Modelling covariance matrices in multivariate dyadic data

Jouni Kuha

LSE Department of Statistics Research Showcase
14.6.2022

The DyLAnIE project

- *Methods for Analysis of Longitudinal Dyadic Data with Applications to Intergenerational Exchanges of Family Support*
- Funded by ESRC and EPSRC, 2017–2021
- Investigators:
  - PI: Fiona Steele
  - co-Is and Research Officers: Irini Moustaki, Chris Skinner, Jouni Kuha, Tania Burchardt, Eleni Karagiannaki, Emily Grundy, Nina Zhang, Siliang Zhang

Substantive research questions

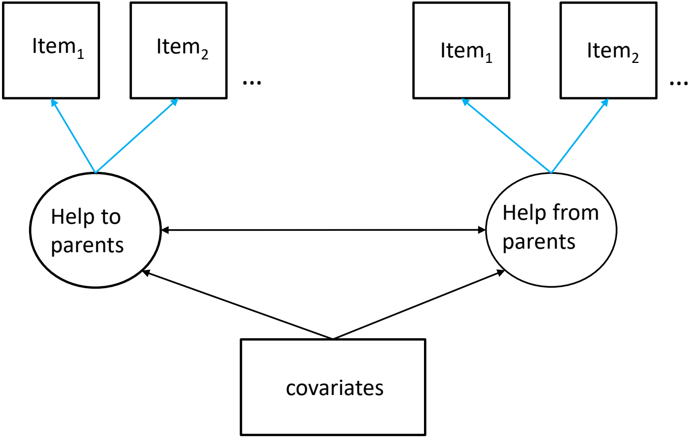
- What factors are associated with giving and receiving support between adult individuals and their non-coresident parents?
- What is the level and nature of reciprocity in these exchanges?

Data

- British Household Panel Survey (1991-2009) and its successor UK Houshold Longitudinal Study (UKHLS; 2010-present)
- Data on exchanges of support collected in the Family Network Module (2001, 2006, 2011, 2013, 2015, 2017)

Paper 1 for today's talk

- Kuha, J., Zhang, S., and Steele, F. (2021). Latent variable models for multivariate dyadic data with zero inflation: Analysis of intergenerational exchanges of family support. `arXiv:2104.11531`



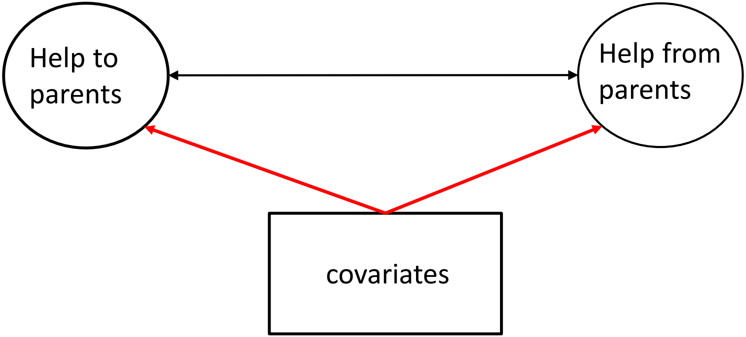
Binary indicators of help given and received

Nowadays, do you regularly or frequently do any of the things listed on this card for your parents not living here?

1. Giving them lifts in your car (if you have one)
2. Shopping for them
3. Providing or cooking meals
4. Helping with basic personal needs like dressing, eating or bathing
5. Washing, ironing or cleaning
6. Dealing with personal affairs e.g. paying bills, writing letters
7. Decorating, gardening or house repairs
8. Financial help

For help received from parents, 4. replaced by Looking after your children.

Focus of Paper 1



Paper 1: Illustrative results

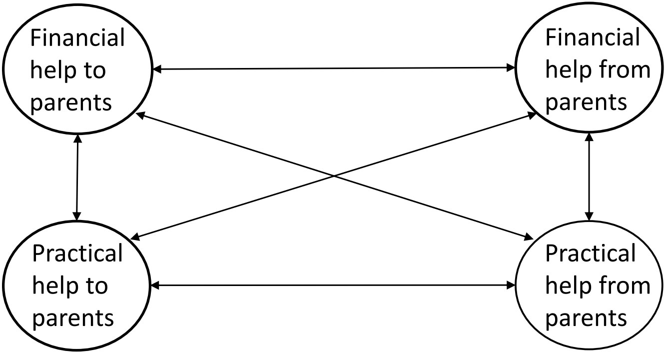
Higher tendency to give help to parents is associated with

- Parent(s) being older and living alone
- Respondent
  - having no siblings or 3 or more siblings
  - having lower household income
  - having no young children at home
  - being single
  - being not employed
- Respondent and parent(s) living near to each other

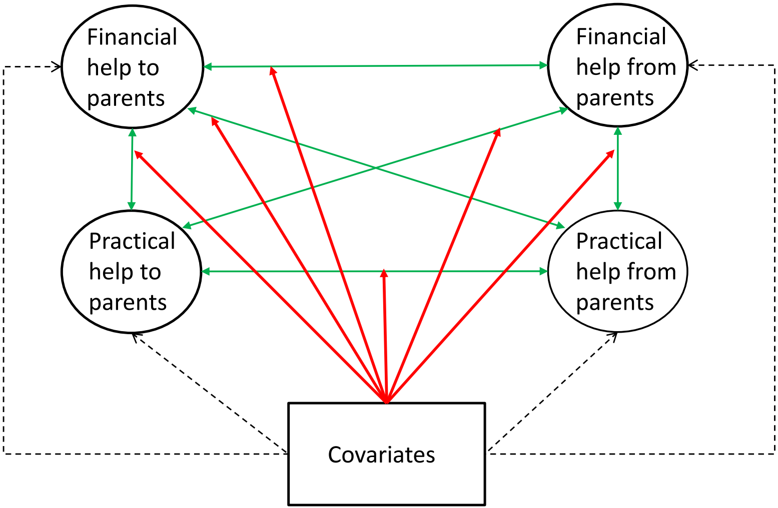
These and ther findings interpreted in terms of *needs* and *capacities* of givers and receivers of help.

Paper 2 for today's talk

- Zhang, S., Kuha, J., and Steele, F. (in progress). Modelling covariance matrices in multivariate dyadic data.



Focus of Paper 2



Modelling (conditional) covariance matrices

- Focus on the correlation matrix $\mathbf{R}(\mathbf{X};\mathbf{B})$ of latent variables $\boldsymbol{\eta}$ given covariates $\mathbf{X}$: $var(\boldsymbol{\eta}|\mathbf{X};\boldsymbol{\sigma},\mathbf{B}) = \text{diag}(\boldsymbol{\sigma})\,\mathbf{R}(\mathbf{X};\mathbf{B})\,\text{diag}(\boldsymbol{\sigma})$
- Broadly, two possible approaches (Pinheiro and Bates 1996):
  - *Unconstrained optimization*: Model a transformation which ensures that the estimated correlation matrix is positive semidefinite
  - *Constrained optimization*: Model the correlations directly, and somehow constrain the estimates so that the matrix remains positive semidefinite.
- We use a constrained approach, with linear models for the correlations

$$\rho_j = \boldsymbol{\beta}_j'\boldsymbol{X}$$

for $j = 1, \ldots, J$, so that $\mathbf{B} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_J')'$.

MCMC estimation of the models

- "Two-step" estimation: Parameters of the *measurement models* for the latent variables are estimated separately and treated as known here.
- MCMC estimation for the rest of the model then has data augmentation structure:
  - Draw (impute) values for the latent variables, treating parameters as known.
  - Draw values for the parameters from their posterior distributions given observed and (imputed) latent variables.

Drawing parameters of models for the correlations

- Goal: Draw an MCMC sample $\mathbf{B}_1, \ldots, \mathbf{B}_M$ such that $\mathbf{R}(\mathbf{X}; \mathbf{B}) \geq 0$...
  - for all $\mathbf{B}$ in the convex hull of $\mathbf{B}_1, \ldots, \mathbf{B}_M$,
  - for all $\mathbf{X}$ in a specified set, e.g. the convex hull of the observed data $\mathbf{X}_1, \ldots, \mathbf{X}_n$.
- This can be achieved through rejection sampling within the MCMC iterations, one scalar element β_{jl} of $\mathbf{B}$ at a time:
  - draw a proposal value for β_{jl} given everything else
  - check if this implies an acceptable $\mathbf{B}_m$
  - update if yes, reject if no

Illustrative results from tentative analysis

- Data: UKHLS wave 9 (2017/18)
- Fitted correlations given covariates, averaged over sample distributions of other covariates

Covariate setting	Marginal correlations GP↔RP	GP↔RF	RP↔GF	GF↔RF	GP↔GF	RP↔RF
Overall	0.38	0.16	0.02	−0.06	0.36	0.20
Age						
35 years	0.53	0.14	0.00	−0.07	0.39	0.31
45 years	0.39	0.18	0.03	−0.08	0.37	0.22
55 years	0.20	0.19	0.06	−0.06	0.32	0.08
Gender						
Female	0.31	0.14	−0.03	−0.10	0.32	0.22
Male	0.47	0.17	0.09	0.00	0.40	0.18
Travel time to the nearest parent						
$>$ 1 hr	0.48	0.01	−0.06	−0.22	0.16	0.02
$\leqslant$ 1 hr	0.34	0.21	0.05	0.00	0.43	0.27
Logarithm of household equivalised income						
25 percentile	0.37	0.15	0.02	−0.06	0.36	0.20
50 percentile	0.38	0.16	0.02	−0.06	0.36	0.20
75 percentile	0.39	0.16	0.03	−0.05	0.36	0.21