The Effect of Missing Out a Component in Multilevel Models with Social Networks Effects

Thomas Suesse [1] David Steel[1] Mark Tranmer [2]

[1]National Institute for Applied Statistics Research Australia
School of Mathematics and Applied Statistics
University of Wollongong

[2]School of Social and Political Sciences
University of Glasgow

Frontiers in Social Statistics Methodology
8 February 2017

Outline

1. Introduction
2. Multilevel models (MLMs)
3. Models for social network (SN) dependencies
4. Extended MLMs - including ML and SN effects
5. Example
6. Conclusions

- Quantitative analysis of social data needs to recognise people are not independent.
- Analysis should account for social contexts and social connections.
- Multilevel modelling now routine – account for group membership – hierarchical, cross classified, and multiple membership models.
- Social network (SN) analysis becoming more common – based on person level networks.
- So far MLM and SN analysis have tended to be used separately, but they can be combined.
- Both account for more complex variance structure than independence across units (people).
- Completely ignoring the variance structure can lead to incorrect estimated SEs and inferences and bias in estimates of model parameters.
- Here we consider models that consider both multilevel and social network effect – and impact of misspecifying the model by omitting a part of the variance structure.

- Standard MLMs for hierarchical groups, e.g.:
  - Students within classes within schools
  - People within households within neighbourhoods.
- Random effects are introduced for each level:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}_2\mathbf{u}^{(2)} + \mathbf{Z}_3\mathbf{u}^{(3)} + \mathbf{e}$$

- Estimation of regression and variance parameters by MLE – need group indicator variables.
- Cross-classified model – groups are not nested – e.g. child is a member of a school and a neighbourhood.
- Multiple membership – a person can belong to more than one group – e.g. patient treated by more than one nurse.
- For the MLM above

$$E(\mathbf{Y}) = \mathbf{X}\beta \quad \mathrm{Var}(\mathbf{Y}) = \sigma_1^2\mathbf{I}_n + \sigma_2^2\mathbf{Z}_2\mathbf{Z}_2^\top + \sigma_3^2\mathbf{Z}_3\mathbf{Z}_3^\top$$

- Much of SN analysis is aimed at modelling the network itself – our interest is in accounting for the dependencies arising from the SN.
- Matrix $\mathbf{W}$ with elements W_{ij} indicating whether units i and j are connected.
- *Network effects model* (NEM) (Leenders, 2002):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{Y} + \mathbf{e}$$

  Also called spatially lagged dependent variable model and is a form of simultaneous autoregressive (SAR) model.
- *Network disturbance model* (NDM) (Leenders, 2002):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu}, \boldsymbol{\nu} = \rho\mathbf{W}\boldsymbol{\nu} + \mathbf{e}$$

  Also called spatial error model.
- Can also consider conditional autoregressive (CAR) models.
- To fit these models we need data to generate $\mathbf{W}$. Even if such data are not available network effects may be present and affecting the analysis.

- Set $\mathbf{A}(\rho) = \mathbf{I}_n - \rho\mathbf{W}$, then for NEM and NDM

$$\text{Var}(\mathbf{Y}) = \sigma_e^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top = \sigma_e^2(\mathbf{A}^\top\mathbf{A})^{-1}$$

- For NEM (as $\mathbf{A}^{-1} \approx \mathbf{I}_n + \rho\mathbf{W}$)

$$E(\mathbf{Y}) = \mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta} \approx \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{X}\boldsymbol{\beta}$$

- For NDM

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

Extended MLMs - including ML and SN effects

- Now combine these models by having a multilevel structure and a SN structure. This leads to three models.
- *M1, Network effect MLM*:

$$\mathbf{Y} = \mathbf{X}\beta + \rho\mathbf{W}\mathbf{Y} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

- *M2, Type I Network disturbance MLM* – SN affects the random and person level error:

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\nu}, \boldsymbol{\nu} = \rho\mathbf{W}\boldsymbol{\nu} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

- *M3, Type II Network disturbance MLM* – SN affects only person level error:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\nu}, \boldsymbol{\nu} = \rho\mathbf{W}\boldsymbol{\nu} + \mathbf{e}$$

- Here **Zu** includes all the levels relevant to the multilevel structure.
- How do we decide which general model to assume?

- For MLMs MLE and REML now firmly established in variety of software, such as MLwiN.
- For SAR models (Lesage and Pace, 2009) without multilevel effects software also available.
- For the extended models we used the general purpose optimisation R-function `optim()` to maximise the log-likelihood.

Omitting part of variance structure

- In some analyses the focus may be on the regression parameters and the variance structure is just something to take into account for efficiency and inference. In other analyses the focus may be on the variance structure and the regression term is just to adjust for covariates. In general both are of interest.
- If rich data that included group and social network membership indicators were available then we could fit these models and decide which terms are statistically significant. In practice this can lead to quite complex models. Fitting of extended MLM that include SN effects is not common
- In practice we may not have all the indicators we need for such an analysis or only have a vague idea of what groupings and networks may be affecting the data.

Omitting part of variance structure

- We know that if we omit regression terms from a model the effect of the omitted variables will go somewhere – some goes to the estimated regression parameters of the explanatory variables that are included and some goes into the estimated residual variance.
- Similar consequences can be expected it we omit some components of the variance structure. For MLMs we know that the variance of an omitted level will inflate the estimated variance for the adjacent levels that are included.
- If we omit the SN effect then some of the effect will go into the estimates of the levels that are included in the MLM, including the individual level. How this play out will depend on how the SN and levels are related.
- If we omit a level then some of its effect will be transmitted to the SN effect, depending on how the two are related.

Example

- Dutch social behaviour study (Houtzager and Baerveldt, 1999).
- Nineteen schools with (simulated) classes
- Network based on giving and receiving emotional support.
- Dependent variable is a measure of delinquent behaviour.
- Explanatory variable – gender at each level.

Table: Regression results for Model $M1$ (fixed effects estimates (s.e.) and variance parameter estimates) based on original network without inter-school connections and subject-level, class-level and school-level covariate 'gender'

	MODEL $M1$							
fitted model	β_0	β_1 (subject)	β_2 (class)	β_3 (school)	ρ	σ_1^2	σ_2^2	σ_3^2
full	1.889 (0.076)	−0.607 (0.004)	−0.735 (0.221)	0.883 (0.529)	0.076	0.845	0.015	0.007
no school	1.892 (0.069)	−0.607 (0.004)	−0.738 (0.253)	0.880 (0.531)	0.078	0.845	0.023	−−
no class	1.887 (0.071)	−0.607 (0.004)	−0.724 (0.156)	0.874 (0.440)	0.077	0.854	−−	0.012
no network	2.002 (0.288)	−0.587 (0.060)	−0.701 (0.471)	0.825 (0.748)	−−	0.855	0.015	0.010
just network	1.889 (0.209)	−0.609 (0.061)	−0.727 (0.397)	0.861 (0.574)	0.099	0.868	−−	−−
just class	2.011 (0.269)	−0.587 (0.060)	−0.706 (0.519)	0.818 (0.751)	−−	0.855	0.026	−−
just school	1.998 (0.278)	−0.587 (0.060)	−0.691 (0.396)	0.822 (0.688)	−−	0.864	−−	0.016
no levels	2.015 (0.208)	−0.587 (0.061)	−0.691 (0.401)	0.796 (0.580)	−−	0.881	−−	−−

Table: Regression results (fixed effects estimates (s.e.) and variance parameter estimates) for Model $M2$ based on original network without inter-school connections and subject-level, class-level and school-level covariate 'gender'

	MODEL $M2$							
fitted model	β_0	β_1 (subjects)	β_2 (class)	β_3 (school)	ρ	σ_1^2	σ_2^2	σ_3^2
full	1.948 (0.086)	−0.586 (0.004)	−0.784 (0.213)	1.008 (0.567)	0.131	0.845	0.013	0.005
no school	1.954 (0.080)	−0.586 (0.004)	−0.786 (0.235)	1.001 (0.564)	0.133	0.845	0.018	−−
no class	1.946 (0.080)	−0.586 (0.004)	−0.774 (0.157)	1.000 (0.485)	0.134	0.853	−−	0.009
no network	2.002 (0.288)	−0.587 (0.060)	−0.701 (0.471)	0.825 (0.748)	−−	0.855	0.015	0.010
just network	1.959 (0.234)	−0.586 (0.065)	−0.787 (0.398)	0.993 (0.620)	0.175	0.859	−−	−−
just class	2.011 (0.269)	−0.587 (0.060)	−0.706 (0.519)	0.818 (0.751)	−−	0.855	0.026	−−
just school	1.998 (0.278)	−0.587 (0.060)	−0.691 (0.396)	0.822 (0.688)	−−	0.864	−−	0.016
no levels	2.015 (0.208)	−0.587 (0.061)	−0.691 (0.401)	0.796 (0.580)	−−	0.881	−−	−−

Table: Regression results (fixed effects estimates (s.e.) and variance parameter estimates) for Model *M*3 based on original network without inter-school connections and subject-level, class-level and school-level covariate 'gender'

	MODEL *M*3							
fitted model	β_0	β_1 (subjects)	β_2 (class)	β_3 (school)	ρ	σ_1^2	σ_2^2	σ_3^2
full	1.965 (0.086)	−0.586 (0.004)	−0.776 (0.218)	0.967 (0.570)	0.134	0.845	0.014	0.007
no school	1.971 (0.077)	−0.586 (0.004)	−0.780 (0.245)	0.963 (0.562)	0.136	0.845	0.020	−−
no class	1.958 (0.080)	−0.585 (0.004)	−0.779 (0.157)	0.982 (0.487)	0.135	0.853	−−	0.012
no network	2.002 (0.288)	−0.587 (0.060)	−0.701 (0.471)	0.825 (0.748)	0.000	0.855	0.015	0.010
just network	1.959 (0.234)	−0.586 (0.065)	−0.787 (0.398)	0.993 (0.620)	0.175	0.859	−−	−−
just class	2.011 (0.269)	−0.587 (0.060)	−0.706 (0.519)	0.818 (0.751)	−−	0.855	0.026	−−
just school	1.998 (0.278)	−0.587 (0.060)	−0.691 (0.396)	0.822 (0.688)	−−	0.864	−−	0.016
no levels	2.015 (0.208)	−0.587 (0.061)	−0.691 (0.401)	0.796 (0.580)	−−	0.881	−−	−−

Conclusions

- A lot of research has focused on the effect of omitting a level in a multi-level model on the standard errors of the fixed effects and the other variance parameters of the multi-level model (Tranmer and Steel, 2001; Berkhof and Kampen, 2004; Moerbeek, 2004; Van Landeghem et al., 2005)
- We combined autoregressive models with multi-level models, this combination led to three different extended multilevel models (termed $M1$, $M2$ and $M3$).
- Irrespective of the extended multilevel model, the results are similar. When a component of the multi-level is ignored then the impact on the network parameter (measuring social dependence) is minimal and usually only the other variance parameters of the other not omitted levels are affected. Similar rules apply in this case as when ignoring a level in a standard multi-level model.

Conclusions

- Ignoring a level also has impact on the standard errors of the fixed effects. Similar conclusions apply as for omitting a level in a multi-level model Tranmer and Steel (2001); Berkhof and Kampen (2004); Moerbeek (2004); Van Landeghem et al. (2005), for example omitting a level leads to increased variance estimates of the flanking level variance estimates and also incorrect standard errors of the fixed effects referring to the omitted and the flanking levels.
- However, when the network is ignored then all other variance parameters are inflated.
- As a consequence, the standard errors of the fixed effects are usually inflated when omitting the network. This has large practical implications.

J. Berkhof and J. K. Kampen. Asymptotic effect of misspecification in the random part of the multilevel model. *Journal of Educational and Behavioral Statistics*, 29(2):201–218, 2004.

Bregje Houtzager and Chris Baerveldt. Just like normal: A social network study of the relation between petty crime and the intimacy of adolescent friendships. *Social Behavior and Personality: an international journal*, 27(2):177–192, 1999.

R. Leenders. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47, 2002.

J. Lesage and R.K. Pace. *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Boca Raton, FL, 2009.

M. Moerbeek. The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*, 39(1):129–149, 2004.

M. Tranmer and D. G. Steel. Ignoring a level in a multilevel model: evidence from uk census data. *Environment and Planning A*, 33(5):941–948, 2001.

G. Van Landeghem, B. De Fraine, and J. Van Damme. The consequence of ignoring a level of nesting in multilevel analysis: A comment. *Multivariate Behavioral Research*, 40(4):423–434, 2005.