

Triadic Network Formation

Chris Muris
McMaster University

Cavit Pakel
University of Oxford

Introduction

- ▶ In link formation models, main focus typically on **dyadic networks**:
 - ▶ employer-employee, student-teacher, social networks, trade networks, ...
- ▶ However, many interesting relationships involve **more than two nodes**:
 - ▶ buyer-seller-product, firm-country-product, platform-user-content creator, ...
- ▶ Granular datasets that cover multiple dimensions increasingly common
 1. **Typically sparse**
 2. **Richer/more complicated forms of heterogeneity**

Example: firm-level production networks

- ▶ Monthly firm-to-firm transaction-level data for hundreds of thousands of firms, potentially observed across multiple years.
- ▶ **Notoriously sparse.**
- ▶ Bernard, Moxnes and Saito (2019): **Firm-level data on 785,939 Japanese firms**
 - ▶ median customer and supplier numbers equal 3
- ▶ Bernard, Moxnes and Ulltveit-Moe (2018): **Annual data on the universe of Norwegian exporters for 2004-2012**
 - ▶ Majority of the firms have few connections...
 - ▶ ...only a few firms have a large number of connections

This paper

- ▶ Triadic link formation in **sparse networks** with **dyad-level fixed effects** and **logistic errors**:

$$Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\}.$$

- ▶ Estimation and inference for β_0
 - ▶ Sparsity $\Rightarrow$ fixed effects not point-identified
 - ▶ Point identification of β_0 obtained via **informative hexads**
 - ▶ **Hexad logit estimator**; consistency and asymptotic normality
- ▶ **The three-way structure with pair-specific effects exhibits complications that are absent from earlier link formation models**

Examples

Genealogy

- **Dyadic link formation** (Graham 2017, Charbonneau 2017, Jochmans 2018):

$$Y_{ij} = \mathbb{1}\{X'_{ij}\beta_0 + A_i + B_j \geq \varepsilon_{ij}\}.$$

- **Triadic link formation with node-level fixed effects** is a *direct* extension:

$$Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_i + B_j + C_k \geq \varepsilon_{ijk}\};$$

considered in the appendix to this paper.

- Node-level heterogeneity provides a more restrictive structure.
- Dyad-level heterogeneity is much less restrictive. In some sense, closer to a fully nonparametric specification for unobserved heterogeneity.

Links vs informative wirings

- ▶ Information accumulates through **informative wirings**: sub-network structures that are specific to the model at hand.
- ▶ Under **node-level heterogeneity**: more links $\Rightarrow$ more informative wirings.
- ▶ Under **dyad-level heterogeneity**: in the limit, one can have infinitely many links...
...but asymptotically zero informative wirings!
- ▶ This is a new phenomenon in this class of models.
- ▶ **Intuition**: Dyad-level heterogeneity is richer; but the type of information it requires is rarer in data.
- ▶ Implications for admissible levels of sparsity, consistency and also for obtaining asymptotic normality.

Literature overview

- ▶ **Binary choice network formation models:** Graham (2016, 2017), Charbonneau (2017), Jochmans (2018) analyse dyadic networks. More recently, Szini (2023) considers distribution regressions in a network setting; Muris, Pakel and Zhang (2025) analyse ordered outcomes in a network structure.

But also note interest in models of the intensive margin; e.g., Bramoullé, Djebbari and Fortin (2009), Davezies, D'Haultfoeuille and Fougère (2009).

Surveys by Graham (2015), de Paula (2017), Graham and de Paula (2020).

- ▶ **Sufficient statistics and incidental parameter problem:** Use of sufficient statistics goes back to Rasch (1960), Andersen (1970), and Chamberlain (1980). Dano, Honoré and Weidner (2025) provide a recent review and also a general result for obtaining sufficient statistics in the binary logit setting. Bonhomme and Dano (2024) consider the functional differencing approach.
- ▶ **Nonlinear three-way models:** Weidner and Zylkin (2021), Yang and Zhang (2023) and Resende, Lecué, Wilner and Choné (2025) consider Poisson count models. Czarnowske and Stammann (2025) consider dense three-way nonlinear panels with dyad-specific effects and propose bias correction methods.

NOTATION AND THE MODEL

Notation

- ▶ **Nodes** i , j and k represent individuals/variables of different types.
- ▶ Our setting is **tripartite**. For simplicity of exposition each **part** has the same cardinality N , i.e. $(i, j, k) \in \{1, \dots, N\}^3$
- ▶ A **triad** (i, j, k) is an ordered triplet consisting of one node from each type.
- ▶ A link connecting the triad (i, j, k) is called a **hyperedge**.
- ▶ $Y_{ijk} = 1$ if a hyperedge exists where i is in position 1, j is in position 2, and k is in position 3. $Y_{ijk} = 0$ otherwise.
- ▶ Y_{ijk} is **not symmetric** in its arguments.
E.g.: considering exporter-importer-product networks, $Y_{\text{FR}, \text{UK}, \text{Cheese}}$ and $Y_{\text{UK}, \text{FR}, \text{Cheese}}$ correspond to different things.

The model

- ▶ Hyperedge formation modelled as

$$Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\},$$

X_{ijk} is a $p \times 1$ vector of observable characteristics, and β_0 is the parameter vector.

- ▶ $F_{ijk} = (A_{ij}, B_{jk}, C_{ik})$ collects dyad-level fixed effects.
- ▶ ε_{ijk} is **logistic** conditional on X_{ijk} and (A_{ij}, B_{jk}, C_{ik}) .
- ▶ Conditional independence of triad-specific errors:

$$\varepsilon_{ijk} \perp\!\!\!\perp \varepsilon_{i'j'k'} | X_{ijk}, X_{i'j'k'}, F_{ijk}, F_{i'j'k'}.$$

The model

- ▶ Let $\mathbb{T}_N$ be the set of all triads for the given network. Let $\mathbf{Y}$ be the collection of all Y_{ijk} across a given network; define $\mathbf{F}$ and $\mathbf{X}$ analogously.
- ▶ Let $\Lambda(k) = \exp(k)/(1 + \exp(k))$.

Assumption 1 (Likelihood)

The conditional likelihood of the network $\mathbf{Y} = \mathbf{y}$ is

$$\mathrm{P}(\mathbf{Y} = \mathbf{y} | \mathbf{X}, \mathbf{F}) = \prod_{(i,j,k) \in \mathbb{T}_N} \mathrm{P}(Y_{ijk} = y_{ijk} | X_{ijk}, F_{ijk}), \quad (1)$$

where

$$\mathrm{P}(Y_{ijk} = y | X_{ijk}, F_{ijk}) = \begin{cases} \Lambda(X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik}) & \text{if } y = 1, \\ 1 - \Lambda(X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik}) & \text{if } y = 0. \end{cases}$$

THE HEXAD LOGIT ESTIMATOR

Identification and Estimation

Incidental parameter issue and conditional likelihood

- ▶ Identification of fixed effects fails due to sparsity.
- ▶ Under logistic errors, β_0 can be point-identified, via a **conditional likelihood approach**; Rasch (1961), Andersen (1970), Chamberlain (1980), Dano, Honoré and Weidner (2025).
- ▶ The conditional likelihood for the full network is available but computationally prohibitive.
- ▶ Classical solution: consider conditional likelihoods for sub-networks.

More on identification

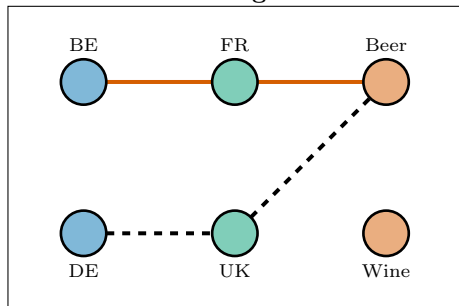
Hexads

- ▶ Our approach is based on **hexad** sub-networks.
- ▶ A **hexad** $\sigma = ((i_1, j_1, k_1), (i_2, j_2, k_2))$ is a collection of six nodes, where each part contributes with two distinct nodes (also called a **2-partite 2-uniform hypergraph**)
- ▶ Identification requires one of two specific hexad wirings, which we call the **informative wirings**.
- ▶ In **dyadic link formation**, Charbonneau (2017), Graham (2017), and Jochmans (2018) use **tetrads** for constructing a sufficient statistic.

More on identification

Hexads

Wiring 1



Wiring 2

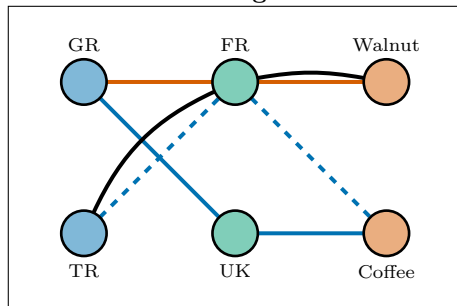
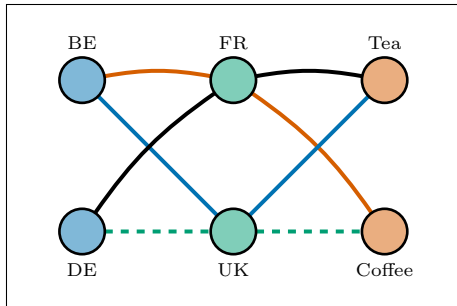
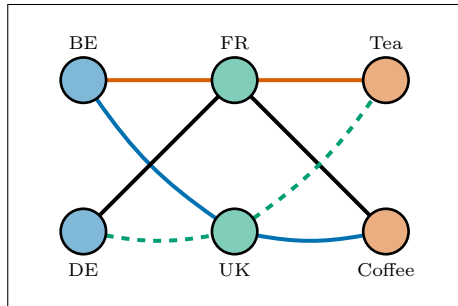
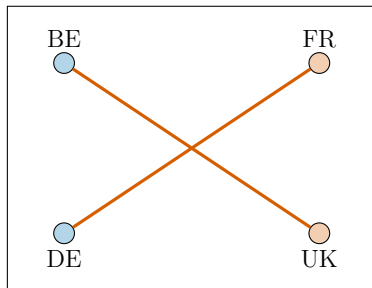
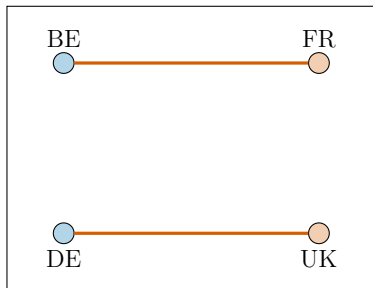


Figure: Two examples of hexad wirings.

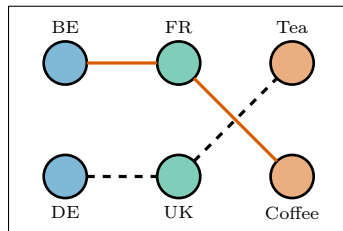
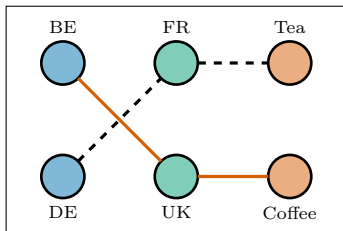
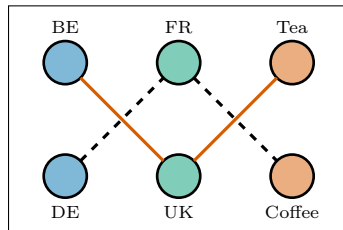
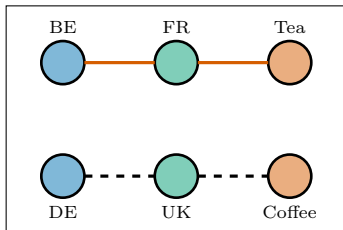
Informative wirings for $Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\}$



Informative wirings for $Y_{ij} = \mathbb{1}\{X'_{ij}\beta_0 + A_i + B_j \geq \varepsilon_{ij}\}$.



Informative wirings for $Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_i + B_j + C_k \geq \varepsilon_{ijk}\}$



Main take-away

- ▶ The informative wirings are used to construct a *conditional likelihood function* free of the fixed effects, enabling estimation of β_0 .
- ▶ Informative wirings under **node-level fixed effects** require **two edges/hyperedges**.
- ▶ Informative wirings under **dyad-level fixed effects** require **four hyperedges**.
- ▶ **Intuitively:** the kind of information required under dyad-level heterogeneity is rarer.
- ▶ Fundamental implications for large sample behaviour.

The Hexad Logit Estimator

- ▶ **Notation:** For a given hexad σ ,
 - ▶ $S_{\sigma,1} = 1$ if it admits the first identifying wiring;
 - ▶ $S_{\sigma,2} = 1$ if it admits the second identifying wiring;
 - ▶ $S_{\sigma,1} = S_{\sigma,2} = 0$ if it admits neither identifying wiring.
- ▶ Informative wirings are mutually exclusive:

$$S_{\sigma} = S_{\sigma,1} + S_{\sigma,2}$$

indicates whether σ admits one of the two informative wirings.

- ▶ Let F_{σ} and X_{σ} be the set of fixed effects and covariates for all triads involved in hexad σ .

The Hexad Logit Estimator

Theorem 1

If Assumption 1 holds, then, for any hexad $\sigma = ((i_1, j_1, k_1), (i_2, j_2, k_2))$,

$$P(S_{\sigma,1} = 1 | S_{\sigma} = 1, \mathbf{F}_{\sigma}, X_{\sigma}; \beta_0) = \Lambda(W'_{\sigma}\beta_0),$$

$$P(S_{\sigma,2} = 1 | S_{\sigma} = 1, \mathbf{F}_{\sigma}, X_{\sigma}; \beta_0) = 1 - \Lambda(W'_{\sigma}\beta_0),$$

where

$$W_{\sigma} = W_{\sigma,1} - W_{\sigma,2},$$

$$W_{\sigma,1} = (X_{i_1 j_1 k_1} + X_{i_1 j_2 k_2} + X_{i_2 j_1 k_2} + X_{i_2 j_2 k_1})$$

$$W_{\sigma,2} = (X_{i_1 j_2 k_1} + X_{i_1 j_1 k_2} + X_{i_2 j_1 k_1} + X_{i_2 j_2 k_2})$$

The Hexad Logit Estimator

- ▶ The conditional log-likelihood function for hexad σ is

$$l_{\sigma}(\beta) = \sum_{c=1}^2 S_{\sigma,c} \ln p_{\sigma,c}(\beta),$$

where $p_{\sigma,c}(\beta) = P(S_{\sigma,c} = 1 | S_{\sigma} = 1, X_{\sigma}; \beta)$.

- ▶ The **hexad logit estimator** of β_0 is then given by

$$\hat{\beta} = \operatorname{argmax}_{b \in \mathcal{B}} \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} l_{\sigma}(b)$$

where $\mathcal{B}$ is the parameter space for β and Σ is the set of all hexads, with $|\Sigma| = N^3(N-1)^3$.

THE HEXAD LOGIT ESTIMATOR

Large Sample Theory

Theory

Assumption 2

The parameter space $\mathcal{B} \subset \mathbb{R}^P$ is compact, and β_0 is in the interior of $\mathcal{B}$. Moreover, the support of the regressors X_{ijk} , $\mathcal{X} \subset \mathbb{R}^P$, is also compact.

- Compactness can be weakened by imposing appropriate moment existence assumptions.

Assumption 3

Let $m_N = |\Sigma| = N^3(N-1)^3$ and $p_N = P(S_\sigma = 1)$, the unconditional probability that the hexad σ is informative. Define also,

$$\overline{W}_\sigma = p_{\sigma,1}(\beta_0)W_{\sigma,1} + p_{\sigma,2}(\beta_0)W_{\sigma,2}.$$

The matrix

$$\Gamma_0 = - \lim_{N \rightarrow \infty} \frac{1}{p_N m_N} \sum_{\sigma \in \Sigma} \mathbb{E} \left(S_\sigma \sum_{c=1}^2 p_{\sigma,c}(\beta_0) (W_{\sigma,c} - \overline{W}_\sigma) (W_{\sigma,c} - \overline{W}_\sigma)' \right),$$

exists and is negative definite.

Theory: key quantities

- ▶ **Link formation probability** is given by $\rho_N = \mathbb{P}(Y_{ijk} = 1)$.
- ▶ **Average node degree** (expected # of hyperedges per node) is $N^2 \rho_N$.
- ▶ **Asymptotic non-emptiness** therefore requires

$$\rho_N = O\left(\frac{1}{N^\delta}\right) \quad \text{where} \quad 0 \leq \delta < 2.$$

- ▶ **The expected number of links** is given by

$$N^3 \rho_N = O(N^{3-\delta}).$$

- ▶ **The number of informative wirings** is of magnitude

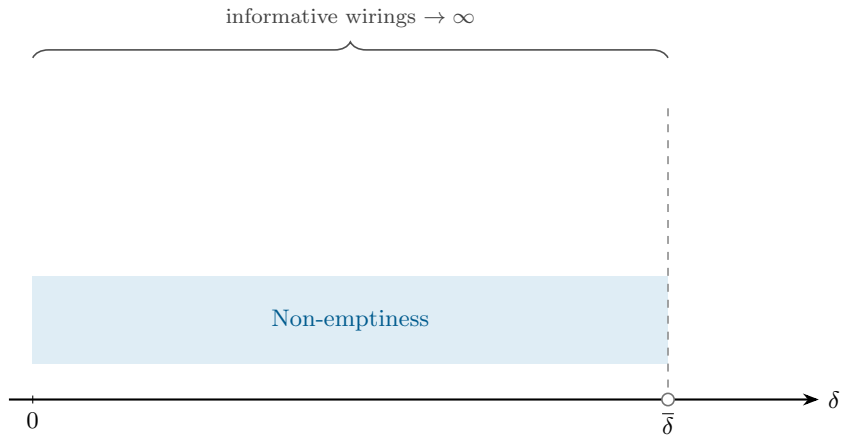
$$O(N^6 \rho_N^4) = O(N^{6-4\delta}).$$

Links vs informative wirings

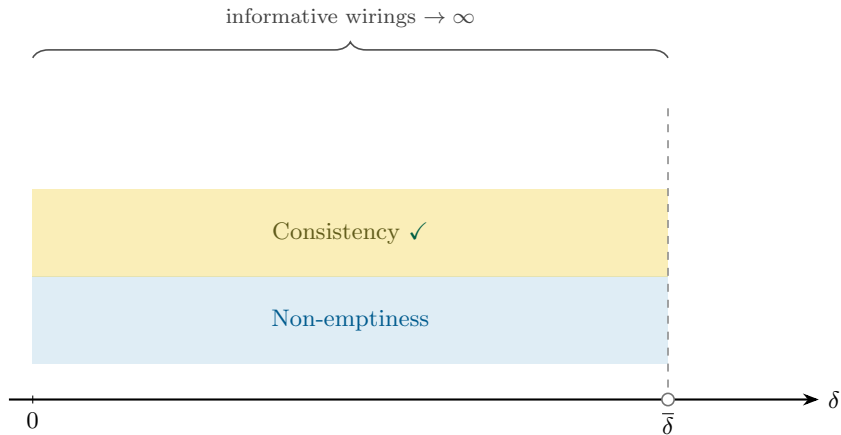
	Dyadic	Triadic I	Triadic II
Heterogeneity format	$A_i + B_j$	$A_i + B_j + C_k$	$A_{ij} + B_{jk} + C_{ik}$
Link probability	$\tilde{\rho}_N = P(Y_{ij} = 1)$	$\bar{\rho}_N = P(Y_{ijk} = 1)$	$\rho_N = P(Y_{ijk} = 1)$
# links	$N^2 \tilde{\rho}_N$	$N^3 \bar{\rho}_N$	$N^3 \rho_N$
# informative wirings	$O(N^4 \tilde{\rho}_N^2)$	$O(N^6 \bar{\rho}_N^2)$	$O(N^6 \rho_N^4)$

- ▶ The triadic model with dyad-level effects can have **infinitely many links yet asymptotically zero informative wirings** (e.g. $\rho_N = 1/N^{8/5}$).
- ▶ This phenomenon is **automatically ruled out in models with node-level heterogeneity**.

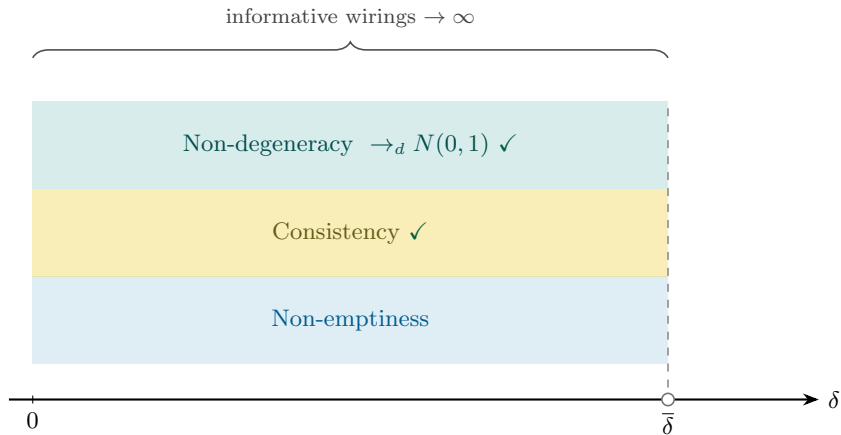
The story across δ regimes: node-level FEs ($A_i + B_j + C_k$)



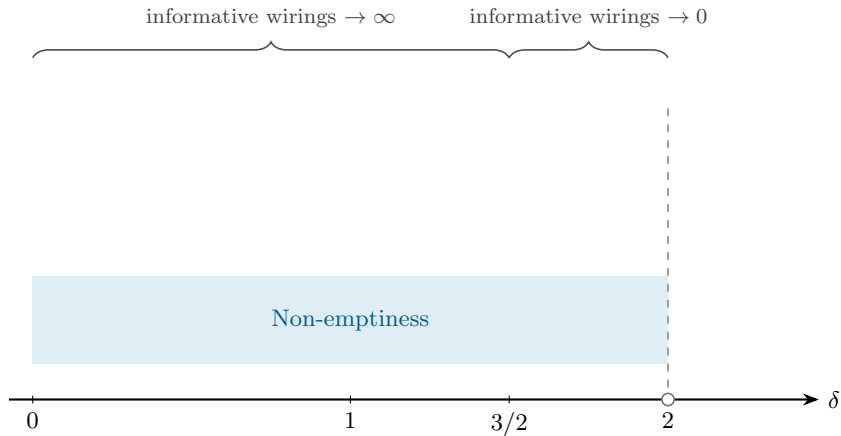
The story across δ regimes: node-level FEs ($A_i + B_j + C_k$)



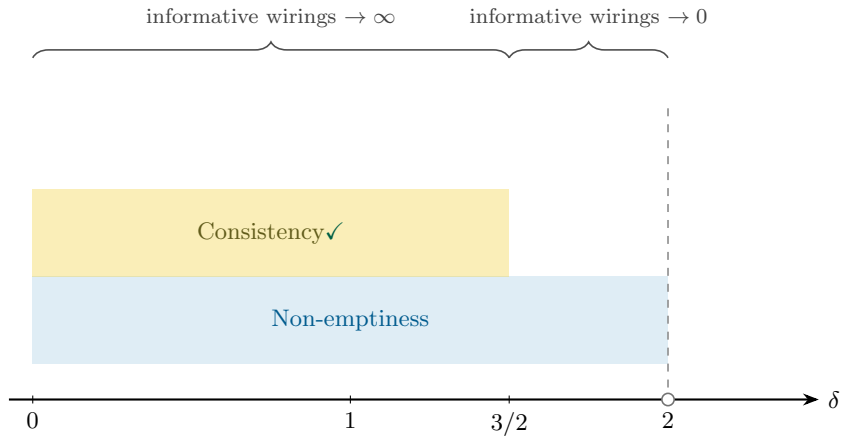
The story across δ regimes: node-level FEs ($A_i + B_j + C_k$)



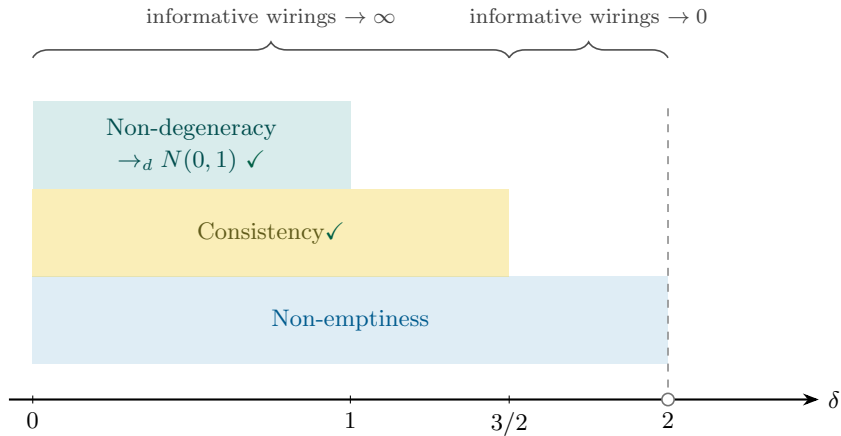
The story across δ regimes: dyad-level FEs ($A_{ij} + B_{jk} + C_{ik}$)



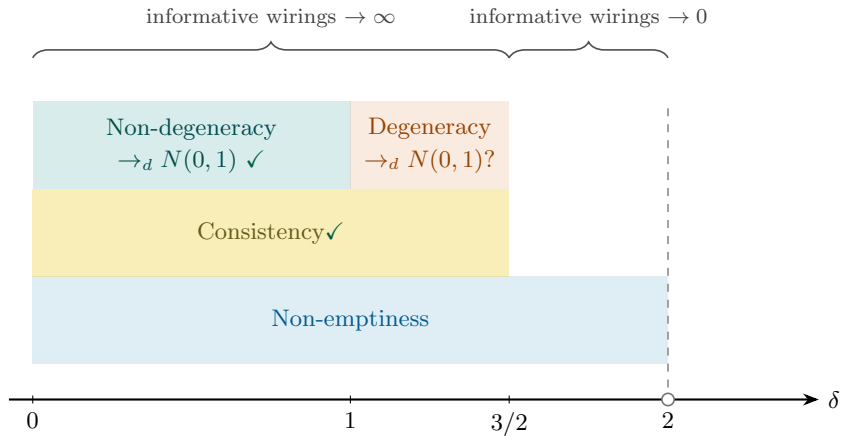
The story across δ regimes: dyad-level FEs ($A_{ij} + B_{jk} + C_{ik}$)



The story across δ regimes: dyad-level FEs ($A_{ij} + B_{jk} + C_{ik}$)



The story across δ regimes: dyad-level FEs ($A_{ij} + B_{jk} + C_{ik}$)



Consistency

Theorem 2 (Consistency)

Let Assumptions 1-3 hold and that $\rho_N = O(1/N^\delta)$ where $0 \leq \delta < 3/2$. Then, $\hat{\beta} - \beta_0 \rightarrow_p 0$.

- ▶ Need more than non-emptiness!
- ▶ **Remember that** the number of informative hexads is

$$O(N^6 \rho_N^4) = O(N^{6-4\delta})$$

and $0 \leq \delta < 3/2$ is the range that guarantees infinitely many informative hexads as $N \rightarrow \infty$.

Asymptotic normality

Theorem 3 (Asymptotic normality)

Let Assumptions 1-3 hold. Assume, furthermore, that $\rho_N = O(1/N^\delta)$ where $0 \leq \delta < 1$. Then,

$$\sqrt{N^3 \rho_N}(\hat{\beta} - \beta_0) \rightarrow_d \mathcal{N}(0, 2^6 \Gamma_0^{-1} \overline{M} \Gamma_0^{-1}).$$

Definitions

- Further restrictions on sparsity:

$$\rho_N = O(1/N^\delta) \quad \text{where } 0 \leq \delta < 1.$$

- Convergence rate will be between $\sqrt{N^3}$ (dense case) and $\sqrt{N^2}$ (sparse case).
- **What happens when $\delta \geq 1$?** The score Z_N behaves analogously to a **degenerate** U-statistic.

Asymptotic normality

- ▶ The score function is given by

$$Z_N = \frac{1}{N^3(N-1)^3} \sum_{\sigma \in \Sigma} s_{\sigma}(\beta_0),$$

where

$$s_{\sigma}(\beta_0) = \sum_{c=1}^2 S_{\sigma,c} \left(W_{\sigma,c} - \overline{W}_{\sigma}(\beta_0) \right).$$

- ▶ The score is essentially a U-statistic, with a complex dependence structure due to overlapping triads across hexads.

Asymptotic normality: details

- We have

$$\text{Var}(Z_N) = \text{Var} \left(\frac{1}{N^3(N-1)^3} \sum_{\sigma \in \Sigma} s_{\sigma}(\beta_0) \right) = C_{3,N} + C_{4,N} + C_{5,N} + C_{6,N},$$

with

$$C_{q,N} = \frac{1}{N^6(N-1)^6} \sum_{[\sigma, \sigma']_q} \sum \mathbb{E}[s_{\sigma}(\beta_0) s_{\sigma'}(\beta_0)]$$

where $\sum_{[\sigma, \sigma']_q}$ denotes summation over all (σ, σ') pairs with q common nodes, $q = 3, 4, 5, 6$.

- For standard U-statistics, the first term is the leading term and provides asymptotic normality.
- For the case at hand, the situation is not necessarily so.

⇒ The probability of two hexads σ and σ' being both informative also enters the calculations, which is the non-standard aspect.

Asymptotic normality: details

- In particular

$$\text{Var}(Z_N) = \underbrace{O\left(\frac{\rho_N^7}{N^3}\right)}_{C_{3,N}} + \underbrace{O\left(\frac{\rho_N^7}{N^4}\right)}_{C_{4,N}} + \underbrace{O\left(\frac{\rho_N^6}{N^5}\right)}_{C_{5,N}} + \underbrace{O\left(\frac{\rho_N^4}{N^6}\right)}_{C_{6,N}}.$$

- Using $\rho_N = O(1/N^\delta)$,

$$\text{Var}(Z_N) = O\left(\frac{1}{N^{3+7\delta}}\right) + O\left(\frac{1}{N^{4+7\delta}}\right) + O\left(\frac{1}{N^{5+6\delta}}\right) + O\left(\frac{1}{N^{6+4\delta}}\right).$$

1. Under $0 \leq \delta < 1$ $C_{3,N}$ dominates
2. Under $1 < \delta < 2$ $C_{6,N}$ dominates
3. At $\delta = 1$ $C_{3,N}$ and $C_{6,N}$ are of equal rate.

- If $C_{3,N}$ dominates, asymptotic normality follows via a Hájek projection argument.
- When $C_{6,N}$ dominates, the score behaves analogously to a **degenerate U-statistic**. In this case, one has a **Gaussian Chaos** distribution.

Asymptotic normality: intuition

- ▶ Notice that we have

$$C_{3,N} = O\left(\frac{\rho_N^8}{N^3 \rho_N}\right) \quad \text{and} \quad C_{6,N} = O\left(\frac{\rho_N^8}{N^6 \rho_N^4}\right).$$

- ▶ The **expected number of hyperedges** is $O(N^3 \rho_N)$.
- ▶ The **expected number of identifying hexads** is $O(N^6 \rho_N^4)$.
- ▶ $C_{3,N}$ **dominates** $\Leftrightarrow$ informative wirings accumulate faster than the number of hyperedges. This happens if $0 \leq \delta < 1$, that is as long as the network is not excessively sparse.
- ▶ $C_{6,N}$ **dominates** $\Leftrightarrow$ informative wirings accumulate much slower than the number of hyperedges.

CONCLUSION

Conclusion

- ▶ We investigate formation of triadic networks with dyad-level fixed effects.
- ▶ We propose the hexad logit estimator and obtain consistency and asymptotic normality.
- ▶ **Accumulation of links** does not necessarily imply **accumulation of informative wirings**.
 - ▶ Richer heterogeneity $\Rightarrow$ lower tolerance for sparsity.
 - ▶ Even under sufficient accumulation of informative wirings, **degeneracy** can arise.
- ▶ These phenomena are **absent from link formation models under node-level heterogeneity**.
- ▶ **Empirical implication:** aggregate the data if $\delta > 3/2$.
- ▶ **Ongoing:** empirical application; limit theory in the degenerate regime. We also have some simulations (not presented today). [Simulations](#)

Bibliography

- Andersen, E. B. (1970). Properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B* 32, 283–301.
- Bernard, A. B., A. Moxnes, and Y. U. Saito (2019). Production networks, geography, and firm performance. *Journal of Political Economy* 127, 639–688.
- Bernard, A. B., A. Moxnes, and K. H. Ulltveit-Moe (2018). Two-sided heterogeneity and trade. *The Review of Economics and Statistics* 100, 424–439.
- Boehm, J., S. Dhingra, and J. Morrow (2022). The comparative advantage of firms. *Journal of Political Economy* 130, 3025–3100.
- Bonhomme, S. and K. Dano (2024). Functional differencing in networks. *Revue Économique* 75, 147–176.
- Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics* 150, 41–55.
- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies* 47, 225–238.
- Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal* 20, S1–S13.

Bibliography (cont.)

- Conconi, P., F. Leone, G. Magerman, and C. Thomas (2024). Multinational networks and trade participation. Working paper.
- Czarnowske, D. and A. Stammann (2025). (Debiased) fixed effects estimation of binary logit models with three-dimensional panel data. Unpublished manuscript.
- Dano, K., B. Honoré, and M. Weidner (2025). Binary choice logit models with general fixed effects for panel and network data. Unpublished manuscript.
- Davezies, L., D'Haultfoeulle, and D. Fougère (2009). Identification of peer effects using group size variation. *Econometrics Journal* 12, 397–413.
- de Paula, A. (2017). Econometrics of network models. In B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson (Eds.), *Advances in Economics and Econometrics: Eleventh World Congress*, pp. 268–323. Cambridge University Press.
- Graham, B. (2015). Fixed-effect regressions on network data. *Annual Review of Economics* 7, 465–485.
- Graham, B. (2016). Homophily and transitivity in dynamic network formation. Working paper.
- Graham, B. and A. de Paula (Eds.) (2020). *The Econometric Analysis of Network Data*. Academic Press.

Bibliography (cont.)

- Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica* 85, 1033–1063.
- Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics* 36, 705–713.
- Kleinman, B., E. Liu, and S. J. Redding (2024). International friends and enemies. Forthcoming in the American Economic Journal: Macroeconomics.
- Muris, C., C. Pakel, and Q. Zhang (2025). Dyadic data with ordered outcome variables. Working paper.
- Rasch, G. (1960). *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence And Attainment Tests*. Nielsen & Lydiche.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 4, Berkeley, pp. 321–333. University of California Press.
- Resende, L., G. Lecué, L. Wilner, and P. Choné (2025). Statistical inference in large multi-way networks. Unpublished manuscript.

Bibliography (cont.)

- Szini, G. M. (2023). A pairwise differencing distribution regression approach for network models. Working paper.
- Weidner, M. and T. Zylkin (2021). Bias and consistency in three-way gravity models. *Journal of International Economics* 132.
- Yang, Y. and H. Zhang (2023). Three-way gravity models with multiplicative unobserved effects. *The Econometrics Journal* 26, 422–443.

More on identification

- ▶ To provide some loose intuition on identification we need to define the degree sequence for a hexad.
- ▶ Let $d(\cdot)$ be the **degree function** for a given node (that is, the number of hyperedges that this node is part of).
- ▶ Then $D(\sigma) = (d(i_1), d(j_1), d(k_1), d(i_2), d(j_2), d(k_2))$ yields the **degree sequence** of the hexad σ .

Dyad-level heterogeneity: some examples

► Firms' diversification decisions

Boehm, Dhingra and Morrow (2022) model the decision of firm j to engage in production in industry k starting from time t , allowing for **firm-time** and **industry-time** fixed effects.

► Multinational corporation network effects

Conconi, Leone, Magerman and Thomas (2024) use data on the universe of Belgian firms to model creation of an import/export link between firm i and country c at time t , including **firm-country**, **firm-time** and **country-time** fixed effects in their specification.

► Does increased economic reliance on a trade partner cause closer political alignment?

Kleinman, Liu and Redding (2024) consider linear models of various variables indicating different types of strategic rivalries and formal alliance types—all **dyad-level fixed effects** included.

More on identification

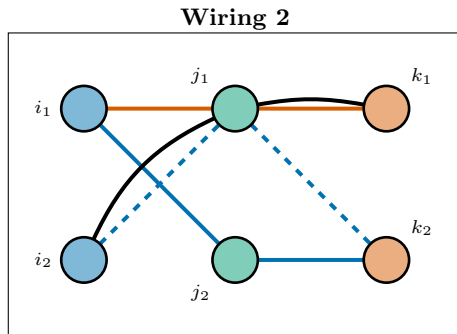
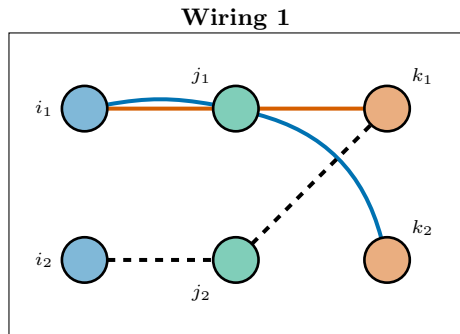


Figure: Two examples of hexad wirings. Wiring 1 has degree sequence $(2, 2, 2, 1, 1, 1)$ whereas Wiring 2 has degree sequence $(2, 3, 2, 2, 1, 2)$.

More on identification

- ▶ **Identification mechanism:** degree sequence acts as a sufficient statistic for the fixed effects.
- ▶ Different hexad wirings with identical degree sequences can be utilised to obtain a conditional likelihood function that is free of fixed effects.
- ▶ **Loose intuition:** Differences in hexad wirings with **identical degree sequences** should be due to differences in observables but not fixed effects.
- ▶ Our identification strategy is based on sub-networks generated by hexads with degree sequences $D(\sigma) = (2, 2, 2, 2, 2, 2)$.

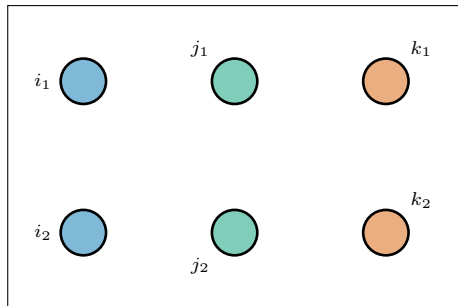
More on identification

- ▶ Under the degree sequence $(2, 2, 2, 2, 2, 2)$ we have two wirings only that allow for construction of a conditional likelihood function.
- ▶ We call these the **informative wirings**.
- ▶ Without loss of generality, consider the hexad $\sigma = ((1, 1, 1), (2, 2, 2))$. Then the two wirings consist of the following hyperedges:
 - ▶ **Wiring 1:** $(1, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1)$;
 - ▶ **Wiring 2:** $(1, 1, 2), (1, 2, 1), (2, 1, 1), (2, 2, 2)$.
- ▶ More generally, denoting the first and second nodes of each part by $-$ and $+$, respectively, in any hexad we look for one of the following patterns:
 - ▶ **Wiring 1:** $(-, -, -), (-, +, +), (+, -, +), (+, +, -)$;
 - ▶ **Wiring 2:** $(-, -, +), (-, +, -), (+, -, -), (+, +, +)$.

Identification intuition

$$Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\}.$$

In any hexad, a particular edge can appear once, twice, or be absent.

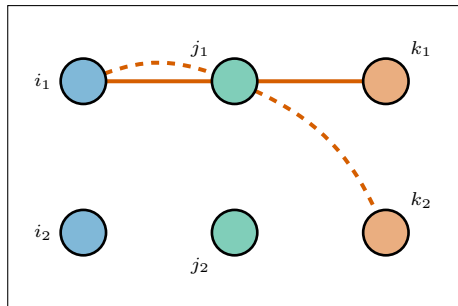


Identification intuition

$$Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\}.$$

Hexads with an edge appearing twice
are not informative.

The given example can be explained by
 $A_{i_1, j_1} = \infty$.

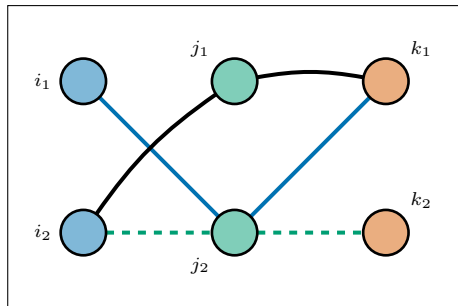


Identification intuition

$$Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\}.$$

Hexads where an edge is completely absent are not informative.

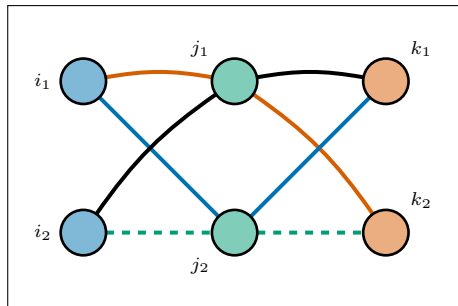
The given example can be explained by $A_{i_1, j_1} = -\infty$.



Identification intuition

$$Y_{ijk} = \mathbb{1}\{X'_{ijk}\beta_0 + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\}.$$

We need every edge to appear exactly once.



Identification

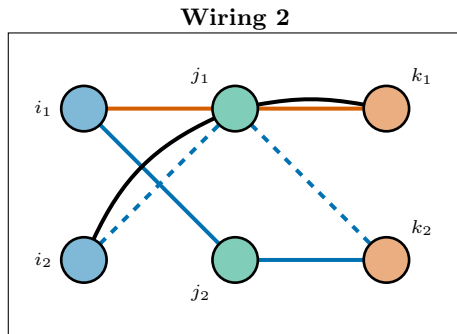
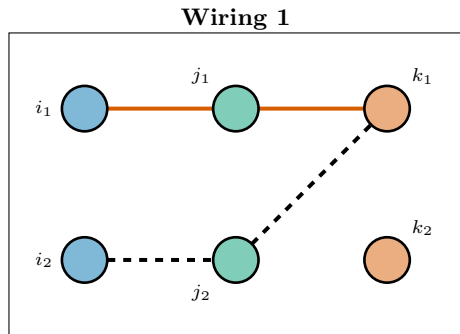


Figure: Two examples of non-informative wirings.

Hájek projection

- ▶ The approach to obtaining asymptotic normality: use a Hájek projection and collapse the problem into the triad dimension where we have conditional independence.
- ▶ Let $\mathcal{I}_{ijk} = (X_{ijk}, F_{ijk}, \varepsilon_{ijk})$ and define the projected score

$$\bar{s}_{ijk} = \mathbb{E}[s_{a,a'|b,b'|c,c'}(\beta_0) | \mathcal{I}_{ijk}]$$

- ▶ In the paper we show that the Hájek projection for the score Z_N is given by

$$Z_N^* = \frac{2^3}{N^3} \sum_{i \neq j \neq k} \bar{s}_{ijk}$$

and is asymptotically equivalent to the score function Z_N (under appropriate rescaling).

- ▶ **Intuition:** it is easier to obtain asymptotic normality for Z_N^* since it is a sum across triads and triad-specific shocks are conditionally independent.

Hájek projection

- Define

$$\overline{M}_N = \frac{1}{\rho_N^7 N^3} \sum_{1 \leq i, j, k \leq N} \mathbb{E}[\overline{s}_{ijk} \overline{s}'_{ijk} | \mathbf{X}, \mathbf{F}],$$

- Rescaling by ρ_N^7 is to account for the fact that the probability of two hexads being both informative and sharing a common hyperedge is $O(\rho_N^7)$.
- The term $\overline{M}$ appearing in the asymptotic variance is given by

$$\overline{M} = \text{plim}_{N \rightarrow \infty} \overline{M}_N.$$

SIMULATION ANALYSIS

[Back to conclusion](#)

Simulation analysis

- The DGP is

$$Y_{ijk} = 1\{X'_{ijk}\beta + A_{ij} + B_{jk} + C_{ik} \geq \varepsilon_{ijk}\},$$

where

$$X_{ijk} \sim iid \mathcal{N}(0, 1) \quad \text{and} \quad \varepsilon_{ijk} \sim iid \text{Logit}.$$

- For each setting we run $K = 1,000$ replications where $\beta_0 = 1$.
- Sparsity of the resulting network is controlled by the DGP for the fixed effects.

Simulation analysis

- ▶ Every node $i, j, k = 1, \dots, N$ has a *tendency parameter*

$$\phi_i = (N - i)/(N - 1).$$

- ▶ Fixed effects are generated as follows,

$$A_{ij} = -c_N(\phi_i + \phi_j)/2,$$

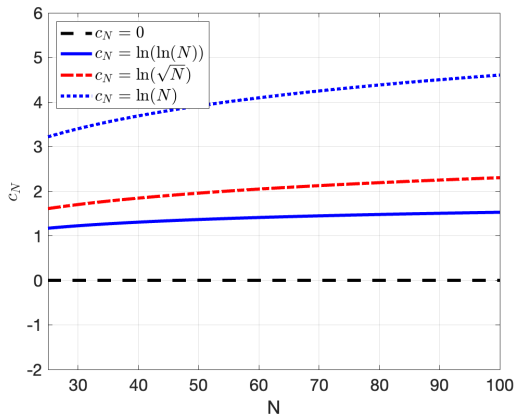
$$B_{jk} = -c_N(\phi_j + \phi_k)/2,$$

$$C_{ik} = -c_N(\phi_i + \phi_k)/2.$$

- ▶ We use

$$c_N \in \{0, \ln(\ln N), \ln(\sqrt{N}), \ln N\}$$

to generate increasing levels of sparsity.



Simulation Results

N	c_N	$\text{mean}(\hat{\beta})$	$\text{std}(\hat{\beta})$	$\text{se}(\hat{\beta})$	RMSE	SE/SD	C_{90}	C_{95}
20	0	1.00	0.04	0.06	0.04	1.43	0.977	0.994
20	$\ln(\ln N)$	1.00	0.06	0.11	0.06	1.79	0.996	1.000
20	$\ln \sqrt{N}$	1.02	0.10	0.23	0.10	2.23	1.000	1.000
30	0	1.00	0.02	0.02	0.02	1.26	0.965	0.988
30	$\ln(\ln N)$	1.00	0.03	0.04	0.03	1.45	0.980	0.997
30	$\ln \sqrt{N}$	1.01	0.04	0.07	0.04	1.80	0.998	1.000
40	0	1.00	0.01	0.01	0.01	1.18	0.952	0.986
40	$\ln(\ln N)$	1.00	0.02	0.02	0.02	1.29	0.963	0.993
40	$\ln \sqrt{N}$	1.00	0.02	0.04	0.02	1.58	0.994	0.999
50	0	1.00	0.01	0.01	0.01	1.11	0.935	0.972
50	$\ln(\ln N)$	1.00	0.01	0.01	0.01	1.16	0.942	0.972
50	$\ln \sqrt{N}$	1.00	0.02	0.02	0.02	1.40	0.984	0.995

Results: Dense Pattern

N	c_N	$\text{mean}(\hat{\beta})$	$\text{std}(\hat{\beta})$	$\text{se}(\hat{\beta})$	RMSE	SE/SD	C_{90}	C_{95}
20	0	1.00	0.04	0.06	0.04	1.43	0.977	0.994
25	0	1.00	0.03	0.04	0.03	1.39	0.976	0.992
30	0	1.00	0.02	0.02	0.02	1.26	0.965	0.988
40	0	1.00	0.01	0.01	0.01	1.18	0.952	0.986
50	0	1.00	0.01	0.01	0.01	1.11	0.935	0.972
60	0	-	-	-	-	-	-	-

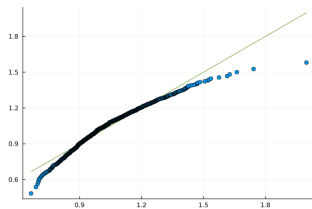
Results: $\ln(\ln N)$ Pattern

N	c_N	$\text{mean}(\hat{\beta})$	$\text{std}(\hat{\beta})$	$\text{se}(\hat{\beta})$	RMSE	SE/SD	C_{90}	C_{95}
20	$\ln(\ln N)$	1.00	0.06	0.11	0.06	1.79	0.996	1.000
25	$\ln(\ln N)$	1.00	0.04	0.06	0.04	1.61	0.988	0.999
30	$\ln(\ln N)$	1.00	0.03	0.04	0.03	1.45	0.980	0.997
40	$\ln(\ln N)$	1.00	0.02	0.02	0.02	1.29	0.963	0.993
50	$\ln(\ln N)$	1.00	0.01	0.01	0.01	1.16	0.942	0.972
60	$\ln(\ln N)$	1.00	0.01	0.01	0.01	1.18	0.938	0.972

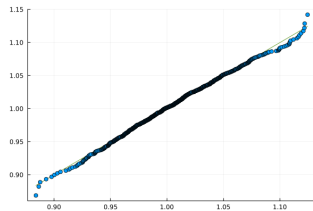
Results: $\ln(\sqrt{N})$ Pattern

N	c_N	$\text{mean}(\hat{\beta})$	$\text{std}(\hat{\beta})$	$\text{se}(\hat{\beta})$	RMSE	SE/SD	C_{90}	C_{95}
20	$\ln \sqrt{N}$	1.02	0.10	0.23	0.10	2.23	1.000	1.000
25	$\ln \sqrt{N}$	1.01	0.06	0.12	0.06	1.95	0.996	1.000
30	$\ln \sqrt{N}$	1.01	0.04	0.07	0.04	1.80	0.998	1.000
40	$\ln \sqrt{N}$	1.00	0.02	0.04	0.02	1.58	0.994	0.999
50	$\ln \sqrt{N}$	1.00	0.02	0.02	0.02	1.40	0.984	0.995
60	$\ln \sqrt{N}$	1.00	0.01	0.02	0.01	1.33	0.972	0.988

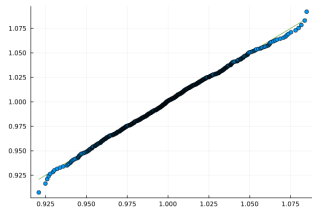
QQ-plots for $N=30$



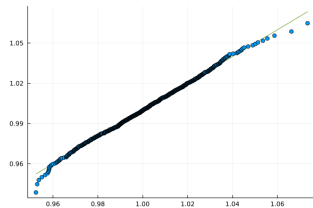
$$c_N = \ln(N)$$



$$c_N = \ln(\sqrt{N})$$

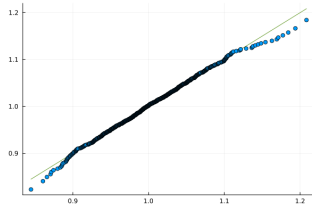


$$c_N = \ln(\ln(N))$$

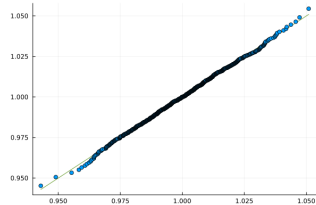


$$c_N = 0$$

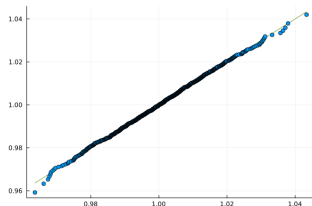
QQ-plots for $N=50$



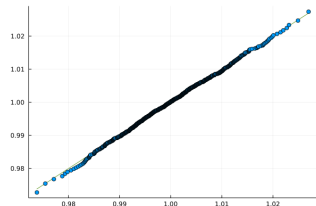
$$c_N = \ln(N)$$



$$c_N = \ln(\sqrt{N})$$



$$c_N = \ln(\ln(N))$$



$$c_N = 0$$