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

\[
y_{ij} = x_{ij}^T \beta + \xi_{ij}
\]

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

\[ y_{ij} = x_{ij}^T \beta + \xi_{ij} \]

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

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

\[
\begin{array}{cccc}
A & B & C & D \\
A & y_{AB} & y_{AC} & y_{AD} \\
B & y_{BA} & y_{BC} & y_{BD} \\
C & y_{CA} & y_{CB} & y_{CD} \\
D & y_{DA} & y_{DB} & y_{DC} \\
\end{array}
\]

\[
Y = \begin{bmatrix} y_{BA} \\ y_{CA} \\ \cdots \\ y_{CD} \end{bmatrix}, \quad X = \begin{bmatrix} x_{BA}^T \\ x_{CA}^T \\ \cdots \\ x_{CD}^T \end{bmatrix}
\]
Goal: inference about $\beta$

- point estimates ($\hat{\beta}$)
- uncertainty estimate ($\hat{\beta} \pm \hat{\text{se}}\{\hat{\beta}\}$)

$\xi_{ij}$ highly structured error

i.e. $\xi_{ij}$ and $\xi_{ik}$ share a node, expect correlation
Linear Regression

• Recall Ordinary Least Squares

\[ \hat{\beta} = \operatorname{argmin}_\beta \| Y - X\beta \|^2_2 = (X^T X)^{-1} X^T Y \]

\[ \operatorname{Var}(\hat{\beta}|X) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} \]

\[ \Sigma = \operatorname{Var}(Y|X) = \operatorname{Var}(\xi) \]

• $X$ is $(n^2 - n) \times p$ matrix of covariates

• $Y$ and $\xi$ are $(n^2 - n)$ vectors of relations and errors

• For inference on $\hat{\beta}$, need an estimate of $\Sigma$
Linear Regression

• Recall normal likelihood

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

\[
\Sigma = \text{Var}(Y|X) = \text{Var}(\xi)
\]

• \(X\) is \((n^2 - n) \times p\) matrix of covariates

• \(Y\) and \(\xi\) are \((n^2 - n)\) vectors of relations and errors

• For inference on \(\widehat{\beta}\), need a model for \(\Sigma\)
Dyadic Clustering

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

\[ \Sigma = \text{Var}(Y|X) = \text{Var}(\xi) \]

- Plug-in estimator

\[ \text{Var}(\hat{\beta}|X) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} \]
Dyadic Clustering

- Assumes that non-overlapping pairs independent

\[
\hat{\xi} = \begin{array}{cccc}
A & B & C & D \\
A & \xi_{AB} & \xi_{AC} & \xi_{AD} \\
B & \xi_{BA} & \xi_{BC} & \xi_{BD} \\
C & \xi_{CA} & \xi_{CB} & \xi_{CD} \\
D & \xi_{DA} & \xi_{DB} & \xi_{DC} \\
\end{array}
\]

\[
\text{Cov}(\xi_{BA}, \xi_{CD}) = 0
\]
Dyadic Clustering

- Model nonzero entries in $\sum$ products of OLS residuals

\[ \xi = \begin{array}{cccc}
A & B & C & D \\
A & \xi_{AB} & \xi_{AC} & \xi_{AD} \\
B & \xi_{BA} & \xi_{BC} & \xi_{BD} \\
C & \xi_{CA} & \xi_{CB} & \xi_{CD} \\
D & \xi_{DA} & \xi_{DB} & \xi_{DC} \\
\end{array} \]

\[ \hat{\text{Cov}}(\xi_{BA}, \xi_{AC}) = e_{BA}e_{AC} \]

\[ e_{AB} := y_{AB} - x_{ij}^T\hat{\beta} \]
Dyadic Clustering

\[ \sum_{DC} = n(n - 1) \times n(n - 1) \]
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

- **$\xi$ jointly exchangeable** if, for any permutation $\pi(.)$,

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

  (akin to homogenous variance assumption)

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

\[ \pi(\{A, B, C, D\}) = \text{Swap } B \text{ and } D \]
Exchangeability

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

- Regardless of $n$
Exchangeability

<table>
<thead>
<tr>
<th>ξ_{BA}</th>
<th>ξ_{CA}</th>
<th>ξ_{DA}</th>
<th>ξ_{AB}</th>
<th>ξ_{CB}</th>
<th>ξ_{DB}</th>
<th>ξ_{AC}</th>
<th>ξ_{BC}</th>
<th>ξ_{AB}</th>
<th>ξ_{CD}</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ^2</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_a</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_c</td>
</tr>
<tr>
<td>φ_b</td>
<td>σ^2</td>
<td>φ_b</td>
<td>φ_d</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_c</td>
</tr>
<tr>
<td>φ_b</td>
<td>φ_b</td>
<td>σ^2</td>
<td>φ_d</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_c</td>
</tr>
<tr>
<td>φ_a</td>
<td>φ_d</td>
<td>φ_d</td>
<td>σ^2</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_c</td>
<td>φ_d</td>
</tr>
<tr>
<td>φ_d</td>
<td>φ_c</td>
<td>φ_b</td>
<td>φ_b</td>
<td>σ^2</td>
<td>φ_d</td>
<td>φ_a</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_c</td>
</tr>
<tr>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_c</td>
<td>φ_d</td>
<td>σ^2</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_c</td>
<td>φ_d</td>
</tr>
<tr>
<td>φ_d</td>
<td>φ_a</td>
<td>φ_d</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_b</td>
<td>σ^2</td>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_b</td>
</tr>
<tr>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_b</td>
<td>σ^2</td>
<td>φ_d</td>
<td>φ_b</td>
</tr>
<tr>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_a</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_d</td>
<td>σ^2</td>
<td>φ_b</td>
</tr>
<tr>
<td>φ_d</td>
<td>φ_a</td>
<td>φ_a</td>
<td>φ_b</td>
<td>φ_d</td>
<td>φ_c</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_b</td>
</tr>
<tr>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_d</td>
<td>φ_d</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_b</td>
<td>φ_b</td>
<td>σ^2</td>
</tr>
</tbody>
</table>
Exchangeable estimator

- Maintain independence assumption from DC

\[ \text{Cov}(\xi_{ij}, \xi_{kl}) = 0 \text{ when } \{i, j\} \cap \{k, l\} = \emptyset \]

- Pool across all relations to estimate 5 nonzero terms in \( \hat{\Sigma}_E \)

- i.e. 1 variance and 4 covariances

- Estimate \( \hat{\sigma}^2, \hat{\phi}_i \) with mean of products of OLS residuals

- Projection of \( \hat{\Sigma}_{DC} \) onto subspace of exchangeable covariance matrices
Exchangeable estimator

- Adds assumption of joint exchangeability of $\xi$ 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)

\[
\xi_{ij} = a_i + b_j + \gamma(i, j) + z_i^T z_j + \epsilon_{ij}
\]

• **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

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

\[ 1_i \sim iid \text{ Bernoulli}(1/2) \]

\[ x_{3i}, x_{4ij} \sim iid \text{ N}(0, 1) \]
IID Errors

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

\[ 1_i 1_j \quad |x_{3i} - x_{3j}| \quad x_{4ij} \]

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

\[ 1_i 1_j \quad \left| x_{3i} - x_{3j} \right| \quad x_{4ij} \]

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.

\[ \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_0) \]

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

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

\[
\sqrt{n} (\hat{V}_E - \text{Var}(\beta)) \xrightarrow{p} 0
\]

\[
\hat{V}_E := (X^T X)^{-1} X^T \hat{\Sigma} E X (X^T X)^{-1}
\]
Theoretical Results

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

\[
\frac{|\text{Bias}(\hat{V}_{DC})|}{|\text{Bias}(\hat{V}_E)|} \geq 1
\]

\[
\hat{V}_{DC} := (X^TX)^{-1} X^T \hat{\Sigma}_{DC} X (X^TX)^{-1}
\]
(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

\[
\begin{array}{cccc}
  t = 1 & t = 2 & t = 3 & t = 4 \\
  \Omega_1 & \Omega_2 & \Omega_2 & \Omega_2 \\
  \Omega_2 & \Omega_1 & \Omega_2 & \Omega_2 \\
  \Omega_2 & \Omega_2 & \Omega_1 & \Omega_2 \\
  \Omega_2 & \Omega_2 & \Omega_2 & \Omega_1 \\
\end{array}
\]
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 $\text{se}\{\hat{\beta}\}$ 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

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

$1_i 1_j$  

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

$x_{4ij}$

### Mean SE

- **Expected percent error, given $X$**
  - Number of actors: 20, 40, 80, 160, 320

### sd( SE )

- **Standard deviation**
  - Number of actors: 20, 40, 80, 160, 320
Standard error comparison - EXCH errors

\[ 1_i 1_j \]

\[ |x_{3i} - x_{3j}| \]

\[ x_{4ij} \]

**Mean SE**

- Expected percent error, given X
  - Number of actors: 20, 40, 80, 160, 320
  - Graphs show comparison across different numbers of actors.

**sd( SE )**

- Standard deviation
  - Number of actors: 20, 40, 80, 160, 320
  - Graphs illustrate standard deviation with different colors indicating different clustering methods.

- Exchangeable
- Dyadic Clustering
- Heteroskedasticity−Consistent
Standard error comparison - Non-exch. errors

$1_i 1_j$

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

$x_{4ij}$

Mean SE

sd( SE )