Mixed Effects Models for Dyadic Data

Peter Hoff [a]

Departments of Statistics and Biostatistics

and the Center for Statistics and the Social Sciences

University of Washington

11-20-03

[a]This research was supported by Office of Naval Research grant N00014-02-1-1011.

Dyadic data

Data is measured on *pairs* of units

- Actors, units, or nodes: $\{1, \ldots, n\}$
- Relations, ties, or edges: $y_{i,j}$ = value of the (directed) relation from i to j.

Data is in the form of an $n \times n$ "sociomatrix" Y, with i, jth entry $y_{i,j}$.

Examples:

- war, trade, political organization;
- communication networks;
- evaluation of friendship, liking, respect;
- transfer of information or disease;
- affiliation or membership to a group;
- physical connection;
- biological/familial relations such as kinship or descent.

Statistical models

$$\Pr(Y|X,\beta)$$

- $Y = \{y_{i,j}\}_{i,j=1}^n$ is the sociomatrix;
- $X = \{x_{i,j}\}_{i,j=1}^n$ is the set of explanatory variables;
- β is a set of unknown parameters, to be estimated.

A simple example: Hansell's (1984) classroom data

- $y_{i,j} = 1$ if i likes j "a lot", 0 else;
- $x_{i,j}$ = categorical variable indicating sex of sender and receiver;
- $\beta = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$ measures the rates of ties.

A naive model : Ordinary logistic regression

$$\Pr(y_{i,j} = 1|x_{i,j}, \beta) = \frac{e^{\beta' x_{i,j}}}{1 + e^{\beta' x_{i,j}}}, \quad y_{i,j}\text{'s are independent given } \beta, x_{i,j}\text{'s.}$$

Classroom data

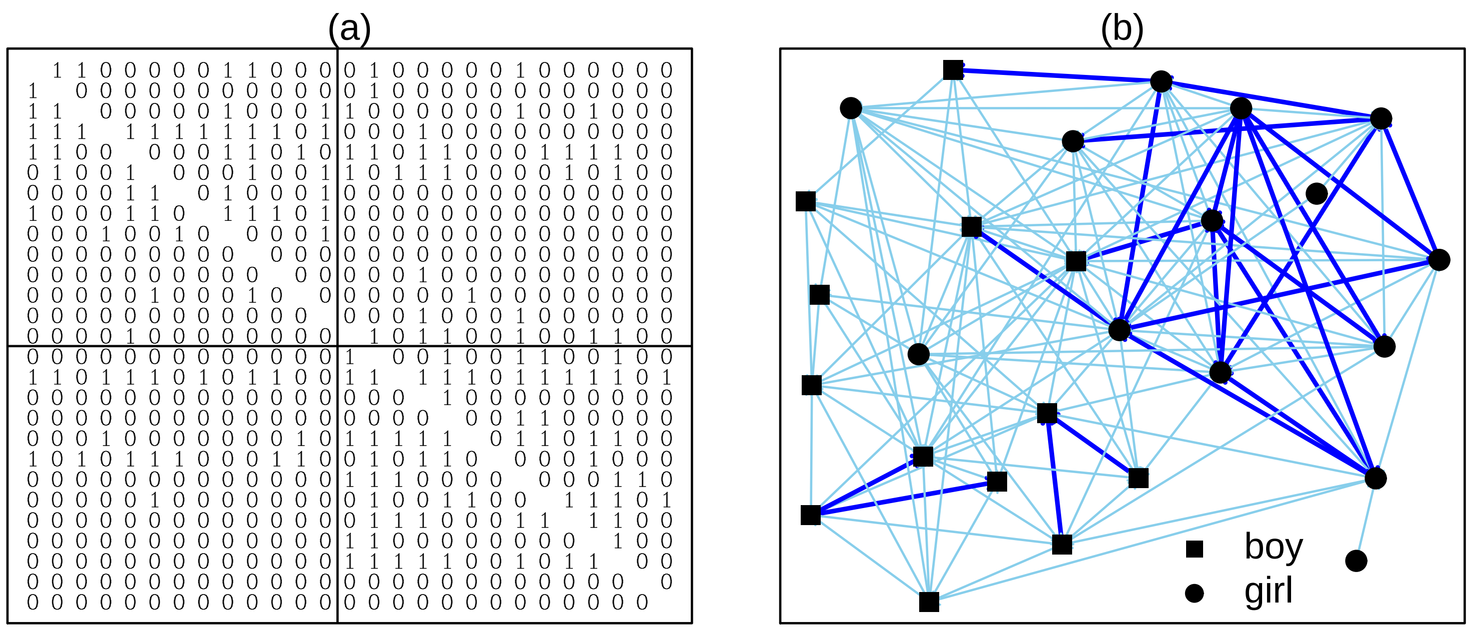


Figure 1: (a) Sociomatrix for friendship data: Rows and columns 1-13 are boys, 14-27 are girls; (b) Graphical representation of friendship data: Dark blue lines are reciprocated ties.

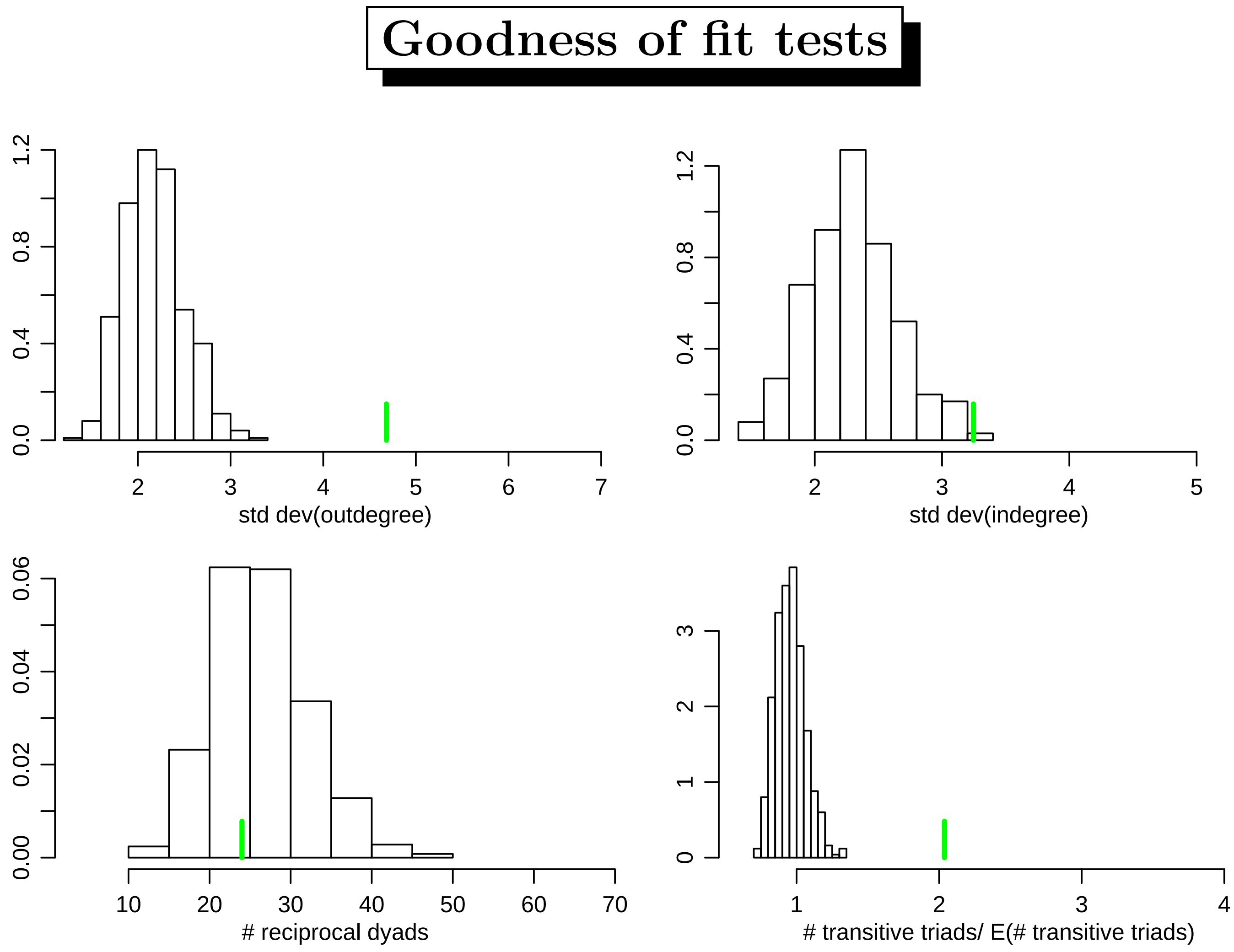


Figure 2: Posterior predictive distributions for four network statistics under the independence model.

Exchangeability, dependence, and random effects

Generalized additive model: $p(y_{i,j}|\beta, x_{i,j}, \xi_{i,j}) = f(\ \beta' x_{i,j} + \xi_{i,j}\ )$

Error structure: $\{\xi_{i,j} : i \neq j\} \stackrel{L}{=} \{\xi_{\pi(i),\pi(j)} : i \neq j\}$

	1	2	3	4	5	⋯
1		−1.06	−0.33	−0.28	−0.98	⋯
2	1.32		0.57	−0.91	0.47	⋯
3	−0.80	1.74		−1.90	−0.60	⋯
4	1.57	−0.05	−1.08		0.21	⋯
5	0.48	2.91	0.39	0.90		⋯
⋮	⋮	⋮	⋮	⋮	⋮	

$= \Pr$

	1	2	3	4	5	⋯
1		−1.06	−0.98	−0.28	−0.33	⋯
2	1.32		0.47	−0.91	0.57	⋯
3	0.48	2.91		0.90	0.39	⋯
4	1.57	−0.05	0.21		−1.08	⋯
5	−0.80	1.74	−0.60	−1.90		⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮

This is "weak row-and-column exchangeability of an array", and implies

$$\xi_{i,j} \stackrel{L}{=} f(\mu, a_i, b_j, c_{i,j})$$

where μ, a's, b's and c's are independent (Aldous 1985).

Orders of dyadic dependence

(a) Within-node dependence:

- $\text{Cov}(y_{i,j}, y_{i,k})$: i as a sender
- $\text{Cov}(y_{i,j}, y_{k,j})$: i as a receiver

(b) Reciprocity: $\text{Cov}(y_{i,j}, y_{j,i})$

(a) and (b) can be modeled (on some scale) using standard random-effects models

$$\begin{aligned} f(E[y_{i,j}]) &= \beta' x_{i,j} + a_i + b_j + \epsilon_{i,j} \\ (a_i, b_i)' &\sim \text{Multivariate Normal}(0, \Sigma_{a,b}) \\ (\epsilon_{i,j}, \epsilon_{j,i})' &\sim \text{Multivariate Normal}(0, \Sigma_\epsilon), \quad \Sigma_\epsilon = \begin{pmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon^2 \\ \rho\sigma_\epsilon^2 & \sigma_\epsilon^2 \end{pmatrix} \end{aligned}$$

- genetic cross data: Yates[1947], Cockerham and Weir [1977].
- social relations: Warner, Kenny and Stoto [1979], Wong [1982], Snijders and Kenny [2003].
- invariant normal model: Andersson and Madsen [1998], Gill and Schwarz [2001], Li [2002], Li and Loken [2002].
- similar to p_1, p_2 models for binary data: Holland and Leinhardt [1981], Lanzega and van Duijn [1997].

(c) Third order dependence: How might $y_{i,k}$ depend on $y_{i,j}$ and $y_{j,k}$ jointly?

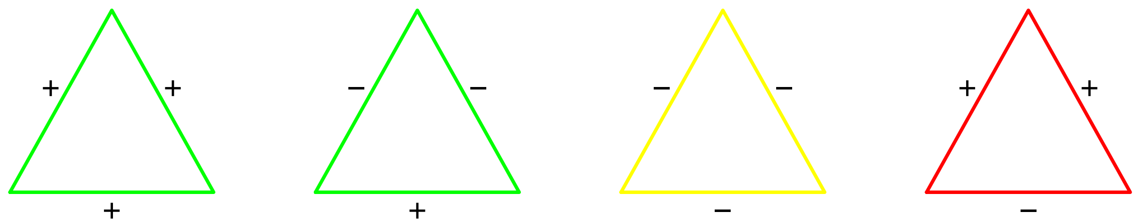
Dyadic dependence continued

For signed responses (or errors) $\gamma_{i,j}$,

Balance: a triad i, j, k is *balanced* if $\gamma_{i,j} \times \gamma_{j,k} \times \gamma_{k,i} > 0$.

Clusterability: a triad i, j, k is *clusterable* if there is not exactly one negative response among $\gamma_{i,j}, \gamma_{j,k}, \gamma_{k,i}$.

These are patterns often seen in residuals.



Random effects model: $y_{i,j} = \beta' x_{i,j} + a_i + b_j + \epsilon_{i,j} + \gamma_{i,j}$

Goal: a (relatively) low-dimensional effect $\gamma_{i,j}$ that captures balance/clusterability/transitivity, perhaps where

$$E(\gamma_{i,j} \times \gamma_{j,k} \times \gamma_{k,i}) > 0$$

Latent variable models

Let $z_1, \ldots, z_n$ be unobserved "latent characteristic vectors" of the nodes, and presume nodes have a larger response to others with similar latent characteristics.

Idea: Let $\gamma_{i,j} = f(z_i, z_j)$, where f is some measure of similarity.

- "Distance" model: $\gamma_{i,j} = -|z_i - z_j|$ (Hoff, Raftery, Handcock 2002).
- "Bilinear effects" model: $\gamma_{i,j} = z_i' z_j$

The bilinear, inner product effects can measure balance and clusterability:

- If the z_i's are scaled to have unit length, then the $\gamma_{i,j}$'s satisfy

$$|\gamma_{i,k} - \gamma_{j,k}| \leq \sqrt{2(1 - \gamma_{i,j})}.$$

- If the z_i's are vectors of independent normal $(0, \sigma_z^2)$ random variables then

$$E(\gamma_{i,j} \times \gamma_{j,k} \times \gamma_{k,i}) = k\sigma_z^6.$$

Such inner product effects have been used Gabriel (1978, 1998) as low-dimensional interactions in factorial models, and are sometimes called bilinear effects or multiplicative interactions.

Mixed Effects Models

Bilinear mixed effects model:

$$
\begin{aligned}
y_{i,j} &= \beta' x_{i,j} + a_i + b_j + z_i' z_j + \epsilon_{i,j} \\
(a_i, b_i)' &\sim \text{multivariate normal}(0, \Sigma_{a,b}), \quad \Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \\
(\epsilon_{i,j}, \epsilon_{j,i})' &\sim \text{multivariate normal}(0, \Sigma_\epsilon), \quad \Sigma_\epsilon = \begin{pmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon^2 \\ \rho\sigma_\epsilon^2 & \sigma_\epsilon^2 \end{pmatrix} \\
z_1, \ldots, z_n &\sim \text{i.i.d. multivariate normal}(0, \sigma_z^2 I) \ .
\end{aligned}
$$

Generalized bilinear mixed effects model:

$$
\begin{aligned}
\theta_{i,j} &= \beta' x_{i,j} + a_i + b_j + z_i' z_j + \epsilon_{i,j} \\
\Pr(Y|\theta) &= \prod_{i \neq j} p(y_{i,j}|\theta_{i,j})
\end{aligned}
$$

In Poisson regression, for example, we model $y_{i,j}|\theta_{i,j}$ as Poisson($e^{\theta_{i,j}}$).

Goal: make inference on β, the variance components Σ_{ab}, Σ_ϵ, and σ_z^2 and possibly on the unit-level effects (a_i, b_i, z_i).

Properties of Moments

If we think of the z_i's as random effects the resulting distribution for the $\gamma_{i,j}$'s has the following moment properties:

- $\mathrm{E}(\gamma_{i,j}) = 0$;
- $\mathrm{E}(\gamma_{i,j}^2) = \text{trace } \Sigma_z^2$;
- $\mathrm{E}(\gamma_{i,j}\gamma_{j,k}\gamma_{k,i}) = \text{trace } \Sigma_z^3$;

with all other second and third order moments equal to zero. Letting $\Sigma_z = \sigma_z^2 I$, this gives the following nonzero second and third order moments for the bilinear random-effects components $\xi_{i,j} = a_i + b_j + \epsilon_{i,j} + z_i' z_j$.

$$
\begin{aligned}
E(\xi_{i,j}^2) &= \sigma_a^2 + 2\sigma_{ab} + \sigma_b^2 + \sigma_\gamma^2 + k\sigma_z^4 & E(\xi_{i,j}\xi_{i,k}) &= \sigma_a^2 \\
E(\xi_{i,j}\xi_{j,i}) &= \rho\sigma_\gamma^2 + 2\sigma_{ab} + k\sigma_z^4 & E(\xi_{i,j}\xi_{k,j}) &= \sigma_b^2 \\
E(\xi_{i,j}\xi_{j,k}\xi_{k,i}) &= k\sigma_z^6 & E(\xi_{i,j}\xi_{k,i}) &= \sigma_{ab}.
\end{aligned}
$$

Goodness of fit continued

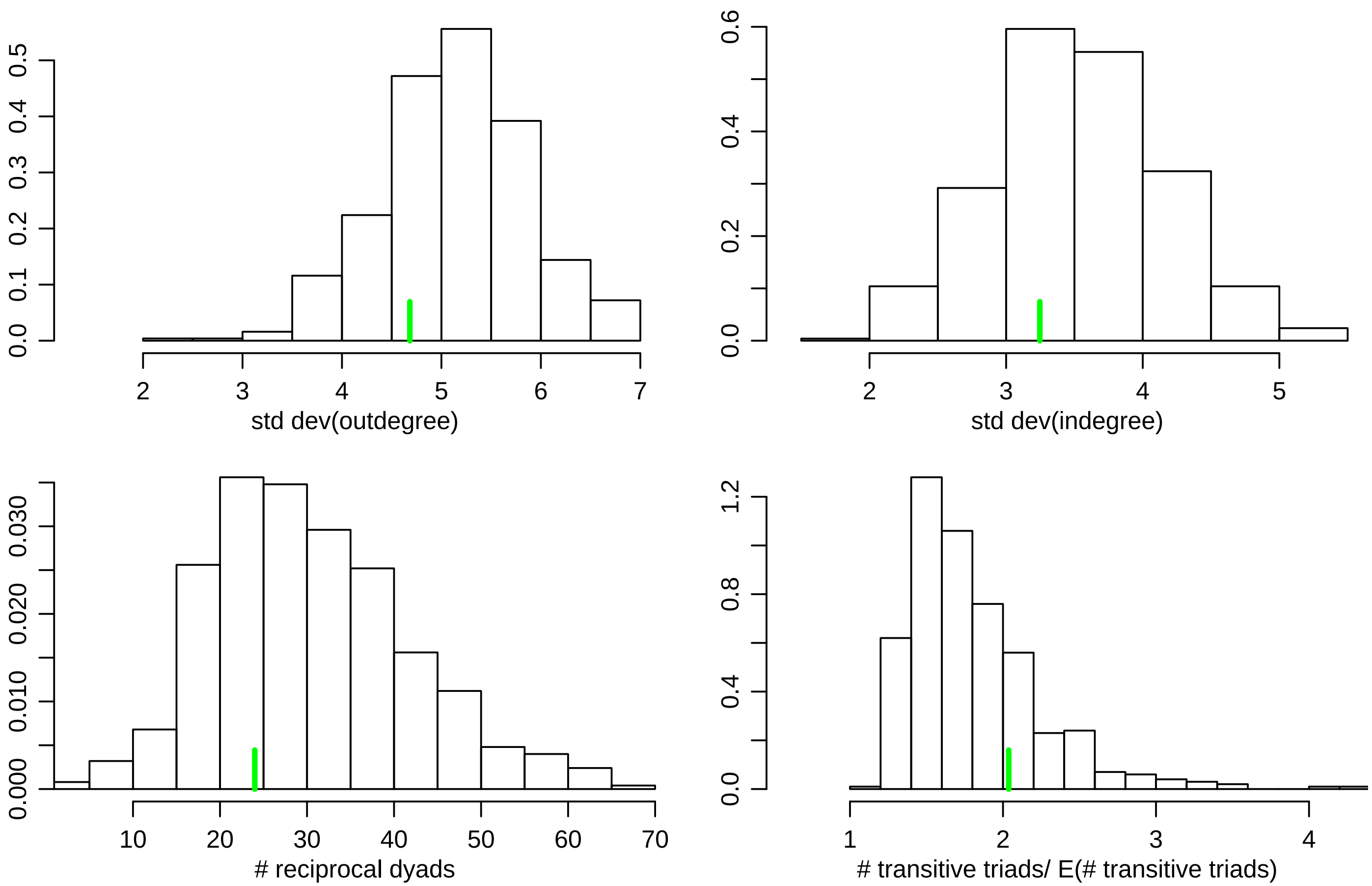


Figure 3: Posterior predictive distributions for four network statistics under the generalized bilinear mixed-effects model.

Bilinear effects example: central Asian international relations

Data: wire reports involving central Asian affairs (Reuters Business Briefing Service on Afghanistan, Armenia, Azerbaijan, and the former Soviet Republics of Central Asia, extracted from the Kansas Event Data Project (`http://www.ku.edu/~keds/project.html`)

Relation: $y_{i,j}$ = total number of "substantial positive actions" taken by country i with target j from 1992 to 1999.

Model: $y_{i,j} \sim \text{Poisson}(e^{\theta_{i,j}})$, where

$$\begin{aligned}
\theta_{i,j} &= \beta_0 + \beta_1 x_{i,j} + \beta_2 x_i + \beta_3 x_j + (a_i + b_j + \epsilon_{i,j} + z_i' z_j) \\
(a_1, b_1), \ldots, (a_n, b_n) &\sim \text{iid MVN}(0, \Sigma_{ab}) \\
(\epsilon_{1,2}, \epsilon_{2,1}), \ldots, (\epsilon_{n-1,n}, \epsilon_{n,n-1}) &\sim \text{iid MVN}(0, \Sigma_\epsilon) \\
z_1, \ldots, z_n &\sim \text{iid MVN}(0, \sigma_z^2 I_{k \times k})
\end{aligned}$$

- $x_{i,j}$ = distance between i, j;
- x_i = log population of country i.

Exploratory plots

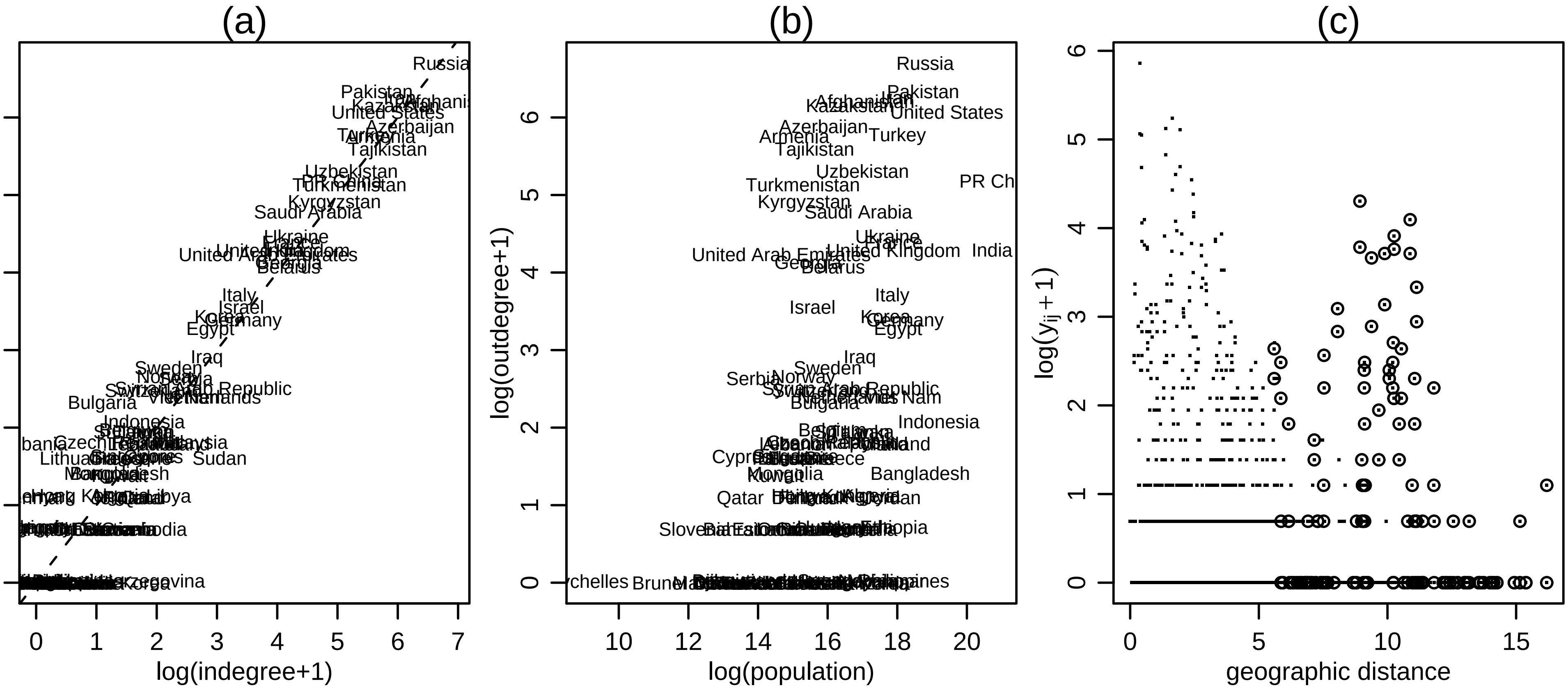


Figure 4: Relationship between (a) outdegree and indegree; (b) outdegree and population; (c) response and geographic distance

Markov chain Monte Carlo

1. Sample $\beta, a, b, \Sigma_{ab}, \Sigma_\epsilon | \theta, z$ (essentially a regression problem).

$$(\theta_{i,j} - z_i' z_j) = \beta' x_{i,j} + a_i + b_j + \epsilon_{i,j}$$

2. For $i = 1, \ldots, n$: sample $z_i | z_{-i}, \theta, \beta, a, b, \Sigma_\epsilon, \sigma_z^2$. This is also a regression problem: Letting $e_{i,j} = \theta_{i,j} - (\beta' x_{i,j} + a_i + b_j)$, we have

$$\begin{aligned} e_{i,1} &= z_i' z_1 + \epsilon_{i,1} \\ e_{i,2} &= z_i' z_2 + \epsilon_{i,2} \\ &\vdots \\ e_{i,n} &= z_i' z_n + \epsilon_{i,n}. \end{aligned}$$

   Then sample σ_z^2 from its full conditional.

3. Update $\theta_{i,j}, \theta_{j,i}$ using a Metropolis-Hastings step:

   (a) Propose $\begin{pmatrix} \theta^*_{i,j} \\ \theta^*_{j,i} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \beta' x_{i,j} + a_i + b_j + z_i' z_j \\ \beta' x_{j,i} + a_j + b_i + z_j' z_i \end{pmatrix}, \Sigma_\epsilon \right)$;

   (b) Accept with probability $\frac{p(y_{i,j}|\theta^*_{i,j}) p(y_{j,i}|\theta^*_{j,i})}{p(y_{i,j}|\theta_{i,j}) p(y_{j,i}|\theta_{j,i})} \wedge 1$.

Selection of k

k	LLP(k)	$\sum_{i<j} \log p(y_{i,j}, y_{j,i} \| \hat{\beta}, \hat{a}, \hat{b}, \hat{Z}, \hat{\Sigma}_\epsilon)$	AIC	$\hat{\sigma}^2_\epsilon$
0	-3558.78	-2432.67	-2638.67	2.38
1	-3351.76	-2317.47	-2623.47	1.66
2	-3078.79	-2214.68	-2620.68	1.23
3	-3076.73	-2127.26	-2633.26	0.87
4	-3077.30	-2038.95	-2644.95	0.54

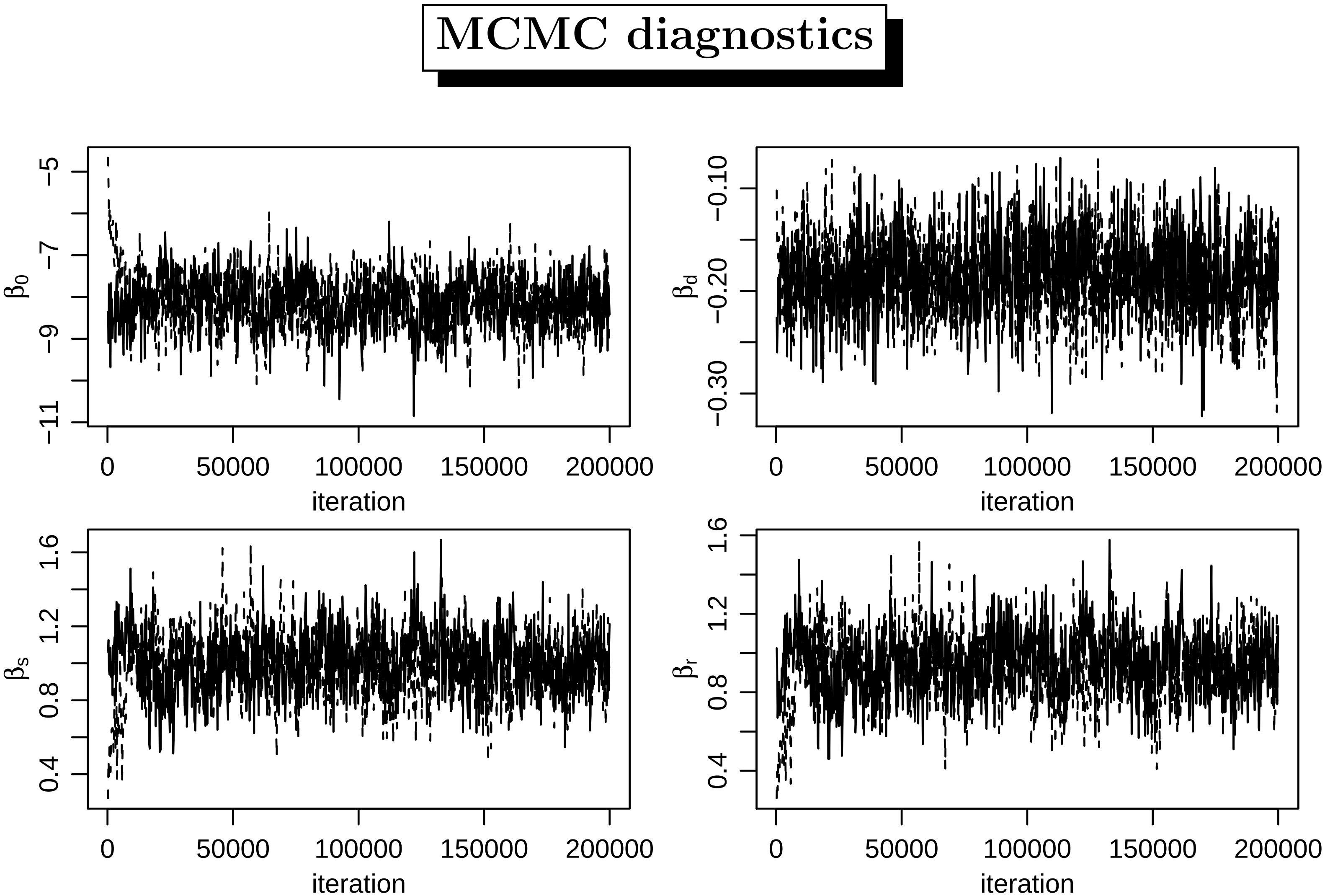


Figure 5: Marginal MCMC output for regression coefficients

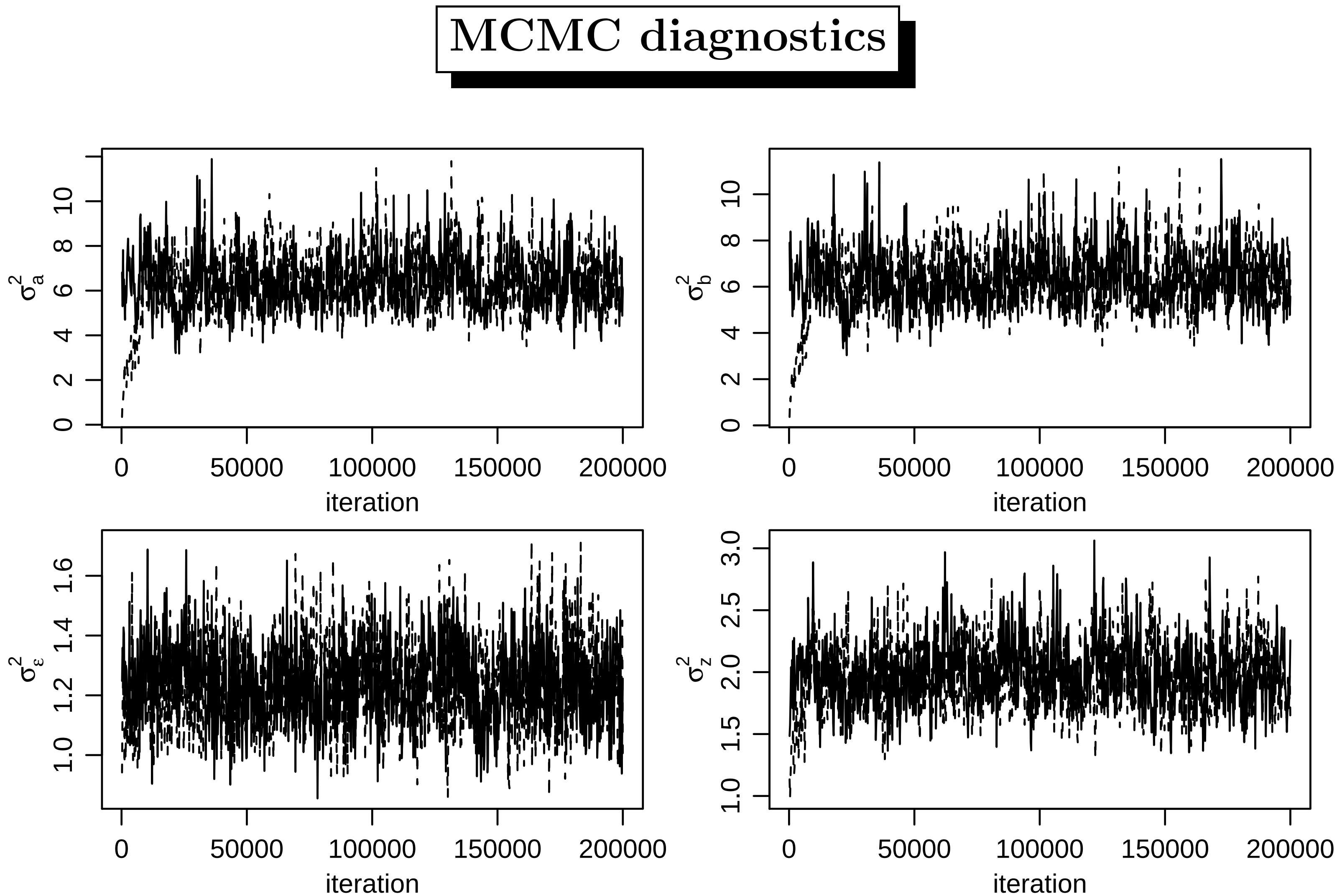


Figure 6: Marginal MCMC output for variance component parameters

Posterior densities

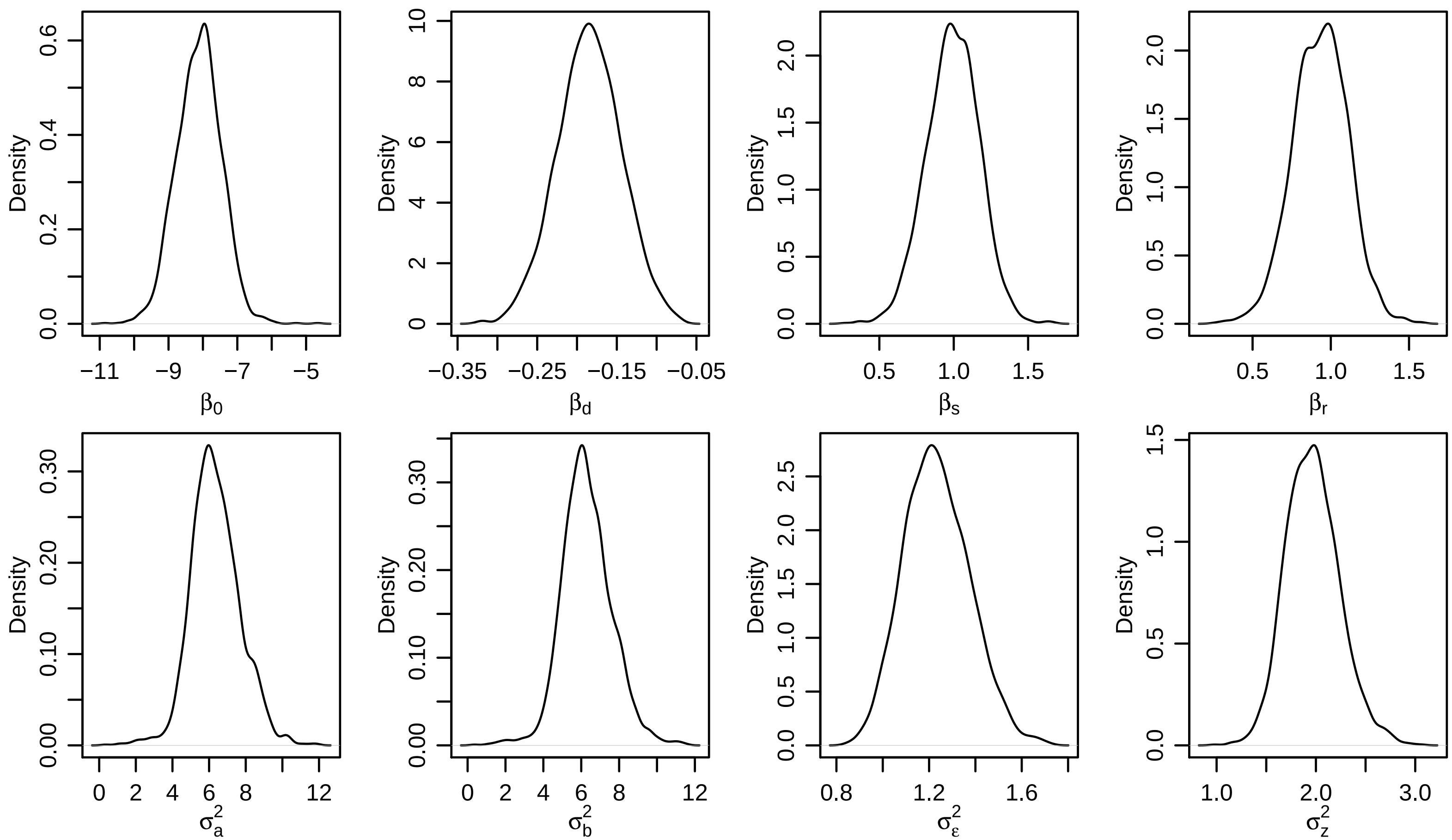
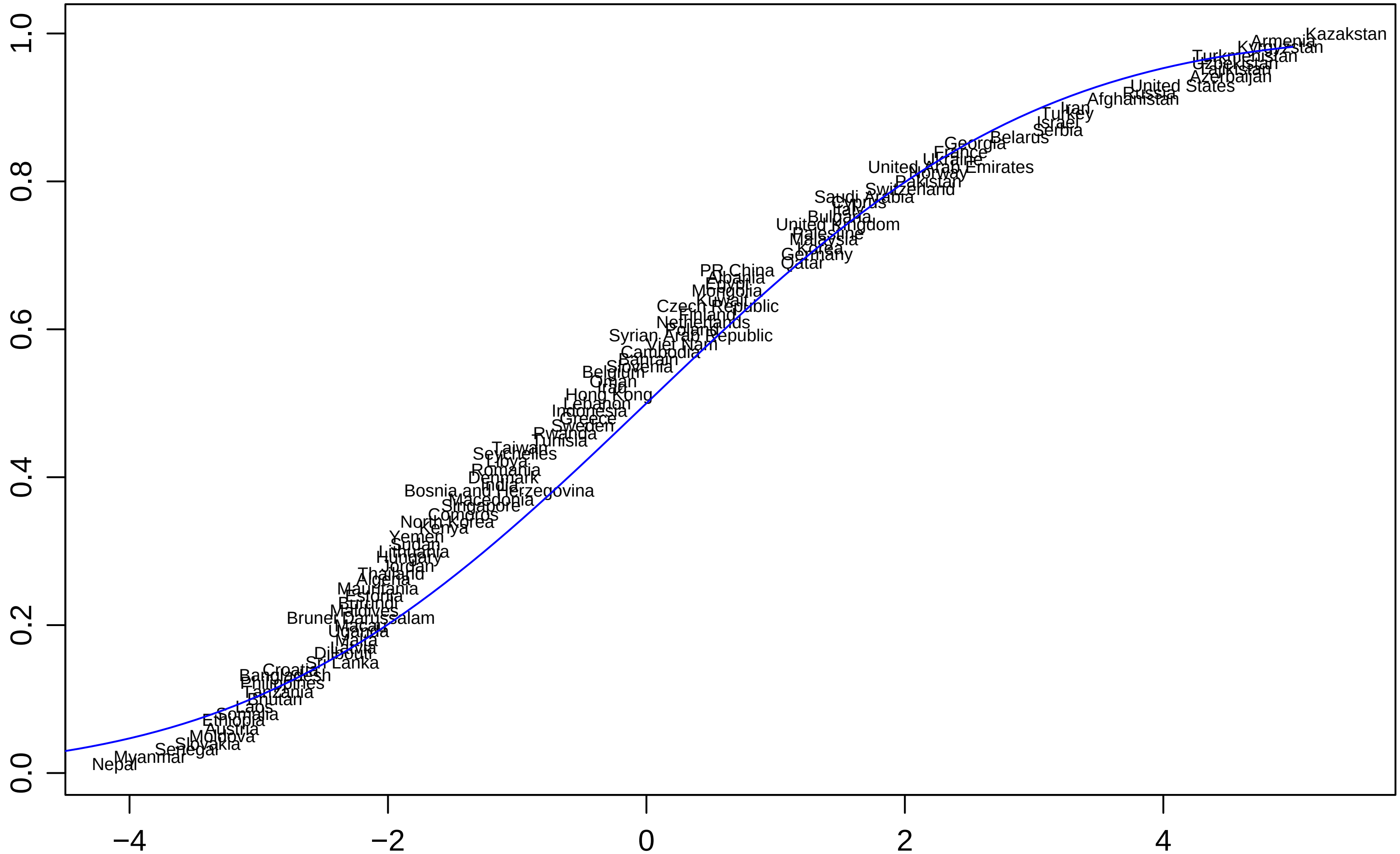
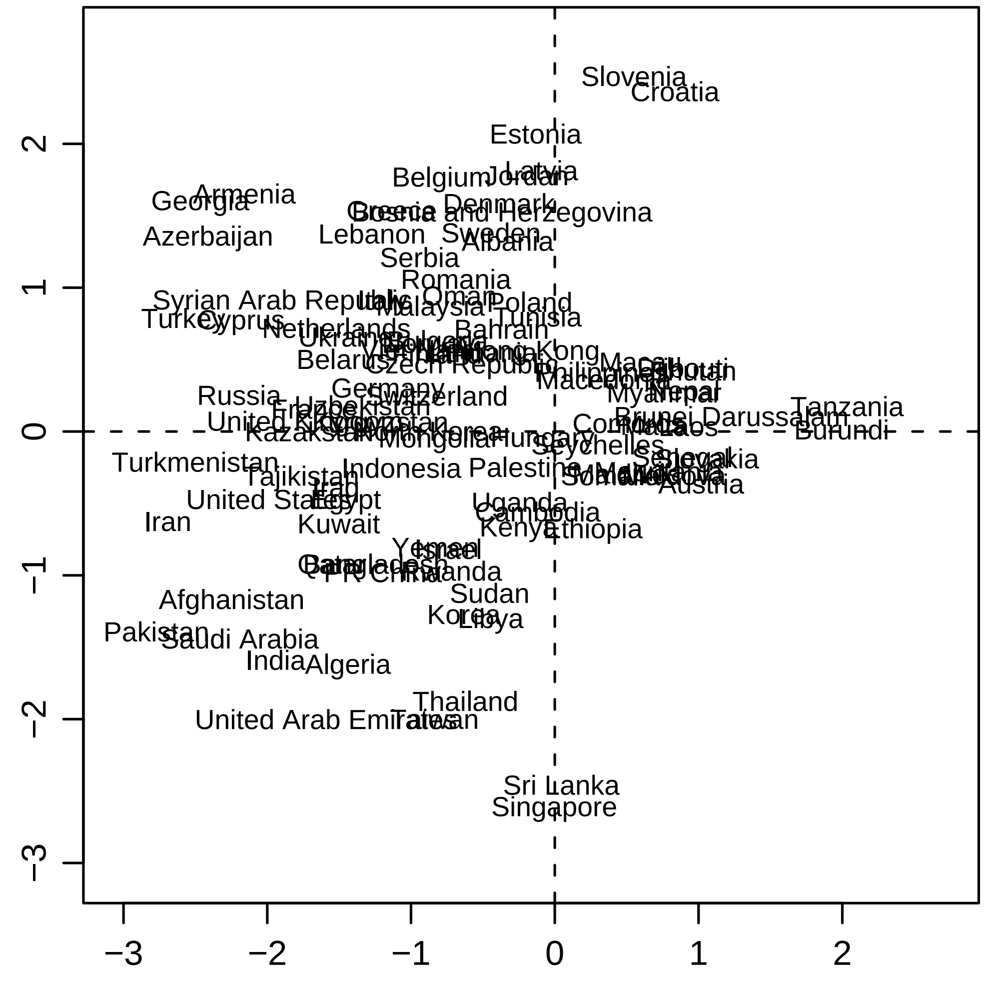


Figure 7: Marginal posterior distributions

Country-specific effects



Latent characteristic vectors



Summary and future work

- The big picture
  - second order dependence can be modeled with linear random effects;
  - some types of third order dependence can be modeled with multiplicative, bilinear effects;
  - accounting for such patterns can improve predictive performance, model fit, etc.
- Model extensions
  - clustering based on stochastic equivalence (Nowicki and Snijders 2001): $\gamma_{i,j} = z_i' R z_j$;
  - multivariate, or time-series network responses;
  - semiparametric approaches.
- Computation
  - feasible estimation methods for large datasets;
  - choice of dimension for the latent space;
  - identification of local and global maximum.