Uncertainty quantification for network regression

Frank W. Marrs
Colorado State University
frank.marrs@colostate.edu

Collaborators

Bailey K. Fosdick
Colorado State University

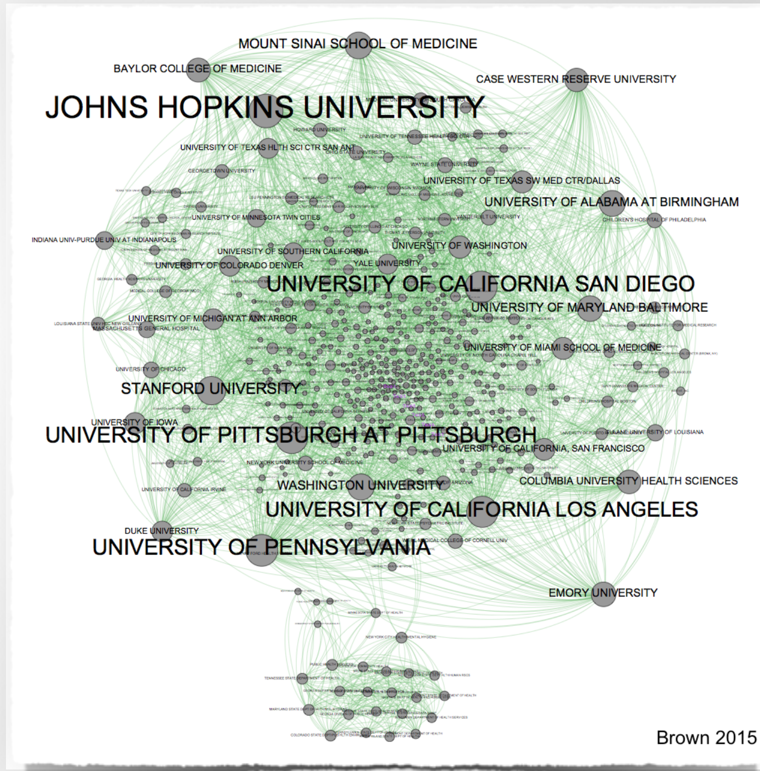
Tyler H. McCormick
University of Washington

Network regression

- response Y: **weighted, directed,** from actor *i* to *j*
- covariates X: individual or pairwise attributes
- Model linear relationship of covariates and response

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \xi_{ij}$$

$$Y = X\boldsymbol{\beta} + \boldsymbol{\xi} \ \in \mathbb{R}^{n(n-1)}$$



Brown 2015

Network regression

Motivation

- **International Trade Data** (Westveld and Hoff, 2011)
- **Informal Risk-Sharing Networks** (Fafchamps and Gubert 2007, Attansio et al 2012, Banerjee et al 2013)
- **International Militarized Disputes** (Russett and Oneal 2011)
- **Friendship Networks** (Goodreau et al 2009, Wimmer and Lewis 2010)
- **Speed Dating Networks** (Fisman et al 2006)

Network regression

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \xi_{ij}$$

$$Y = X\boldsymbol{\beta} + \boldsymbol{\xi} \;\; \in \mathbb{R}^{n(n-1)}$$

- response *Y*: **weighted, directed,** from actor *i* to *j*
- covariates *X*: individual or pairwise attributes

$$\begin{array}{c|cccc|} & A & B & C & D \\ \hline A & \blacksquare & y_{AB} & y_{AC} & y_{AD} \\ B & y_{BA} & \blacksquare & y_{BC} & y_{BD} \\ C & y_{CA} & y_{CB} & \blacksquare & y_{CD} \\ D & y_{DA} & y_{DB} & y_{DC} & \blacksquare \\ \hline \end{array}$$

$$Y = \begin{bmatrix} y_{BA} \\ y_{CA} \\ \cdots \\ y_{CD} \end{bmatrix} \qquad X = \begin{bmatrix} \mathbf{x}_{BA}^T \\ \mathbf{x}_{CA}^T \\ \cdots \\ \mathbf{x}_{CD}^T \end{bmatrix}$$

Network regression

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \xi_{ij}$$

- **Goal**: inference about $\boldsymbol{\beta}$
  - point estimates $(\widehat{\boldsymbol{\beta}})$
  - uncertainty estimate $(\widehat{\boldsymbol{\beta}} \pm \widehat{\mathrm{se}}\{\widehat{\boldsymbol{\beta}}\})$
- ξ_{ij} highly structured error
  - i.e. ξ_{ij} and ξ_{ik} share a node, expect correlation

Linear Regression

- Recall Ordinary Least Squares

$$\widehat{\boldsymbol{\beta}} = \text{argmin}_\beta ||Y - X\boldsymbol{\beta}||_2^2 = (X^TX)^{-1}X^TY$$

$$Var(\widehat{\boldsymbol{\beta}}|X) = (X^TX)^{-1}X^T\Sigma X(X^TX)^{-1}$$

$$\Sigma = Var(Y|X) = Var(\boldsymbol{\xi})$$

- X is $(n^2 - n)$ x p matrix of covariates
- Y and $\boldsymbol{\xi}$ are $(n^2 - n)$ vectors of relations and errors
- For inference on $\widehat{\boldsymbol{\beta}}$, need an estimate of Σ

Linear Regression

- Recall normal likelihood

$$\ell(Y|\boldsymbol{\beta}, \Sigma) \propto -\frac{1}{2} log(|\Sigma|) - \frac{1}{2} \boldsymbol{\xi}^T \Sigma^{-1} \boldsymbol{\xi}$$

$$\Sigma = Var(Y|X) = Var(\boldsymbol{\xi})$$

- X is $(n^2 - n)$ x p matrix of covariates
- Y and $\boldsymbol{\xi}$ are $(n^2 - n)$ vectors of relations and errors
- For inference on $\widehat{\boldsymbol{\beta}}$, need a model for Σ

Dyadic Clustering

- Fafchamps and Gubert 2007
- Non-parametric approach
- Estimate every nonzero entry in

$$\Sigma = Var(Y|X) = Var(\boldsymbol{\xi})$$

- Plug-in estimator

$$Var(\widehat{\boldsymbol{\beta}}|X) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1}$$

Dyadic Clustering

- Assumes that non-overlapping pairs independent

$$
\xi = \begin{array}{c|c|c|c|c}
 & A & B & C & D \\
\hline
A & \blacksquare & \xi_{AB} & \xi_{AC} & \xi_{AD} \\
\hline
B & \boxed{\xi_{BA}} & \blacksquare & \xi_{BC} & \xi_{BD} \\
\hline
C & \xi_{CA} & \xi_{CB} & \blacksquare & \boxed{\xi_{CD}} \\
\hline
D & \xi_{DA} & \xi_{DB} & \xi_{DC} & \blacksquare
\end{array}
$$

$$
\widehat{Cov}(\color{blue}{\xi_{BA}}, \color{red}{\xi_{CD}}) = 0
$$

Dyadic Clustering

- Model nonzero entries in Σ products of OLS residuals

$$
\xi = \begin{array}{c|cccc}
 & A & B & C & D \\
\hline
A & \blacksquare & \xi_{AB} & \boxed{\xi_{AC}} & \xi_{AD} \\
B & \boxed{\xi_{BA}} & \blacksquare & \xi_{BC} & \xi_{BD} \\
C & \xi_{CA} & \xi_{CB} & \blacksquare & \boxed{\xi_{CD}} \\
D & \xi_{DA} & \xi_{DB} & \xi_{DC} & \blacksquare
\end{array}
$$

$$\widehat{Cov}(\xi_{BA}, \xi_{AC}) = e_{BA} e_{AC}$$

$$e_{AB} := y_{AB} - \mathbf{x}_{ij}^T \widehat{\boldsymbol{\beta}}$$

Dyadic Clustering

$$\widehat{\Sigma}_{DC} =$$

$$n(n-1) \times n(n-1)$$

	BA	CA	DA	AB	CB	DB	AC	BC	DC	AD	BD	CD
BA	$e_{BA}e_{BA}$	$e_{CA}e_{BA}$	$e_{DA}e_{BA}$	$e_{AB}e_{BA}$	$e_{CB}e_{BA}$	$e_{DB}e_{BA}$	$e_{AC}e_{BA}$	$e_{BC}e_{BA}$		$e_{AD}e_{BA}$	$e_{BD}e_{BA}$	
CA	$e_{BA}e_{CA}$	$e_{CA}e_{CA}$	$e_{DA}e_{CA}$	$e_{AB}e_{CA}$	$e_{CB}e_{CA}$		$e_{AC}e_{CA}$	$e_{BC}e_{CA}$	$e_{DC}e_{CA}$	$e_{AD}e_{CA}$		$e_{CD}e_{CD}$
DA	$e_{BA}e_{DA}$	$e_{CA}e_{DA}$	$e_{DA}e_{DA}$	$e_{AB}e_{DA}$		$e_{DB}e_{DA}$	$e_{AC}e_{DA}$		$e_{DC}e_{DA}$	$e_{AD}e_{DA}$	$e_{BD}e_{DA}$	$e_{CD}e_{DA}$
AB	$e_{BA}e_{AB}$	$e_{CA}e_{AB}$	$e_{DA}e_{AB}$	$e_{AB}e_{AB}$	$e_{CB}e_{AB}$	$e_{DB}e_{AB}$	$e_{AC}e_{AB}$	$e_{BC}e_{AB}$		$e_{AD}e_{AB}$	$e_{BD}e_{AB}$	
CB	$e_{BA}e_{CB}$	$e_{CA}e_{CB}$		$e_{AB}e_{CB}$	$e_{CB}e_{CB}$	$e_{DB}e_{CB}$	$e_{AC}e_{CB}$	$e_{BC}e_{CB}$	$e_{DC}e_{CB}$		$e_{BD}e_{CB}$	$e_{CD}e_{CB}$
DB	$e_{BA}e_{DB}$		$e_{DA}e_{DB}$	$e_{AB}e_{DB}$	$e_{CB}e_{DB}$	$e_{DB}e_{DB}$		$e_{BC}e_{DB}$	$e_{DC}e_{DB}$	$e_{AD}e_{DB}$	$e_{BD}e_{DB}$	$e_{CD}e_{DB}$
AC	$e_{BA}e_{AC}$	$e_{CA}e_{AC}$	$e_{DA}e_{AC}$	$e_{AB}e_{AC}$	$e_{CB}e_{AC}$		$e_{AC}e_{AC}$	$e_{BC}e_{AC}$	$e_{DC}e_{AC}$	$e_{AD}e_{AC}$		$e_{CD}e_{AC}$
BC	$e_{BA}e_{BC}$	$e_{CA}e_{BC}$		$e_{AB}e_{BC}$	$e_{CB}e_{BC}$	$e_{DB}e_{BC}$	$e_{AC}e_{BC}$	$e_{BC}e_{BC}$	$e_{DC}e_{BC}$		$e_{BD}e_{BC}$	$e_{CD}e_{BC}$
DC		$e_{CA}e_{DC}$	$e_{DA}e_{DC}$		$e_{CB}e_{DC}$	$e_{DB}e_{DC}$	$e_{AC}e_{DC}$	$e_{BC}e_{DC}$	$e_{DC}e_{DC}$	$e_{AD}e_{DC}$	$e_{BD}e_{DC}$	$e_{CD}e_{DC}$
AD	$e_{BA}e_{AD}$	$e_{CA}e_{AD}$	$e_{DA}e_{AD}$	$e_{AB}e_{AD}$		$e_{DB}e_{AD}$	$e_{AC}e_{AD}$		$e_{DC}e_{AD}$	$e_{AD}e_{AD}$	$e_{BD}e_{AD}$	$e_{CD}e_{AD}$
BD	$e_{BA}e_{BD}$		$e_{DA}e_{BD}$	$e_{AB}e_{BD}$	$e_{CB}e_{BD}$	$e_{DB}e_{BD}$		$e_{BC}e_{BD}$	$e_{DC}e_{BD}$	$e_{AD}e_{BD}$	$e_{BD}e_{BD}$	$e_{CD}e_{BD}$
CD		$e_{CA}e_{CD}$	$e_{DA}e_{CD}$		$e_{CB}e_{CD}$	$e_{DB}e_{CD}$	$e_{AC}e_{CD}$	$e_{BC}e_{CD}$	$e_{DC}e_{CD}$	$e_{AD}e_{CD}$	$e_{BD}e_{CD}$	$e_{CD}e_{CD}$

Dyadic Clustering

- **Issues:**
  - More estimates than data points, $O(n^3) > O(n^2)$
  - No sharing of information
  - Singular with probability 1
- Can we add a reasonable assumption to improve the estimate?

Exchangeability

- **Intuition**: Node labeling on errors uninformative
- $\boldsymbol{\xi}$ *jointly exchangeable* if, for any permutation $\pi(.)$,

$$\mathbb{P}\left(\{\xi_{ij} : i \neq j, 1 \leq i, j \leq n\}\right) = \mathbb{P}\left(\{\xi_{\pi(i)\pi(j)} : i \neq j, 1 \leq i, j \leq n\}\right)$$

(akin to homogenous variance assumption)

- Many network models are exchangeable: e.g. latent space, stochastic block, etc.

Exchangeability

$\pi(\{A, B, C, D\})$ = Swap *B* and *D*

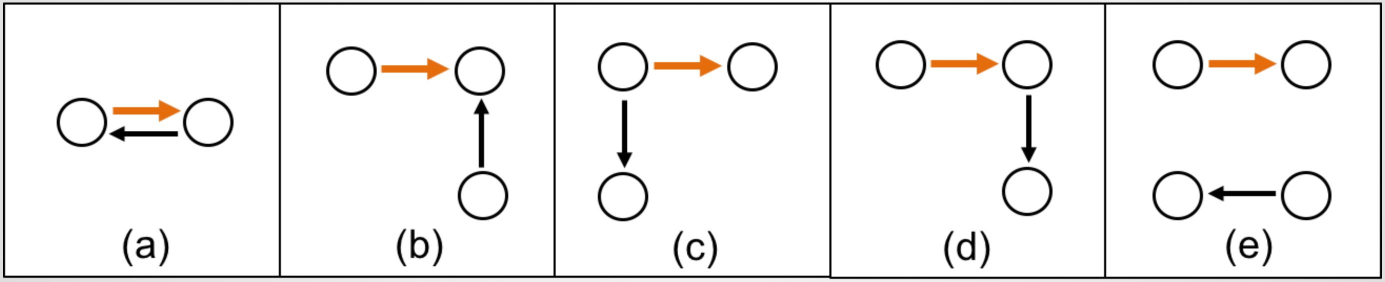
$$
\begin{array}{c|cccc}
 & A & B & C & D \\
\hline
A & & \xi_{AB} & \xi_{AC} & \xi_{AD} \\
B & \xi_{BA} & & \xi_{BC} & \xi_{BD} \\
C & \xi_{CA} & \xi_{CB} & & \xi_{CD} \\
D & \xi_{DA} & \xi_{DB} & \xi_{DC} &
\end{array}
\quad \overset{\mathbb{P}}{=} \quad
\begin{array}{c|cccc}
 & A & D & C & B \\
\hline
A & & \xi_{AD} & \xi_{AC} & \xi_{AB} \\
D & \xi_{DA} & & \xi_{DC} & \xi_{DB} \\
C & \xi_{CA} & \xi_{CD} & & \xi_{CB} \\
B & \xi_{BA} & \xi_{BD} & \xi_{BC} &
\end{array}
$$

Original $\boldsymbol{\xi}$ Permuted $\boldsymbol{\xi}$

Exchangeability

- **Major contribution**: Prove covariance matrix of jointly exchangeable vector $\boldsymbol{\xi}$ has 5 covariances and 1 variance

- Regardless of n



Exchangeability

	A	B	C	D
A		ξ_{AB}	ξ_{AC}	ξ_{AD}
B	ξ_{BA}		ξ_{BC}	ξ_{BD}
C	ξ_{CA}	ξ_{CB}		ξ_{CD}
D	ξ_{DA}	ξ_{DB}	ξ_{DC}	

	ξ_{BA}	ξ_{CA}	ξ_{DA}	ξ_{AB}	ξ_{CB}	ξ_{DB}	ξ_{AC}	ξ_{BC}	ξ_{DC}	ξ_{AD}	ξ_{BD}	ξ_{CD}
ξ_{BA}	σ^2	ϕ_b	ϕ_b	ϕ_a	ϕ_d	ϕ_d	ϕ_d	ϕ_c		ϕ_d	ϕ_c	
ξ_{CA}	ϕ_b	σ^2	ϕ_b	ϕ_d	ϕ_c		ϕ_a	ϕ_d	ϕ_d	ϕ_d		ϕ_c
ξ_{DA}	ϕ_b	ϕ_b	σ^2	ϕ_d		ϕ_c	ϕ_d		ϕ_c	ϕ_a	ϕ_d	ϕ_d
ξ_{AB}	ϕ_a	ϕ_d	ϕ_d	σ^2	ϕ_b	ϕ_b	ϕ_c	ϕ_d		ϕ_c	ϕ_d	
ξ_{CB}	ϕ_d	ϕ_c		ϕ_b	σ^2	ϕ_b	ϕ_d	ϕ_a	ϕ_d		ϕ_d	ϕ_c
ξ_{DB}	ϕ_d		ϕ_c	ϕ_b	ϕ_b	σ^2		ϕ_d	ϕ_c	ϕ_d	ϕ_a	ϕ_d
ξ_{AC}	ϕ_d	ϕ_a	ϕ_d	ϕ_c	ϕ_d		σ^2	ϕ_b	ϕ_b	ϕ_c		ϕ_d
ξ_{BC}	ϕ_c	ϕ_d		ϕ_d	ϕ_a	ϕ_d	ϕ_b	σ^2	ϕ_b		ϕ_c	ϕ_d
ξ_{DC}		ϕ_d	ϕ_c		ϕ_d	ϕ_c	ϕ_b	ϕ_b	σ^2	ϕ_d	ϕ_d	ϕ_a
ξ_{AD}	ϕ_d	ϕ_d	ϕ_a	ϕ_c		ϕ_d	ϕ_c		ϕ_d	σ^2	ϕ_b	ϕ_b
ξ_{BD}	ϕ_c		ϕ_d	ϕ_d	ϕ_d	ϕ_a		ϕ_c	ϕ_d	ϕ_b	σ^2	ϕ_b
ξ_{CD}		ϕ_c	ϕ_d		ϕ_c	ϕ_d	ϕ_d	ϕ_d	ϕ_a	ϕ_b	ϕ_b	σ^2

Exchangeable estimator

- Maintain independence assumption from DC

$$Cov(\xi_{ij}, \xi_{kl}) = 0 \text{ when } \{i, j\} \cap \{k, l\} = \varnothing$$

- Pool across all relations to estimate 5 nonzero terms in $\widehat{\Sigma}_E$
    - i.e. 1 variance and 4 covariances
- Estimate $\widehat{\sigma}^2$, $\widehat{\phi}_i$ with mean of products of OLS residuals
- Projection of $\widehat{\Sigma}_{DC}$ onto subspace of exchangeable covariance matrices

Exchangeable estimator

- Adds assumption of joint exchangeability of ξ to DC estimator
- Shares information: should see reduced variability
- Should see improved performance when assumption is reasonable
  - Covariates explain all variability except for exchangeable structure
  - Heterogeneities small relative to variability across 5 parameters
- Subsumes ALL exchangeable networks modeled with random effects, such as Latent Factor Model of Hoff (2005, 2007)
- Fast, direct estimation of covariance matrices

Latent Factor Model of Hoff (2005)

$$\xi_{ij} = a_i + b_j + \gamma_{(ij)} + z_i^T z_j + \epsilon_{ij}$$

- **Issues:**
  - Parametric model
  - Random effects model
  - May be slow to estimate

Simulation study

- Generate data for networks of size *n*
- Estimate coefficients using OLS
- Estimate standard errors with *exchangeable, dyadic clustering,* and *heteroskedasticity consistent* estimators

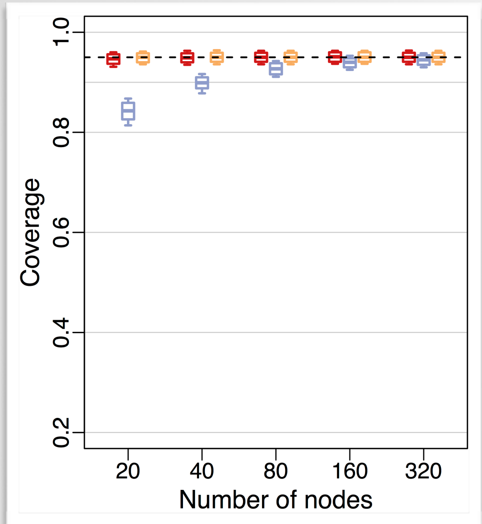
$$y_{ij} = \beta_1 + \beta_2 \mathbf{1}_i \mathbf{1}_j + \beta_3 |x_{3i} - x_{3j}| + \beta_4 x_{4ij} + \xi_{ij}$$

$$\mathbf{1}_i \sim_{iid} \mathrm{Bernoulli}(1/2)$$

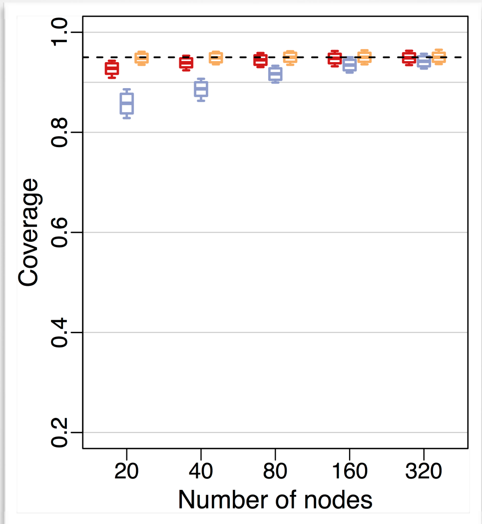
$$x_{3i}, x_{4ij} \sim_{iid} \mathrm{N}(0, 1)$$

IID Errors

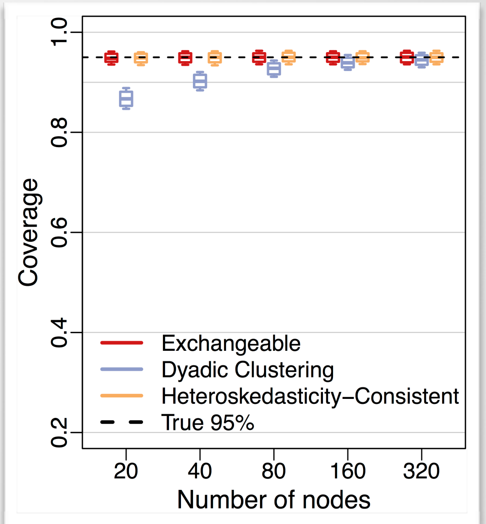
$\mathbf{1}_i\mathbf{1}_j$



$|x_{3i} - x_{3j}|$



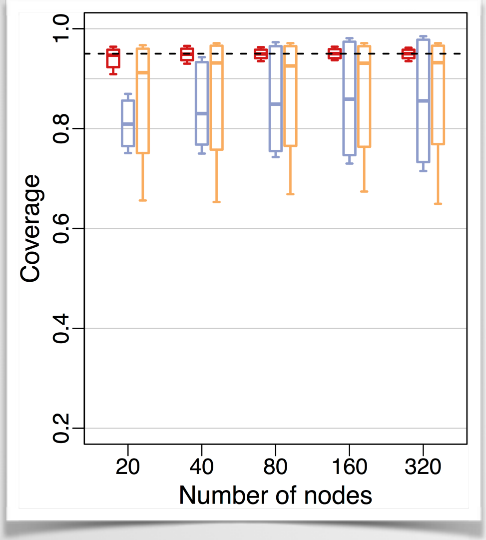
x_{4ij}



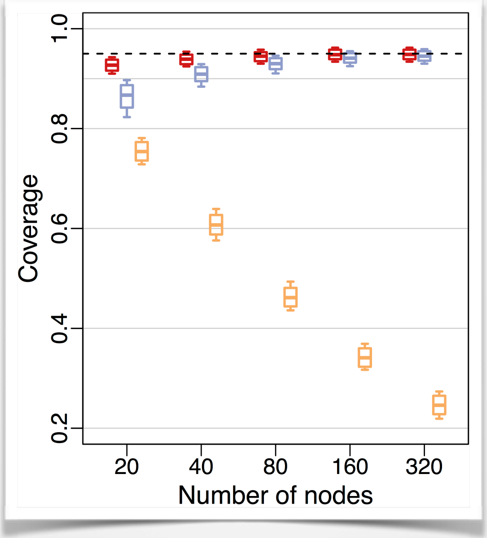
Probability true coefficient pertaining to each covariate is in 95% confidence interval

Exchangeable Errors

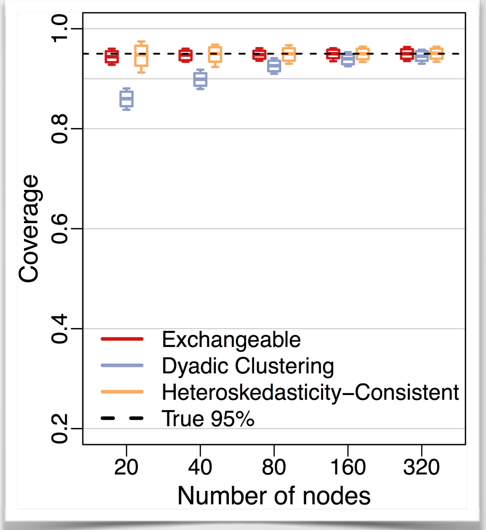
$\mathbf{1}_i\mathbf{1}_j$



$|x_{3i} - x_{3j}|$



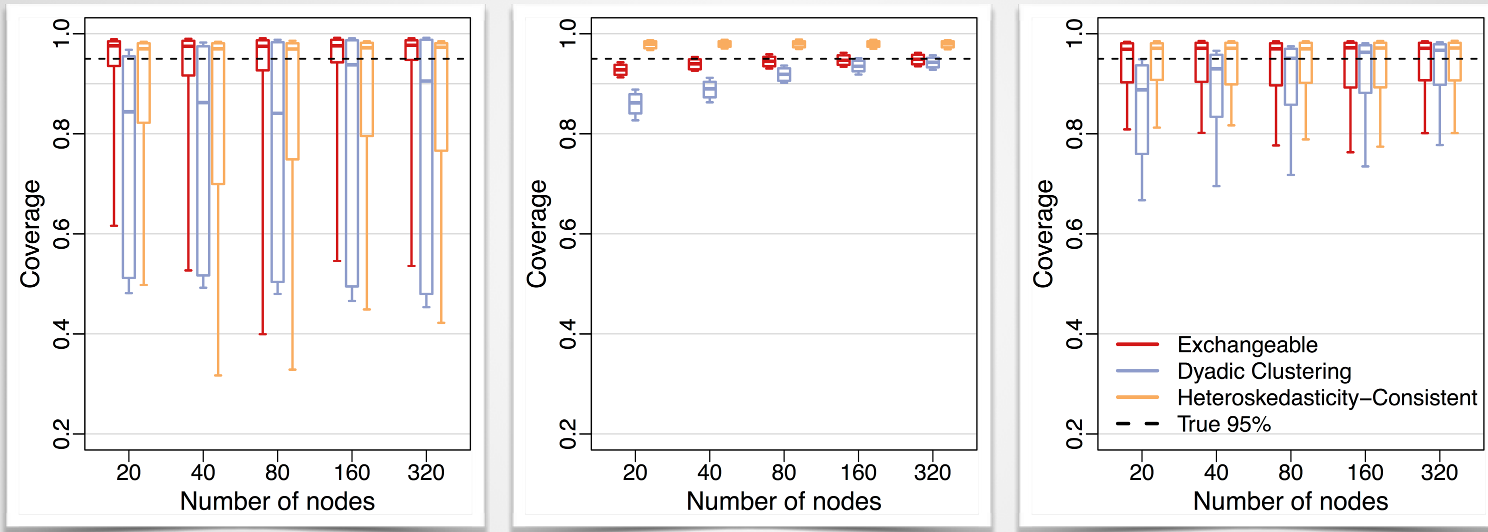
x_{4ij}



Probability true coefficient pertaining to each covariate is in 95% confidence interval

Nonexchangeable Errors

$\mathbf{1}_i\mathbf{1}_j$ $|x_{3i} - x_{3j}|$ x_{4ij}



Probability true coefficient pertaining to each covariate is in 95% confidence interval

Theoretical Results

- ***Theorem:*** OLS is consistent and asymptotically normal under exchangeable error structure.

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}(0, V_0)$$

$$V_0 = (\phi_b + \phi_c + 2\phi_d) E[x_{jk} x_{jk}^T]^{-1}$$

Theoretical Results

- ***Theorem:*** Exchangeable estimator is consistent.

$$\sqrt{n}(\widehat{V}_E - Var(\boldsymbol{\beta})) \xrightarrow{p} 0$$

$$\widehat{V}_E := (X^T X)^{-1} X^T \widehat{\Sigma}_E X (X^T X)^{-1}$$

Theoretical Results

- ***Theorem:*** The bias of the exchangeable estimator is less than that of the dyadic clustering estimator (for centered simple linear regression).

$$\frac{|\mathrm{Bias}(\widehat{V}_{DC})|}{|\mathrm{Bias}(\widehat{V}_{E})|} \geq 1$$

$$\widehat{V}_{DC} := (X^T X)^{-1} X^T \widehat{\Sigma}_{DC} X (X^T X)^{-1}$$

Theoretical Results

(Previously discussed)

- ***Theorem:*** Covariance matrix of exchangeable errors has at most 6 unique terms.
- ***Theorem:*** Dyadic Clustering estimate of covariance matrix of the errors is singular with probability 1.

(Still to prove)

- ***Conjecture:*** Exchangeable estimate of covariance matrix of errors is invertible with probability 1.
- ***Conjecture:*** Precision matrix of exchangeable errors has at most 6 unique terms *in the same pattern!* (leads to 6x6 inversion regardless of n)
- ***Conjecture:*** Exchangeable estimator is asymptotically efficient for normally-distributed exchangeable errors.

International trade example

- 25 countries over 20 years
- “gravity model” of trade
- 8 covariates
  - Nodal
  - Edge
- Fit with Iteratively-reweighted least squares/GEE
- Working covariance has 10 terms
  - 5 at same time, 5 at different times

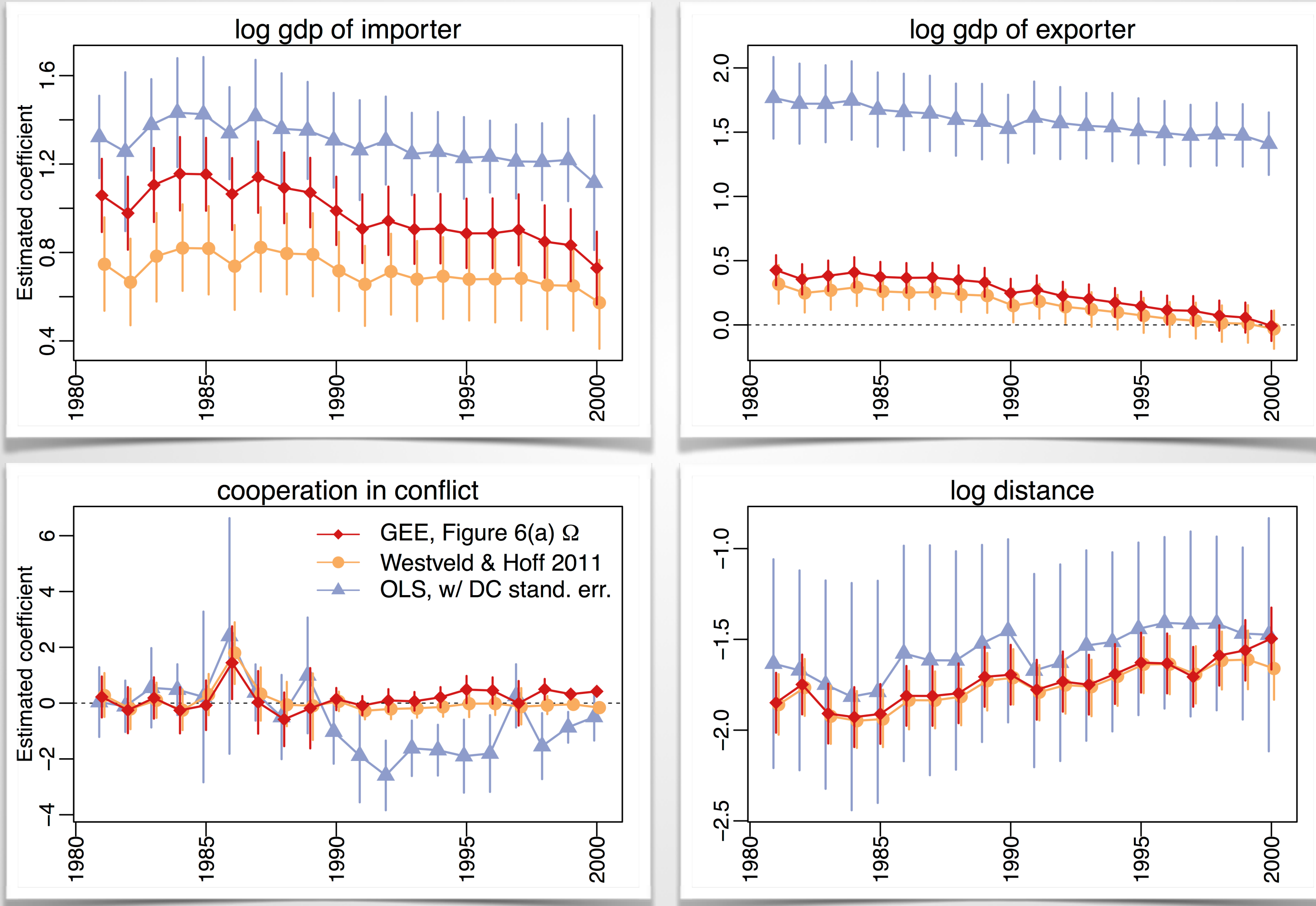
International trade example

- Westveld and Hoff fit Bayesian regression model
- Stationary working covariance model
  - Appx 30 parameters
  - Complex hierarchical model with many modeling decisions

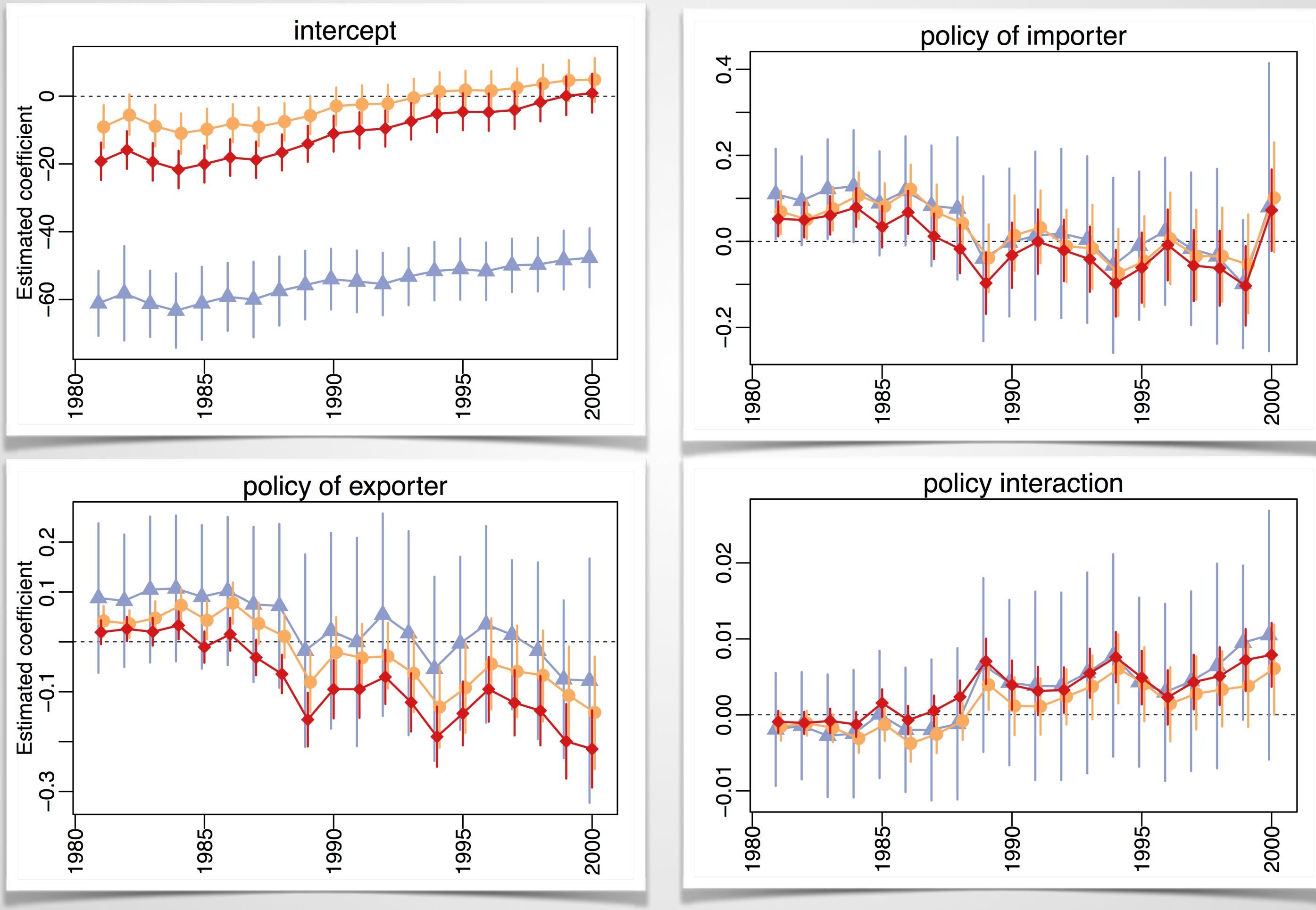
International trade example

	$t=1$	$t=2$	$t=3$	$t=4$
$t=1$	Ω_1	Ω_2	Ω_2	Ω_2
$t=2$	Ω_2	Ω_1	Ω_2	Ω_2
$t=3$	Ω_2	Ω_2	Ω_1	Ω_2
$t=4$	Ω_2	Ω_2	Ω_2	Ω_1

International trade example



International trade example



Summary

- Dyadic clustering approach may be noisy
- Exchangeable covariance matrix
  - 6 unique terms, one of which we assume is zero
- Many common network models are jointly exchangeable
  - i.e. Latent Factor Model of Hoff (2005)
- Estimates of $\mathrm{se}\{\widehat{\boldsymbol{\beta}}\}$ based on exchangeable error structure perform well
  - simulations and trade data

Future work

- Prove conjectures
- Extend approach to binary data
- Test for exchangeability
- Principled extensions to heterogeneous cases

Thank you!

Frank Marrs

Colorado State University

frank.marrs@colostate.edu

http://www.stat.colostate.edu/~marrs

Marrs, F.W., McCormick, T.H., and Fosdick, B.K. (2017) "Standard errors for regression on relational data with exchangeable errors", arXiv:1701.05530. [Preprint]

Simulation study

- Error structures:

    1. IID errors

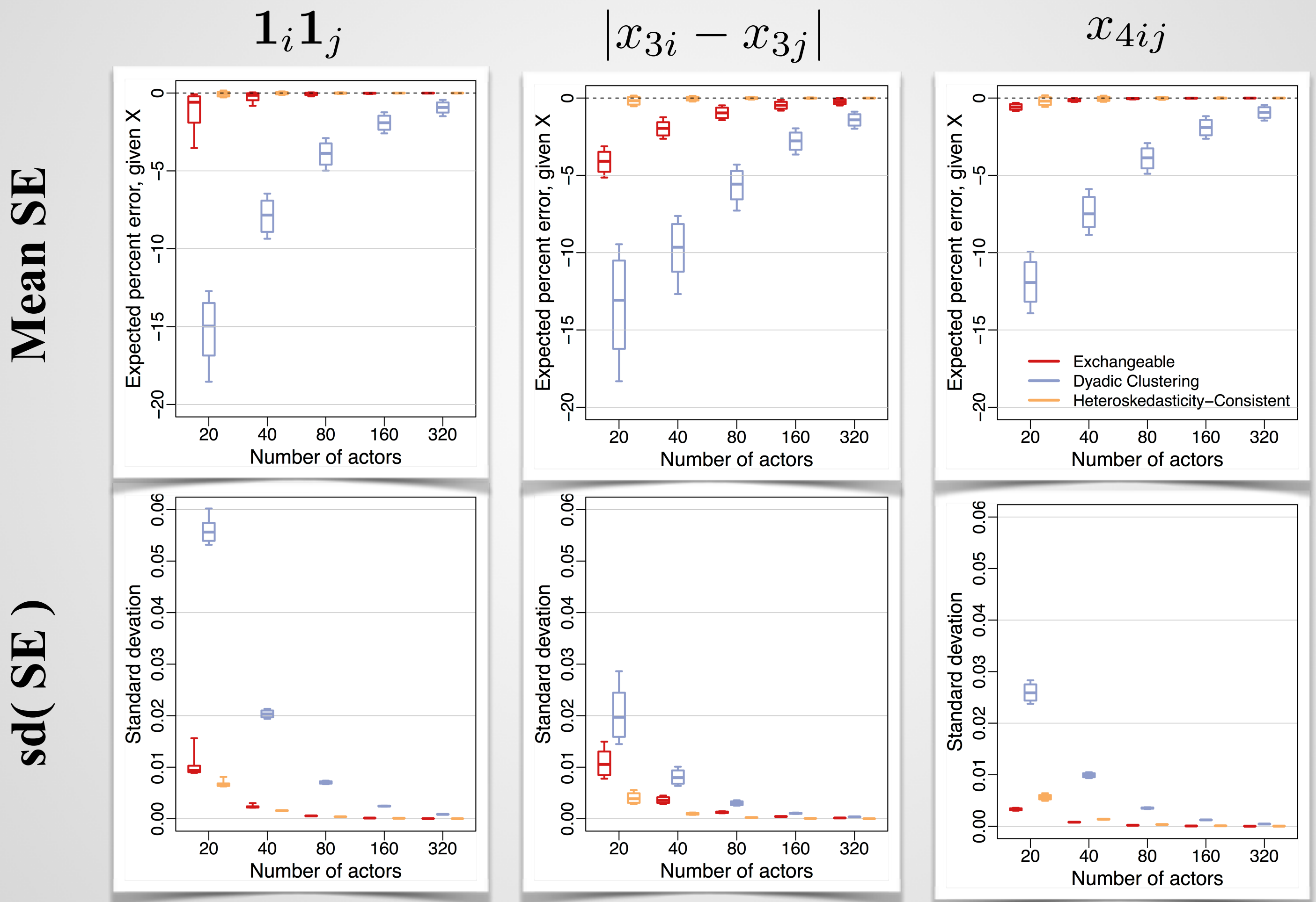
    2. Exchangeable errors (latent distance model of Hoff)

    3. Non-exchangeable error structure

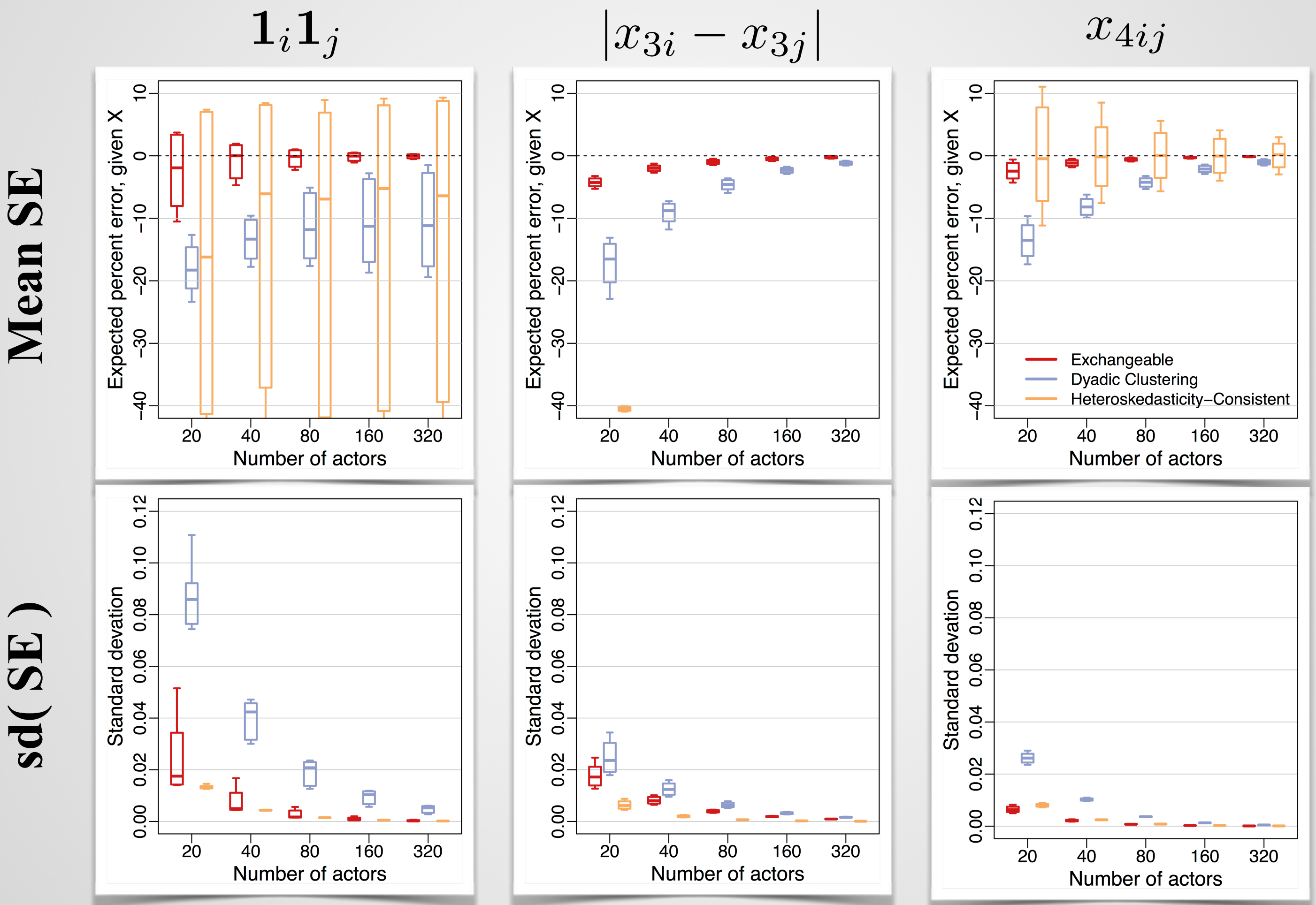
- 1,000 error draws for each of 500 X draws

- Fit OLS and estimate standard errors using DC, E estimators

Standard error comparison - IID errors



Standard error comparison - EXCH errors



Standard error comparison - Non-exch. errors