Statistical Rethinking
Winter 2019

Lecture 18 / Week 9

Further Adventures in Covariance

Non-centered random chimps

```
m14.3 <- ulam(
    alist(
        L ~ binomial(1,p),
        logit(p) <- g[tid] + alpha[actor,tid] + beta[block_id,tid],

        # adaptive priors - non-centered
        transpars> matrix[actor,4]:alpha <-
                compose_noncentered( sigma_actor , L_Rho_actor , z_actor ),
        transpars> matrix[block_id,4]:beta <-
                compose_noncentered( sigma_block , L_Rho_block , z_block ),
        matrix[4,actor]:z_actor ~ normal( 0 , 1 ),
        matrix[4,block_id]:z_block ~ normal( 0 , 1 ),

        # fixed priors
        g[tid] ~ normal(0,1),
        vector[4]:sigma_actor ~ dexp(1),
        cholesky_factor_corr[4]:L_Rho_actor ~ lkj_corr_cholesky( 2 ),
        vector[4]:sigma_block ~ dexp(1),
        cholesky_factor_corr[4]:L_Rho_block ~ lkj_corr_cholesky( 2 )
    ) , data=dat , chains=4 , cores=4 , log_lik=TRUE )
```

Non-centered random chimps

```
m14.3 <- ulam(
    alist(
        L ~ binomial(1,p),
        logit(p) <- g[tid] + alpha[actor,tid] + beta[block_id,tid],

        # adaptive priors - non-centered
        transpars> matrix[actor,4]:alpha <-
                compose_noncentered( sigma_actor , L_Rho_actor , z_actor ),
        transpars> matrix[block_id,4]:beta <-
                compose_noncentered( sigma_block , L_Rho_block , z_block ),
        matrix[4,actor]:z_actor ~ normal( 0 , 1 ),
        matrix[4,block_id]:z_block ~ normal( 0 , 1 ),

        # fixed priors
        g[tid] ~ normal(0,1),
        vector[4]:sigma_actor ~ dexp(1),
        cholesky_factor_corr[4]:L_Rho_actor ~ lkj_corr_cholesky( 2 ),
        vector[4]:sigma_block ~ dexp(1),
        cholesky_factor_corr[4]:L_Rho_block ~ lkj_corr_cholesky( 2 )
    ) , data=dat , chains=4 , cores=4 , log_lik=TRUE )
```

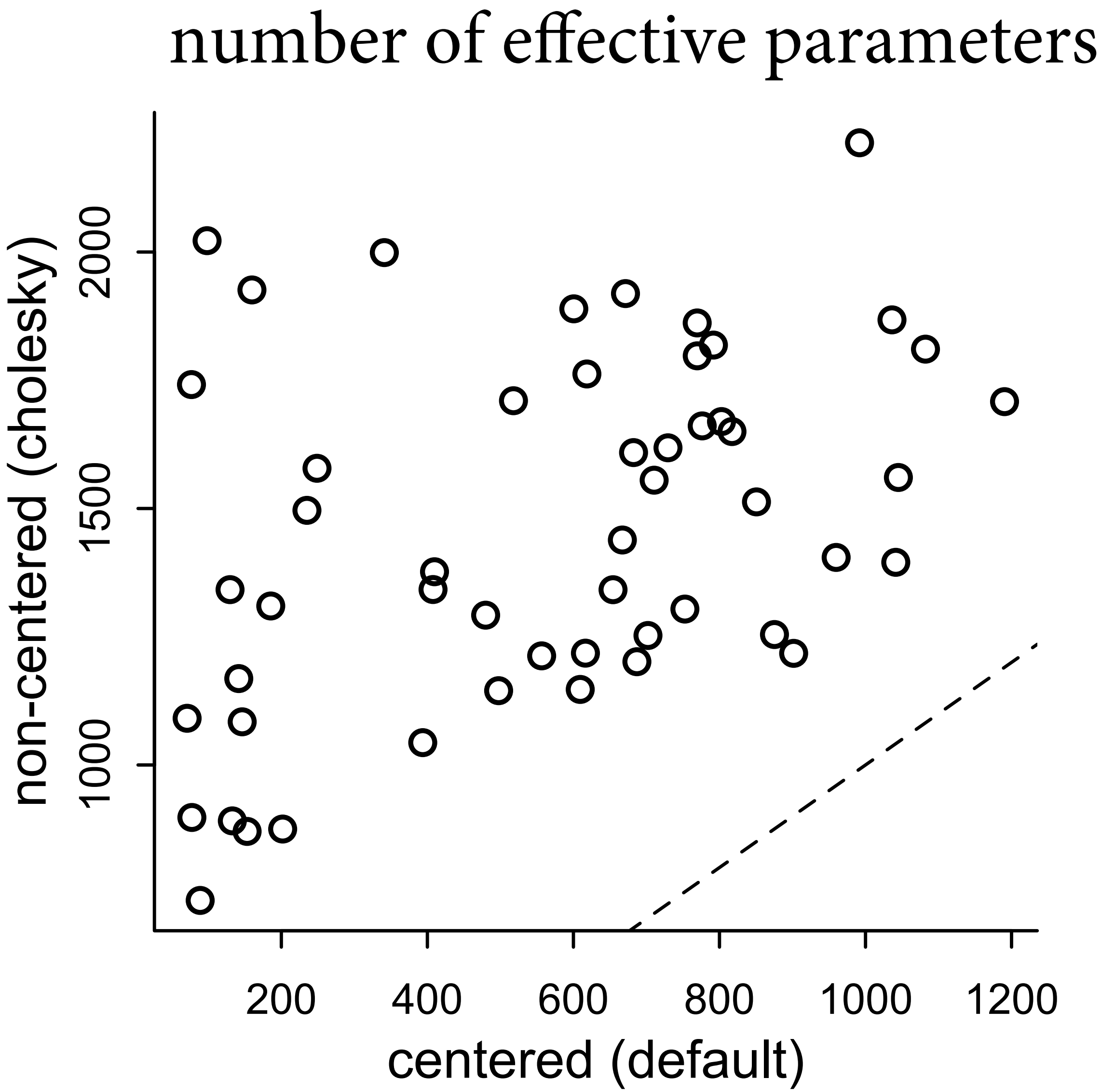
Non-centered random chimps

```
m14.3 <- ulam(
    alist(
        L ~ binomial(1,p),
        logit(p) <- g[tid] + alpha[actor,tid] + beta[block_id,tid],

        # adaptive priors - non-centered
        transpars> matrix[actor,4]:alpha <-
                compose_noncentered( sigma_actor , L_Rho_actor , z_actor ),
        transpars> matrix[block_id,4]:beta <-
                compose_noncentered( sigma_block , L_Rho_block , z_block ),
        matrix[4,actor]:z_actor ~ normal( 0 , 1 ),
        matrix[4,block_id]:z_block ~ normal( 0 , 1 ),

        # fixed priors
        g[tid] ~ normal(0,1),
        vector[4]:sigma_actor ~ dexp(1),
        cholesky_factor_corr[4]:L_Rho_actor ~ lkj_corr_cholesky( 2 ),
        vector[4]:sigma_block ~ dexp(1),
        cholesky_factor_corr[4]:L_Rho_block ~ lkj_corr_cholesky( 2 )
    ) , data=dat , chains=4 , cores=4 , log_lik=TRUE )
```

Non-centered random chimps



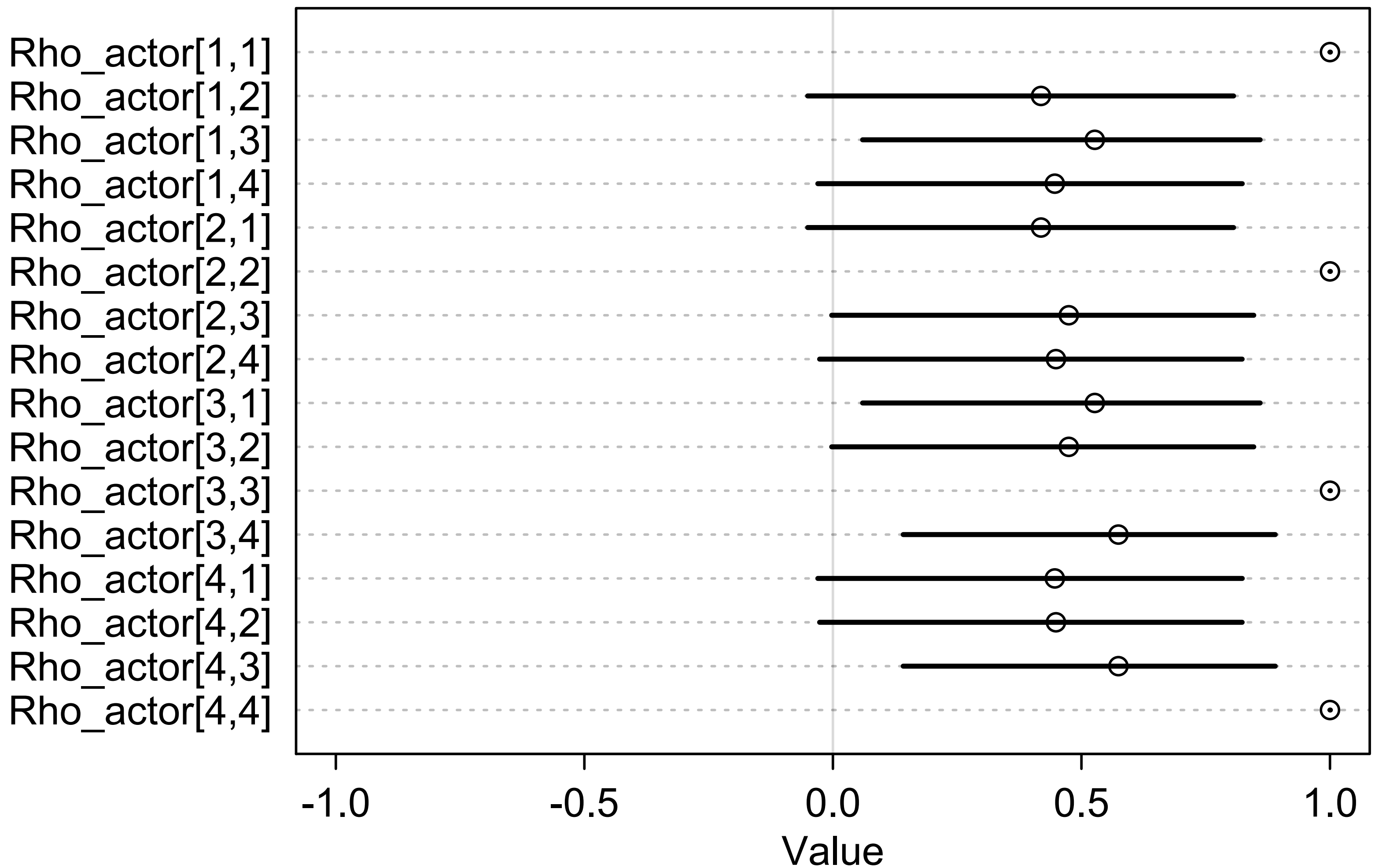
Random chimpanzees

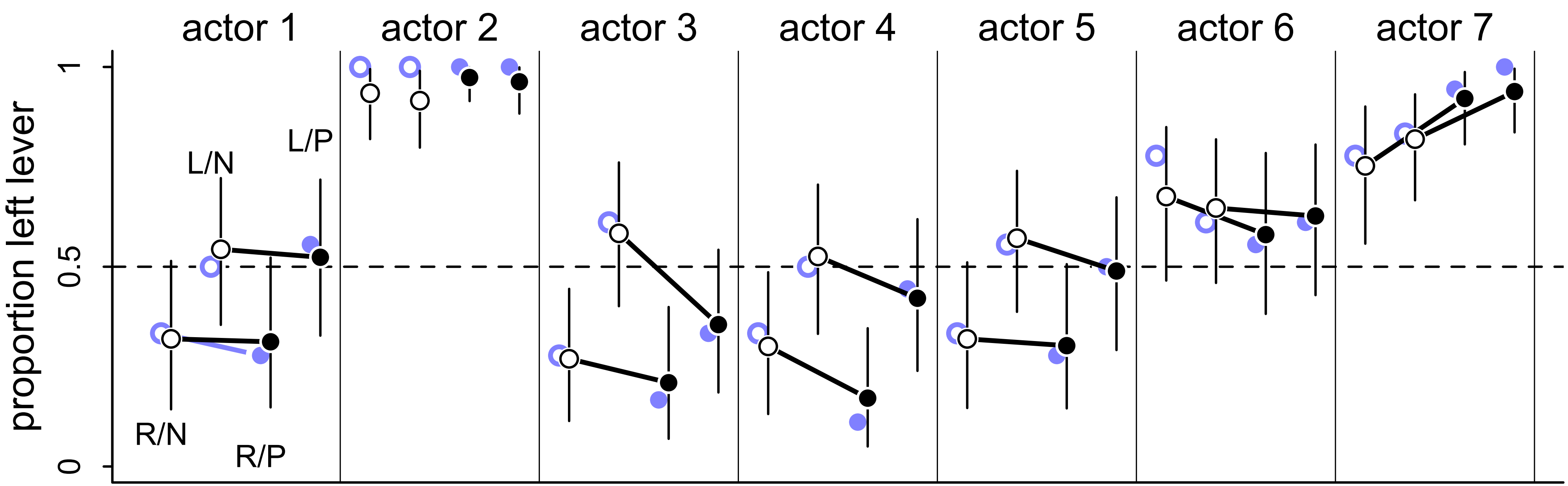
```
precis( m14.3 , depth=2 , pars=c("sigma_actor","sigma_block") )
```

R code 14.21

```
                mean   sd 5.5% 94.5% n_eff Rhat
sigma_actor[1] 1.39 0.49 0.80  2.24   906    1
sigma_actor[2] 0.92 0.38 0.44  1.64  1060    1
sigma_actor[3] 1.86 0.57 1.14  2.89  1191    1
sigma_actor[4] 1.59 0.66 0.86  2.81  1148    1
sigma_block[1] 0.40 0.32 0.03  0.98  1033    1
sigma_block[2] 0.44 0.36 0.04  1.10   944    1
sigma_block[3] 0.30 0.27 0.02  0.79  1606    1
sigma_block[4] 0.47 0.38 0.03  1.15  1073    1
```

Correlations





	mean	sd	5.5%	94.5%	n_eff	Rhat
sigma_actor[1]	1.39	0.49	0.80	2.24	906	1
sigma_actor[2]	0.92	0.38	0.44	1.64	1060	1
sigma_actor[3]	1.86	0.57	1.14	2.89	1191	1
sigma_actor[4]	1.59	0.66	0.86	2.81	1148	1

Multilevel horoscopes

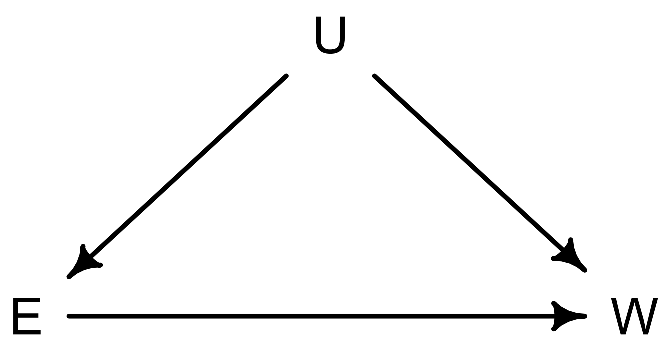
- **Think about the causal model first**
- Begin with “empty” model with varying intercepts on relevant clusters
- Standardize predictors
- Use regularizing priors (simulate)
- Add in predictors and vary their slopes
- Can drop varying effects with tiny sigmas
- Consider two sorts of posterior prediction
  - Same units: What happened in these data?
  - New units: What might we expect for new units?
- **Your knowledge of domain trumps all**

Adventures in covariance

- Many possibilities arise from using multi-variate Gaussian distributions
- Models of unobserved confounds: Instrumental variables, Mendelian randomization
- Models of social relations, networks
- Factor analysis (item-response theory)
- "Animal model" — heritability of phenotype
- Phylogenetic regressions
- Spatial autocorrelation

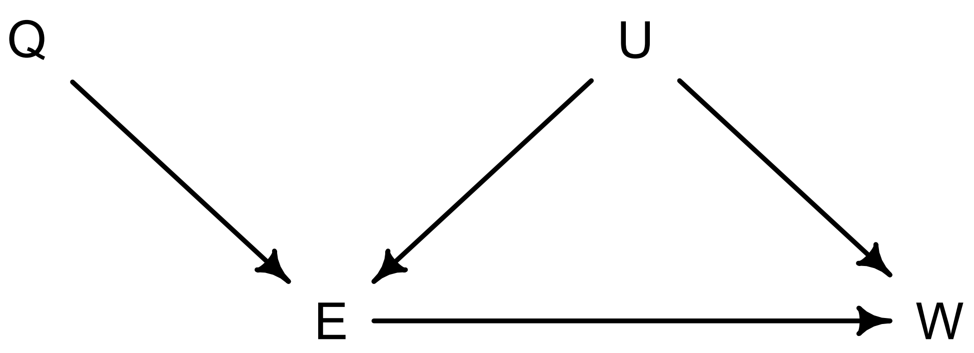
Instrumental variables

- Imagine trying to estimate influence of education on wages — lots of unmeasured confounds.



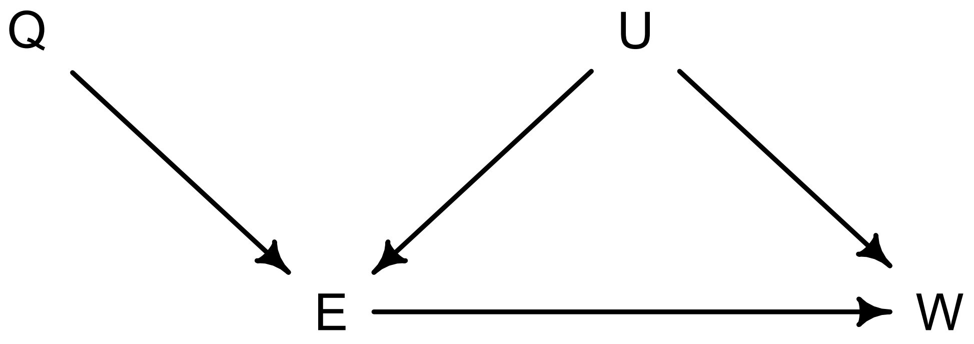
Instrumental variables

- Instrument: A variable that influences exposure (E) but not outcome (W)
- Here: Birthday position in year (Q). People born earlier in year consume less education.
  - Start school later (biologically)
  - Eligible to quit school earlier (biologically)



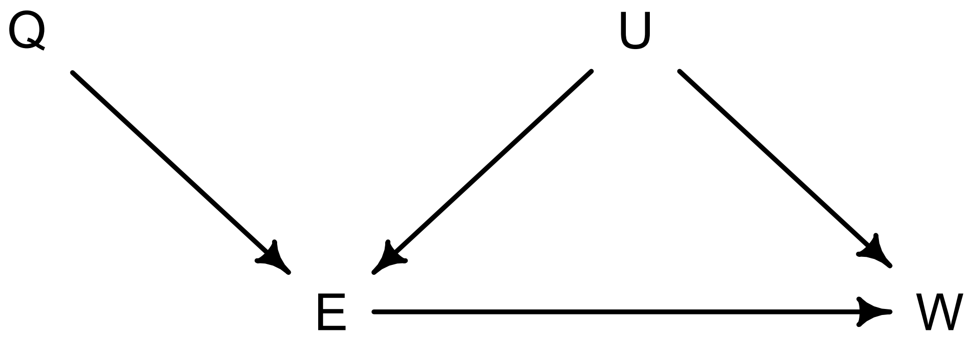
Instrumental variables

- Instrument: A variable that influences exposure (E) but not outcome (W)
- How could this help us?
- Gives us information about U
- E and W correlated, due to U
- Q helps us measure that correlation



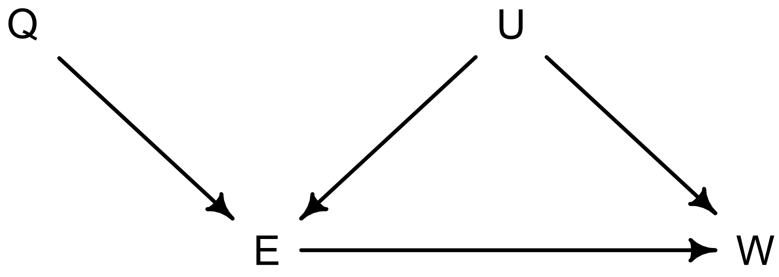
Instrumental variables

- Example:
- People born in 1st quarter (Q1) of year consume 10 years of education on average
- A specific person born in Q1 consumed 12 years
- Gives us information about unmeasured U

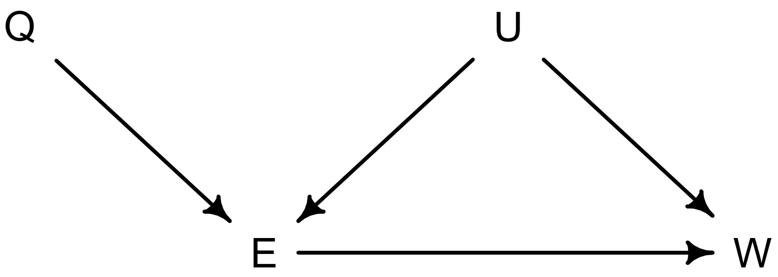


Instrumental variables

- Another perspective:
- Q is a "natural experiment"
- Q assigns E, as if by experimenter giving education pills
- But individuals are uncooperative and don't always take their pills => imperfect randomization
- Many (most?) real "experiments" are actually like this, have *intent to treat*



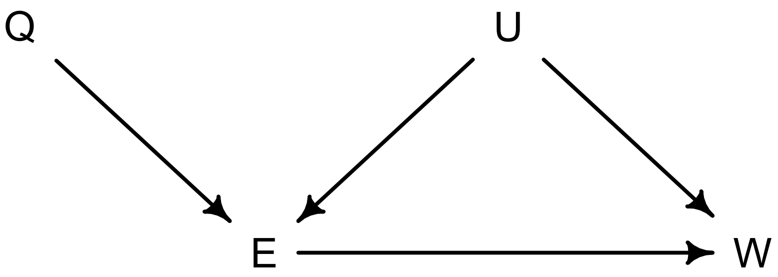
Simulated instrument



$$W_i \sim \text{Normal}(\mu_{\text{W},i}, \sigma_{\text{W}})$$
$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i + U_i$$

[Wage model]

Simulated instrument



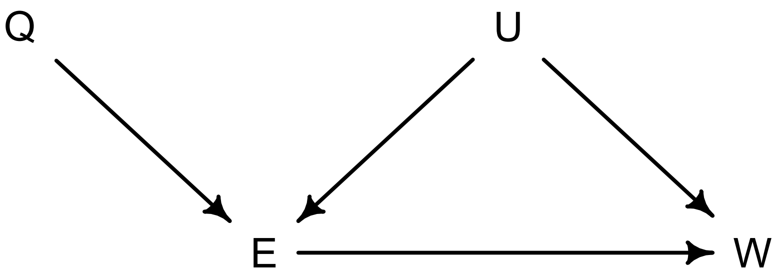
$$W_i \sim \text{Normal}(\mu_{\text{W},i}, \sigma_{\text{W}})$$
$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i + U_i$$

[Wage model]

$$E_i \sim \text{Normal}(\mu_{\text{E},i}, \sigma_{\text{E}})$$
$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i + U_i$$

[Education model]

Simulated instrument



$$W_i \sim \text{Normal}(\mu_{\text{W},i}, \sigma_{\text{W}})$$

[Wage model]

$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i + U_i$$

$$E_i \sim \text{Normal}(\mu_{\text{E},i}, \sigma_{\text{E}})$$

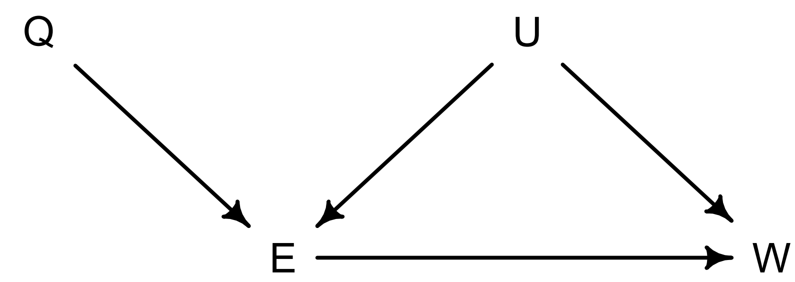
[Education model]

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i + U_i$$

$$Q_i \sim \text{Bernoulli}(0.25)$$

[Birth model]

Simulated instrument



$$W_i \sim \text{Normal}(\mu_{\text{W},i}, \sigma_{\text{W}})$$ [Wage model]

$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i + U_i$$

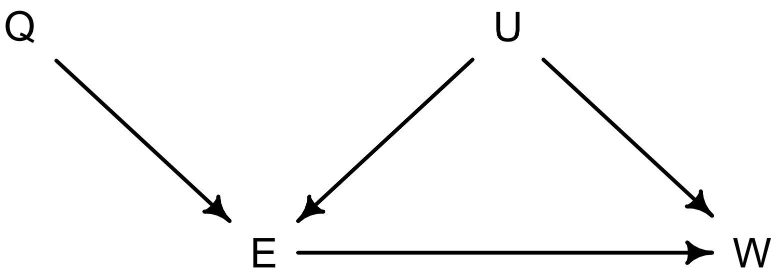
$$E_i \sim \text{Normal}(\mu_{\text{E},i}, \sigma_{\text{E}})$$ [Education model]

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i + U_i$$

$$Q_i \sim \text{Bernoulli}(0.25)$$ [Birth model]

$$U_i \sim \text{Normal}(0, 1)$$ [Confound model]

Simulated instrument




```
set.seed(73)
N <- 500
U_sim <- rnorm( N )
Q_sim <- sample( 1:4 , size=N , replace=TRUE )
E_sim <- rnorm( N , U_sim + Q_sim )
W_sim <- rnorm( N , U_sim + 0*E_sim )
dat_sim <- list(
    W=standardize(W_sim) ,
    E=standardize(E_sim) ,
    Q=standardize(Q_sim) )
```

Simulated instrument

- E —> W confounded

R code 14.24

```
m14.4 <- ulam(
    alist(
        W ~ dnorm( mu , sigma ),
        mu <- aW + bEW*E,
        aW ~ dnorm( 0 , 0.2 ),
        bEW ~ dnorm( 0 , 0.5 ),
        sigma ~ dexp( 1 )
    ) , data=dat_sim , chains=4 , cores=4 )
precis( m14.4 )
```

```
      mean   sd  5.5% 94.5% n_eff Rhat
aW    0.00 0.04 -0.07  0.07  2028    1
bEW   0.39 0.04  0.32  0.45  2032    1
sigma 0.93 0.03  0.88  0.97  1999    1
```

Instrumentality

- Think of pairs of (W,E) values as sampled from a common distribution with some covariance structure:

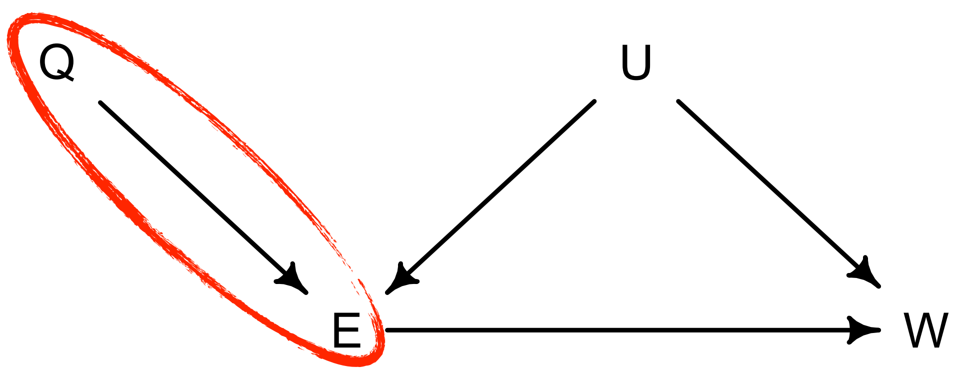
$$
\begin{aligned}
\begin{pmatrix} W_i \\ E_i \end{pmatrix} &\sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right) \\
\mu_{\text{W},i} &= \alpha_{\text{W}} + \beta_{\text{EW}} E_i \\
\mu_{\text{E},i} &= \alpha_{\text{E}} + \beta_{\text{QE}} Q_i
\end{aligned}
$$

Instrumentality

$$\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)$$

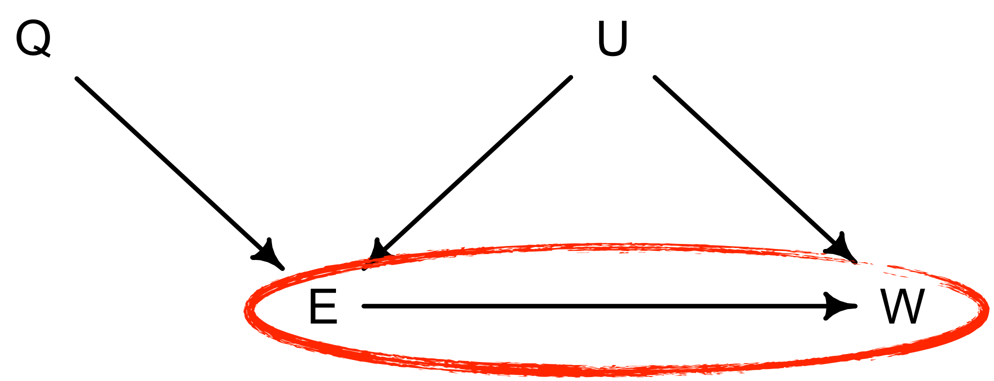
$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i$$

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i$$



Instrumentality

$$\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)$$
$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i$$
$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i$$

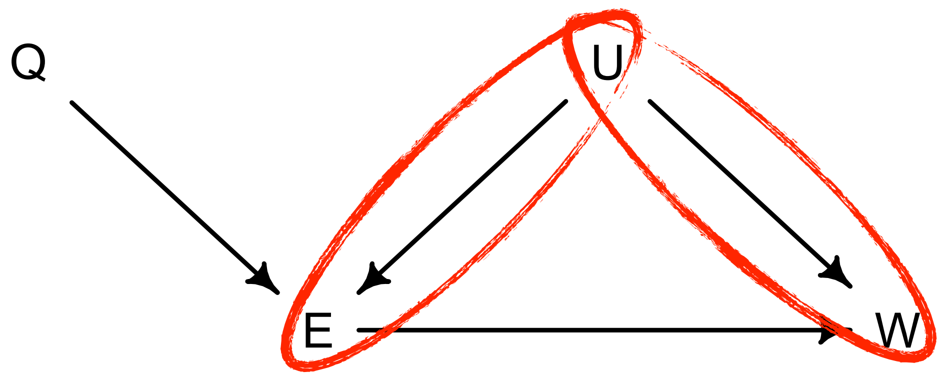


Instrumentality

$$\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)$$

$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i$$

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i$$



$$
\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)
$$

$$
\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i
$$

$$
\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i
$$

R code
14.25

```
m14.5 <- ulam(
    alist(
        c(W,E) ~ multi_normal( c(muW,muE) , Rho , Sigma ),
        muW <- aW + bEW*E,
        muE <- aE + bQE*Q,
        c(aW,aE) ~ normal( 0 , 0.2 ),
        c(bEW,bQE) ~ normal( 0 , 0.5 ),
        Rho ~ lkj_corr( 2 ),
        Sigma ~ exponential( 1 )
    ), data=dat_sim , chains=4 , cores=4 )
precis( m14.5 , depth=3 )
```

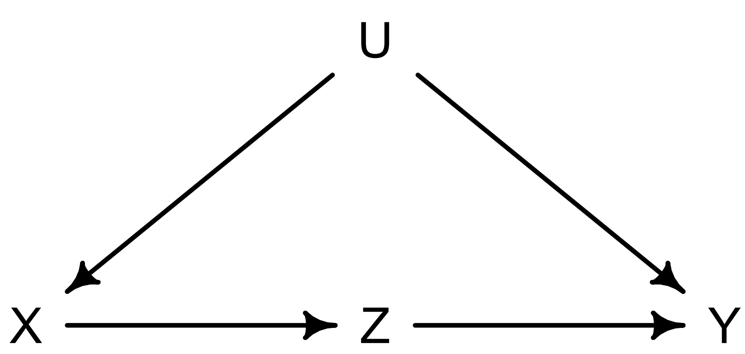
R code 14.25

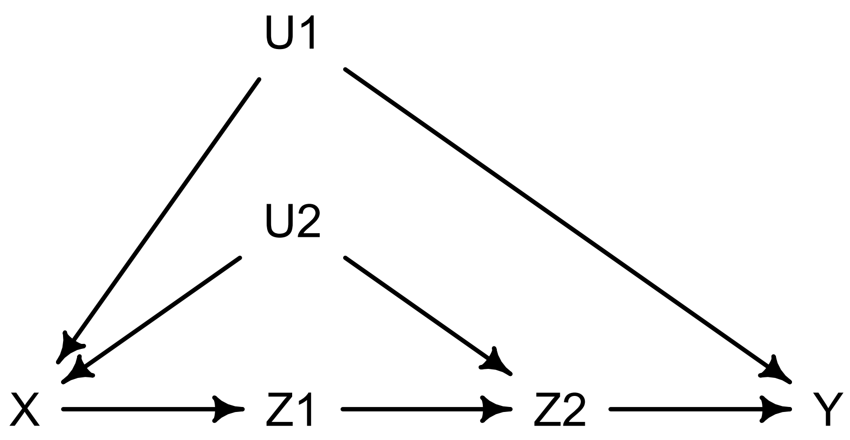
```
m14.5 <- ulam(
    alist(
        c(W,E) ~ multi_normal( c(muW,muE) , Rho , Sigma ),
        muW <- aW + bEW*E,
        muE <- aE + bQE*Q,
        c(aW,aE) ~ normal( 0 , 0.2 ),
        c(bEW,bQE) ~ normal( 0 , 0.5 ),
        Rho ~ lkj_corr( 2 ),
        Sigma ~ exponential( 1 )
    ), data=dat_sim , chains=4 , cores=4 )
precis( m14.5 , depth=3 )
```

```
           mean   sd  5.5% 94.5% n_eff Rhat
aE         0.00 0.03 -0.05  0.05  1158    1
aW         0.00 0.04 -0.07  0.07  1400    1
bQE        0.63 0.03  0.58  0.69  1557    1
bEW       -0.03 0.07 -0.14  0.08  1010    1
Rho[1,1]   1.00 0.00  1.00  1.00   NaN  NaN
Rho[1,2]   0.53 0.05  0.45  0.60   987    1
Rho[2,1]   0.53 0.05  0.45  0.60   987    1
Rho[2,2]   1.00 0.00  1.00  1.00  1714    1
Sigma[1]   1.01 0.04  0.95  1.08  1028    1
Sigma[2]   0.77 0.03  0.73  0.81  1478    1
```

Other doors

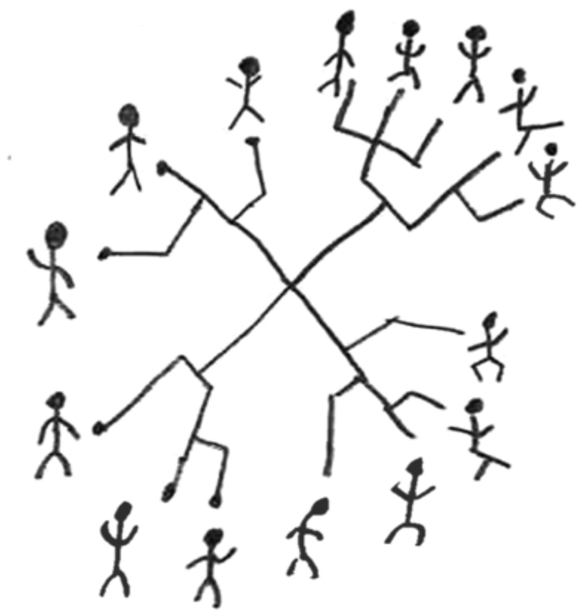
- In principle, many idiosyncratic ways to de-confound inference, if you analyze the graph correctly ("do-calculus")
- Another well-known tool: **Front-door criterion**





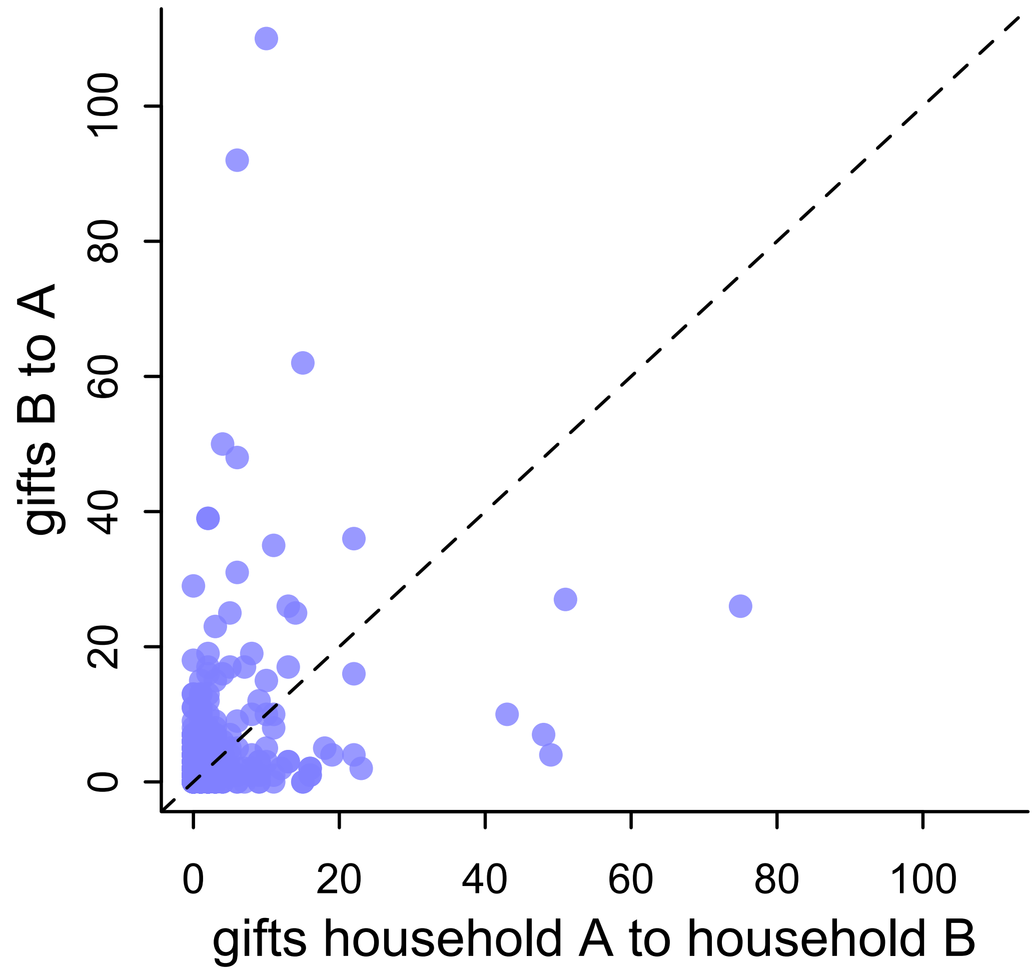
Social Relations Models

- Context: Dyadic interactions between units
- Common in social sciences, animal behavior
- How to separate general behavior from specific dyadic relationships?
- Social Relations Models (SRM) one approach — require custom covariance structure
- Really just a custom varying effects model



Nicaragua households

- data(KosterLeckie)
- 25 households
- 300 dyads
  > combn(1:25,2)
- Gift correlation 0.24



Nicaragua households

- Outcome: Count of gifts from A –> B
- Lots of predictors, but we'll ignore those for now
- Instead use varying effects to measure structure

$$y_{A \to B} \sim \text{Poisson}(\lambda_{AB})$$
$$\log \lambda_{AB} = \alpha + g_A + r_B + d_{AB}$$

average giving

giving offset for A

receiving offset for A

dyad offset A–>B

Nicaragua households

$$y_{A \to B} \sim \text{Poisson}(\lambda_{AB})$$
$$\log \lambda_{AB} = \alpha + g_A + r_B + d_{AB}$$

$$\begin{pmatrix} g_i \\ r_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_g^2 & \sigma_g \sigma_r \rho_{gr} \\ \sigma_g \sigma_r \rho_{gr} & \sigma_r^2 \end{pmatrix} \right)$$

Nicaragua households

$$y_{A \to B} \sim \text{Poisson}(\lambda_{AB})$$
$$\log \lambda_{AB} = \alpha + g_A + r_B + d_{AB}$$

$$\begin{pmatrix} g_i \\ r_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_g^2 & \sigma_g \sigma_r \rho_{gr} \\ \sigma_g \sigma_r \rho_{gr} & \sigma_r^2 \end{pmatrix} \right)$$

$$\begin{pmatrix} d_{ij} \\ d_{ji} \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \sigma_d^2 \rho_d \\ \sigma_d^2 \rho_d & \sigma_d^2 \end{pmatrix} \right)$$

Nicaragua households

$$\begin{pmatrix} d_{ij} \\ d_{ji} \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \sigma_d^2 \rho_d \\ \sigma_d^2 \rho_d & \sigma_d^2 \end{pmatrix} \right)$$

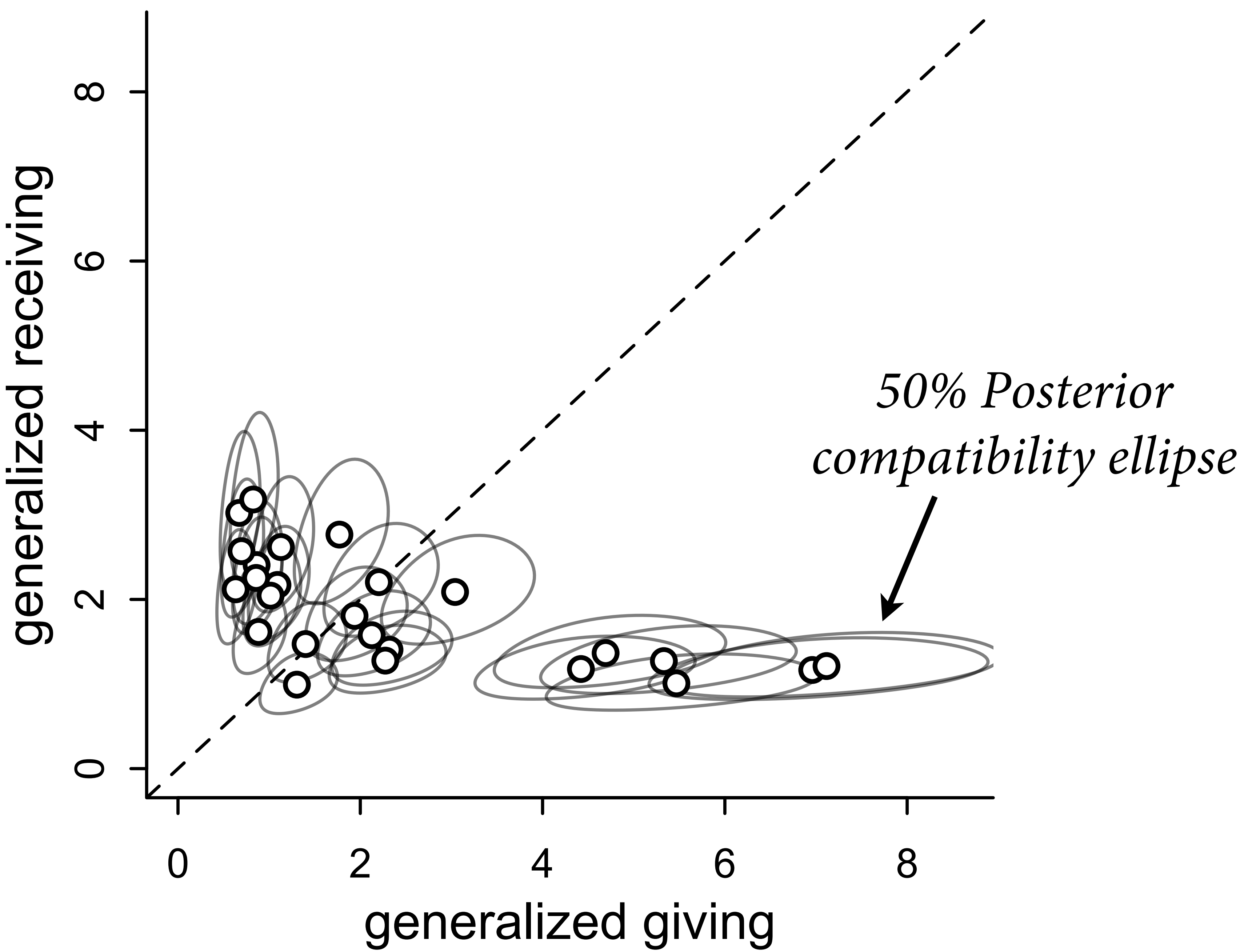
Dyad is symmetric (A/B just labels), so variance same for both variables

Nicaragua households

- Model code in text
- Only trick is copying sigma_d
- Consider general g/r effects first:

```
precis( m14.4 , depth=3 , pars=c("Rho_gr","sigma_gr") )
```

```
              mean   sd  5.5% 94.5% n_eff Rhat
Rho_gr[1,1]   1.00 0.00  1.00  1.00   NaN  NaN
Rho_gr[1,2]  -0.40 0.19 -0.71 -0.08  1423 1.00
Rho_gr[2,1]  -0.40 0.19 -0.71 -0.08  1423 1.00
Rho_gr[2,2]   1.00 0.00  1.00  1.00  3939 1.00
sigma_gr[1]   0.83 0.14  0.64  1.07  2252 1.00
sigma_gr[2]   0.42 0.09  0.28  0.58  1055 1.00
```



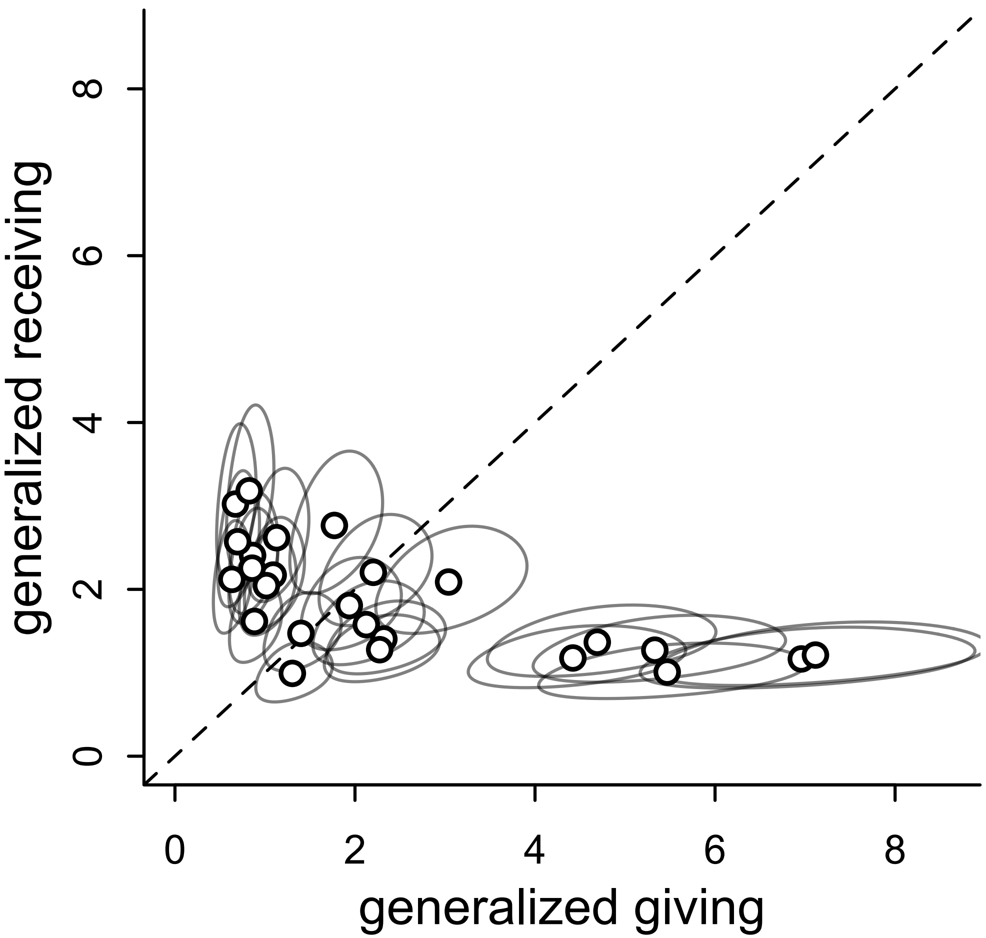
50% Posterior
compatibility ellipse
generalized receiving
generalized giving
0
2
4
6
8
0
2
4
6
8

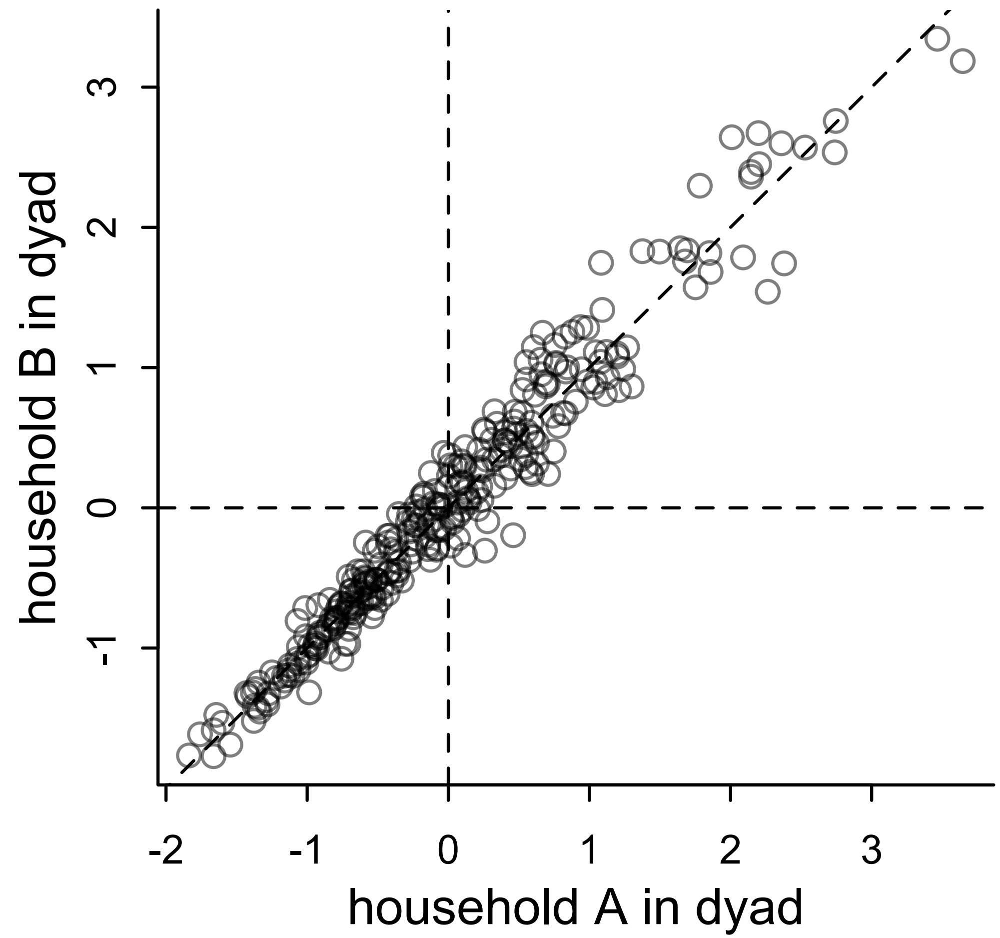
Nicaragua households

- Now consider dyad-specific effects:

```
precis( m14.4 , depth=3 , pars=c("Rho_d","sigma_d") )
```

```
            mean   sd 5.5% 94.5% n_eff Rhat
Rho_d[1,1] 1.00 0.00 1.00  1.00   NaN  NaN
Rho_d[1,2] 0.88 0.03 0.82  0.93  1072 1.01
Rho_d[2,1] 0.88 0.03 0.82  0.93  1072 1.01
Rho_d[2,2] 1.00 0.00 1.00  1.00   NaN  NaN
sigma_d    1.10 0.06 1.02  1.20  1345 1.00
```





Conditioning on general giving/receiving, gifts are very balanced. Role of zeros?

Homework

- Bangladesh contraception again
- Next week: Gaussian processes, measurement error, missing data, ***horoscopes***