setting, these are averages over functions of a single potential link (i, j) in the network. We will additionally show that averages over multiple links also provide interesting objects of interest; for example, averages over dyads $\{(i, j), (j, i)\}$, triads (groups of three nodes), or other network patterns. As an example, we focus on how these objects can be used to

[7]Of course, W^{-2}, W^{-3} and similar terms also appear in high-order parts of the expansion. The validity of the large-network approximation $W_{(k)}^{-1} \approx \bar{W}^{-1}$ is still necessary for the jackknife to be consistent and asymptotically normal. The weighting scheme simply aims to improve the finite-sample properties of the estimator.

construct tests of the assumption of independent link formation stated in (1), but they may have wider relevance in empirical work. Estimation of the many fixed effect parameters means that these averages also suffer from an incidental parameter problem. We show that the jackknife can be used to bias-correct average effects estimates and obtain correct inference.

3.1 Averages over single observations

A simple fixed effect average may be expressed as

$$\begin{aligned} \delta &= E\big[\Delta_N(\beta_0, \phi_0)\big], \\ \Delta_N &= \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} m(X_{ij}, \beta, \alpha_i + \gamma_j), \end{aligned} \tag{11}$$

where the expectation is taken over the joint distribution of covariates X_{ij} and fixed effects (α_i, γ_j), and the function m represents the effect of interest. Here we specify two possible parameters of interest, δ the population average, and Δ_N, the sample average; this choice will affect the asymptotic distribution of the estimator, a point we return to in Section 4. As earlier, we will impose that the fixed effects enter the function m in an additively separable way, as $\pi_{ij} = \alpha_i + \gamma_j$; this will imply that the choice of fixed effect normalization will not affect the estimator.

Example. ***Marginal effect***

As an example, consider a binary outcome model with $P(Y_{ij} = 1|X, \beta, \alpha, \gamma) = F(\beta X_{ij} + \alpha_i + \gamma_j)$. We may be interested in estimated the average partial effect of the covariate, in which case we would have

$$m(X_{ij}, \beta, \alpha_i + \gamma_j) = \beta \frac{\partial}{\partial X} F(\beta X_{ij} + \alpha_i + \gamma_j).$$

Alternatively, we may be interested in the average partial effect at some fixed value of $X_{ij} = x$, in which case, $m(X_{ij}, \beta, \alpha_i + \gamma_j) = \beta \frac{\partial}{\partial X} F(\beta x + \alpha_i + \gamma_j)$.

Example. ***Counterfactual change***

Alternatively, assume that we estimate the conditional mean function $E[Y_{ij}|X_{ij}, \beta, \phi] = h(\beta X_{ij} + \alpha_i + \gamma_j)$. We may be interested in the counterfactual change in predicted outcome from a change in the value of the covariate X_{ij} from x_0 to x_1, e.g. the effect of entering or exiting a trade agreement. In this case

$$m(X_{ij}, \beta, \alpha_i, \gamma_j) = h(\beta x_1 + \alpha_i + \gamma_j) - h(\beta x_0 + \alpha_i + \gamma_j).$$

The average effect in (11) can be estimated by plugging in estimates of the model parameters

$$\widehat{\Delta}_N = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} m(X_{ij}, \widehat{\beta}, \widehat{\alpha}_i, \widehat{\gamma}_j).$$

As with estimates of the parameters themselves, the average effect estimate $\widehat{\Delta}_N$ is asymptotically biased, that is, $N(\widehat{\Delta}_N - \Delta_N)$ converges to a normal distribution that is not centered at zero. The asymptotic bias in $\widehat{\Delta}_N$ stems from three sources: (i) bias in the common parameter estimates $\widehat{\beta}$, (ii) averaging over a nonlinear function of noisy fixed effect estimates (a Jensen inequality type bias), and (iii) correlation between the fixed effect errors and $m(X_{ij}, \beta, \alpha_i, \gamma_j)$.

The average effect estimator can be bias-corrected using the jackknife in an almost identical way to the bias correction of $\widehat{\beta}$. Let

$$\widehat{\Delta}_{(k)} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} m(X_{ij}, \widehat{\beta}_{(k)}, \widehat{\alpha}_{(k),i}, \widehat{\gamma}_{(k),j})$$

be the average effect estimate that uses the parameter estimates in the k-th leave-out sample (that is, the sample which we drop observations $(i, j) \in \mathcal{I}_k$). A jackknife bias-corrected estimator is

$$\widehat{\Delta}_J = (N-1)\widehat{\Delta}_N - (N-2)\frac{1}{N-1} \sum_{k=1}^{N-1} \widehat{\Delta}_{(k)}. \tag{12}$$

We show in Section 4 that the bias-corrected estimator is asymptotically normal with mean zero. Note that there is no need to use a bias-corrected estimate of $\widehat{\beta}$ in the construction of the average effects. The jackknife takes care of the bias generated by bias in $\widehat{\beta}$ as well as the other sources of bias in a single step.

3.2 Specification testing

The parameter in (11) is an average over a function that depends on a single observation in the network. In some cases, we may be interested in averages over functions that depend on multiple observations such as patterns depending on pairs of nodes (dyads), groups of three or four nodes (triads or tetrads), or other structures. Although these averages may be of interest in their own right, they prove particularly useful in developing specification tests, and we focus on this case. In this section, we show that like simple average effects, these averages also suffer from the incidental parameters bias problem, but can be bias corrected using the jackknife approach.

Let λ be a set of observations in the network; for example, $\lambda = \{(i,j),(j,i)\}$ collects the two observations within a dyad, and $\lambda = \{(i,j),(j,k),(k,i)\}$ collects a sequence of potential links between three nodes. Let Λ_N be the set of all possible λ formed by permuting the nodes for a network of size N. We consider averages of the form

$$\Delta_N = \frac{1}{|\Lambda_N|}\sum_{\lambda} m(Y_\lambda, X_\lambda, \beta, \pi_\lambda), \tag{13}$$

where $Y_\lambda = \{Y_{ij}\}_{(i,j)\in\lambda}$, $X_\lambda = \{X_{ij}\}_{(i,j)\in\lambda}$, and $\pi_\lambda = \{\alpha_i + \gamma_j\}_{(i,j)\in\lambda}$ collect the outcomes, covariates and fixed effects for the observations in λ. These generalize the averages in (11) in two ways: (i) they allow for averages over functions of multiple observations in the network, and (ii) they allow the function m to depend on the outcome variable Y_{ij}.

One important application of the type of averages in (13) is to specification testing. The model presented in this paper assumes that decisions are made bilaterally, that is, agents i and j decide on Y_{ij} independently of other outcomes in the network. In some settings, we may expect that decision making has some strategic aspect, in that an agent's utility from a link depends on the presence (or strength) of other links. One way to model such a phenomenon is to include network statistics in the utility function. For example, imagine that i sends a link to j according to

$$Y_{ij} = \mathbb{1}\big\{\delta S_{ij} + \beta' X_{ij} + \alpha_i + \gamma_j \geq \varepsilon_{ij}\big\}, \tag{14}$$

where S_{ij} is the value of some network statistic, and ε_{ij} are independent shocks. Models of this form generally result in multiple equilibria, which raises a number of difficulties for estimation. However, under the null hypothesis $H_0 : \delta = 0$, the model is the dyadic link formation model considered in this paper and can be consistently estimated. This suggests that a test statistic based on S_{ij} may be useful for testing the null hypothesis of the dyadic model (i.e. assumption (1)) against an alternative of the form (14). One possible test statistic is

$$T_N = \frac{1}{N(N-1)} \sum_i \sum_{j \neq i} (Y_{ij} - p_{ij}) S_{ij}, \tag{15}$$

where $p_{ij} = E[Y_{ij}|X, \beta, \alpha, \gamma]$. The statistic tests the ability of the network statistic S_{ij} to predict the model errors $Y_{ij} - p_{ij}$. Under the null model, $E[(Y_{ij} - p_{ij})S_{ij}] = 0$ and so values of T_N far from zero suggest the presence of unexplained strategic interactions. Graham and Pelican (2020) consider a similar setup in the case of a logit model and derive the locally best conditional similar test for the null hypothesis $H_0 : \delta = 0$ in (14). The resulting statistic is similar to (15), although this certainly does not imply any optimality of the test proposed here. The motivation for the statistic suggested here is heuristic, but has the advantage of being applicable to a wide set of models (e.g. models that do not admit a sufficient statistic) and has the simplicity of using asymptotic critical values. Some examples of potential test statistics in this framework may be useful.

Example. Reciprocity in link formation

Consider a model in which links are reciprocal, that is the presence of a link from j to i increases the utility of the reverse link from i to j. In this case we could let $S_{ij} = Y_{ji}$ in (15), which gives the statistic

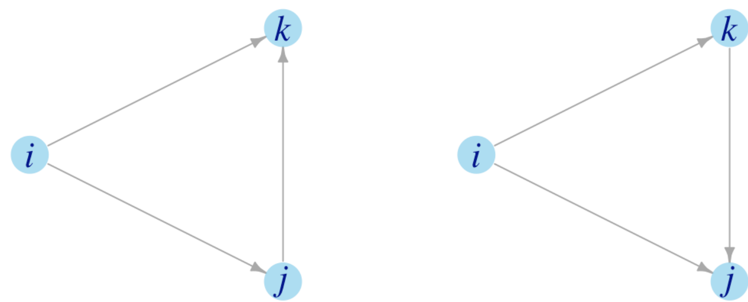
$$T_N = \frac{1}{N(N-1)} \sum_i \sum_{j \neq i} (Y_{ij} - p_{ij}) Y_{ji} \tag{16}$$

This statistic measures whether the average prediction error for reciprocal links differs from the average prediction error for non-reciprocal links. Note that reciprocity is allowed for under the assumptions of this paper, so that a rejection of the null hypothesis of no reciprocity does not affect the interpretation of model estimates.

Example. Transitivity

Consider a *triad,* three nodes in the network (i, j, k). Under the dyadic model, linking decisions are independent across the three pairs of nodes in the triad, but in many settings it may be reasonable to think that the existence of links between two of these pairs may affect the formation of the links in the third. For example, imagine that i has formed a directed link to j, and j has formed a link to k. We may expect the existence of the indirect path from i to k (passing through j) to increase the likelihood of observing the direct link from i to k. This linking structure, $i \to j, i \to k$,$k \to j$ (shown in the right diagram of Figure 2), is known as a *transitive triangle.*

Figure 2: Transitive triangles



There are six potential transitive triangles that may exist within any triad; the figure above shows two of these (in which i is the sender of two links). Additional links may exist within the triad - these do not affect the existence of a transitive triangle.

Transitivity in linking is a feature of many models of strategic network formation and we may test the null hypothesis that linking decisions are dyadic against an alternative in which transitivity exists by choosing $S_{ij} = \frac{1}{N-2}\sum_{k\neq\{i,j\}} Y_{ik}Y_{kj}$ and using the statistic

$$T_N = \frac{1}{N(N-1)(N-2)} \sum_{i} \sum_{j\neq i} \sum_{k\neq\{i,j\}} (Y_{ij} - p_{ij})Y_{ik}Y_{kj}. \tag{17}$$

Note that in both examples, and more generally, the outcome Y_{ij} need not be binary. The test applies equally to non-binary outcomes, for example, in a trade network the presence of large export flows of a particular good from i to k and from k to j may reduce the expected direct exports of that good from i to j.

The framework in (14) generates just one possible set of specification tests, and many others statistics of the form (13) are possible. One alternative method is to compare the observed frequency of some possible subgraph configuration with the expected frequency under the assumed dyadic model. Such a test, for the case of transitive and cyclic triangles, was proposed by Dzemski (2019), who also derived an analytical bias correction for the statistic in a binary outcome model with normal disturbances. The statistic suggested by Dzemski (2019) is of the form

$$T_N = \frac{1}{N(N-1)(N-2)} \sum_{i=1}^{N} \sum_{j\neq i} \sum_{k\neq\{i,j\}} \Big(Y_{ij}Y_{ik}Y_{kj} - p_{ij}p_{ik}p_{kj}\Big),$$

where $Y_{ij}Y_{ik}Y_{kj}$ is an indicator for a transitive triangle and $p_{ij}p_{ik}p_{kj}$ is the probability of observing such a triangle when all three links are independent. Many test statistics of this form could be derived in the same way, by taking some function of network outcomes between multiple agents and comparing it to its expectation under the dyadic model. For example, an alternative statistic for reciprocity would be $\frac{1}{N(N-1)}\sum_i \sum_{j\neq i}(Y_{ij}Y_{ji} - p_{ij}p_{ji})$. A key advantage of the jackknife procedure proposed in this paper is that it applies to such a wide variety of statistics. Given the plethora of potential test statistics, an analysis of their power properties would certainly be useful, although is beyond the scope of this paper.

Like estimates of the common parameter β, the test statistics discussed above suffer from an incidental parameter bias. Although the infeasible test statistics are mean zero under a correctly specified model, the feasible versions, which replace parameters $\beta_0, \alpha_0, \gamma_0$ with their estimated values, have an asymptotic bias that leads to incorrect inference. However, we show in Section 4 that statistics of the form in (13), may be jackknife bias corrected.

To describe the jackknife bias correction for these statistics, denote the number of observations contained in λ as r. Let $\mathbb{1}_\lambda^k = \prod_{(i,j)\in\lambda} \mathbb{1}_{ij}^k$ be an indicator that is *zero* whenever any of the observations in λ are included in the k-th leave-out set $\mathcal{I}_k$. Define the leave-out estimate

$$\widehat{\Delta}_{(k)} = \frac{N-1}{N-r-1}\frac{1}{|\Lambda_N|}\sum_\lambda m(Y_\lambda, X_\lambda, \widehat{\beta}_{(k)}, \widehat{\pi}_{\lambda,(k)}) \times \mathbb{1}_\lambda^k,$$

where $\widehat{\beta}_{(k)}, \widehat{\pi}_{\lambda,(k)}$ are parameter estimates from the k-th leave-out estimation. The factor $\frac{N-1}{N-r-1}$ accounts for the fact that m_λ is dropped from the average whenever any of the r observations it is a function of are dropped.[8] A jackknife bias-corrected estimator is again given by

$$\widehat{\Delta}_J = (N-1)\widehat{\Delta}_N - (N-2)\frac{1}{N-1}\sum_k \widehat{\Delta}_{(k)}. \tag{18}$$

In Section 4, we show that the bias-corrected statistic is asymptotically normal and well centered. In the case of the specification test statistics discussed above, we have $\bar{E}[\Delta_N] = 0$ and so $N\widehat{\Delta}_J \Rightarrow N(0, V_\Delta)$, where the form of the variance is shown in Theorem 2. This allows us to test hypotheses in the usual way, comparing $N\widehat{\Delta}_J/\sqrt{V_\Delta}$ to the quantiles of a standard normal distribution. Further details on the implementation of these tests are

[8]This jackknife differs from (12) by jackknifing the average itself, not just the parameter estimates. This is because m here may depend on the outcome Y_{ij}. In settings where m does not depend on outcomes, or it is separable between outcomes and parameters, the simpler jackknife in (12) can be used. For example, when $m_\lambda = Y_{ij}Y_{ik}Y_{kj} - p_{ij}p_{ik}p_{kj}$, the average over the second term ($p_{ij}p_{ik}p_{kj}$) may be jackknifed separately and the average over the first term ($Y_{ij}Y_{ik}Y_{kj}$) left as is.

discussed below.

4 Asymptotic theory

We consider an asymptotic framework in which a single network of N agents grows in size, i.e. $N \to \infty$. Recall that the parameters of interest maximize the objective function in (2) for some function $\ell_{ij} = \ell(Y_{ij}, X_{ij}, \beta, \alpha_i + \gamma_j)$ of the observables $Z_{ij} = (Y_{ij}, X_{ij})$ and additive unobserved fixed effects $\alpha_i + \gamma_j$. The asymptotic theory relies on expansions of the objective function, for which we require certain differentiability and moment conditions. We let $\phi' = (\alpha', \gamma')$ denote the $2N \times 1$ vector of fixed effect parameters and $\pi_{ij} = \alpha_i + \gamma_j$ represent the additive index through which they enter the objective function. We denote derivatives of the function ℓ with respect to parameters by $\partial_\beta \ell_{ij}(\beta, \alpha, \gamma) = \partial \ell_{ij}(\beta, \alpha, \gamma)/\partial \beta$, $\partial_{\pi^q} \ell_{ij}(\beta, \alpha, \gamma) = \partial^q \ell_{ij}(\beta, \alpha, \gamma)/\partial \pi^q$ etc. When evaluating these objects at the true parameter values, we simply write $\partial_{\pi^q} \ell_{ij}$ and so on.

4.1 Asymptotic analysis for the common parameters

The results below are derived under the following set of assumptions. Proofs are provided in the Appendix.

Assumption 1. *Let $\varepsilon > 0$. For every (i,j) let $\mathcal{B}_{\varepsilon,ij}$ be a subset of $\mathbb{R}^{\dim \beta + 1}$ that contains an ε-neighborhood of $(\beta_0, \pi_{0,ij})$ for all N.*

(i) Conditional on (X, α, γ), dyads are independent, that is,

$$Y_{ij} \perp\!\!\!\perp Y_{kl} | X, \beta, \alpha, \gamma \qquad \forall (k,l) \notin \{(i,j), (j,i)\}.$$

(ii) For all i, j and N we have that $\bar{E}[\partial_\beta \ell_{ij}] = \bar{E}[\partial_\pi \ell_{ij}] = 0$. For all N, the objective function $\mathcal{L}$ is strictly concave over $\mathbb{R}^{\dim \beta + 2N}$, and the matrix $\bar{\mathcal{H}} = -\partial_{\phi\phi'} \bar{\mathcal{L}}$ is positive definite.

(iii) For all (i,j), the function $(\beta, \pi) \mapsto \ell_{ij}(\beta, \pi)$ is five times continuously differentiable over $\mathcal{B}_{\varepsilon,ij}$ almost surely. For all (i,j), the partial derivatives of ℓ_{ij} with respect to the elements of (β, π) up to fifth order are bounded in absolute value uniformly over $(\beta, \pi) \in \mathcal{B}_{\varepsilon,ij}$ by a function $M(Z_{ij}) > 0$ a.s., where $\max_{i,j} \bar{E}[M(Z_{ij})^{16}]$ is a.s. uniformly bounded over N.

(iv) Let

$$\bar{W}_N = -\frac{1}{N}\big(\partial_{\beta\beta'}\bar{\mathcal{L}} - (\partial_{\phi\beta'}\bar{\mathcal{L}})(\partial_{\phi\phi'}\bar{\mathcal{L}})^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big)$$
$$\bar{\Omega}_N = \bar{E}\big[\big(\partial_\beta\mathcal{L} - (\partial_{\beta\phi'}\bar{\mathcal{L}})(\partial_{\phi\phi'}\bar{\mathcal{L}})^{-1}(\partial_\phi\mathcal{L})\big)^2\big].$$

The limits $plim_{N\to\infty}\bar{W}_N = W$ *and* $plim_{N\to\infty}\bar{\Omega}_N = \Omega$ *exist for* W *and* Ω *positive definite matrices.*

(v) There exist constants b_{min} *and* b_{max} *such that for all* $(\beta,\pi) \in \mathcal{B}_{\varepsilon,ij}$

$$0 < b_{min} \leq -\bar{E}[\partial_{\pi^2}\ell_{ij}(\beta,\pi)] \leq b_{max}$$

a.s. uniformly over i, j *and* N.

Assumption 1(i) specifies that outcomes depend on dyad specific variables only, and not on other features of the network. Conditional on the observed covariates and fixed effects, the outcome Y_{ij} is independent of other outcomes in the network, with the exception of Y_{ji}. Note that unconditionally outcomes are allowed to be dependent, through dependence across covariates X_{ij} and the fixed effects.

Assumption (ii) contains the parametric restriction of the model and requires that the true parameters $\beta_0, \alpha_0, \gamma_0$ are solutions to the first-order equations of the objective function. Concavity of the objective function ensures that the population problem has a unique solution. This is satisfied in many common nonlinear models, including the class of regression models with log-concave densities (as well as censored and truncated versions of these models), which includes probit, logit, ordered probit, Tobit, gamma and beta models among others (see Pratt (1981), Newey and McFadden (1994)).

Assumption 1(iii) provides basic smoothness conditions for the objective function. The derivative and moment conditions are required to ensure the validity of the asymptotic expansions to high enough order to establish the properties of the jackknife procedure, and to ensure that remainder terms are well bounded. Analysis of the jackknife requires higher order expansions than are required for characterization of the analytical bias and first-order asymptotic properties of the estimator, and so Assumption 1(iii) is somewhat stronger than the equivalent assumption employed in Fernández-Val and Weidner (2016).

Assumption 1(iv) ensures that the variance of $\widehat{\beta}$ is non-degenerate. The term $\bar{W}_N$ is the Hessian matrix for the common parameters β, after concentrating out the fixed effect parameters, while $\bar{\Omega}_N$ is the variance of the score for β (again after concentrating out the fixed effect parameters). To describe estimators of these terms, let

$$\Xi_{ij} = -\frac{1}{N}\sum_s \sum_{t\neq s} \left(\bar{\mathcal{H}}^{-1}_{(\alpha\alpha)is} + \bar{\mathcal{H}}^{-1}_{(\gamma\alpha)jt} + \bar{\mathcal{H}}^{-1}_{(\alpha\gamma)it} + \bar{\mathcal{H}}^{-1}_{(\gamma\gamma)st}\right)\bar{E}[\partial_{\beta\pi}\ell_{st}],$$

and define $D_\beta \ell_{ij} = \partial_\beta \ell_{ij} - \Xi_{ij}(\partial_\pi \ell_{ij})$. This term is the score for β after partialling out the fixed effect parameters. Estimators of the variance terms can be created in the usual way by plugging in estimates of the model parameters, i.e.

$$\begin{aligned}\widehat{W}_N &= -\frac{1}{N}\left(\partial_{\beta\beta'}\widehat{\mathcal{L}} - (\partial_{\phi\beta'}\widehat{\mathcal{L}})(\partial_{\phi\phi'}\widehat{\mathcal{L}})^{-1}(\partial_{\beta\phi}\widehat{\mathcal{L}})\right), \\ \widehat{\Omega}_N &= \frac{1}{N(N-1)}\sum_i \sum_{j<i}(\widehat{D_\beta\ell_{ij}} + \widehat{D_\beta\ell_{ji}})^2,\end{aligned} \tag{19}$$

where the terms $\partial_{\beta\beta'}\widehat{\mathcal{L}}$, $\widehat{D_\beta\ell_{ij}}$ etc. are evaluated at the estimates $\widehat{\beta}$, $\widehat{\alpha}$, $\widehat{\gamma}$. Note that the estimator $\widehat{\Omega}_N$ allows for correlation between the Y_{ij} and Y_{ji} outcomes.

Finally, Assumption 1(v) ensures that the Hessian for the fixed effect parameters is positive definite. This requires sufficient variation in the outcomes across both dimensions – i.e., variation in Y_{ij} over j (for fixed sender i) and over i (for fixed receiver j). In the binary outcome case, it implies that the model generates a *dense network*, one in which the number of links formed by each node tends to infinity as the size of the network grows. The assumption may not be reasonable in all empirical settings – in simulations we investigate the robustness of the estimator to sparsity in finite samples. The density assumption can be avoided in settings where sufficient statistics for the incidental parameters exists, such as the conditional logit framework, since estimation of the fixed effects is avoided. This comes at the expense of no longer being able to estimate average effects or counterfactual outcomes.

We now state the main theorem of the paper, on the asymptotic distribution of the jackknife bias-corrected estimator.

Theorem 1. *Let Assumption 1 hold and let $\widehat{\beta}_J$ be the jackknife bias-corrected estimator (8), the leave-l-out estimator (9) or the weighted jackknife (10). Let $V_N = \bar{W}_N^{-1}\bar{\Omega}_N\bar{W}_N^{-1}$ and assume that $V = plim_{N\to\infty} V_N$ exists and is positive definite. Then,*

$$N(\widehat{\beta}_J - \beta_0) \Rightarrow \mathcal{N}(0, V).$$

Let $\widehat{V}_N = \widehat{W}_N^{-1}\widehat{\Omega}_N\widehat{W}_N^{-1}$ be an estimator of the asymptotic variance, where $\widehat{W}_N$ and $\widehat{\Omega}_N$ are the plug-in estimators shown in (19). Then $\widehat{V}_N \to V$.

The jackknife estimator is asymptotically normally distributed and unbiased. It also has the same asymptotic variance as the non-bias-corrected estimator in (2). The variance is the usual sandwich form one, and is easily computed. In the case of maximum likelihood we will have that $\bar{W}_N = \bar{\Omega}_N$ so that the variance simplifies to $V_N = \bar{W}_N^{-1}$. In general this will not be true, for example the researcher may wish to allow for correlation between Y_{ij} and Y_{ji} by clustering $\bar{\Omega}_N$ at the dyad level as in (19).

4.2 Asymptotic analysis for fixed effect averages

Here we present asymptotic results for averages of functions that may take more than one observation as arguments. This structure will cover a number of interesting cases such as standard average effects, averages over dyads, triads or other structures in the network, and specification tests. Recall that λ is a set of observations (i,j), and Λ_N is the collection of all such sets formed by permuting the nodes in λ. We let $Y_\lambda = \{Y_{ij}\}_{(i,j)\in\lambda}$, $X_\lambda = \{X_{ij}\}_{(i,j)\in\lambda}$, and $\pi_\lambda = \{\alpha_i + \gamma_j\}_{(i,j)\in\lambda}$ collect the outcomes, covariates and fixed effects for the observations in λ. The function of interest is $m_\lambda = m(Y_\lambda, X_\lambda, \beta, \pi_\lambda)$, which is a function of $(Y_{ij}, X_{ij}, \alpha_i + \gamma_j)$ for each $(i,j) \in \lambda$.

It will be useful to define three separate quantities

$$\delta = E\big[\Delta_N\big], \quad \bar{\Delta}_N = \bar{E}\big[\Delta_N\big], \text{ and } \Delta_N = \frac{1}{|\Lambda_N|}\sum_{\lambda\in\Lambda_N} m_\lambda.$$

Here, Δ_N is the average effect computed in the observed sample (at the true parameter values), $\bar{\Delta}_N$ is the expectation of the average effects conditional on the distribution of covariates and fixed effects in the observed sample, while δ_N is the population expectation. We can decompose the estimation error $\widehat{\Delta}_N - \delta$ into three sources

$$\widehat{\Delta}_N - \delta = (\widehat{\Delta}_N - \Delta_N) + (\Delta_N - \bar{\Delta}_N) + (\bar{\Delta}_N - \delta). \tag{20}$$

The first term, $\widehat{\Delta}_N - \Delta_N$, represents variation caused by estimation of the parameters in the model, including fixed effects. The next term, $\Delta_N - \bar{\Delta}_N$, is variation of the sample outcomes m_λ around their conditional expectations $\bar{m}_\lambda = \bar{E}[m_\lambda]$. In the case that m is a function of the data only through X_λ, i.e. it does not depend on outcomes Y_λ, we will have $\bar{m}_\lambda = m_\lambda$ and this second term will vanish. Finally, $\bar{\Delta}_N - \delta$ captures differences in the distribution of covariates and fixed effects in the observed network, relative to the population. In the case that the full network is observed, or whenever $\bar{\Delta}_N = 0$ as is the case for specification tests, we will have that $\bar{\Delta}_N = \delta_N$.

The results below will rely on Assumption 1, as well as additional restrictions on the choices of λ and m.

Assumption 2. *Let λ be a set of r observations (i, j) containing p distinct agents. For every λ, let $\mathcal{B}_{\varepsilon,\lambda}$ be a subset of $\mathbb{R}^{\dim \beta + r}$ that contains an ε-neighborhood of $(\beta_0, \pi_{0,\lambda})$ for all N, with $\varepsilon > 0$.*

(i) The number of observations and unique agents in λ, r and p, are fixed over N. The set Λ_N contains all $\frac{N!}{(N-p)!}$ permutations of agents in the set of observations λ.

(ii) The function m depends on (α, γ) only through $\pi_\lambda = \{\alpha_i + \gamma_j\}_{(i,j)\in\lambda}$. For all λ, the function $(\beta, \pi_\lambda) \mapsto m(Z_\lambda, \beta, \pi_\lambda)$ is five times continuously differentiable over $\mathcal{B}_{\varepsilon,\lambda}$ a.s. For all λ, the partial derivatives of m with respect to the elements of (β, π_λ) up to fifth order are bounded in absolute value uniformly over $(\beta, \pi_\lambda) \in \mathcal{B}_{\varepsilon,\lambda}$ by a function $M(Z_\lambda) > 0$ a.s., and $\max_\lambda \bar{E}[M(Z_\lambda)^{16}]$ is a.s. uniformly bounded over N.

(iii) We have that $0 < \min_\lambda E[m_\lambda^2] - E[m_\lambda]^2 \leq \max_\lambda E[m_\lambda^2] - E[m_\lambda]^2 < \infty$ uniformly over N

Assumption 2 (i) restricts m to be a function of a fixed number of edges in the network, which allows us to construct leave-out sets to bias correct using the jackknife. It also assumes that Δ_N is an average of all possible arrangements of the nodes in λ, ensuring that averaging occurs over all dimensions. For example, an average of the form $\frac{1}{N}\sum_{j\neq i} m(X_{ij}, \alpha_i, \gamma_j)$ is not allowed since we are only averaging over the receiver dimension, while holding i fixed.

Assumption 2 (ii) is analogous to Assumption 1 (iii), and imposes the same differentiability and moment conditions on m as are imposed on ℓ. This allows for asymptotic expansions of $\widehat{\Delta}_N$ to be derived, in the same way as for $\widehat{\beta}$. Finally, (iii) ensures that the unconditional second moments of m are well defined.

The asymptotic distribution depends on the choice of target parameter, either a conditional or population average. The following theorem states the asymptotic result for the jackknife bias-corrected estimator of the conditional fixed effect average $\bar{\Delta}_N$.

Theorem 2. *Let Assumptions 1 and 2 hold, and let $\widehat{\Delta}_J$ be the jackknife bias-corrected estimator (18). Then*

$$N(\widehat{\Delta}_J - \bar{\Delta}_N) \Rightarrow \mathcal{N}(0, V_\Delta).$$

If we additionally assume that $E\big[(m_\lambda - \bar{m}_\lambda)(m_{\lambda'} - \bar{m}_{\lambda'})\big] \neq 0$ for sets λ and λ' that share exactly one observation in common, then the asymptotic variance is

$$V_\Delta = \lim_{N\to\infty} \frac{1}{N(N-1)} \sum_i \sum_{j<i} \bar{E}\big[(h_{ij} + s_{ij})^2\big]$$

where $h_{ij} = -N(\partial_\theta \bar{\Delta}_N)(\partial_{\theta\theta'}\bar{\mathcal{L}})^{-1}\big((\partial_\theta \ell_{ij}) + (\partial_\theta \ell_{ji})\big)$, for $\theta' = (\beta, \alpha', \gamma')$, and $s_{ij} = \frac{(N-p)!}{(N-2)!}\sum_{\lambda\in\tilde{\Lambda}_{ij}}(m_\lambda - \bar{m}_\lambda)$, with $\tilde{\Lambda}_{ij} = \{\lambda : (i,j) \in \lambda \text{ or } (j,i) \in \lambda\}$.

If we have either $m_\lambda = \bar{m}_\lambda$ or $E\big[(m_\lambda - \bar{m}_\lambda)(m_{\lambda'} - \bar{m}_{\lambda'})\big] = 0$ for sets λ and λ' that share exactly one observation in common, then the asymptotic variance is

$$V_\Delta = \lim_{N\to\infty} \frac{1}{N^2} \sum_i \sum_{j<i} \bar{E}\big[h_{ij}^2\big].$$

In either case, let $\widehat{V}_\Delta$ be the plug-in estimator for V_Δ that replaces the unknown θ with estimates $\widehat{\theta}$. Then $\widehat{V}_\Delta \to V_\Delta$.

Some explanation for the form of the variance may be useful. The two terms h_{ij} and s_{ij} relate to the first two components of (20). The first component $\widehat{\Delta}_J - \Delta_N$ contains variation from estimation of the common parameters β and fixed effects α, γ. In the Appendix it is shown that this term can be approximated using the delta-method

$$\begin{aligned} N(\widehat{\Delta}_J - \Delta_N) &= -N(\partial_\theta \bar{\Delta}_N)(\partial_{\theta\theta'}\bar{\mathcal{L}})^{-1}(\partial_\theta \ell_{ij}) + o_p(1) \\ &= \frac{1}{N-1}\sum_i \sum_{j\neq i} h_{ij} + o_p(1). \end{aligned}$$

Note that, replacing the jackknife estimate $\widehat{\Delta}_J$ with the standard estimator $\widehat{\Delta}_N$ would result in additional terms appearing in the above first-order approximation, related to the incidental parameter bias. The second component is

$$N(\Delta_N - \bar{\Delta}_N) = \frac{N}{|\Lambda_N|} \sum_{\lambda} (m_\lambda - \bar{m}_\lambda).$$

The variance of this term depends on the conditional covariance between m_λ and $m_{\lambda'}$ for distinct sets λ and λ'. Note that if λ and λ' share no dyads in common then they are conditionally independent. The variance of this term depends on the condition $E\big[(m_\lambda - \bar{m}_\lambda)(m_{\lambda'} - \bar{m}_{\lambda'})\big] \neq 0$ for sets λ and λ' share exactly one observation in common. Under this condition, the variance of $\sum_\lambda (m_\lambda - \bar{m}_\lambda)$ is dominated by covariances between m_λ and $m_{\lambda'}$ for λ and λ' that share exactly one dyad in common; although λ with two or more common dyads also contribute to the variance, there are an order of magnitude fewer such combinations, and so these represent smaller order contributions that do not appear in the asymptotic variance. In settings where $E\big[(m_\lambda - \bar{m}_\lambda)(m_{\lambda'} - \bar{m}_{\lambda'})\big] = 0$ for λ and λ' that share exactly one observation in common, $\Delta_N - \bar{\Delta}_N$ is a degenerate U-statistic and its variance is asymptotically of smaller order than the variance from parameter estimation, i.e. $\widehat{\Delta}_J - \Delta_N$, and so may be ignored.

Theorem 2 shows how we may construct confidence sets for the parameter of interest $\bar{\Delta}_N$. When the object of interest is the unconditional average δ, the convergence of the estimator will be dominated by variation from the third component in (20), $\bar{\Delta}_N - \delta$. To describe the statistic in this setting, it is useful to use its U-statistic representation. We will additionally assume that $X_{ij} = h(X_i, X_j)$. This condition appears in other work on dyadic models, for example in Graham (2017). When X_{ij} measures the similarity (or difference) between i and j in some measure, we will commonly have $X_{ij} = d(X_i - X_j)$, for d some distance function. Alternatively, if X_{ij} captures common membership in some group, we may have $X_{ij} = X_i X_j$ where X_i is an indicator for i's membership.

To give $\bar{\Delta}_N$ a U-statistic representation, we first sum together all $\bar{m}_\lambda$ which share the same set of agents. Since there are p agents in each λ, this gives $p!$ different λ that can be created from a given set of agents. We denote these sets of unordered agents by τ. We have, for

$\tilde{m} = \bar{m} - E[m]$

$$\begin{aligned}
\bar{\Delta}_N - \delta &= \frac{1}{N \cdots (N-p+1)} \sum_{\lambda} \tilde{m}_\lambda \\
&= \frac{p!}{N \cdots (N-p+1)} \sum_{\tau} \Big(\frac{1}{p!} \sum_{\lambda \in \tau} \tilde{m}_\lambda \Big) \\
&= \binom{N}{p}^{-1} \sum_{\tau} u_\tau,
\end{aligned}$$

where $u_\tau = \frac{1}{p!} \sum_{\lambda \in \tau} \tilde{m}_\lambda$. The term u_τ is a symmetric function of $\{\beta, X_i, \alpha_i, \gamma_i\}$ for p agents i. Assuming that the $\{X_i, \alpha_i, \gamma_i\}$ are i.i.d. over agents, $\bar{\Delta}_N - \delta$ is a U-statistic of order p and we may apply standard theory on such statistics to compute its asymptotic distribution.

Theorem 3. *Let Assumptions 1 and 2 hold. Additionally, assume that $X_{ij} = h(X_i, X_j)$ where X_i is an observed agent-specific characteristic, and also that $(\alpha_i, \gamma_i, X_i)$ is i.i.d. over i. Let*

$$\Sigma_1 = Cov\big(u_\tau, u_{\tau'}\big),$$

for τ, τ' such that $\tau \cap \tau' = \{i\}$. Then, for $V_\delta = p^2 \Sigma_1$

$$\sqrt{N}(\widehat{\Delta}_J - \delta) \Rightarrow \mathcal{N}(0, V_\delta).$$

Additionally, the variance estimator $\widehat{V_\delta}$ in (21) is consistent, i.e. $\widehat{V_\delta} \to V_\delta$.

The convergence rate in Theorem 3 is slower than the rate in Theorem 2. While m_λ and $m_{\lambda'}$ are *conditionally* independent when λ and λ' share no dyads in common, the two are *unconditionally* independent only when they share no agents in common. Since there are many more sets λ that share a single agent i than share a dyad (i, j), the variance of $\bar{\Delta}_N - \delta$ is an order of magnitude larger than that of $\widehat{\Delta}_N - \Delta_N$, and so the convergence rate is slower.

Similarly to Theorem 2, the variance is dominated by covariances between sets λ that share exactly one node in common. The term Σ_1 is the covariance between u_τ and $u_{\tau'}$ when τ and

τ' share exactly one agent in common. A consistent estimator for this quantity is given by

$$
\begin{aligned}
\widehat{V}_\delta &= \frac{1}{N}\sum_i \tilde{\mu}_i^2, \\
\tilde{\mu}_i &= \frac{(N-p)!}{(N-1)!}\sum_{\lambda: i\in\lambda} (\widehat{\bar{m}}_\lambda - \widehat{\mu}), \\
\widehat{\mu} &= \frac{(N-p)!}{N!}\sum_\lambda \widehat{m}_\lambda,
\end{aligned} \tag{21}
$$

where $\tilde{\mu}_i$ is the average over all sets λ containing agent i, $\widehat{\bar{m}}_\lambda$ is a plug-in estimator for $\bar{m}_\lambda$, and $\widehat{\mu} = \sum_\lambda \widehat{\bar{m}}_\lambda$ is the overall mean.

Since the rate of convergence in Theorem 3 is $N^{-1/2}$, there is in fact no asymptotic bias generated by the incidental parameters. The bias from the estimation of the fixed effects is of order N^{-1}, which is smaller than the variation in the sampled distribution of fixed effects around is population distribution. Nonetheless, we would still recommend bias correcting estimators as it is likely to improve the finite sample properties of inference, in terms of correct centering of confidence sets, with little or no cost in terms of additional variance. In the panel data setting, Fernández-Val and Weidner (2016) report such improvements in simulations.

4.3 Examples

Theorems 2 and 3 suggest that the construction of confidence sets for estimates and hypothesis testing may be performed in the usual way, by using the asymptotic normal approximations. Standard plug-in estimates for the variance expressions may be used. Here we provide some examples of how these results can be used.

Example 1. ***Average marginal effect in a probit model***

The average marginal effect in the probit model can be estimated using

$$
\widehat{\Delta} = \frac{1}{N(N-1)}\sum_i \sum_{j\neq i} \beta\varphi(\beta' X_{ij} + \alpha_i + \gamma_j),
$$

where φ is the standard normal density function. In this setting we have $\lambda = (i,j)$ and $m_{ij} = \bar{m}_{ij}$ so that $s_{ij} = 0$. The variance in Theorem 2 is then the standard delta-method variance of

$$V_\Delta = (\partial_\theta \bar{\Delta}_N)(\partial_{\theta\theta'}\bar{\mathcal{L}})^{-1}\bar{\Omega}(\partial_{\theta\theta'}\bar{\mathcal{L}})^{-1}(\partial_\theta \bar{\Delta}_N),$$
$$\bar{\Omega} = \frac{1}{N(N-1)}\sum_{i<j}\bar{E}\big[(\partial_\theta \ell_{ij} + \partial_\theta \ell_{ji})(\partial_\theta \ell_{ij} + \partial_\theta \ell_{ji})'\big],$$

and standard plug-ing estimators may be used. A $(1-\alpha)$-per cent confidence set could be constructed for Δ_N using $\widehat{\Delta}_J \pm c_{1-\alpha/2} V_\Delta^{1/2}/N$, where $c_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

If instead we are interested in inference with respect to the population parameter δ, Theorem 3 states that we may compute the asymptotic variance as

$$\widehat{V}_\delta = \frac{1}{N}\sum_i \Big(\frac{1}{N-1}\sum_{j\neq i}(\beta\varphi_{ij} + \beta\varphi_{ji} - 2\widehat{\Delta})\Big)^2.$$

A $(1-\alpha)$-per cent confidence set for δ is $\widehat{\Delta}_J \pm c_{1-\alpha/2}\widehat{V}_\delta^{1/2}/\sqrt{N}$.

Example 2. ***Testing transitivity in a probit model***

Recall that a statistic for testing the presence of transitivity is

$$\widehat{\Delta}_N = \frac{1}{N(N-1)(N-2)}\sum_i \sum_{j\neq i}\sum_{k\neq\{i,j\}}(Y_{ij} - p_{ij})Y_{ik}Y_{kj},$$

where $p_{ij} = E[Y_{ij}|X,\beta,\alpha,\gamma]$. Fitting Y_{ij} with a probit regression, we have $p_{ij} = \Phi(\beta' X_{ij} + \alpha_i + \gamma_j)$. Since $\bar{m}_\lambda = \bar{E}\big[(Y_{ij} - p_{ij})Y_{ik}Y_{kj}\big] = 0$, we have $\bar{\Delta}_N = \delta = 0$ so we may determine the asymptotic distribution of the test statistic using Theorem 2. We have

$$s_{ij} = \frac{1}{N-2}\sum_{k\neq\{i,j\}}\Big((Y_{ij} - p_{ij})Y_{ik}Y_{kj} + (Y_{ji} - p_{ji})Y_{jk}Y_{ki}$$
$$+ (Y_{ik} - p_{ik})Y_{ij}Y_{jk} + (Y_{jk} - p_{jk})Y_{ji}Y_{ik}$$
$$+ (Y_{kj} - p_{kj})Y_{ki}Y_{ij} + (Y_{ki} - p_{ki})Y_{kj}Y_{ji}\Big).$$

From the likelihood for a probit model, we have $\ell_{ij} = Y_{ij}\log(p_{ij}) + (1 - Y_{ij})\log(1 - p_{ij})$, which gives $\partial_\pi \ell_{ij} = H_{ij}(Y_{ij} - p_{ij})$ and $\partial_\beta \ell_{ij} = H_{ij}(Y_{ij} - p_{ij})X_{ij}$ for $H_{ij} = \varphi_{ij}/p_{ij}(1 - p_{ij})$.

Also,

$$\partial_\beta \bar{\Delta}_N = -\frac{1}{N(N-1)(N-2)} \sum_i \sum_{j \neq i} \sum_{k \neq \{i,j\}} \varphi_{ij} p_{ik} p_{kj} X_{ij},$$

$$\partial_{\alpha_i} \bar{\Delta}_N = -\frac{1}{N(N-1)(N-2)} \sum_{j \neq i} \sum_{k \neq \{i,j\}} \varphi_{ij} p_{ik} p_{kj},$$

$$\partial_{\gamma_i} \bar{\Delta}_N = -\frac{1}{N(N-1)(N-2)} \sum_{j \neq i} \sum_{k \neq \{i,j\}} \varphi_{ji} p_{jk} p_{ki},$$

from which we can construct h_{ij} using the estimated Hessian matrix and the formula given in Theorem 2.

5 Empirical example

We illustrate the jackknife procedure on a data set consisting of a directed network of export volumes between 136 countries (136×135 country pair observations) in 1990. The data are taken from Santos Silva and Tenreyro (2006), and additional details on their construction can be found in their paper. The outcome variable Y_{ij} is the value of exports from country i to country j. We also use several covariates to capture homophily in trade relationships, which include: *log distance*, the log of the distance between the capitals of the countries; *border*, an indicator of whether the countries share a common border; *language*, an indicator for whether the countries share a language; *colonial*, and indicator for whether either country had colonized the other at some point in history; and *trade agreement*, an indicator for the presence of a joint preferential trade agreement between the two countries.

5.1 Zero-inflated binomial model

Burger et al. (2009) propose a zero-inflated negative binomial model. The value of trade between i and j is given by the product of two variables $Y_{ij} = z_{ij} Y^*_{ij}$, where $z_{ij} \in \{0, 1\}$ is a binary decision to enter into a trading relationship, while Y^*_{ij} is the value of exports that will be realized, conditional on $z_{ij} = 1$. The binary decision is modeled using as a probit function, while the latent outcome Y^*_{ij} is modeled as a negative binomial variable.

In this example, the objective function is given by

$$f(Y_{ij}|X_{ij},\theta) = \mathbb{1}\{Y_{ij}=0\}p_{ij} + (1-p_{ij})g(Y_{ij}|X_{ij},\theta)$$

where $\theta = (\beta, \alpha, \gamma, \nu)$, and

$$\begin{aligned} p_{ij} &= \Phi(X_{ij}'\beta^z + \alpha_i^z + \gamma_j^z) \\ g(Y_{ij}|X_{ij},\theta) &= \frac{\Gamma(Y_{ij}+\nu)}{\Gamma(\nu)Y_{ij}!}\Big(\frac{\nu}{\nu+\mu_{ij}}\Big)^{\nu}\Big(\frac{\mu_{ij}}{\nu+\mu_{ij}}\Big)^{Y_{ij}} \\ \mu_{ij} &= \exp(X_{ij}'\beta^y + \alpha_i^y + \gamma_j^y) \end{aligned}$$

The parameter ν captures the degree of overdispersion in the model for Y_{ij}^*, with $\nu \to \infty$ resulting in a model with equal mean and variance (as in the Poisson), while smaller ν lead to greater degrees of dispersion.

Estimates of the parameters in the model are presented in Table 1. Most variables change by only small amounts after bias correction. However, the effect of sharing a common border on the probability of engaging in zero trade changes significantly after bias correction; while the maximum likelihood estimate suggests that common borders are important for link formation, the bias corrected estimate is no longer significant. This suggests that the sharing a common border has little effect on the likelihood of engaging in trade, but does affect the volume of trade. The results also suggests a substantial impact of trade agreements, both on the probability of engaging in trade and on the volume of trade, a result that is robust to bias correction. The overdispersion parameter ν is less than a half, suggesting a significant amount of overdispersion, i.e. export volumes have far greater variation across country pairs than suggested by a Poisson model.

The rightmost column in the table reports the difference between the MLE and jackknife bias-corrected estimators in terms of their standard errors. For a number of variables in the model of export volumes, the change in estimate is around three-quarters of the standard deviation or more, which has an important impact on inference. To give some idea of the scale of these biases, a bias of three-quarters of a standard error is enough for a five per cent test two reject around 12 per cent of the time (more than twice nominal size), while bias of 1.5 standard errors leads to a rejection rate of more than 30 per cent.

Table 1: Zero-inflated negative binomial model

	Coefficients			
	MLE	Jackknife	SE	(Bias/SE)
Zero model				
log distance	0.721	0.721	0.029	0.00
border	0.628	0.157	0.120	3.93
language	-0.330	-0.306	0.053	0.45
colonial	-0.305	-0.282	0.056	0.41
trade agreement	-1.168	-1.126	0.180	0.24
Volume model				
log distance	-1.243	-1.218	0.033	0.77
border	0.437	0.483	0.129	0.36
language	0.405	0.418	0.068	0.18
colonial	0.399	0.335	0.073	0.88
trade agreement	1.055	0.960	0.131	0.73
ν	0.492	0.459	0.008	4.38

Table 2 contains estimates of the average effect of a regressor on expected export volume, conditional on non-zero trade, over the distribution of regressors and fixed effects for trading country pairs. That is, we calculate (for $n_1 = \sum_i \sum_{j \neq i} \mathbb{1}\{Y_{ij} > 0\}$)

$$\Delta_N = \frac{1}{n_1} \sum_i \sum_{j \neq i} \mathbb{1}\{Y_{ij} > 0\} \beta_{dist} \exp(X_{ij}' \beta^y + \alpha_i^y + \gamma_j^y)$$

for the continuous regressor *log distance* and

$$\Delta_N = \frac{1}{n_1} \sum_i \sum_{j \neq i} \mathbb{1}\{Y_{ij} > 0\} \big(\exp(\beta^{y\prime} X_{ij}^{(1)} + \alpha_i^y + \gamma_j^y) - \exp(\beta^{y\prime} X_{ij}^{(0)} + \alpha_i^y + \gamma_j^y) \big)$$

for binary regressors, where $X_{ij}^{(1)}$ sets the binary regressor of interest to one for all (i, j) and $X_{ij}^{(0)}$ sets it to zero (leaving other regressors unchanged). Again, the jackknife bias correction has an important impact on two of the effects; for example, the effect of a trade agreement on expected export volumes decreases by about a quarter (more than a one standard error change in magnitude). Note that, as is the case here, a small bias in the coefficient on some variable does not necessarily imply low bias in the corresponding marginal effect.

Table 2: Zero-inflated negative binomial model

	Average effects			
	MLE	Jackknife	SE	(Bias/SE)
log distance	-116.2	-113.8	9.08	0.26
border	47.5	50.2	16.95	0.16
language	43.4	41.0	8.64	0.28
colonial	44.4	31.0	10.00	1.34
trade agreement	140.2	107.1	28.10	1.18

5.2 Testing for strategic interactions

To demonstrate the use of the jackknife for specification testing, we implement the test in (15) in a binary model for the probability of country i exporting to country j. We test for two types of strategic interaction: reciprocity, using $S_{ij} = Y_{ji}$; and transitivity, using $S_{ij} = \frac{1}{N-2} \sum_{k \neq \{i,j\}} Y_{ik} Y_{kj}$. In both cases we construct the statistics by estimating p_{ij} using a probit model and jackknife the statistic. Standard errors are computed using the expressions in Theorem 2. Table 3 presents the values of the statistics as well as the standardized values $t_N = NT_N/\sqrt{V_T}$ and $t_J = NT_J/\sqrt{V_T}$ for both tests.

For the reciprocity statistic, the jackknife bias correction appears to have little effect. The statistic rejects the null of no reciprocity strongly suggesting that the existence of an export relationship from i to j increases the likelihood of i also importing from j. This is perhaps not surprising. It is important to note that the model considered in this paper allows for reciprocity so that this conclusion has no impact on the validity of model estimates. The presence of reciprocity does suggest that standard errors should be clustered at the dyad level to account for correlation between the outcomes Y_{ij} and Y_{ji}.

Table 3: Strategic interaction tests

	NT_N	NT_J	t_N	t_J
Reciprocity	3.160	3.249	16.96	17.43
Transitivity	-0.094	0.012	-6.75	0.88

In contrast, the jackknife bias correction appears to have an important effect on the transitivity statistic. The uncorrected statistic leads to a rejection of the null, and the conclusion that indirect paths of trade (exports paths from i to j through a third-party country) are associated with a lower probability of a direct export relationship than is expected given the model. However, the jackknifed statistic is close to zero so that we do not reject the null hypothesis that trade decisions are bilateral in nature.

5.3 Comparison with conditional logit estimator

As a comparison with an existing approach to the incidental parameters problem in networks, we consider estimating a logit model for the probability that country i exports to country j. Under the logit specification, it is possible to estimate the common parameters in the model by first removing the fixed effects. This approach has been suggested by Graham (2017) for an undirected network, and Jochmans (2018) for a directed network. The conditional logit estimator works by forming difference-in-differences style contrasts among sets of four nodes (tetrads). Let

$$z_{ij,kl} = \frac{(Y_{ik} - Y_{il}) - (Y_{jk} - Y_{jl})}{2}$$
$$r_{ij,kl} = (X_{ik} - X_{il}) - (X_{jk} - X_{jl})$$

Given the logistic specification

$$P(Y_{ij} = 1|X, \beta, \alpha, \gamma) = \frac{\exp(\beta' X_{ij} + \alpha_i + \gamma_j)}{1 + \exp(\beta' X_{ij} + \alpha_i + \gamma_j)}$$

and conditional independence of outcomes across dyads, we may write

$$P(z_{ij,kl} = 1|z_{ij,kl} \in \{-1, 1\}, X, \beta, \alpha, \gamma) = \frac{\exp(\beta' r_{ij})}{1 + \exp(\beta' r_{ij})} \tag{22}$$

That is, conditional on the event $z_{ij,kl} \in \{-1, 1\}$, the outcome $z_{ij,kl}$ follows a logit model without any fixed effect parameters. This allows us to estimate the common parameter β by a standard logit regression of $z_{ij,kl}$ on $r_{ij,kl}$ in the subsample of $z_{ij,kl} \in \{-1, 1\}$. Estimates from this model, as well as the jackknife bias-corrected estimates, are shown in Table 4.

Interestingly, although the jackknife bias correction suggests little bias in the original logit parameter estimates, the conditional logit estimates are significantly different. For example, while the coefficient on distance is similar across the MLE and jackknife estimates (-1.34 and -1.31 respectively), the conditional logit estimate differs by more than three standard errors.

There may be a number of reasons for such a discrepancy between the estimates. One

Table 4: Coefficient estimates for logit model

	MLE	Jackknife	Conditional logit
log distance	-1.341	-1.305	-1.125
	(0.062)	(0.062)	(0.059)
border	-1.192	-1.157	0.866
	(0.265)	(0.265)	(0.268)
language	0.590	0.573	0.488
	(0.105)	(0.105)	(0.104)
colonial	0.509	0.493	0.579
	(0.110)	(0.110)	(0.106)
trade agreement	2.057	1.998	1.653
	(0.407)	(0.407)	(0.349)

Standard errors are shown in parentheses. For the MLE and jackknife, the standard errors are those shown in Theorem 1, while for the conditional logit standard errors are computed using the asymptotic distribution in Jochmans (2018).

possibility is that the model is misspecified in the sense that the correct link function is not the logistic function. In this case, maximum likelihood estimates a set of pseudo parameters that represent the parameters that minimize the Kullback-Leibler distance between the logit model and the true model (White, 1982). Since the jackknife theory does not rely on any information equalities, the jackknife bias correction is asymptotically unbiased for these pseudo parameters. In contrast, the conditional logit estimator estimates a different set of pseudo parameters, those that minimize the KL distance conditional on $z_{ijkl} \in \{-1, 1\}$, for a logit regression of z_{ijkl} on r_{ijkl}. To the best of our knowledge, there is no reason to suspect that these two sets of pseudo parameters would coincide.

An alternative explanation is that the assumption of independence between dyads is violated in the data. In this case, the likelihood function for the conditional logit will be incorrect, since the identity (22) will no longer hold. In this setting the bias correction given by the jackknife estimator is also likely to be incorrect since it will not account for bias terms generated by the dependence across dyads.

6 Simulations

Here I demonstrate the effectiveness of the jackknife in simulations. I repeat the simulation design of Dzemski (2019), which has also been used in a number of other network papers.

Table 5: Fixed effect distributions

Name	C_N^l	C_N^u	Density[(a)]
bal	$-\log\log N$	$\log\log N$	0.50
llog	$-\log\log N$	0	0.19
slog	$-\log^{1/2} N$	0	0.12
log	$-\log N$	0	0.03

[a] Values are the average density over 100 simulations in a network of $N = 50$ nodes.

The binary outcome Y_{ij} is determined by

$$Y_{ij} = 1\{\theta X_{ij} + \alpha_i + \gamma_j > \varepsilon_{ij}\}$$

where $\theta = 1$ and $\varepsilon_{ij} \sim N(0,1)$. Individual i is characterized by the binary scalar $X_i = 1 - 2 \times 1\{i \text{ is odd}\}$, and the homophily variable is given by $X_{ij} = X_i X_j$, i.e. it is one for pairs with the same sign and minus one for pairs with opposing signs. The fixed effects are given by

$$\alpha_i = \gamma_i = C_N^l - \frac{N-i}{N-1}(C_N^u - C_N^l)$$

which is a sequence from C_N^l to C_N^u. The value of (C_N^l, C_N^u) is intended to control the sparsity of the network, and we consider four choices, shown in Table 5.

In the balanced setting ('bal'), fixed effects range between $\pm\log\log N$, generating a dense network in which around half of all links are formed. Subsequent settings feature increasingly sparse networks in which some nodes remain well connected, while others make few links. In the sparsest setting ('log') only around 3 per cent of all links are formed.

I simulate the model 1000 times and compute the MLE, analytical bias-corrected estimate, and both the standard and weighted jackknife bias-corrected estimates for each fixed effect distribution.

Table 6 presents the bias, standard deviation, and rejection rates for a 5 per cent test of the null hypothesis $\beta_0 = 1$ for each estimator. As expected, the MLE is biased, with the size of the bias increasing in the sparsity of the network. In each case, the bias is around one standard deviation in size, resulting in substantial over-rejection. In contrast, the jackknife estimator is approximately unbiased in the first two designs. In the sparsest design the jackknife estimator does not appear to perform well, and shows similar bias to the MLE. The weighted jackknife performs well, even in the sparse designs; it is unbiased in the first

Table 6: Simulation results

	$\bar{N}^{(b)}$	Bias (mean) $\hat{\theta}_{MLE}$	$\hat{\theta}_{BC}$	$\hat{\theta}_J$	$\hat{\theta}_{wJ}$	SD $\hat{\theta}_{MLE}$	$\hat{\theta}_{BC}$	$\hat{\theta}_J$	$\hat{\theta}_{wJ}$	Rejection (5%) $\hat{\theta}_{MLE}$	$\hat{\theta}_{BC}$	$\hat{\theta}_J$	$\hat{\theta}_{wJ}$
bal	50	0.06	0.00	-0.01	0.00	0.04	0.04	0.04	0.04	0.29	0.03	0.03	0.03
llog	50	0.07	0.01	-0.01	0.00	0.06	0.05	0.06	0.05	0.27	0.04	0.04	0.04
slog	50	0.11	0.02	-0.03	0.01	0.10	0.09	0.13	0.12	0.26	0.05	0.05	0.05
log	29	0.51	~	-1.23	0.04	0.60	~	1.37	0.42	0.27	0.23	0.67	0.29

[a] The top panel is $N = 50$, lower panel is $N = 70$; $\bar{N}$ is the average number of nodes in the connected network

[b] In sparse settings the analytical correction generated some extreme outliers. Median and 95-5 percentile range are reported in Table 7.

Table 7: Simulation results - median and range

	$\bar{N}^{(b)}$	Bias (median) $\hat{\theta}_{MLE}$	$\hat{\theta}_{BC}$	$\hat{\theta}_J$	$\hat{\theta}_{wJ}$	95-5 range $\hat{\theta}_{MLE}$	$\hat{\theta}_{BC}$	$\hat{\theta}_J$	$\hat{\theta}_{wJ}$
bal	50	0.06	0.00	-0.01	0.00	0.14	0.13	0.13	0.13
llog	50	0.07	0.00	-0.01	0.00	0.19	0.18	0.17	0.18
slog	50	0.10	0.01	-0.03	0.00	0.27	0.25	0.23	0.25
log	29	0.23	0.11	-0.95	-0.06	1.55	1.38	4.24	1.13

[a] top panel is $N = 50$, lower panel is $N = 70$

[b] $\bar{N}$ is the average number of nodes in the connected network

three designs, and removes the majority of the bias even in the sparsest design. In each case, the jackknife estimators have smaller standard deviation than the MLE (with the exception of the standard jackknife in the sparsest design). Rejection rates are at or below the nominal level in three of the four settings, although in the sparsest setting all estimators over reject.

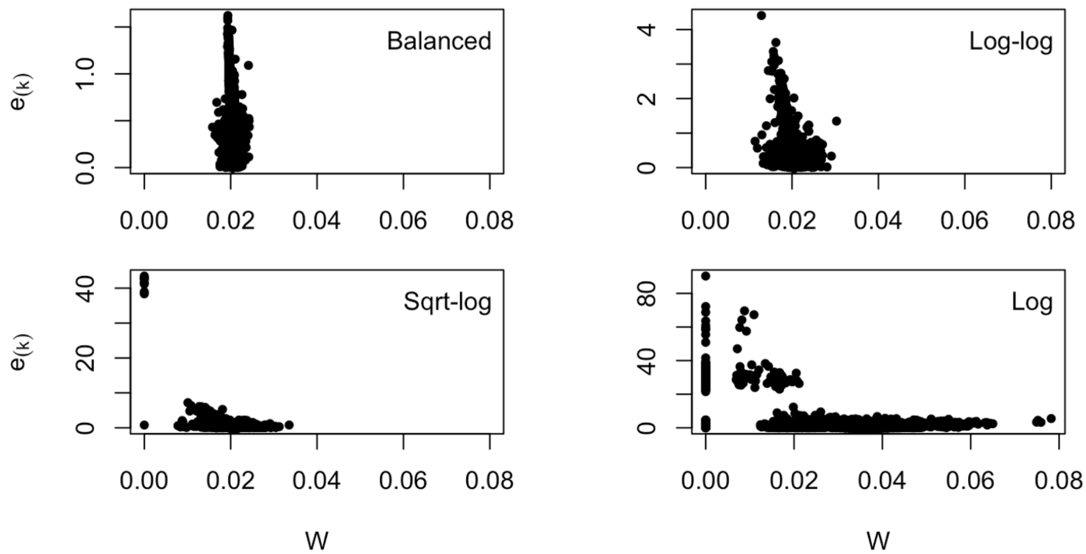
Weighted jackknife

To help explain the properties of the weighted jackknife, we investigate how the weight given to each leave-out estimate $\widehat{\beta}_{(k)}$ correlates with the contribution of that estimate to the total error of the jackknife estimate. Define

$$e_{(k)} = (N-1)\widehat{\beta}_{MLE} - (N-2)\widehat{\beta}_{(k)} - \beta_0$$

as the contribution to the error $\widehat{\beta}_J - \beta_0$ from a single leave-out estimate $\widehat{\beta}_{(k)}$. The total error of the standard jackknife is simply the average of these errors $\widehat{\beta}_J - \beta_0 = \frac{1}{N-1}\sum_k e_{(k)}$, while the weighted jackknife gives differing weights to different leave-out estimates. Figure 3 plots

Figure 3: Weights ($\widehat{W}_{(k)}$) versus estimates ($\widehat{\beta}_{(k)}$) in leave-out samples



The y-axis is which measures the absolute error of the jackknife estimator using the single leave-out estimate. The x-axis is the weight given to that leave-out estimate in the weighted jackknife.

the absolute value of these errors against their weights in the weighted jackknife across the four simulation designs.

For the balanced design (upper left panel), which is the densest network, the weights are concentrated around $\frac{1}{N-1} \approx 0.02$ with very little variation. In this case, the weighted jackknife is almost identical to the standard jackknife, as is clear in Tables 6 and 7. As the level of sparsity in the network increases, so does the dispersion in weights and estimates across the leave-out samples. In the sparsest design, $C_N^l = -\log N$ (lower right panel), the weights vary considerably, with many close to zero. Note that the sparser designs also exhibit large outliers, with extremely large errors, but that these outliers typically receive weights close to zero. It is this feature that appears to drive the success of the weighted jackknife over the standard jackknife in the sparser designs ($C_N^l = -\{\log^{1/2} N, \log N\}$) in Tables 6 and 7.

Simulations for leave-l-out style jackknife

Table 8 reports the results of simulations of the leave-l-out style jackknife in a network of $N = 101$ nodes. The model is the same as that in the previous section, with fixed effect distributions given in Table 5. Results are shown for $l = \{5, 10, 20, 50\}$, which corresponds to $20, 10, 5$, and 2 leave-out estimates in each case. The leave-l-out jackknife performs well,

Table 8: Simulation results $N = 101$

$l =$	Bias (mean)				SD				Rejection (5%)			
	5	10	20	50	5	10	20	50	5	10	20	50
Standard jackknife												
bal	0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02	0.05	0.06	0.04	0.05
llog	0.00	0.00	0.00	-0.01	0.03	0.03	0.03	0.03	0.05	0.05	0.03	0.06
slog	-0.01	-0.01	-0.01	-0.01	0.04	0.04	0.04	0.04	0.05	0.04	0.05	0.09
log	-0.09	-0.08	-0.13	-0.53	0.21	0.24	0.32	0.65	0.13	0.15	0.27	0.75
Weighted jackknife												
bal	0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02	0.05	0.05	0.04	0.05
llog	0.00	0.00	0.00	0.00	0.03	0.03	0.03	0.03	0.05	0.05	0.03	0.06
slog	0.00	0.00	0.00	-0.01	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.10
log	0.01	0.02	0.02	-0.53	0.13	0.16	0.20	0.65	0.10	0.10	0.20	0.75

even for l as large as 50 in the dense network scenarios. In the sparsest scenario, the jackknife performs reasonably well for l as large as 20, but appears to break down for $l = 50$. For large $l = 50$, half of all observations in the network are dropped for each leave-out estimation, which appears to create some issues when the network is sparse.

Comparison with other jackknife bias corrections

Finally, we compare the jackknife suggested in this paper to previous suggestions. Specifically, Cruz-Gonzalez et al. (2017) suggest two jackknife bias corrections for network models of the type considered in this paper.[9] The first bias correction is based on a split-sample approach. Divide the agents into two halves, A_1 and A_2. Define $\widehat{\beta}_{\alpha,\gamma/2} = \frac{1}{2}(\widehat{\beta}_{\alpha,\gamma\in A_1} + \widehat{\beta}_{\alpha,\gamma\in A_2})$, where $\widehat{\beta}_{\alpha,\gamma\in A_1}$ is the estimator that uses only observations in which the receiver is in the first set of agents A_1, and $\widehat{\beta}_{\alpha,\gamma\in A_2}$ uses only observations in which the receiver is in A_2. Similarly, define $\widehat{\beta}_{\alpha/2,\gamma} = \frac{1}{2}(\widehat{\beta}_{\alpha\in A_1,\gamma} + \widehat{\beta}_{\alpha\in A_2,\gamma})$ as the average of the two estimators that split the sample based on sending agents. A split-sample jackknife is given by

$$\widehat{\beta}_{ss} = 3\widehat{\beta}_N - \widehat{\beta}_{\alpha/2,\gamma} - \widehat{\beta}_{\alpha,\gamma/2}.$$

The second bias correction is based on dropping all observations associated with a particular

[9]Cruz-Gonzalez et al. (2017) suggest these jackknife corrections for the network model although do not prove their validity. Establishing the validity of the jackknife corrections requires the higher-order asymptotic expansions that are derived in this paper.

Table 9: Comparison of jackknife corrections

	$\bar{N}^{(b)}$	Bias (mean)				SD				Rejection (5%)			
		$\widehat{\theta}_J$	$\widehat{\theta}_{wJ}$	$\widehat{\theta}_d$	$\widehat{\theta}_{ss}$	$\widehat{\theta}_J$	$\widehat{\theta}_{wJ}$	$\widehat{\theta}_d$	$\widehat{\theta}_{ss}$	$\widehat{\theta}_J$	$\widehat{\theta}_{wJ}$	$\widehat{\theta}_d$	$\widehat{\theta}_{ss}$
bal	50	-0.01	0.00	0.00	-0.02	0.04	0.04	0.04	0.04	0.03	0.03	0.05	0.06
llog	50	-0.01	0.00	-0.02	-0.04	0.06	0.05	0.05	0.10	0.04	0.04	0.03	0.11
slog	50	-0.03	0.00	-0.09	-0.26	0.11	0.09	0.25	0.31	0.06	0.05	0.07	0.43
log	29	-1.23	0.05	-1.18	-1.51	1.38	0.47	1.55	1.42	0.25	0.28	0.64	0.61

$\widehat{\theta}_J$ and $\widehat{\theta}_{wJ}$ are the jackknife and weighted jackknife proposed in this paper. $\widehat{\theta}_d$ and $\widehat{\theta}_{ss}$ are the 'double' and 'ss2' jackknife estimators proposed in Cruz-Gonzalez et al. (2017).

agent. Let $\widehat{\beta}_{(i)}$ be the estimate using only observations in the sub-network that excludes agent i. Cruz-Gonzalez et al. (2017) define the 'double' correction as

$$\widehat{\beta}_d = N\widehat{\beta}_N - (N-1)\frac{1}{N}\sum_i \widehat{\beta}_{(i)}.$$

Both jackknife corrections differ from $\widehat{\beta}_J$ in (8), only $\widehat{\beta}_J$ preserves the distribution of fixed effects in each leave-out estimate. This appears to be an important property for the jackknife to perform well in settings with substantial unobserved heterogeneity. Table 9 reports results from simulations of the same model discussed above for the standard and weighted jackknife estimators as well as the split-sample and double corrections suggested by Cruz-Gonzalez et al. (2017). As is clear from the results, although the split-sample and double corrections appear to work well in the densest network DGP, they do not perform well in settings with more heterogeneity and sparser networks. In particular, the split-sample correction has much larger variance, and removes less bias than the leave-one-out style corrections (Hughes and Hahn (2020) derive higher-order bias and variance expressions that explain this phenomenon in the panel setting).

7 Conclusion

This paper presents a new method for bias correcting nonlinear dyadic network models with fixed effects. We provide a novel formulation of the jackknife method that applies to networks with both sender and receiver fixed effects. The jackknife method provides an 'off-the-shelf' procedure for bias correction that is easy to apply, and applicable to a wide set of models.

It allows for discrete multivalued and continuous outcome variables, and is able to obtain estimates of average effects and counterfactual outcomes.

In addition, we show how the jackknife can be used to bias correct averages of functions that depend on multiple observations, including dyads, triads, and tetrads in the network. These averages can be used to produce a wide array of test statistics for the presence of strategic interactions in the network, such as reciprocity or transitivity. In simulations, we show that the jackknife performs well, even in relatively low density networks, and outperforms previous suggestions for jackknife procedures.

There are a number of interesting areas in which the work in this paper might be usefully extended. The jackknife procedure developed in this paper applies to data on a single observation of a network. It would be interesting to extend these results to networks observed over multiple time periods, perhaps with the addition of time fixed effects. It is expected that a similar jackknife procedure could also be used in this setting, with appropriate splitting across the time dimension to account for dynamics in the network. The jackknife procedures proposed in this paper may also be useful in the interactive fixed effect model of Chen et al. (2021), and establishing their validity in this setting would also be useful.

References

Anderson, J. E. and Van Wincoop, E. 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192.

Bloch, F. and Jackson, M. O. 2007. The Formation of Networks with Transfers Among Players. *Journal of Econometric Theory*, 133(1):83–110.

Burger, M., van Oort, F., and Linders, G. J. 2009. On the Ppecification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. *Spatial Economic Analysis*, 4 (2):167–190.

Candelaria, L. E. 2020. A Semiparametric Network Formation Model with Unobserved Linear Heterogeneity.

Charbonneau, K. B. 2017. Multiple Fixed Effects in Binary Response Panel Data Models. *The Econometrics Journal*, 20(3):1–13.

Chen, M., Fernández-Val, I., and Weidner, M. 2021. Nonlinear Factor Models for Network and Panel Data. *Journal of Econometrics*, 220:296–324.

Cruz-Gonzalez, M., Fernández-Val, I., and Weidner, M. 2017. Bias Corrections for Probit and Logit Models with Two-way Fixed Effects. *The Stata Journal*, 17(3):517–545.

de Paula, Á. 2020. Econometrics of Network Models. *Annual Review of Economics*, 12: 775–799.

Dhaene, G. and Jochmans, K. 2015. Split-panel Jackknife Estimation of Fixed-effect Models. *Review of Economic Studies*, 82(3):991–1030.

Dzemski, A. 2019. An Empirical Model of Dyadic Link Formation in a Network with Unobserved Heterogeneity. *Review of Economics and Statistics*, 101(5):763–776.

Fernández-Val, I. and Weidner, M. 2016. Individual and Time Effects in Nonlinear Panel Models with Large N, T. *Journal of Econometrics*, 192(1):291–312.

Gao, W. Y. 2020. Nonparametric Identification in Index Models of Link Formation. *Journal of Econometrics*, 215(2):399–413.

Graham, B. S. 2017. An Econometric Model of Network Formation With Degree Heterogeneity. *Econometrica*, 85(4):1033–1063.

Graham, B. S. 2020. Network Data. *Handbook of Econometrics*, 7.

Graham, B. S. and Pelican, A. 2020. An Optimal Test for Strategic Interaction in Social and Economic Network Formation Between Heterogeneous Agents.

Hahn, J. and Newey, W. K. 2004. Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica*, 72:1295–1319.

Helpman, E., Melitz, M., and Rubinstein, Y. 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics*, 123(2):441–487.

Holland, P. W. and Leinhardt, S. 1981. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373):33–50.

Hughes, D. W. and Hahn, J. 2020. The Higher-order Variance of Bias-corrected Panel Estimators.

Jochmans, K. 2018. Semiparametric Analysis of Network Formation. *Journal of Business and Economic Statistics*, 36(4):705–713.

Linders, G.-J. and de Groot, H. L. F. 2006. Estimation of the Gravity Equation in the Presence of Zero Flows.

Newey, W. K. and McFadden, D. 1994. Chapter 36: Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics*, 4:2111–2245.

Neyman, J. and Scott, E. L. 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 16:1–32.

Pratt, J. W. 1981. Concavity of the Log Likelihood. *Journal of the American Statistical Association*, 76:103–106.

Rose, A. K. 2004. Do We Really Know That the WTO Increases Trade? *American Economic Review*, 94(1):98–114.

Santos Silva, J. M. C. and Tenreyro, S. 2006. The Log of Gravity. *Review of Economic Studies*, 88(4):641–658.

Toth, P. *Semiparametric Estimation in Network Formation Models with Homophily and Degree Heterogeneity*. PhD thesis, 2017.

van der Vaart, A. W. *Asymptotic Statistics*. Cambridge University Press, 1998.

White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50 (1):1–25.

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. 2019. Statistical Inference in a Directed Network Model With Covariates. *Journal of the American Statistical Association*, 114 (526).

Zeleneev, A. 2020. Identification and Estimation of Network Models with Nonparameteric Unobserved Heterogeneity.

A Notation and norms

The notation in the appendices follows that in the main paper. That is, denote partial derivatives of the objective function using subscripts, so that $\partial_\beta \mathcal{L}(\beta,\phi)$ denotes $\partial \mathcal{L}(\beta,\phi)/\partial\beta$ and so on, where $\phi' = (\alpha', \gamma')$. When functions are evaluated at β_0, ϕ_0 the dependence on these arguments is dropped. We also use ℓ_{ij} as shorthand for the function $\ell(Z_{ij}, \cdot)$, with $Z_{ij} = (Y_{ij}, X_{ij})$. Let

$$\mathcal{S}(\beta,\phi) = \partial_\phi \mathcal{L}(\beta,\phi), \qquad \mathcal{S}_\beta(\beta,\phi) = \partial_\beta \mathcal{L}(\beta,\phi)$$

be the first derivatives of the objective function. We also write

$$\mathcal{H}(\beta,\phi) = -\partial_{\phi\phi'} \mathcal{L}(\beta,\phi)$$

for the negative of the Hessian with respect to the fixed effects, a $2N \times 2N$ matrix.

We follow Fernández-Val and Weidner (2016) (FVW) in using the Euclidean norm $\|\cdot\|$ for $\dim\beta$ vectors, and the norm induced by the Euclidean norm for matrices and tensors, i.e.

$$\|\partial_{\beta\beta\beta}\mathcal{L}(\beta,\phi)\| = \max_{u,v\in\mathbb{R}^{\dim\beta}:\|u\|=1,\|v\|=1} \|\sum_{k,l=1}^{\dim\beta} u_k v_l \partial_{\beta\beta_k\beta_l}\mathcal{L}(\beta,\phi)\|$$

In the proofs we sometimes take β to be a scalar to simplify notation, although the results apply to any vector of fixed size. Since the number of fixed effect parameters in the model grows with N, the choice of norm for $\dim \phi$ vectors and matrices is important. Following FVW, we choose the ℓ_q-norm for $\dim \phi$ vectors and the corresponding induced norms for matrices and tensors

$$\|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\phi)\| = \max_{u,v\in\mathbb{R}^{\dim\phi}:\|u\|=1,\|v\|=1} \|\sum_{k,l=1}^{\dim\beta} u_k v_l \partial_{\phi\phi_k\phi_l}\mathcal{L}(\beta,\phi)\|_q$$

See FVW for more details on these norms.

We define the sets $\mathcal{B}(r,\beta_0) = \{\beta : \|\beta-\beta_0\| \leq r\}$, for $r > 0$, and $\mathcal{B}_q(r,\phi_0) = \{\phi : \|\phi-\phi_0\|_q \leq r\}$.

B Asymptotic expansions

B.1 Verifying Assumption B.1 in Fernández-Val and Weidner (2016)

The results in the paper are based on asymptotic expansions of the objective function. Fernández-Val and Weidner (2016) (FVW) derive expansions for a general class of M-estimators with multiple incidental parameters, which includes the model studied here. In order to determine the properties of the jackknife estimator, I extend the expansions in FVW to higher-order, which requires additional conditions on the number of moments and derivatives of the objective function that exist. Derivations of higher-order terms and their bounds are quite long and largely similar to the derivations in FVW, and so are provided in the Supplementary Appendix. This appendix contains analysis based on the first-order expansions and focuses on results related to the jackknife, which are of most interest. To do so, I begin by verifying the conditions in Assumption B.1. of FVW, which will allow us to use some important results in that paper, including consistency of the common parameter and vector of fixed effects.

As is the paper, we use $\bar{E}$ to denote expectation conditional on exogenous covariates and fixed effects. Let $\bar{\mathcal{H}} = \bar{E}[\mathcal{H}]$, $\tilde{\mathcal{H}} = \mathcal{H} - \bar{\mathcal{H}}$, and similarly for other conditional expectations and their residuals. As in FVW, we may write

$$\bar{\mathcal{H}} = \begin{bmatrix} \bar{\mathcal{H}}^*_{\alpha\alpha} & \bar{\mathcal{H}}^*_{\alpha\gamma} \\ \bar{\mathcal{H}}^*_{\gamma\alpha} & \bar{\mathcal{H}}^*_{\gamma\gamma} \end{bmatrix} + \frac{1}{N} b v_N v_N'$$

where $\bar{\mathcal{H}}^*_{\alpha\alpha}$ and $\bar{\mathcal{H}}^*_{\gamma\gamma}$ are the diagonal matrices with elements

$$\left(\bar{\mathcal{H}}^*_{\alpha\alpha}\right)_{ii} = -\frac{1}{N-1} \sum_{j \neq i} \bar{E}[\partial_{\pi^2} \ell_{ij}]$$

$$\left(\bar{\mathcal{H}}^*_{\gamma\gamma}\right)_{ii} = -\frac{1}{N-1} \sum_{j \neq i} \bar{E}[\partial_{\pi^2} \ell_{ji}]$$

and $\bar{\mathcal{H}}^*_{\alpha\gamma} = (\bar{\mathcal{H}}^*_{\gamma\alpha})'$ has off-diagonal entries $\left(\bar{\mathcal{H}}^*_{\alpha\gamma}\right)_{ij} = -\bar{E}[\partial_{\pi^2} \ell_{ij}]/(N-1)$ and zeroes in diagonal entries. As in Lemma D.1 of FVW and Lemma A.1 in Dzemski (2019), we can show that $\bar{\mathcal{H}}^{-1}$ is dominated by its diagonal elements

$$\|\bar{\mathcal{H}}^{-1} - diag\left(\bar{\mathcal{H}}^*_{\alpha\alpha}, \bar{\mathcal{H}}^*_{\gamma\gamma}\right)^{-1}\|_{max} = O_p(N^{-1}) \tag{23}$$

We now demonstrate that the conditions in Assumption B.1 of FVW hold; a similar proof is provided by Dzemski (2019). Given the stronger moment conditions in Assumption 1, we can choose $q = 16$, $\epsilon = 1/(32 + 2\nu)$, $r_\beta = \log(N)N^{-1/4}$, and $r_\phi = N^{-3/16}$.

Assumption (i) holds trivially since $N = T$ in the network setting, while (ii) follows from Assumption 1.

For part (iv), by (23), the condition holds by Assumption 1 (v). Condition (v) follows similarly to the proofs in FVW for the panel case. For example, we have by Assumption 1 (iii) for $q \leq 16$

$$
\begin{aligned}
&\bar{E}\Big[\sup_{\beta\in\mathcal{B}(r_\beta,\beta_0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi_0)} \frac{1}{N}\sum_i \Big|\frac{1}{N-1}\sum_{j\neq i} \partial_{\beta_k \pi}\ell_{ij}\Big|^q\Big] \\
&\leq \bar{E}\Big[\sup_{\beta\in\mathcal{B}(r_\beta,\beta_0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi_0)} \frac{1}{N}\sum_i \Big(\frac{1}{N-1}\sum_{j\neq i} |\partial_{\beta_k \pi}\ell_{ij}|\Big)^q\Big] \\
&\leq \bar{E}\Big[\frac{1}{N}\sum_i \frac{1}{N-1}\sum_{j\neq i} M(Z_{ij})^q\Big] \\
&= O_p(1)
\end{aligned}
$$

and so

$$
\begin{aligned}
\|\partial_{\beta_k \alpha}\mathcal{L}\|_q &= \Big(\sum_i \Big|\frac{1}{N-1}\sum_{j\neq i} \partial_{\beta_k \pi}\ell_{ij}\Big|^q\Big)^{1/q} \\
&= O_p(N^{1/q})
\end{aligned}
$$

and similarly for $\|\partial_{\beta_k \gamma}\mathcal{L}\|_q$, and hence $\|\partial_{\beta_k \phi}\mathcal{L}\|_q$. For part (vi) the proofs in the panel case again carry over, for example see the proof in Lemma A.2 of Dzemski (2019). Inspection of that proof shows that we can get the bound $\|\tilde{\mathcal{H}}\| = O_p(N^{2\epsilon-\frac{1}{2}})$, for $\epsilon > \frac{1}{32}$, which will be useful in the higher-order approximations.

Given Assumptions B.1 (i), (ii), (iv), (v), and (vi) we can apply Theorem B.3 in FVW to establish consistency of the estimates for common parameters and fixed effects, and to establish the bounds

$$
\begin{aligned}
\sup_{\beta\in\mathcal{B}(r_\beta,\beta_0)} \|\widehat{\phi}(\beta) - \phi_0\|_q &= o_p(r_\phi) \\
\|\widehat{\beta} - \beta_0\| &= O_p(N^{-1/2})
\end{aligned}
\tag{24}
$$

B.2 Asymptotic expansions

The next lemma gives the asymptotic expansion for the estimated fixed effects and common parameters. It is in part a restatement of Theorem B.1 in Fernández-Val and Weidner (2016), but the remainder terms are split into two parts. The expressions for the remainder terms are given in the Supplementary Appendix, and are used to derive the properties of the jackknife bias correction. The proof of Lemma 1 is provided in the Supplementary Appendix.

Lemma 1. *Let Assumption 1 hold. Then,*

$$
\begin{aligned}
\widehat{\phi} - \phi_0 &= \mathcal{H}^{-1}\mathcal{S} \\
&\quad - W^{-1}\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L})\big(\mathcal{S}_\beta + (\partial_{\beta\phi}\mathcal{L})'\mathcal{H}^{-1}\mathcal{S}\big) \\
&\quad + W^{-1}\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L})\frac{1}{2}\sum_g \big((\partial_{\beta\phi\phi_g}) + (\partial_{\phi\phi'\phi_g}\mathcal{L})\big)[\mathcal{H}^{-1}\mathcal{S}]_g \mathcal{H}^{-1}\mathcal{S} \\
&\quad + \frac{1}{2}\mathcal{H}^{-1}\sum_g (\partial_{\phi\phi'\phi_g}\mathcal{L})[\mathcal{H}^{-1}\mathcal{S}]_g \mathcal{H}^{-1}\mathcal{S} \\
&\quad + R_\phi + \tilde{R}_\phi \\
&= \mathcal{H}^{-1}\mathcal{S} - W^{-1}\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L})\big(U^{(0)} + U^{(1)}\big) \\
&\quad + \frac{1}{2}\mathcal{H}^{-1}\sum_g (\partial_{\phi\phi'\phi_g}\mathcal{L})[\mathcal{H}^{-1}\mathcal{S}]_g \mathcal{H}^{-1}\mathcal{S} \\
&\quad + R_\phi + \tilde{R}_\phi
\end{aligned}
$$

and

$$
N\bar{W}_N(\widehat{\beta} - \beta_0) = U^{(0)} + U^{(1)} + R_\beta + \tilde{R}_\beta
$$

where

$$
\begin{aligned}
U^{(0)} &= (\partial_\beta \mathcal{L}) + (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S} \\
U^{(1)} &= (\partial_{\beta\phi'}\tilde{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} \\
&\quad + \frac{1}{2}\sum_{g=1}^{\dim\phi} \big((\partial_{\beta\phi\phi_g}\bar{\mathcal{L}}) + (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})\big)\big[\bar{\mathcal{H}}^{-1}\mathcal{S}\big]_g \bar{\mathcal{H}}^{-1}\mathcal{S}
\end{aligned}
$$

and the remainders satisfy $\|R_\phi\| = o_p(1)$, $\|R_\beta\| = o_p(1)$ *and* $\|\tilde{R}_\phi\| = o_p(N^{-1})$, $\|\tilde{R}_\beta\| = o_p(N^{-1})$.

C Jackknife results for β

Here we allow for a more general construction of the leave-out sets, but impose two important conditions.

Condition 1. Let $\mathcal{I}_k$ for $k = 1, \dots N_k$ be a partition of the observations in a network of size N. Define $\mathbb{1}^k_{ij} = \mathbb{1}\{(i,j) \notin \mathcal{I}_k\}$ as an indicator that the observation (i,j) is *not* included in the k-th set. We impose the following conditions on the sets:

(i) $\sum_{k=1}^{N_k}(1 - \mathbb{1}^k_{ij}) = 1$ for all (i,j)

(ii) $\sum_{j \neq i}(1 - \mathbb{1}^k_{ij}) = \sum_{i \neq j}(1 - \mathbb{1}^k_{ij}) = 1$ for all i, j, and k.

Condition 1 imposes two important constrains on the sets $\mathcal{I}_k$: (i) that they are mutually exclusive, such that every edge (i,j) appears in exactly one of the sets, and (ii) that each set contains exactly one observation related to each of the N sender fixed effects α and each of the N receiver fixed effects γ. The first condition ensures that all observations are used equally in the jackknife, and is important for showing that the asymptotic variance of the estimator is not affected by the jackknife. The second condition ensures that each fixed effect parameter is affected equally in the leave-out sets, and that $\bar{\mathcal{H}}_{(k)}$ is well-defined and positive definite. We assume in the proofs below that $N_k = N - 1$ and that we are using the leave-one-out jackknife. All of the results still hold for the leave-l-out style jackknife for fixed l.

Some additional notation will also be useful for studying the jackknife estimates. Let A be some statistic and $A_{(k)}$ the same statistic in the k-th leave-out sample. We define the jackknife operator $\mathbf{J}$ as

$$\mathbf{J}[A] = A_J = (N-1)A - (N-2)\frac{1}{N-1}\sum_{k=1}^{N-1} A_{(k)}$$

Additionally, we define a set of indicators that count the number of unique leave-out sets $\mathcal{I}_k$ that a group of edges $(i_1, j_1), \dots, (i_t, j_t)$ are contained in. Let

$$I^r_{(i_1,j_1),\dots,(i_t,j_t)} = \begin{cases} 1 & (i_1,j_1),\dots,(i_t,j_t) \text{ span exactly } r \text{ of the sets } \mathcal{I}_k \\ 0 & \text{otherwise} \end{cases}$$

Assume that the researcher randomizes the ordering of the N node labels by choosing, with equal probability a labelling from the $N!$ possible orderings. Since the ordering of nodes is random, the indicator 1_{ij}^k is a random variable, with mean equal to the probability that $(i,j) \notin \mathcal{I}_k$ across each of the possible orderings. There are $N \times (N-2)!$ ways to order the N nodes keeping $(i,j) \in \mathcal{I}_k$ for any fixed k, so that $E[1_{ij}^k] = \frac{N-2}{N-1}$. The following lemma states that the expectation, over the randomness induced by the ordering of nodes, of sums in the leave-out sets is equal to that in the full sample.

Lemma 2. *Let 1_{ij}^k satisfy Condition 1, and define the sums over random variable A_{ij} in the full-sample and k-th leave-out sample*

$$A = \frac{1}{N-1}\sum_i \sum_{j \neq i} A_{ij}$$
$$A_{(k)} = \frac{1}{N-2}\sum_i \sum_{j \neq i} A_{ij} 1_{ij}^k$$

Then, we have that $\bar{E}[A] = \bar{E}[A_{(k)}]$ for any $k = 1, \ldots, N-1$.

Proof. The proof follows simply from the fact that $E[1_{ij}^k] = \frac{N-2}{N-1}$, where the expectation is taken over the randomness induced by the random ordering of nodes. Note that the ordering of nodes is independent of any other randomness in the sample, so that

$$\begin{aligned}\bar{E}[A_{ij}1_{ij}^k] &= E[A_{ij}1_{ij}^k | X_{ij}, \alpha_i, \gamma_j] \\ &= E[A_{ij}|X_{ij}, \alpha_i, \gamma_j] E[1_{ij}^k] \\ &= \frac{N-2}{N-1} E[A_{ij}|X_{ij}, \alpha_i, \gamma_j]\end{aligned}$$

Then we have

$$\begin{aligned}\bar{E}[A_{(k)}] &= \frac{1}{N-2}\sum_i \sum_{j \neq i} E[A_{ij}|X_{ij}, \alpha_i, \gamma_j] \frac{N-2}{N-1} \\ &= \bar{E}[A]\end{aligned}$$

as required. □

An important corollary of 2 is that $\bar{\mathcal{H}} = \bar{\mathcal{H}}_{(k)}$ and $\bar{W} = \bar{W}_{(k)}$.

C.1 Jackknifing first-order expansion terms

We next use this result to show the impact of the jackknife operator on the first-order expansion of $N(\widehat{\beta}-\beta)$. From the expansion in Lemma 1, we have that a first-order approximation is given by

$$N\bar{W}_N(\widehat{\beta}-\beta) = \left(U^{(0)} + U^{(1)}\right) + o_p(1)$$

where

$$\begin{aligned} U^{(0)} &= (\partial_\beta \mathcal{L}) + (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S} \\ U^{(1)} &= (\partial_{\beta\phi'}\tilde{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} \\ &\quad + \frac{1}{2}\sum_{g=1}^{\dim\phi}\left((\partial_{\beta\phi\phi_g}\bar{\mathcal{L}}) + (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})\right)\left[\bar{\mathcal{H}}^{-1}\mathcal{S}\right]_g\bar{\mathcal{H}}^{-1}\mathcal{S} \end{aligned}$$

The next lemmas demonstrate the effect of the jackknife on general sums of the forms in $U^{(0)}$ and $U^{(1)}$.

Lemma 3. *Let 1_{ij}^k satisfy Condition 1. For A_{ij} a mean-zero random variable, let*

$$\begin{aligned} A_i &= \frac{1}{N-1}\sum_{s\neq i} A_{is} \\ A_{(k),i} &= \frac{1}{N-2}\sum_{s\neq i} A_{is}1_{is}^k \end{aligned}$$

Define the jackknifed version of A_i as

$$A_{J,i} = \mathbf{J}[A_i] = (N-1)A_i - (N-2)\frac{1}{N-1}\sum_{k=1}^{N-1}A_i^k$$

Then, $A_{J,i} = A_i$.

Proof. First note that by Condition 1, $\sum_k 1_{is}^k = N-2$. Then, $\sum_k A_{(k),i} = \sum_{s\neq i} A_{is}$ so that

$$\begin{aligned} A_{J,i} &= (N-1)A_i - (N-2)\frac{1}{N-1}\sum_{k=1}^{N-1} A_{(k),i} \\ &= (N-1)A_i - (N-2)\frac{1}{N-1}\sum_{s\neq i} A_{is} \\ &= \frac{1}{N-1}\sum_{s\neq i} A_{is} = A_i \end{aligned}$$

□

Lemma 4. *Let 1_{ij}^k satisfy Condition 1. For A_{ij} a mean-zero random variable with bounded fourth moment, let*

$$\begin{aligned} \mathbf{A} &= \frac{1}{N-1}\big(\{\sum_{s\neq i} A_{is}\}_{i=1,\ldots,N}, \{\sum_{s\neq j} A_{sj}\}_{j=1,\ldots,N}\big) \\ &= (\mathbf{A}_\alpha, \mathbf{A}_\gamma) \\ \mathbf{A}_k &= \frac{1}{N-2}\big(\{\sum_{s\neq i} A_{is}1_{is}^k\}_{i=1,\ldots,N}, \{\sum_{s\neq j} A_{sj}1_{sj}^k\}_{j=1,\ldots,N}\big) \\ &= (\mathbf{A}_{\alpha,k}, \mathbf{A}_{\gamma,k}) \end{aligned}$$

and let $\mathbf{B}$ and $\mathbf{B}_k$ be similarly defined vectors of sums involving mean-zero random variables B_{ij}. Assume that (A_{ij}, B_{ij}) are independent of (A_{st}, B_{st}) for $(i,j) \notin \{(s,t),(t,s)\}$. Define the jackknifed term

$$\mathcal{J}_0 = (N-1)\mathbf{A}'M\mathbf{B} - \frac{N-2}{N-1}\sum_k \mathbf{A}'_{(k)}M\mathbf{B}_{(k)}$$

where M is a non-random matrix that has $O_p(1)$ elements on its diagonal and $O_p(N^{-1})$ off-diagonal terms. Then we have: (i) $\bar{E}\big[\mathcal{J}_0\big] = o_p(1)$, and (ii) $\mathcal{J}_0 = o_p(1)$.

Proof. The most common choice of M will be $\bar{\mathcal{H}}^{-1}$, which satisfies the conditions for M by Assumption 1 and the approximation property in (23). We show the proof using $\bar{\mathcal{H}}^{-1}$, but note that it holds for any M satisfying the conditions stated above. We have

$$\begin{aligned}\mathbf{A}'\bar{\mathcal{H}}^{-1}\mathbf{B} &= \mathbf{A}_\alpha'\bar{\mathcal{H}}_{\alpha\alpha}^{-1}\mathbf{B}_\alpha + \mathbf{A}_\alpha'\bar{\mathcal{H}}_{\alpha\gamma}^{-1}\mathbf{B}_\gamma \\ &\quad + \mathbf{A}_\gamma'\bar{\mathcal{H}}_{\gamma\alpha}^{-1}\mathbf{B}_\alpha + \mathbf{A}_\gamma'\bar{\mathcal{H}}_{\gamma\gamma}^{-1}\mathbf{B}_\gamma\end{aligned}$$

Recall that $I^1_{(ij)(st)}$ is one whenever (i,j) and (s,t) are contained in the same $\mathcal{I}_k$, and so $\sum_k 1^k_{ij}1^k_{st} = (N-2)I^1_{(ij)(st)} + (N-3)(1-I^1_{(ij)(st)})$.

$$\begin{aligned}\mathbf{A}_\alpha'\bar{\mathcal{H}}_{\alpha\alpha}^{-1}\mathbf{B}_\alpha &= \sum_{i,j}\mathbf{A}_{\alpha,i}(\bar{\mathcal{H}}_{\alpha\alpha}^{-1})_{ij}\mathbf{B}_{\alpha,j} \\ &= \frac{1}{(N-1)^2}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}(\bar{\mathcal{H}}_{\alpha\alpha}^{-1})_{ij}A_{is}B_{jt} \\ \frac{1}{N-1}\sum_k \mathbf{A}_{k,\alpha}'\bar{\mathcal{H}}_{\alpha\alpha}^{-1}\mathbf{B}_{k,\alpha} &= \frac{1}{(N-1)(N-2)^2}\sum_k\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}(\bar{\mathcal{H}}_{\alpha\alpha}^{-1})_{ij}A_{is}B_{jt}1^k_{is}1^k_{jt} \\ &= \frac{1}{(N-1)(N-2)}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}(\bar{\mathcal{H}}_{\alpha\alpha}^{-1})_{ij}A_{is}B_{jt}I^1_{(is)(jt)} \\ &\quad + \frac{N-3}{(N-1)(N-2)^2}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}(\bar{\mathcal{H}}_{\alpha\alpha}^{-1})_{ij}A_{is}B_{jt}(1-I^1_{(is)(jt)})\end{aligned}$$

Then, we have

$$\begin{aligned}\mathcal{J}_{\alpha\alpha} =&(N-1)\mathbf{A}_\alpha'\bar{\mathcal{H}}_{\alpha\alpha'}^{-1}\mathbf{B}_\alpha - \frac{N-2}{N-1}\sum_k \mathbf{A}_{\alpha,k}'\bar{\mathcal{H}}_{\alpha\alpha'}^{-1}\mathbf{B}_{\alpha,k} \\ =&\frac{1}{(N-1)(N-2)}\sum_i\sum_j\sum_{s\neq i}\sum_{t\neq j}(\bar{\mathcal{H}}_{\alpha\alpha'}^{-1})_{ij}\big(A_{is}B_{jt}\big)(1-I^1_{(is)(jt)})\end{aligned}$$

Similar computations for the other three elements gives

$$\begin{aligned}\mathcal{J}_0 =&(N-1)\mathbf{A}'\bar{\mathcal{H}}^{-1}\mathbf{B} - \frac{N-2}{N-1}\sum_k \mathbf{A}_k'\bar{\mathcal{H}}^{-1}\mathbf{B}_k \\ =&\frac{1}{(N-1)(N-2)}\sum_i\sum_j\sum_{s\neq i}\sum_{t\neq j}\Big\{(\bar{\mathcal{H}}_{\alpha\alpha'}^{-1})_{ij} + (\bar{\mathcal{H}}_{\alpha\gamma'}^{-1})_{it} \\ &+ (\bar{\mathcal{H}}_{\gamma\alpha'}^{-1})_{sj} + (\bar{\mathcal{H}}_{\gamma\gamma'}^{-1})_{st}\Big\}\big(A_{is}B_{jt}\big)(1-I^1_{(is)(jt)})\end{aligned}$$

Now, since $\bar{E}[A_{is}B_{jt}(1 - I^1_{(is)(jt)})] \neq 0$ only when $(j,t) = (s,i)$, we have

$$\begin{aligned}\bar{E}[\mathcal{J}_0] =& \frac{1}{(N-1)(N-2)} \sum_i \sum_{s \neq i} \Big\{ (\bar{\mathcal{H}}^{-1}_{\alpha\alpha'})_{is} + (\bar{\mathcal{H}}^{-1}_{\alpha\gamma'})_{ii} \\ & + (\bar{\mathcal{H}}^{-1}_{\gamma\alpha'})_{ss} + (\bar{\mathcal{H}}^{-1}_{\gamma\gamma'})_{si} \Big\} \big(A_{is}B_{si}\big) \\ =& o_p(1)\end{aligned}$$

since, for for $i \neq s$, we have $(\bar{\mathcal{H}}^{-1}_{\alpha\alpha'})_{is} = O_p(N^{-1})$ and $(\bar{\mathcal{H}}^{-1}_{\gamma\gamma'})_{si} = O_p(N^{-1})$, while $(\bar{\mathcal{H}}^{-1}_{\alpha\gamma'})_{ii}$ and $(\bar{\mathcal{H}}^{-1}_{\gamma\alpha'})_{ss}$ are both $O_p(N^{-1})$ also.

Next, let $\Gamma_{ijst} = (\bar{\mathcal{H}}^{-1}_{\alpha\alpha'})_{ij} + (\bar{\mathcal{H}}^{-1}_{\alpha\gamma'})_{it} + (\bar{\mathcal{H}}^{-1}_{\gamma\alpha'})_{sj} + (\bar{\mathcal{H}}^{-1}_{\gamma\gamma'})_{st}$. We have

$$\Gamma_{ijst} = \begin{cases} O_p(1) & i = j \text{ or } s = t \\ O_p(N^{-1}) & \text{otherwise} \end{cases}$$

We can decompose $\mathcal{J}_0$ as

$$\begin{aligned}\mathcal{J}_0 = & \frac{1}{(N-1)(N-2)} \sum_i \sum_j \Big(\sum_{s<i} \sum_{t<j} \Gamma_{ijst} A_{is} B_{jt} + \sum_{s<i} \sum_{t>j} \Gamma_{ijst} A_{is} B_{jt} \\ & + \sum_{s>i} \sum_{t<j} \Gamma_{ijst} A_{is} B_{jt} + \sum_{s>i} \sum_{t>j} \Gamma_{ijst} A_{is} B_{jt} \Big) \\ & = \mathcal{J}_{0,11} + \mathcal{J}_{0,12} + \mathcal{J}_{0,21} + \mathcal{J}_{0,22}\end{aligned}$$

Then we have

$$
\begin{aligned}
\bar{E}[\mathcal{J}_{0,11}^2] &= \frac{1}{(N-1)^2(N-2)^2}\sum_i\sum_j\sum_k\sum_l\sum_{s<i}\sum_{t<j}\sum_{p<k}\sum_{q<l}\Gamma_{ijst}\Gamma_{klpq} \\
&\times E\big[A_{is}B_{jt}A_{kp}B_{lq}\big](1-I^1_{(is)(jt)})(1-I^1_{(kp)(lq)}) \\
&= \frac{1}{(N-1)^2(N-2)^2}\sum_i\sum_j\sum_{s<i}\sum_{t<j}\Gamma^2_{ijst}\bar{E}[A^2_{is}B^2_{jt}](1-I^1_{(is)(jt)}) \\
&+ \frac{1}{(N-1)^2(N-2)^2}\sum_i\sum_j\sum_{s<i}\sum_{t<j}\Gamma_{ijst}\Gamma_{stij}\bar{E}\big[A_{is}B_{is}A_{jt}B_{jt}\big](1-I^1_{(is)(jt)}) \\
&= \frac{1}{(N-1)^2(N-2)^2}\sum_i\sum_{s<i}\sum_{t<i}\Gamma^2_{iist}\bar{E}[A^2_{is}]\bar{E}[B^2_{it}](1-I^1_{(is)(it)}) \\
&+ \frac{1}{(N-1)^2(N-2)^2}\sum_i\sum_{j\neq i}\sum_{s<(i\wedge j)}\Gamma^2_{ijss}\bar{E}[A^2_{is}]\bar{E}[B^2_{js}](1-I^1_{(is)(js)}) \\
&+ \frac{1}{N^2(N-1)^2(N-2)^2}\sum_i\sum_{j\neq i}\sum_{s<i}\sum_{t<j,t\neq s}N^2\Gamma^2_{ijst}\bar{E}[A^2_{is}]\bar{E}[B^2_{jt}](1-I^1_{(is)(jt)}) \\
&+ \frac{1}{(N-1)^2(N-2)^2}\sum_i\sum_{s<i}\sum_{t<i}\Gamma_{iist}\Gamma_{stii}\bar{E}[A_{is}B_{is}]\bar{E}\big[A_{it}B_{it}\big](1-I^1_{(is)(it)}) \\
&+ \frac{1}{(N-1)^2(N-2)^2}\sum_i\sum_{j\neq i}\sum_{s\neq(i\wedge j)}\Gamma_{ijss}\Gamma_{ssij}\bar{E}[A_{is}B_{is}]\bar{E}\big[A_{js}B_{js}\big](1-I^1_{(is)(js)}) \\
&+ \frac{1}{N^2(N-1)^2(N-2)^2}\sum_i\sum_{j\neq i}\sum_{s<i}\sum_{t<j,t\neq s}N^2\Gamma_{ijst}\Gamma_{stij}\bar{E}[A_{is}B_{is}]\bar{E}\big[A_{jt}B_{jt}\big](1-I^1_{(is)(jt)}) \\
&= O_p(N^{-1})
\end{aligned}
$$

Where the last line follows from the properties of Γ_{ijst}. The same result holds for $\bar{E}[\mathcal{J}_{0,12}^2]$, $\bar{E}[\mathcal{J}_{0,21}^2]$, and $\bar{E}[\mathcal{J}_{0,22}^2]$, hence $\mathcal{J}_0 = O_p(N^{-1/2}) = o_p(1)$. □

The following lemma derives the forms of $\mathbf{J}[U^{(0)}]$ and $\mathbf{J}[U^{(1)}]$.

Lemma 5. *Let Assumption 1 hold. Then*

$$
\begin{aligned}
\mathbf{J}[U^{(0)}] &= U^{(0)} \\
\mathbf{J}[U^{(1)}] &= o_p(1)
\end{aligned}
$$

Proof. For $U^{(0)} = (\partial_\beta\mathcal{L}) + (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S}$ we can appeal to Lemma 3 with $A_i = (\partial_\beta\mathcal{L})_i = \frac{1}{N-1}\sum_{j\neq i}\partial_\beta\ell_{ij}$ and with $A_i = \mathcal{S}_i = \frac{1}{N-1}\sum_{j\neq i}\partial_\pi\ell_{ij}$. Note that the jackknife operator is linear so that, since $(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}$ are fixed across leave-out samples (by Lemma 2),

$$\mathbf{J}[\partial_\beta \mathcal{L}] = \sum_i \mathbf{J}[(\partial_\beta \mathcal{L})_i]$$
$$\mathbf{J}\big[(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S}\big] = (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathbf{J}[\mathcal{S}]$$

Now, for $U^{(1)}$ we can appeal to Lemma 4. For the first term, $(\partial_{\beta\phi'}\tilde{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S}$ we set $A_{ij} = \partial_{\beta\pi}\tilde{\ell}_{ij}$ and $B_{ij} = \partial_\pi \ell_{ij}$. For the second term, $(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}$ we note that

$$(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}} = \big((\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}_{\cdot,\alpha}, (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}_{\cdot,\gamma}\big)$$

where the i-th element of $(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}_{\cdot,\alpha}$ is

$$\begin{aligned}
(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}_{\cdot,\alpha_i} &= (\partial_{\beta\alpha'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\tilde{\mathcal{H}}_{\alpha\alpha_i} + (\partial_{\beta\alpha'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\alpha\gamma}\tilde{\mathcal{H}}_{\gamma\alpha_i} \\
&\quad + (\partial_{\beta\gamma'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\gamma\alpha}\tilde{\mathcal{H}}_{\alpha\alpha_i} + (\partial_{\beta\gamma'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\gamma\gamma}\tilde{\mathcal{H}}_{\gamma\alpha_i} \\
&= -(\partial_{\beta\alpha'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\alpha\alpha_i}\frac{1}{N-1}\sum_{j\neq i}\partial_{\pi^2}\tilde{\ell}_{ij} - \frac{1}{N-1}\sum_{j\neq i}(\partial_{\beta\alpha'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\alpha\gamma_j}\partial_{\pi^2}\tilde{\ell}_{ij} \\
&\quad - (\partial_{\beta\gamma'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\gamma\alpha_i}\frac{1}{N-1}\sum_{j\neq i}\partial_{\pi^2}\tilde{\ell}_{ij} - \frac{1}{N-1}\sum_{j\neq i}(\partial_{\beta\gamma'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\gamma\gamma_j}\partial_{\pi^2}\tilde{\ell}_{ij}
\end{aligned}$$

and similarly for elements of $(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}_{\cdot,\gamma}$. So we can let

$$A_{ij} = \big((\partial_{\beta\alpha'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\alpha\alpha_i} + (\partial_{\beta\alpha'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\alpha\gamma_j} + (\partial_{\beta\gamma'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\gamma\alpha_i} + (\partial_{\beta\gamma'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}_{\gamma\gamma_j}\big)\partial_{\pi^2}\tilde{\ell}_{ij}$$

and $B_{ij} = \partial_\pi \ell_{ij}$ in Lemma 4. Note that we have A_{ij} and B_{ij} mean zero and the moment condition also holds by assumption. For the final term, we begin by demonstrating that both $\bar{\mathcal{H}}^{-1}\sum_g(\partial_{\beta\phi'\phi_g}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}$and $\bar{\mathcal{H}}^{-1}\sum_g(\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})\big[\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big]_g\bar{\mathcal{H}}^{-1}$are matrices that satisfy the requirements for M in Lemma 4. The proof for the second term is shown, with the result for the first term following nearly identically. Firstly,

$$\begin{aligned}
&\sum_g(\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})\big[\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big]_g \\
&\qquad = \sum_{s,t}(\partial_{\phi\phi'\alpha_s}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{st}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) + \sum_{s,t}(\partial_{\phi\phi'\gamma_s}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\gamma\alpha}\big)_{st}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) \\
&\qquad + \sum_{s,t}(\partial_{\phi\phi'\alpha_s}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\gamma}\big)_{st}(\partial_{\beta\gamma_t}\bar{\mathcal{L}}) + \sum_{s,t}(\partial_{\phi\phi'\gamma_s}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\gamma\gamma}\big)_{st}(\partial_{\beta\gamma_t}\bar{\mathcal{L}})
\end{aligned}$$

Taking the first of these terms, the (i,j) element is given by

$$\Big[\sum_{s,t}(\partial_{\phi\phi'\alpha_s}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{st}(\partial_{\beta\alpha_t}\bar{\mathcal{L}})\Big]_{ij}$$
$$= \sum_t(\partial_{\phi_i\phi'_j\alpha_t}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{tt}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) + \frac{1}{N}\sum_t\sum_{s\neq t}(\partial_{\phi_i\phi'_j\alpha_s}\bar{\mathcal{L}})\big(N\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{st}(\partial_{\beta\alpha_t}\bar{\mathcal{L}})$$

Now if $\phi_i = \phi_j = \alpha_i$ then

$$\sum_t(\partial_{\phi_i\phi'_j\alpha_t}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{tt}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) = (\partial_{\alpha_i^3}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{ii}(\partial_{\beta\alpha_i}\bar{\mathcal{L}}) = O_p(1)$$

and if $\phi_i = \phi_j = \gamma_i$ then

$$\sum_t(\partial_{\phi_i\phi'_j\alpha_t}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{tt}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) = \frac{1}{N-1}\sum_{t\neq i}(\partial_{\alpha_t\gamma_i^2}\bar{\ell}_{ti})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{tt}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) = O_p(1)$$

Finally, if $i \neq j$ then we have either 0, or

$$\sum_t(\partial_{\phi_i\phi'_j\alpha_t}\bar{\mathcal{L}})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{tt}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) = \frac{1}{N-1}(\partial_{\pi\alpha_t\gamma_i}\bar{\ell}_{ti})\big(\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\big)_{tt}(\partial_{\beta\alpha_t}\bar{\mathcal{L}}) = O_p(N^{-1})$$

Identical results apply to the other elements in $\sum_g(\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})\big[\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big]_g$ and hence we can conclude that the matrix has $O_p(1)$ diagonal elements and $O_p(N^{-1})$ off-diagonal elements. It then follows that the same is true of $\bar{\mathcal{H}}^{-1}\sum_g(\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})\big[\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big]_g\bar{\mathcal{H}}^{-1}$. Then, we can apply Lemma 4 with $\mathbf{A} = \mathbf{B} = \mathcal{S}$ to give the result. □

Lemma 6. *Let Assumption 1 hold, and let $\widehat{\beta}_J$ be the either the jackknife, leave-l-out jackknife, or weighted jackknife estimator. Then, a first-order approximation to the estimator is given by*

$$\bar{W}_N N(\widehat{\beta}_J - \beta_0) = U^{(0)} + o_p(1)$$

where $U^{(0)} = (\partial_\beta\mathcal{L}) + (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S}$.

Proof. Recall from Lemma 1 that

$$N\bar{W}_N(\widehat{\beta} - \beta_0) = U^{(0)} + U^{(1)} + R_\beta + \tilde{R}_\beta$$

Since $\bar{W}_N$ is fixed across leave-out samples (Lemma 2), we can focus on the jackknife operator applied to the RHS. By Lemma 5 we have that $\mathbf{J}[U^{(0)} + U^{(1)}] = U^{(0)} + o_p(1)$, while in the Supplementary Appendix (S.4) it is shown that $\mathbf{J}[R_\beta] = o_p(1)$. Finally,

$$\begin{aligned}\mathbf{J}[\tilde{R}_\beta] &= (N-1)\tilde{R}_\beta - (N-2)\frac{1}{N-1}\sum_k \tilde{R}_{\beta,(k)} \\ &= o_p(1)\end{aligned}$$

since each remainder term in the above is $o_p(N^{-1})$. □

C.2 Approximation of W_N

The next two results show that the sample version of the Hessian for the common parameters β is consistent, and that it is approximately the same across leave-out samples.

Lemma 7. *Let Assumption 1 hold. Then, for* $\epsilon \geq \frac{1}{32}$

$$\begin{aligned}\|W_N - \bar{W}_N\| &= O_p(N^{-\frac{1}{2}+2\epsilon}) \\ \|W_{N,(k)} - \bar{W}_N\| &= O_p(N^{-\frac{1}{2}+2\epsilon})\end{aligned}$$

Let $\tilde{W}_N = W_N - \bar{W}_N$, then

$$\tilde{W}_N = \frac{1}{N}\partial_{\beta\beta}\tilde{\mathcal{L}} + \frac{1}{N}\big((\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L}) - (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big)$$

The first term is $\frac{1}{N}\partial_{\beta\beta}\tilde{\mathcal{L}} = O_p(N^{-1})$, since

$$\begin{aligned}\frac{1}{N^2}\bar{E}\big[(\partial_{\beta\beta}\tilde{\mathcal{L}})^2\big] &= \frac{1}{N^2(N-1)^2}\sum_{i,s}\sum_{j\neq i}\sum_{t\neq s}\bar{E}\big[(\partial_{\beta\beta'}\tilde{\ell}_{ij})(\partial_{\beta\beta'}\tilde{\ell}_{st})\big] \\ &= \frac{1}{N^2(N-1)^2}\sum_i\sum_{j\neq i}\big(\bar{E}\big[(\partial_{\beta\beta'}\tilde{\ell}_{ij})^2\big] + \bar{E}\big[(\partial_{\beta\beta'}\tilde{\ell}_{ij})(\partial_{\beta\beta'}\tilde{\ell}_{ji})\big]\big) \\ &= O_p(N^{-2})\end{aligned}$$

For the remaining term, we can decompose it as

$$\begin{aligned}
&\frac{1}{N}(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L}) - \frac{1}{N}(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}}) \\
&\qquad = \frac{1}{N}(\partial_{\beta\phi'}\tilde{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}}) + \frac{1}{N}(\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\tilde{\mathcal{L}}) \\
&\qquad + \frac{1}{N}(\partial_{\beta\phi'}\tilde{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\tilde{\mathcal{L}}) + \frac{1}{N}(\partial_{\beta\phi'}\mathcal{L})\big(\mathcal{H}^{-1} - \bar{\mathcal{H}}^{-1}\big)(\partial_{\beta\phi}\mathcal{L})
\end{aligned}$$

By Assumption B.1 of FVW we have

$$\begin{aligned}
\frac{1}{N}\|(\partial_{\beta\phi'}\tilde{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\| &\leq O_p(N^{-1/2}) \\
\frac{1}{N}\|(\partial_{\beta\phi'}\tilde{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\tilde{\mathcal{L}})\| &\leq O_p(N^{-1})
\end{aligned}$$

Also,

$$\begin{aligned}
\frac{1}{N}\|(\partial_{\beta\phi'}\mathcal{L})\big(\mathcal{H}^{-1} - \bar{\mathcal{H}}^{-1}\big)(\partial_{\beta\phi}\mathcal{L})\| &\leq \frac{1}{N}\|\partial_{\beta\phi'}\mathcal{L}\|^2\|\mathcal{H}^{-1} - \bar{\mathcal{H}}^{-1}\| \\
&= O_p(N^{-\frac{1}{2}+2\epsilon})
\end{aligned}$$

So we may write $\tilde{W}_N = O_p(N^{-\frac{1}{2}+2\epsilon})$. For the leave-out term note that the moment bounds in Assumption 1 (iii) imply identical bounds in the leave-out samples, simply by replacing $\partial_{\beta\pi}\ell_{ij}$ with $(\partial_{\beta\pi}\ell_{ij})1^k_{ij}\frac{N-1}{N-2}$ since we can condition on the node labels so that $1^k_{ij}\frac{N-1}{N-2}$ is simply an $O(1)$ constant. We can therefore apply the bounds in Assumption B.1 of FVW to the leave-out sample, as well as $\|\tilde{\mathcal{H}}_{(k)}\| = O_p(N^{-\frac{1}{2}+2\epsilon})$, and hence $\|\mathcal{H}^{-1}_{(k)} - \bar{\mathcal{H}}^{-1}\| = O_p(N^{-\frac{1}{2}+2\epsilon})$. Then, similar steps to the above proof also give $\|W_{N,(k)} - \bar{W}_N\| = O_p(N^{-\frac{1}{2}+2\epsilon})$.

Lemma 8. *Let Assumption 1 hold, then for all k*

$$\begin{aligned}
\|\widehat{W}_N - W_N\| &\to 0 \\
\|\widehat{W}_{N,(k)} - W_{N,(k)}\| &\to 0
\end{aligned}$$

Proof. We prove the first statement, since the proof of the second is identical. A first-order Taylor expansion of $\widehat{W}_N$ gives

$$\begin{aligned}
\widehat{W}_N = W_N(\widehat{\beta}, \widehat{\phi}) &= W_N(\beta_0, \phi_0) + \partial_\beta W_N(\bar{\beta}, \bar{\phi})(\widehat{\beta} - \beta_0) \\
&+ \partial_{\phi'} W_N(\bar{\beta}, \bar{\phi})(\widehat{\phi} - \phi_0)
\end{aligned}$$

where $\bar{\beta}$ and $\bar{\phi}$ are intermediate values between (β_0, ϕ_0) and $(\widehat{\beta}, \widehat{\phi})$. From Assumption 1 and the bounds in Assumption B.1 of FVW, we have that W_N is differentiable with derivatives that are $O_p(1)$, since

$$\begin{aligned}
\partial_\beta W_N &= \frac{1}{N}\partial_{\beta\beta\beta}\mathcal{L} + \frac{2}{N}(\partial_{\beta\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L}) \\
&- \frac{1}{N}(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L}) \\
&= O_p(1)
\end{aligned}$$

where $\frac{1}{N}(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L}) = O_p(1)$ follows simply by expansion of the matrix product and the properties of $\mathcal{H}^{-1}$ in (23). Similarly,

$$\begin{aligned}
(\partial_{\phi'} W_N)(\widehat{\phi} - \phi_0) &= \frac{1}{N}(\partial_{\beta\beta\phi'}\mathcal{L})(\widehat{\phi} - \phi_0) \\
&+ \frac{2}{N}(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi\phi'}\mathcal{L})(\widehat{\phi} - \phi_0) \\
&- \frac{1}{N}\sum_{g=1}^{2N}(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_g}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L})(\widehat{\phi}_g - \phi_g) \\
&\leq 2\|\widehat{\phi} - \phi\|_\infty \left| \frac{1}{N(N-1)}\sum_i \sum_{j\neq i}(\partial_{\beta\beta\pi}\ell_{ij}) \right| \\
&+ \|\widehat{\phi} - \phi\|_\infty \left| \frac{2}{N}\sum_{g=1}^{2N}(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi\phi_g}\mathcal{L}) \right| \\
&+ \|\widehat{\phi} - \phi\|_\infty \left| \frac{1}{N}\sum_{g=1}^{2N}(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_g}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta\phi}\mathcal{L}) \right| \\
&= O_p(1) \times \|\widehat{\phi} - \phi\|_\infty
\end{aligned}$$

Then, since $\|\widehat{\beta} - \beta_0\| \to 0$ and $\|\widehat{\phi} - \phi\|_\infty \to 0$ by (24), we get the result. □

C.3 Proof of Theorem 1

From Lemma 6 we have that

$$\begin{aligned}\bar{W}_N N(\widehat{\beta}_J - \beta_0) &= (\partial_\beta \mathcal{L}) + (\partial_{\beta\phi'}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}\mathcal{S} + o_p(1) \\ &= \frac{1}{N-1}\sum_i \sum_{j<i} \left(D_\beta \ell_{ij} + D_\beta \ell_{ji}\right)\end{aligned}$$

where $D_\beta \ell_{ij} = \partial_\beta \ell_{ij} - \partial_\pi \ell_{ij} \Xi_{ij}$ for

$$\begin{aligned}\Xi_{ij} &= -\frac{1}{N-1}\sum_s \sum_{t\neq s} \Gamma_{ijst}\bar{E}[\partial_{\beta\pi}\ell_{st}] \\ \Gamma_{ijst} &= (\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{is} + (\bar{\mathcal{H}}^{-1}_{\gamma\alpha})_{js} + (\bar{\mathcal{H}}^{-1}_{\alpha\gamma})_{it} + (\bar{\mathcal{H}}^{-1}_{\gamma\gamma})_{jt}\end{aligned}$$

The result then follows from a standard CLT argument, noting that $\left(D_\beta \ell_{ij} + D_\beta \ell_{ji}\right)$ are independent over (i, j). For the weighted jackknife, we use the fact that Lemmas 7 and 8, and the triangle inequality give $\|\bar{W}_J - \bar{W}_N\| = o_p(1)$ and hence

$$\|\bar{W}_J^{-1}\widehat{W}_{(k)} - I\| \leq \|\bar{W}_J^{-1}\|\|\widehat{W}_{N,(k)} - \bar{W}_J\| = o_p(1)$$

so that

$$\begin{aligned}\frac{1}{N-1}\sum_k \bar{W}_J^{-1}\widehat{W}_{(k)}\widehat{\beta}_{(k)} &= \frac{1}{N-1}\sum_k \widehat{\beta}_{(k)} + \frac{1}{N-1}\sum_k \left(\bar{W}_J^{-1}\widehat{W}_{(k)} - I\right)\widehat{\beta}_{(k)} \\ &= \frac{1}{N-1}\sum_k \widehat{\beta}_{(k)} + o_p(1)\end{aligned}$$

and hence the weighted jackknife is equal to the standard jackknife to first-order.

To show consistency of the plug-in estimator $\widehat{\Omega}_N$, we note that by Assumption 1 (iii), $D_\beta \ell_{ij}$ has a first-order Taylor approximation that is a continuously differentiable function of the parameters. Then by the continuous mapping theorem and the consistency results $\|\widehat{\beta} - \beta_0\| \to 0$ and $\|\widehat{\phi} - \phi\|_\infty \to 0$ in (24), $\widehat{\Omega}_N \to \Omega$ as required (see for example Lemma S.1 in FVW).

D Jackknife results for average effects

We begin by stating a first-order asymptotic expansion for the average effect estimator that will be used in the proof of Theorem 2. The proof of this result is provided in the Supplementary Appendix.

Lemma 9. *Let Assumptions 1 and 2 hold. Then*

$$\begin{aligned}
N(\widehat{\Delta}_N - \Delta_N) &= (\partial_\beta \Delta_N) N(\widehat{\beta} - \beta) + N(\partial_{\phi'} \Delta_N)(\widehat{\phi} - \phi) \\
&+ \frac{1}{2} N(\widehat{\phi} - \phi)'(\partial_{\phi\phi'} \Delta_N)(\widehat{\phi} - \phi) + R^1_\Delta + \tilde{R}^1_\Delta \\
&= \Big[(\partial_\beta \bar{\Delta}_N) - (\partial_{\phi'} \bar{\Delta}_N)\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi} \bar{\mathcal{L}})\Big] \bar{W}_N^{-1}\big(U^{(0)} + U^{(1)}\big) \\
&+ N(\partial_{\phi'} \bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\mathcal{S} \\
&+ N(\partial_{\phi'} \tilde{\Delta}_N)\bar{\mathcal{H}}^{-1}\mathcal{S} - N(\partial_{\phi'} \bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} \\
&+ \frac{1}{2} N \mathcal{S}'\bar{\mathcal{H}}^{-1}\Big((\partial_{\phi\phi'} \bar{\Delta}_N) + \sum_g (\partial_{\phi\phi'\phi_g} \bar{\mathcal{L}})[(\partial_{\phi'} \bar{\Delta}_N)\bar{\mathcal{H}}^{-1}]_g\Big)\bar{\mathcal{H}}^{-1}\mathcal{S} \\
&+ R_\Delta + \tilde{R}_\Delta
\end{aligned}$$

where $\|R^1_\Delta\| = o_p(1)$, $\|\tilde{R}^1_\Delta\| = o_p(N^{-1})$, $\|R_\Delta\| = o_p(1)$, *and* $\|\tilde{R}_\Delta\| = o_p(N^{-1})$.

In order to establish an equivalent asymptotic expansion for the leave-out estimators $\widehat{\Delta}_{(k)}$ we need to determine the value of expectations in the leave-out samples. The next lemma does this, and is analogous to Lemma 2, which states the same result for averages over single observations.

Lemma 10. *Let* 1^k_{ij} *satisfy Condition 1, and define* $1^k_\lambda = \prod_{(i,j)\in\lambda} 1^k_{ij}$ *for* λ *a set of* r *observations* (i, j). *Then, for sums*

$$\begin{aligned}
A &= \frac{1}{|\Lambda_N|} \sum_\lambda A_\lambda \\
A_{(k)} &= \frac{N-1}{N-r-1} \frac{1}{|\Lambda_N|} \sum_\lambda A_\lambda 1^k_\lambda
\end{aligned}$$

we have

$$\bar{E}[A_{(k)}] = \bar{E}[A]\big(1 + O(N^{-2})\big)$$

Proof. We will prove that $E[1_\lambda^k] = \frac{N-r-1}{N-1} + O(N^{-2})$ from which the statement in the lemma follows since

$$\begin{aligned}
\bar{E}[A_{(k)}] &= \frac{N-1}{N-r-1}\frac{1}{|\Lambda_N|}\sum_\lambda \bar{E}[A_\lambda]E[1_\lambda^k] \\
&= \frac{1}{|\Lambda_N|}\sum_\lambda \bar{E}[A_\lambda] + \frac{N-1}{N-r-1}\frac{1}{|\Lambda_N|}\sum_\lambda \bar{E}[A_\lambda] \times O(N^{-2}) \\
&= \bar{E}[A]\big(1 + O(N^{-2})\big)
\end{aligned}$$

Let I_λ be equal to the number of leave-out sets $\mathcal{I}_k$ spanned by the observations in λ. Since any observation is equally likely to appear in any leave-out set,

$$E[1_\lambda^k] = \sum_{s=1}^{r} E[1_\lambda^k | I_\lambda = s] P(I_\lambda = s)$$

We begin by arguing that $P(I_\lambda = r) = 1 - O(N^{-1})$, that is, the sets λ do not contain multiple observations from within the same leave-out set with probability approaching one. It is sufficient to consider sets λ that contain $p = 2r$ unique agents, since this case represents the most likely scenario for λ containing observations in the same leave-out set as within a leave-out set no two senders can be the same, and no two receivers may be the same. There are $\frac{N!}{(N-2r)!}$ possible choices for the ordered set of agents in λ. Then, there are *at most* $N(N-1)(N-2)(N-4)\cdots(N-2r+2)(N-3r+2)$ orderings of agents in λ which span r different leave-out sets ($N(N-1)$ choices for the first observation (i,j), then $(N-2)(N-4)$ choices for the second observation (s,t) since (s,t) cannot belong in the same leave-out set as (i,j), and so on). This gives

$$P(I_\lambda = r) \geq \frac{N(N-1)(N-2)(N-4)\cdots(N-2r+2)(N-3r+2)}{N(N-1)\cdots(N-2r+1)}$$

which is the product of $2r$ ratios each of which is equal to $1 - O(N^{-1})$, which implies $P(I_\lambda^r = 1) = 1 - O(N^{-1})$.

Next, note that whenever $I_\lambda = s$, we have $1_\lambda^k = 1$ only if $\mathcal{I}_k$ is not one of the s leave-out sets spanned by λ. This happens with probability $\binom{N-2}{s}/\binom{N-1}{s} = \frac{N-s-1}{N-1}$.

$$\begin{aligned} E[1_\lambda^k] &= \sum_{s=1}^{r} \frac{N-s-1}{N-1} P(I_\lambda = s) \\ &= \frac{N-r-1}{N-1} \\ &+ \sum_{s=1}^{r-1} \frac{r-s}{N-1} P(I_\lambda = s) \end{aligned}$$

and so

$$\frac{1}{N-1}\big(1 - P(I_\lambda = r)\big) \leq E[1_\lambda^k] - \frac{N-r-1}{N-1} \leq \frac{r-1}{N-1}\big(1 - P(I_\lambda = r)\big)$$

Since we have $P(I_\lambda = r) = 1 - O(N^{-1})$, we can conclude $E[1_\lambda^k] = \frac{N-r-1}{N-1} + O(N^{-2})$. □

Lemma 11. *Let λ be a set of r observations (i, j) involving p unique agents, and Λ_N be the collection of all such λ formed by permuting the agents in λ. Then, under Assumption 2,*

(i) $\partial_\phi \bar{\Delta}_N$ has $O_p(N^{-1})$ elements

(ii) $\partial_{\alpha\alpha'} \bar{\Delta}_N$, $\partial_{\alpha\gamma'} \bar{\Delta}_N$, $\partial_{\gamma\alpha'} \bar{\Delta}_N$, and $\partial_{\gamma\gamma'} \bar{\Delta}_N$ each have $O_p(N^{-1})$ diagonal elements and $O_p(N^{-2})$ off-diagonal elements

Proof. Let λ_α denote the set of p_α sender agents in the observations within λ, and λ_γ the set of p_γ receiving agents. There are $|\Lambda_N| = \frac{N!}{(N-p)!}$ ways of selecting the p agents in λ. Among these permutations, agent i is a sender $p_\alpha \frac{(N-1)!}{(N-p)!}$ times, while node j is the receiver $p_\gamma \frac{(N-1)!}{(N-p)!}$ times. Using this, the first derivatives of $\bar{\Delta}_N$ with respect to the fixed effects are

$$\begin{aligned} \partial_{\alpha_i} \bar{\Delta}_N &= \frac{1}{|\Lambda_N|} \sum_{\lambda: i \in \lambda_\alpha} \partial_{\alpha_i} \bar{m}_\lambda = O_p(N^{-1}) \\ \partial_{\gamma_i} \bar{\Delta}_N &= \frac{1}{|\Lambda_N|} \sum_{\lambda: i \in \lambda_\gamma} \partial_{\gamma_i} \bar{m}_\lambda = O_p(N^{-1}) \end{aligned}$$

where the $O_p(N^{-1})$ statements come from the fact that $p_\alpha \frac{(N-1)!}{(N-p)!} / \frac{N!}{(N-p)!} = p_\alpha / N$. An identical result applies to the diagonal elements of $\partial_{\phi\phi'} \bar{\Delta}_N$, i.e. $\partial_{\alpha_i\alpha_i} \bar{\Delta}_N = O_p(N^{-1})$ and $\partial_{\gamma_j\gamma_j} \bar{\Delta}_N = O_p(N^{-1})$ since they are sums over the same sets of λ. Also, if the presence of i as a sender agent implies that i is also a receiver in λ (e.g. the cyclic triangle $\{(i, j), (j, k), (k, i)\}$) then it will be the case that $\partial_{\alpha_i\gamma_i} \bar{\Delta}_N = O_p(N^{-1})$ also (if this is not true it will be lower order).

Next, consider the off-diagonal components of $\partial_{\phi\phi'}\bar{\Delta}_N$. If $p_\alpha = 1$ then $\partial_{\alpha_i\alpha_j}\bar{\Delta}_N = 0$, otherwise, there are $\binom{p_\alpha}{2}\frac{(N-2)!}{(N-p)!}$ permutations that contain both i and j as senders. Similarly, for $p_\gamma \geq 2$, there are $\binom{p_\gamma}{2}\frac{(N-2)!}{(N-p)!}$ permutations that contain both i and j as receivers. Finally, there are *at most* $p_\alpha p_\gamma \frac{(N-2)!}{(N-p)!}$ permutations in which i is a sender and j a receiver (this is an upper bound since with i in a particular sender position, not all receiver positions may be valid for j). This, along with Assumption 2, gives the results

$$\begin{aligned}
\partial_{\alpha_i\alpha_j}\bar{\Delta}_N &= O_p(N^{-2}) \\
\partial_{\alpha_i\gamma_j}\bar{\Delta}_N &= O_p(N^{-2}) \\
\partial_{\gamma_i\gamma_j}\bar{\Delta}_N &= O_p(N^{-2})
\end{aligned}$$

which demonstrates the lemma. □

Lemma 12. *Let 1^k_{ij} satisfy Condition 1 and let $1^k_\lambda = \prod_{(i,j)\in\lambda} 1^k_{ij}$. Let A_{ij} be a mean-zero random variable with bounded fourth moment, and define*

$$\begin{aligned}
\mathbf{A} &= \frac{1}{N-1}\big(\{\sum_{s\neq i} A_{is}\}_{i=1,\dots,N}, \{\sum_{s\neq j} A_{sj}\}_{j=1,\dots,N}\big) \\
&= (\mathbf{A}_\alpha, \mathbf{A}_\gamma) \\
\mathbf{A}_k &= \frac{1}{N-2}\big(\{\sum_{s\neq i} A_{is}1^k_{is}\}_{i=1,\dots,N}, \{\sum_{s\neq j} A_{sj}1^k_{sj}\}_{j=1,\dots,N}\big) \\
&= (\mathbf{A}_{\alpha,k}, \mathbf{A}_{\gamma,k})
\end{aligned}$$

and let $\mathbf{B}$ and $\mathbf{B}_k$ be defined as

$$\begin{aligned}
\mathbf{B} &= \frac{N}{|\Lambda_N|}\big(\{\sum_{s\neq i}\sum_{\lambda\in\Lambda_{is}} B_\lambda\}_{i=1,\dots,N}, \{\sum_{s\neq j}\sum_{\lambda\in\Lambda_{sj}} B_\lambda\}_{i=1,\dots,N}\big) \\
&= (\mathbf{B}_\alpha, \mathbf{B}_\gamma) \\
\mathbf{B}_k &= \frac{N-1}{N-r-1}\frac{N}{|\Lambda_N|}\big(\{\sum_{s\neq i}\sum_{\lambda\in\Lambda_{is}} B_\lambda 1^k_\lambda\}_{i=1,\dots,N}, \{\sum_{s\neq j}\sum_{\lambda\in\Lambda_{sj}} B_\lambda 1^k_\lambda\}_{i=1,\dots,N}\big) \\
&= (\mathbf{B}_{\alpha,k}, \mathbf{B}_{\gamma,k})
\end{aligned}$$

for mean zero B_λ with bounded fourth moment. Assume that A_{ij} is independent of A_{st} for $(i,j) \notin \{(s,t),(t,s)\}$, and independent of B_λ whenever λ does not contain either (i,j) or

(j, i). *Define the jackknifed term*

$$\mathcal{J}_0 = (N-1)\mathbf{A}'M\mathbf{B} - \frac{N-2}{N-1}\sum_k \mathbf{A}'_{(k)}M\mathbf{B}_{(k)}$$

where M is a non-random matrix that has $O_p(1)$ elements on its diagonal and $O_p(N^{-1})$ off-diagonal terms. Then we have:

(i) $\bar{E}\big[\mathcal{J}_0\big] = o_p(1)$,

(ii) $\mathcal{J}_0 = o_p(1)$.

Proof. The most common choice of M will be $\bar{\mathcal{H}}^{-1}$, which satisfies the conditions for M by Assumption 1 and (23). We show the proof using $\bar{\mathcal{H}}^{-1}$, but note that it holds for any M satisfying the conditions stated above. We have

$$\begin{aligned}\mathbf{A}'\bar{\mathcal{H}}^{-1}\mathbf{B} = \mathbf{A}'_\alpha\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\mathbf{B}_\alpha + \mathbf{A}'_\alpha\bar{\mathcal{H}}^{-1}_{\alpha\gamma}\mathbf{B}_\gamma \\ + \mathbf{A}'_\gamma\bar{\mathcal{H}}^{-1}_{\gamma\alpha}\mathbf{B}_\alpha + \mathbf{A}'_\gamma\bar{\mathcal{H}}^{-1}_{\gamma\gamma}\mathbf{B}_\gamma\end{aligned}$$

Let $\Lambda_{is} = \{\lambda : (i, s) \in \lambda\}$ be the set of λ containing observation (i, s). The full sample and leave-out versions of the first term are

$$\begin{aligned}\mathbf{A}'_\alpha\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\mathbf{B}_\alpha &= \sum_{i,j}\mathbf{A}_{\alpha,i}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}\mathbf{B}_{\alpha,j} \\ &= \frac{N}{(N-1)|\Lambda_N|}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in\Lambda_{jt}}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_\lambda \\ \frac{1}{N-1}\sum_k \mathbf{A}'_{k,\alpha}\bar{\mathcal{H}}^{-1}_{\alpha\alpha}\mathbf{B}_{k,\alpha} & \\ &= \frac{N}{(N-r-1)(N-2)|\Lambda_N|}\sum_k\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in\Lambda_{jt}}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_\lambda 1^k_{is}1^k_\lambda \\ &= \frac{N}{(N-r-1)(N-2)|\Lambda_N|}\sum_k\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in(\Lambda_{jt}\cap\Lambda_{is})}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_\lambda 1^k_\lambda \\ &+ \frac{N}{(N-r-1)(N-2)|\Lambda_N|}\sum_k\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in(\Lambda_{jt}\setminus\Lambda_{is})}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_\lambda 1^k_{is}1^k_\lambda\end{aligned}$$

Let I_λ be equal to the number of leave-out sets $\mathcal{I}_k$ spanned by the r observations in λ so that $\sum_k 1^k_\lambda = N - I_\lambda - 1$. As shown in the proof of Lemma 10, $|\{\lambda : I_\lambda < r\}|/|\Lambda_N| \to 0$, that is, the fraction of sets λ that contain two or more observations in the same $\mathcal{I}_k$ is a vanishingly small. Using this, we have

$$\begin{aligned}
\mathcal{J}_{\alpha\alpha} =&(N-1)\mathbf{A}'_{\alpha}\bar{\mathcal{H}}^{-1}_{\alpha\alpha'}\mathbf{B}_{\alpha} - \frac{N-2}{N-1}\sum_k \mathbf{A}'_{\alpha,k}\bar{\mathcal{H}}^{-1}_{\alpha\alpha'}\mathbf{B}_{\alpha,k} \\
=&\frac{N}{|\Lambda_N|}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in(\Lambda_{jt}\cap\Lambda_{is})}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_{\lambda}\Big(1-\frac{\sum_k 1^k_{\lambda}}{N-r-1}\Big) \\
&+\frac{N}{|\Lambda_N|}\sum_k\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in(\Lambda_{jt}\setminus\Lambda_{is})}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_{\lambda}\Big(1-\frac{\sum_k 1^k_{is}1^k_{\lambda}}{N-r-1}\Big) \\
=&\frac{N}{|\Lambda_N|}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\substack{\lambda\in(\Lambda_{jt}\cap\Lambda_{is})\\ I_\lambda<r}}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_{\lambda}\Big(1-\frac{N-I_\lambda-1}{N-r-1}\Big) \\
&+\frac{N}{|\Lambda_N|}\sum_k\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in(\Lambda_{jt}\setminus\Lambda_{is})}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{ij}A_{is}B_{\lambda}\Big(1-\frac{N-I_\lambda-2}{N-r-1}\Big)
\end{aligned}$$

Letting $\Gamma_{ijst} = (\bar{\mathcal{H}}^{-1}_{\alpha\alpha'})_{ij} + (\bar{\mathcal{H}}^{-1}_{\alpha\gamma'})_{it} + (\bar{\mathcal{H}}^{-1}_{\gamma\alpha'})_{sj} + (\bar{\mathcal{H}}^{-1}_{\gamma\gamma'})_{st}$, similar computations for the other three elements gives

$$\begin{aligned}
\mathcal{J}_0 =&(N-1)\mathbf{A}'\bar{\mathcal{H}}^{-1}\mathbf{B} - \frac{N-2}{N-1}\sum_k \mathbf{A}'_k\bar{\mathcal{H}}^{-1}\mathbf{B}_k \\
=&\frac{N}{|\Lambda_N|}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\substack{\lambda\in(\Lambda_{jt}\cap\Lambda_{is})\\ I_\lambda<r}}\Gamma_{ijst}A_{is}B_{\lambda}\Big(1-\frac{N-I_\lambda-1}{N-r-1}\Big) \\
&\frac{N}{|\Lambda_N|}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in(\Lambda_{jt}\cap\Lambda_{si}\setminus\Lambda_{is})}\Gamma_{ijst}A_{is}B_{\lambda}\Big(1-\frac{N-I_\lambda-2}{N-r-1}\Big) \\
&+\frac{N}{|\Lambda_N|}\sum_{i,j}\sum_{s\neq i}\sum_{t\neq j}\sum_{\lambda\in(\Lambda_{jt}\setminus(\Lambda_{is}\cup\Lambda_{si}))}\Gamma_{ijst}A_{is}B_{\lambda}\Big(1-\frac{N-I_\lambda-2}{N-r-1}\Big) \\
&= \mathcal{J}_{0,1}+\mathcal{J}_{0,2}+\mathcal{J}_{0,3}
\end{aligned}$$

Next, recall that $\Gamma_{ijst} = O_p(1)$ whenever $i=j$ or $s=t$, and is $O_p(N^{-1})$ otherwise. Note that $\bar{E}[\mathcal{J}_{0,3}] = 0$ since λ does not contain (i,s) or (s,i). Taking expectations we get

$$
\begin{aligned}
\bar{E}[\mathcal{J}_0] =& \frac{N}{|\Lambda_N|}\sum_i\sum_{s\neq i}\sum_{\lambda\in\Lambda_{is}:I_\lambda<r}\Gamma_{iiss}\bar{E}[A_{is}B_\lambda]\Big(\frac{I_\lambda - r}{N-r-1}\Big)\\
&+\frac{N}{|\Lambda_N|}\sum_i\sum_{s\neq i}\sum_{t\neq\{i,s\}}\sum_{\lambda\in(\Lambda_{it}\cap\Lambda_{is}):I_\lambda<r}\Gamma_{iist}\bar{E}[A_{is}B_\lambda]\Big(\frac{I_\lambda - r}{N-r-1}\Big)\\
&+\frac{N}{|\Lambda_N|}\sum_i\sum_{j\neq i}\sum_{s\neq\{i,j\}}\sum_{\lambda\in(\Lambda_{js}\cap\Lambda_{is}):I_\lambda<r}\Gamma_{ijss}\bar{E}[A_{is}B_\lambda]\Big(\frac{I_\lambda - r}{N-r-1}\Big)\\
&+\frac{N}{|\Lambda_N|}\sum_i\sum_{j\neq i}\sum_{s\neq i}\sum_{t\neq\{s,j\}}\sum_{\lambda\in(\Lambda_{jt}\cap\Lambda_{is}):I_\lambda<r}\Gamma_{ijst}\bar{E}[A_{is}B_\lambda]\Big(\frac{I_\lambda - r}{N-r-1}\Big)\\
&+\frac{N}{|\Lambda_N|}\sum_i\sum_{s\neq i}\sum_{t\neq\{i,s\}}\sum_{\lambda\in((\Lambda_{it}\cap\Lambda_{si})\setminus\Lambda_{is})}\Gamma_{iist}\bar{E}[A_{is}B_\lambda]\Big(\frac{I_\lambda - r+1}{N-r-1}\Big)\\
&+\frac{N}{|\Lambda_N|}\sum_i\sum_{j\neq i}\sum_{s\neq\{i,j\}}\sum_{\lambda\in((\Lambda_{js}\cap\Lambda_{si})\setminus\Lambda_{is})}\Gamma_{ijss}\bar{E}[A_{is}B_\lambda]\Big(\frac{I_\lambda - r+1}{N-r-1}\Big)\\
&+\frac{N}{|\Lambda_N|}\sum_i\sum_{j\neq i}\sum_{s\neq i}\sum_{t\neq\{j,s\}}\sum_{\lambda\in((\Lambda_{jt}\cap\Lambda_{si})\setminus\Lambda_{is})}\Gamma_{ijst}\bar{E}[A_{is}B_\lambda]\Big(\frac{I_\lambda - r+1}{N-r-1}\Big)\\
=& o_p(1)
\end{aligned}
$$

since, $\frac{I_\lambda - r}{N-r-1} = O(N^{-1})$, $\frac{I_\lambda - r+1}{N-r-1} = O(N^{-1})$, $\frac{N}{|\Lambda_N|}|\lambda\in\Lambda_{is}:I_\lambda<r| = O(N^{-2})$, $\frac{N}{|\Lambda_N|}|\lambda\in(\Lambda_{it}\cap\Lambda_{is}):I_\lambda<r| = O(N^{-3})$, and so on applying the results on the size of sets Λ_{ij}, $I_\lambda<r$, and Γ_{ijst}.

Then,

$$\bar{E}[\mathcal{J}_{0,3}^2] = \frac{N^2}{|\Lambda_N|^2} \sum_i \sum_j \sum_k \sum_l \sum_{s\neq i} \sum_{t\neq j} \sum_{p\neq k} \sum_{q\neq l} \sum_{\lambda\in(\Lambda_{jt}\backslash(\Lambda_{is}\cup\Lambda_{si}))} \sum_{\lambda\in(\Lambda_{ql}\backslash(\Lambda_{pk}\cup\Lambda_{kp}))}$$

$$\Gamma_{ijst}\Gamma_{klpq}\bar{E}[A_{is}B_\lambda A_{pk}B_{\lambda'}]\Big(\frac{I_\lambda - r+1}{N-r-1}\Big)\Big(\frac{I_{\lambda'} - r+1}{N-r-1}\Big)$$

$$= \frac{N^2}{|\Lambda_N|^2} \sum_i \sum_j \sum_l \sum_{s\neq i} \sum_{t\neq j} \sum_{q\neq l} \sum_{\lambda\in(\Lambda_{jt}\backslash(\Lambda_{is}\cup\Lambda_{si}))} \sum_{\lambda'\in((\Lambda_{ql}\cap(\Lambda_{jt}\cup\Lambda_{tj}))\backslash(\Lambda_{is}\cup\Lambda_{si}))}$$

$$\Gamma_{ijst}\Gamma_{ilsq}\bar{E}[A_{is}(A_{is}+A_{si})]\bar{E}[B_\lambda B_{\lambda'}]\Big(\frac{I_\lambda - r+1}{N-r-1}\Big)\Big(\frac{I_{\lambda'} - r+1}{N-r-1}\Big)$$

$$+ \frac{N^2}{|\Lambda_N|^2} \sum_i \sum_j \sum_k \sum_l \sum_{s\neq i} \sum_{t\neq j} \sum_{p\neq k} \sum_{q\neq l} \sum_{\lambda\in(\Lambda_{jt}\cap(\Lambda_{pk}\cup\Lambda_{kp})\backslash(\Lambda_{is}\cup\Lambda_{si}))} \sum_{\lambda\in(\Lambda_{ql}(\Lambda_{is}\cup\Lambda_{si})\backslash(\Lambda_{pk}\cup\Lambda_{kp}))}$$

$$\Gamma_{ijst}\Gamma_{klpq}\bar{E}[A_{is}B_{\lambda'}]\bar{E}[A_{pk}B_\lambda]\Big(\frac{I_\lambda - r+1}{N-r-1}\Big)\Big(\frac{I_{\lambda'} - r+1}{N-r-1}\Big)$$

Note that $\frac{N}{|\Lambda_N|}|\lambda \in (\Lambda_{jt}\backslash(\Lambda_{is}\cup\Lambda_{si}))| = O(N^{-1})$, while $\frac{N}{|\Lambda_N|}|\lambda' \in ((\Lambda_{ql}\cap(\Lambda_{jt}\cup\Lambda_{tj}))\backslash(\Lambda_{is}\cup\Lambda_{si}))|$ is $O(N^{-1})$ if (q,l) equals (t,j) or (j,t), $O(N^{-2})$ if either $q\in\{t,j\}$ or $l\in\{t,j\}$ and $O(N^{-3})$ otherwise. Also, $\Big(\frac{I_\lambda-r+1}{N-r-1}\Big)\Big(\frac{I_{\lambda'}-r+1}{N-r-1}\Big) = O(N^{-2})$. Combining these facts with, $\Gamma_{ijst} = O_p(1)$ whenever $i=j$ or $s=t$, and $O_p(N^{-1})$ otherwise gives $\bar{E}[\mathcal{J}_{0,3}^2] = o_p(1)$.

An almost identical analysis applies to $\mathcal{J}_{0,1}$ and $\mathcal{J}_{0,2}$, giving the result $\mathcal{J}_0 = o_p(1)$. □

The next lemma states the first-order approximation for the jackknife bias-corrected average effect estimator.

Lemma 13. *Let Assumptions 1 and 2 hold and let $\widehat{\Delta}_J$ be the jackknife bias-corrected estimator in (12). Then,*

$$N(\widehat{\Delta}_J - \Delta_N) = \Big[(\partial_\beta\bar{\Delta}_N) - N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\Big]\bar{W}_N^{-1}U^{(0)} + N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\mathcal{S} + o_p(1)$$

Proof. We can write an expansion for the leave-out estimate

$$\begin{aligned}N(\widehat{\Delta}_{(k)} - \Delta_N) = &\Big[(\partial_\beta \bar{\Delta}_N) - (\partial_{\phi'} \bar{\Delta}_N)\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\Big]\bar{W}_N^{-1}\big(U_{(k)}^{(0)} + U_{(k)}^{(1)}\big) \\ &+ N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\mathcal{S}_{(k)} \\ &+ N(\partial_{\phi'}\tilde{\Delta}_{(k)})\bar{\mathcal{H}}^{-1}\mathcal{S}_{(k)} - N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}_{(k)}\bar{\mathcal{H}}^{-1}\mathcal{S}_{(k)} \\ &+ \frac{1}{2}N\mathcal{S}_{(k)}'\bar{\mathcal{H}}^{-1}\Big((\partial_{\phi\phi'}\bar{\Delta}_N) + \sum_g (\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})[(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}]_g\Big)\bar{\mathcal{H}}^{-1}\mathcal{S}_{(k)} \\ &+ R_{(k),\Delta} + \tilde{R}_{(k),\Delta}\end{aligned}$$

where $R_{(k),\Delta} = o_p(1)$ is the version of R_Δ in the leave-out sample and $\tilde{R}_{(k),\Delta} = o_p(N^{-1})$ is the leave-out version of $\tilde{R}_\Delta$ combined with the error from replacing terms like $\partial_\beta \bar{\Delta}_{(k)}$ with $\partial_\beta \bar{\Delta}_N$ (i.e. applying the result in Lemma 10). Using the expansion for the leave-out estimate, we can apply the jackknife operator to each line above.

For the first term, we apply the result in Lemma 5 to give

$$\begin{aligned}&\mathbf{J}\Big[\big((\partial_\beta \bar{\Delta}_N) - N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big)\bar{W}_N^{-1}\big(U^{(0)} + U^{(1)}\big)\Big] \\ &\quad = \big((\partial_\beta \bar{\Delta}_N) - N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\big)\bar{W}_N^{-1}U^{(0)} + o_p(1)\end{aligned}$$

Similarly, Lemma 3 implies that jackknifing the second term gives $N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\mathcal{S}$.

For the third term, we note that by Lemma 11 we can apply Lemma 4 with $M = \bar{\mathcal{H}}^{-1}$, $A = \mathcal{S}$, and $B = N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}$, and apply Lemma 12 with $M = \bar{\mathcal{H}}^{-1}$, $A = \mathcal{S}$, and either $B = N(\partial_{\phi'}\tilde{\Delta}_N)$.

For the fourth term, we first show that

$$N\bar{\mathcal{H}}^{-1}\Big((\partial_{\phi\phi'}\bar{\Delta}_N) + \sum_g (\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})[(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}]_g\Big)\bar{\mathcal{H}}^{-1}$$

satisfies the conditions for M in Lemma 4, from which we will be able to conclude that the jackknifed term will be $o_p(1)$. The above expression is non-random (conditional on exogenous regressors and fixed effects) and so we must demonstrate that it is a $2N \times 2N$ matrix with $O_p(1)$ diagonal elements, and $O_p(N^{-1})$ off-diagonal elements. Note that if two $2N \times 2N$ matrices both have $O_p(1)$ diagonal elements and $O_p(N^{-1})$ off-diagonal elements, then their product also shares this property. Since this is true of $\bar{\mathcal{H}}^{-1}$(see Lemma D.1 in Fernández-Val and Weidner (2016)), it remains to demonstrate this fact for the terms $N(\partial_{\phi\phi'}\bar{\Delta}_N)$ and

$N\sum_g (\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})[(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}]_g$.

By Lemma 11 we have that $N(\partial_{\phi\phi'}\bar{\Delta}_N)$ has $O_p(1)$ diagonal elements and $O_p(N^{-1})$ off-diagonal, with the possible exception of the elements $\partial_{\alpha_i\gamma_i'}\bar{\Delta}_N$. However, this still implies that $N\bar{\mathcal{H}}^{-1}(\partial_{\phi\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}$ satisfies the condition. For $N\sum_g (\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})[(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}]_g$, diagonal elements are given by (for $i \leq N$)

$$
\begin{aligned}
&\Big[N\sum_g (\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})[(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}]_g\Big]_{ii} \\
&= N(\partial_{\alpha_i\alpha_i'\phi}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\phi}\bar{\Delta}_N) \\
&= N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{st}(\partial_{\alpha_i\alpha_i\alpha_s}\bar{\mathcal{L}})(\partial_{\alpha_t}\bar{\Delta}_N) + N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\alpha\gamma})_{st}(\partial_{\alpha_i\alpha_i\alpha_s}\bar{\mathcal{L}})(\partial_{\gamma_t}\bar{\Delta}_N) \\
&+ N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\gamma\alpha})_{st}(\partial_{\alpha_i\alpha_i\gamma_s}\bar{\mathcal{L}})(\partial_{\alpha_t}\bar{\Delta}_N) + N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\gamma\gamma})_{st}(\partial_{\alpha_i\alpha_i\gamma_s}\bar{\mathcal{L}})(\partial_{\gamma_t}\bar{\Delta}_N) \\
&= \frac{N}{N-1}\sum_t\sum_{j\neq i}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{it}(\partial_{\pi^3}\bar{\ell}_{ij})(\partial_{\alpha_t}\bar{\Delta}_N) + \frac{N}{N-1}\sum_t\sum_{j\neq i}(\bar{\mathcal{H}}^{-1}_{\alpha\gamma})_{st}(\partial_{\pi^3}\bar{\ell}_{ij})(\partial_{\gamma_t}\bar{\Delta}_N) \\
&+ \frac{N}{N-1}\sum_t\sum_{s\neq i}(\bar{\mathcal{H}}^{-1}_{\gamma\alpha})_{st}(\partial_{\pi^3}\bar{\ell}_{is})(\partial_{\alpha_t}\bar{\Delta}_N) + \frac{N}{N-1}\sum_t\sum_{s\neq i}(\bar{\mathcal{H}}^{-1}_{\gamma\gamma})_{st}(\partial_{\pi^3}\bar{\ell}_{is})(\partial_{\gamma_t}\bar{\Delta}_N)
\end{aligned}
$$

which is $O_p(1)$ since by Lemma 11 and Lemma D.1 in Fernández-Val and Weidner (2016), and similarly for $i > N$. Off-diagonal components can be shown similarly, e.g. for $i < N$ and $j > N$ we have

$$
\begin{aligned}
&\Big[N\sum_g (\partial_{\phi\phi'\phi_g}\bar{\mathcal{L}})[(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}]_g\Big]_{ij} \\
&= N(\partial_{\alpha_i\gamma_j'\phi}\bar{\mathcal{L}})\bar{\mathcal{H}}^{-1}(\partial_{\phi}\bar{\Delta}_N) \\
&= N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{st}(\partial_{\alpha_i\gamma_j\alpha_s}\bar{\mathcal{L}})(\partial_{\alpha_t}\bar{\Delta}_N) + N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\alpha\gamma})_{st}(\partial_{\alpha_i\gamma_j\alpha_s}\bar{\mathcal{L}})(\partial_{\gamma_t}\bar{\Delta}_N) \\
&+ N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\gamma\alpha})_{st}(\partial_{\alpha_i\gamma_j\gamma_s}\bar{\mathcal{L}})(\partial_{\alpha_t}\bar{\Delta}_N) + N\sum_{s,t}(\bar{\mathcal{H}}^{-1}_{\gamma\gamma})_{st}(\partial_{\alpha_i\gamma_j\gamma_s}\bar{\mathcal{L}})(\partial_{\gamma_t}\bar{\Delta}_N) \\
&= \frac{N}{N-1}\sum_t(\bar{\mathcal{H}}^{-1}_{\alpha\alpha})_{it}(\partial_{\pi^3}\bar{\ell}_{ij})(\partial_{\alpha_t}\bar{\Delta}_N) + \frac{N}{N-1}\sum_t(\bar{\mathcal{H}}^{-1}_{\alpha\gamma})_{st}(\partial_{\pi^3}\bar{\ell}_{ij})(\partial_{\gamma_t}\bar{\Delta}_N) \\
&+ \frac{N}{N-1}\sum_t(\bar{\mathcal{H}}^{-1}_{\gamma\alpha})_{jt}(\partial_{\pi^3}\bar{\ell}_{ij})(\partial_{\alpha_t}\bar{\Delta}_N) + \frac{N}{N-1}\sum_t(\bar{\mathcal{H}}^{-1}_{\gamma\gamma})_{jt}(\partial_{\pi^3}\bar{\ell}_{ij})(\partial_{\gamma_t}\bar{\Delta}_N)
\end{aligned}
$$

which is $O_p(N^{-1})$. Finally, in the Supplementary Appendix (S.4) it is shown that $\mathbf{J}[R_\Delta] = o_p(1)$, and by $\tilde{R}_{\Delta,(k)} = o_p(N^{-1})$ for each k (and in the full sample) we have that $(N -$

2)$\tilde{R}_{\Delta,(k)} = o_p(1)$ and $(N-1)\tilde{R}_{\Delta} = o_p(1)$ so that the jackknifed version of this term is also $o_p(1)$. □

D.1 Proof of Theorem 2

We can decompose $\widehat{\Delta}_J - \bar{\Delta}_N$ into

$$\widehat{\Delta}_J - \bar{\Delta}_N = (\widehat{\Delta}_J - \Delta_N) + (\Delta_N - \bar{\Delta}_N)$$

From Lemma 13 we have

$$\begin{aligned} N(\widehat{\Delta}_J - \Delta_N) = &\Big[(\partial_\beta \bar{\Delta}_N) - (\partial_{\phi'} \bar{\Delta}_N)\bar{\mathcal{H}}^{-1}(\partial_{\beta\phi}\bar{\mathcal{L}})\Big]\bar{W}_N^{-1}U^{(0)} \\ &+ N(\partial_{\phi'}\bar{\Delta}_N)\bar{\mathcal{H}}^{-1}\mathcal{S} + o_p(1) \end{aligned}$$

Some tedious matrix algebra shows that this expression is equivalent to

$$\begin{aligned} N(\widehat{\Delta}_J - \Delta_N) &= -N(\partial_\theta \bar{\Delta}_N)(\partial_{\theta\theta'}\bar{\mathcal{L}})^{-1}(\partial_\theta \mathcal{L}) \\ &= \big(-N(\partial_\theta \bar{\Delta}_N)(\partial_{\theta\theta'}\bar{\mathcal{L}})^{-1}\big)\frac{1}{N-1}\sum_i \sum_{j\neq i} \partial_\theta \ell_{ij} \\ &= \frac{1}{N-1}\sum_i \sum_{j\neq i} \tilde{h}_{ij} \end{aligned}$$

where $\theta' = (\beta, \phi')$, and $\tilde{h}_{ij} = -N(\partial_\theta \bar{\Delta}_N)(\partial_{\theta\theta'}\bar{\mathcal{L}})^{-1}\partial_\theta \ell_{ij}$. Next, let $\tilde{s}_{ij} = \frac{1}{|\Lambda_{ij}|}\sum_{\lambda\in\Lambda_{ij}}(m_\lambda - \bar{m}_\lambda)$, where $\Lambda_{ij} = \{\lambda : (i,j) \in \lambda\}$. Then

$$\begin{aligned} N(\Delta_N - \bar{\Delta}_N) &= \frac{N}{|\Lambda_N|}\sum_\lambda (m_\lambda - \bar{m}_\lambda) \\ &= \frac{1}{r}\frac{N}{|\Lambda_N|}\sum_i \sum_{j\neq i} \sum_{\lambda\in\Lambda_{ij}} (m_\lambda - \bar{m}_\lambda) \\ &= \frac{1}{N-1}\sum_i \sum_{j\neq i} \tilde{s}_{ij} \end{aligned}$$

since $|\Lambda_{ij}| = r\frac{(N-2)!}{(N-p)!}$. We then have

$$N(\widehat{\Delta}_J - \bar{\Delta}_N) = \frac{1}{N-1}\sum_i \sum_{j\neq i}(\tilde{h}_{ij} + \tilde{s}_{ij}) + o_p(1)$$

which is asymptotically normal by a standard CLT. We now determine the variance of this term.

$$Var\Big(\frac{1}{N-1}\sum_i \sum_{j\neq i}(\tilde{h}_{ij} + \tilde{s}_{ij})\Big) = \frac{1}{(N-1)^2}\sum_i \sum_{j\neq i}\sum_s \sum_{t\neq s}\bar{E}[(\tilde{h}_{ij} + \tilde{s}_{ij})(\tilde{h}_{st} + \tilde{s}_{st})]$$

To compute this, first note that $\sum_s \sum_{t\neq s}\bar{E}[\tilde{h}_{ij}\tilde{h}_{st}] = \bar{E}[\tilde{h}_{ij}(\tilde{h}_{ij} + \tilde{h}_{ji})]$. Also,

$$\begin{aligned}
\sum_s \sum_{t\neq s}\bar{E}[\tilde{h}_{ij}\tilde{s}_{st}] &= \frac{1}{r}\frac{(N-p)!}{(N-2)!}\sum_s \sum_{t\neq s}\sum_{\lambda\in\Lambda_{st}}\bar{E}\big[\tilde{h}_{ij}(m_\lambda - \bar{m}_\lambda)\big] \\
&= \frac{(N-p)!}{(N-2)!}\sum_\lambda \bar{E}\big[\tilde{h}_{ij}(m_\lambda - \bar{m}_\lambda)\big] \\
&= \frac{(N-p)!}{(N-2)!}\sum_{\lambda\in(\Lambda_{ij}\cup\Lambda_{ji})}\bar{E}\big[\tilde{h}_{ij}(m_\lambda - \bar{m}_\lambda)\big] \\
&= \bar{E}[\tilde{h}_{ij}s_{ij}]
\end{aligned}$$

where $s_{ij} = \frac{(N-p)!}{(N-2)!}\sum_{\lambda\in(\Lambda_{ij}\cup\Lambda_{ji})}(m_\lambda - \bar{m}_\lambda)$.

Let $D(\lambda)$ be the set of dyads formed from the observations in λ, i.e. if $(i,j)\in\lambda$ then $(i,j),(j,i)\in D(\lambda)$. Assume that λ and λ' both contain the observations (i,j),(i,k). There are $O(N^{p-3})$ sets λ containing the corresponding dyads, so that there are $O(N^{2p-3})$ such λ,λ' pairs ($O(N^3)$ choices of (i,j,k) and $O(N^{p-3})$ choices for each of λ and λ'). Similarly, there are $O(N^{2p-4})$ λ, λ' pairs that share two dyads made up for four agents, (i,j),(k,l).

Using this,

$$
\begin{aligned}
&\frac{1}{(N-1)^2}\Big(\frac{1}{r}\frac{(N-p)!}{(N-2)!}\Big)^2\sum_i\sum_{j\neq i}\sum_{\lambda\in\Lambda_{ij}}\sum_s\sum_{t\neq s}\sum_{\lambda'\in\Lambda_{st}}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad=\frac{N^2}{|\Lambda_N|^2}\sum_\lambda\sum_{\lambda'}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad=\frac{N^2}{|\Lambda_N|^2}\sum_\lambda\sum_{\lambda':|D(\lambda)\cap D(\lambda')|=2}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad+\frac{N^2}{|\Lambda_N|^2}\sum_\lambda\sum_{\lambda':|D(\lambda)\cap D(\lambda')|>2}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad=\frac{N^2}{|\Lambda_N|^2}\sum_i\sum_{j<i}\sum_{\lambda\in(\Lambda_{ij}\cup\Lambda_{ji})}\sum_{\lambda':|D(\lambda)\cap D(\lambda')|=\{(i,j),(j,i)\}}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad+\frac{N^2}{|\Lambda_N|^2}\sum_\lambda\sum_{\lambda':|D(\lambda)\cap D(\lambda')|>2}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad=\frac{N^2}{|\Lambda_N|^2}\sum_i\sum_{j<i}\sum_{\lambda\in(\Lambda_{ij}\cup\Lambda_{ji})}\sum_{\lambda'\in(\Lambda_{ij}\cup\Lambda_{ji})}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad-\frac{N^2}{|\Lambda_N|^2}\sum_i\sum_{j<i}\sum_{\lambda\in(\Lambda_{ij}\cup\Lambda_{ji})}\sum_{\substack{\lambda'\in(\Lambda_{ij}\cup\Lambda_{ji}):\\|D(\lambda)\cap D(\lambda')|>2}}\bar{E}\big[(m_\lambda-\bar{m}_\lambda)(m_{\lambda'}-\bar{m}_{\lambda'})\big]\\
&\quad+O_p(N^{-1})\\
&\quad=\frac{1}{(N-1)^2}\sum_i\sum_{j<i}\bar{E}[s_{ij}^2]+o_p(1)
\end{aligned}
$$

This implies that

$$
\begin{aligned}
Var\Big(\frac{1}{N-1}\sum_i\sum_{j\neq i}(h_{ij}+\tilde{s}_{ij})\Big)&=\frac{1}{(N-1)^2}\sum_i\sum_{j<i}\Big(\bar{E}[(\tilde{h}_{ij}+\tilde{h}_{ji})^2]\\
&\quad+2\bar{E}[(\tilde{h}_{ij}+\tilde{h}_{ji})s_{ij}]+\bar{E}[s_{ij}^2]\Big)+o_p(1)\\
&=\frac{1}{(N-1)^2}\sum_i\sum_{j<i}\bar{E}\big[(h_{ij}+s_{ij})^2\big]+o_p(1)
\end{aligned}
$$

for $h_{ij}=\tilde{h}_{ij}+\tilde{h}_{ji}$. The asymptotic variance of $N(\widehat{\Delta}_J-\bar{\Delta}_N)$ is given by the limit of this expression. Assumptions 1 (iii) and 2 (ii) guarantee that both h_{ij} and s_{ij} have first-order approximations that are continuously differentiable in the parameters, so that the continuous mapping theorem and the consistency results $\|\widehat{\beta}-\beta_0\| \to 0$ and $\|\widehat{\phi}-\phi\|_\infty \to 0$ in (24), imply consistency of the plug-in estimator for V_Δ (see for example Lemma S.1 in FVW).

D.2 Proof of Theorem 3

We begin with a U-statistic representation of $\bar{\Delta}_N$, which will allow us to apply standard asymptotic results on U-statistics. We have defined m to be a function of the sets λ, which depend on an *ordered* set of p agents. For example the transitive triangle $\lambda = \{(i,j),(i,k),(k,j)\}$ depends on the agents $\{i,j,k\}$ in a non-symmetric manner. We first rewrite $\bar{\Delta}_N$ to be a sum over functions that are symmetric in agents. Denote the set of agents in λ by $N(\lambda)$, and let $\eta = \{i_1,\ldots,i_p\}$ be some set of p agents. Then we may define $\tau_\eta = \{\lambda : N(\lambda) = \eta\}$ as the collection of all λ that contain the same set of agents. Note that $|\tau_\eta| = p!$. We have, for $\tilde{m} = \bar{m} - E[m]$

$$
\begin{aligned}
\bar{\Delta}_N - \delta &= \frac{1}{N\cdots(N-p+1)}\sum_{\lambda}\tilde{m}_\lambda \\
&= \frac{p!}{N\cdots(N-p+1)}\sum_{\tau}\Big(\frac{1}{p!}\sum_{\lambda\in\tau}\tilde{m}_\lambda\Big) \\
&= \binom{N}{p}^{-1}\sum_{\tau}u_\tau
\end{aligned}
$$

where $u_\tau = \frac{1}{p!}\sum_{\lambda\in\tau}\tilde{m}_\lambda$. The variable u_τ is symmetric function of $\{\beta, X_i, \alpha_i, \gamma_i\}$ for p agents i. For example, there are $3! = 6$ possible transitive triangles using agents $\{i,j,k\}$ so that u is the average of the function m evaluated at these 6 different triangles. Assuming that the $\{X_i, \alpha_i, \gamma_i\}$ are i.i.d. over agents, $\bar{\Delta}_N - \delta$ is a U-statistic of order p and we apply standard theory on such statistics to compute its asymptotic distribution. As in Theorem 12.3 in van der Vaart (1998) we have

$$\sqrt{N}(\bar{\Delta}_N - \delta) \Rightarrow N(0, p^2\zeta_1)$$

where, for τ and τ' sharing exactly one agent in common,

$$\zeta_1 = Cov(u_\tau, u_{\tau'})$$

An estimator of ζ_1 is

$$\begin{aligned}\sqrt{N}\binom{N}{p}^{-1}\sum_{\tau}u_\tau &= \frac{1}{\sqrt{N}}\sum_i \binom{N-1}{p-1}^{-1}\sum_{\tau:i\in\tau}u_\tau \\ &= \frac{1}{\sqrt{N}}\sum_i t_i\end{aligned}$$

The variance of t_i is given by

$$\begin{aligned}Var(t_i) &= \binom{N-1}{p-1}^{-2}\sum_{\tau:i\in\tau}\sum_{\tau':i\in\tau'}E[u_\tau u_{\tau'}] \\ &= \binom{N-1}{p-1}^{-2}\sum_{\tau:i\in\tau}\sum_{\tau':\tau\cap\tau'=\{i\}}E[u_\tau u_{\tau'}] \\ &+ \binom{N-1}{p-1}^{-2}\sum_{\tau:i\in\tau}\sum_{\substack{\tau':i\in\tau' \\ |\tau\cap\tau'|>1}}E[u_\tau u_{\tau'}] \\ &= \binom{N-1}{p-1}^{-2}\sum_{\tau:i\in\tau}\sum_{\tau':\tau\cap\tau'=\{i\}}E[u_\tau u_{\tau'}] + o(1) \\ &= \frac{(N-p)!(p-1)!}{(N-1)!}\frac{(N-p)!}{(N-2p+1)!(p-1)!}\zeta_1 + o(1) \\ &= \zeta_1 + o(1)\end{aligned}$$

To explain the final line, there are $\binom{N-1}{p-1}$ ways to choose τ containing i, and $\binom{N-p}{p-1}$ ways to choose the remaining $p-1$ agents in τ' so that τ and τ' share only agent i in common. The first term is therefore $O(1)$. Now assume τ and τ' share two agents in common (one of which is i). There are $N-1$ choices for the second common agent, $\binom{N-2}{p-2}$ ways to choose τ containing i and the second common agent, and $\binom{N-p}{p-2}$ ways to choose the remaining agents in τ'. This implies that the sum for τ and τ' with two agents in common is $O(N^{-1})$. Similarly, the sums for three agents in common are $O(N^{-2})$ and so on.

This implies that the variance of t_i converges to ζ_1. We can alternatively express this variance as

$$
\begin{aligned}
Var(t_i) &= \binom{N-1}{p-1}^{-2} \sum_{\tau: i\in\tau} \sum_{\tau': i\in\tau'} E[u_\tau u_{\tau'}] \\
&= \binom{N-1}{p-1}^{-2} \sum_{\tau: i\in\tau} \sum_{\tau': i\in\tau'} \frac{1}{p!} \sum_{\lambda\in\tau} \frac{1}{p!} \sum_{\lambda'\in\tau} E[\tilde{m}_\lambda \tilde{m}_{\lambda'}] \\
&= \frac{1}{p!^2} \binom{N-1}{p-1}^{-2} \sum_{\lambda: i\in\lambda} \sum_{\lambda': i\in\lambda'} E[\tilde{m}_\lambda \tilde{m}_{\lambda'}] \\
&= \frac{1}{p^2} E\Big[\Big(\frac{(N-p)!}{(N-1)!} \sum_{\lambda: i\in\lambda} \tilde{m}_\lambda\Big)^2\Big]
\end{aligned}
$$

An estimator of $p^2\zeta_1^2$ is therefore given by

$$
\begin{aligned}
\widehat{V}_\delta &= \frac{1}{N} \sum_i \tilde{\mu}_i^2 \\
\tilde{\mu}_i &= \frac{(N-p)!}{(N-1)!} \sum_{\lambda: i\in\lambda} (\widehat{\bar{m}}_\lambda - \widehat{\mu}) \\
\widehat{\mu} &= \frac{(N-p)!}{N!} \sum_\lambda \widehat{\bar{m}}_\lambda
\end{aligned}
$$

and $\widehat{\bar{m}}_\lambda$ is a plug-in estimator for $\bar{m}_\lambda$. Assumption 2 and consistency of parameters $\|\widehat{\phi} - \phi_0\|_\infty \to 0$, $\|\widehat{\beta} - \beta_0\| \to 0$ ensures consistency of $\widehat{m}_\lambda$ and hence consistency of $\widehat{V}_\delta$.