

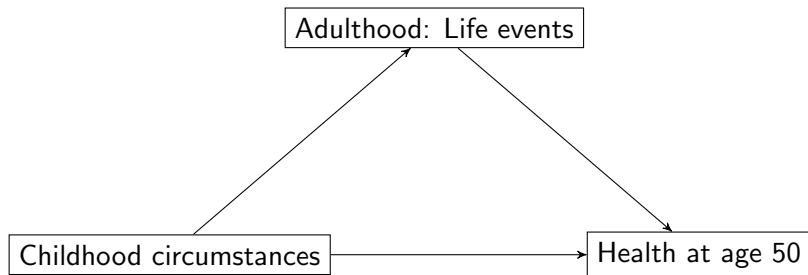
# A multilevel SEM to estimate the effects of childhood SECs on midlife health

Yajing Zhu

Department of Statistics  
London School of Economics and Political Science, UK

UCL, 9 March 2018

# Substantive Research Question



**Figure 1 :** A general joint modelling framework to explore the potential pathways between childhood circumstances, life events and health in mid-life.

# Review of previous work

## Main interest

How to include latent summaries of childhood SEC as predictors of a distal outcome?

- 1-step approach
  - Problem: unintended circular relationship.
- naive 3-step approaches (modal class, pseudo class)
  - Problem: misclassification, underestimated/overestimated standard errors.
- Advanced 3-step approaches (ML)

# A general 3-step ML approach I

- 1-LV: Vermunt (2010), Asparouhov and Muthén (2014).
- Multiple LVs: Zhu et al. (2017): generalisation & robustness test.

## Steps

- Step 1: Estimate separate latent class models for categorical predictors.
- Step 2: Calculate misclassification probabilities.
- Step 3: Estimate models of interest, with categorical LVs as predictors.

## A general 3-step ML approach II

- Notation:  $Us$  (indicators for  $Cs$ ),  $Ms$  (most likely class membership),  $Z$  (distal outcome).
- Assumption:  $C_1 \perp\!\!\!\perp M_2 | C_2$ ;  $C_2 \perp\!\!\!\perp M_1 | C_1$ ,  $Z \perp\!\!\!\perp Us | C_1, C_2$ .

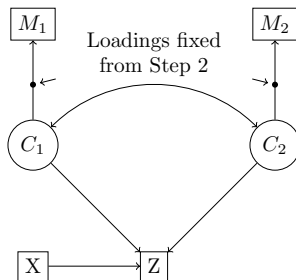


Figure 2 : The 3-step approach with two latent categorical variables  $C_1$  and  $C_2$ .

## Extension: RE discrete-time EHA I

*Childhood SECs* → *time-to-event outcome*

- Two associated survival processes: time to first partnership formation, time to recurrent partnership dissolutions (selection).
- Joint model: allow for association through an individual-specific random effect term (but with differential effects).
- Data: NCDS 1958 with co-residential partnership records on duration, partnership type, outcome (e.g. separated, married).

## Extension: RE discrete-time EHA II

### Discrete-time survival data

- Denote by  $y_{ij}$  the duration of episode  $j$  of individual  $i$ , which is fully observed if an event occurs ( $\delta_{ij} = 1$ ) and right-censored if not ( $\delta_{ij} = 0$ ).
- Data restructuring: convert the observed data  $(y_{ij}, \delta_{ij})$  to a sequence of binary responses  $(y_{tij})$ , indicating whether an event has occurred in time interval  $[t, t + 1)$ .
- Discrete-time hazard function:  $h_{tij} = Pr(y_{tij} = 1 | y_{t-1,ij} = 0)$ .
- Multilevel model: dissolutions are recurrent.

## Extension: RE discrete-time EHA III

Step 3 is a random effects logit model, allowing for a log-linear structure between LVs.

$$\log \left( \frac{h_{tij}}{1 - h_{tij}} \right) = \alpha_t + \beta^T \mathbf{X}_{tij} + \sum_{k_1=1}^{K_1-1} \tau_{k_1}^{C_1} I(C_{1i} = k_1) + \sum_{k_2=1}^{K_2-1} \tau_{k_2}^{C_2} I(C_{2i} = k_2) + u_i$$

- $\alpha_t$  is the baseline hazard function
- $\mathbf{X}_{tij}$  is the vector of time-varying and time-invariant predictors
- $\tau_{k_1}^{C_1}$  and  $\tau_{k_2}^{C_2}$  are the class-specific coefficients of LVs
- $u_i \sim N(0, \sigma_u^2)$  is the individual-specific unobserved random effect



## Extension: structural equation models

Recall the substantive research question:

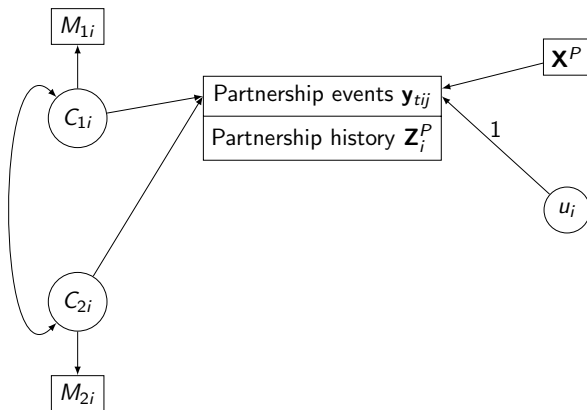


Figure 3 : A general path diagram of a multilevel SEM with factorised individual-level random effects.

## Extension: structural equation models

Recall the substantive research question:

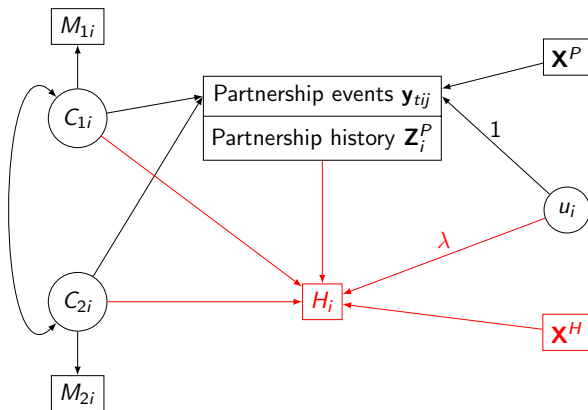


Figure 3 : A general path diagram of a multilevel SEM with factorised individual-level random effects.

# Advantages of the framework

- LCA: measurement error of a set of measurements, longitudinal typology over 4 childhood waves, FIML (only CMs with no information in ALL 4 waves are dropped)
- Joint modelling handles endogeneity of  $\mathbf{Z}_i^{(P)}$  in the health model.
- Allow for differential effects ( $\lambda$ s) of a common set of individual-specific unobservables ( $u_i$ ) on the hazard of union formation, separation and later health.
- Generalisability:
  - Can handle data with complex structures (e.g. multilevel, longitudinal, mixed response types)
  - Multivariate health outcome, multiple related processes (e.g. dropout, health, partnership)  $\rightarrow$  better identification of  $\sigma_u^2$  (factor model)
  - Different cross-process residual structures  $\rightarrow$  sensitivity tests.

# Substantive findings I

- Outcome: binary general health status at age 50.
- LCA for each dimension of childhood SECs  $\rightarrow$  entropy  $> 0.7$ .  
Associated LVs: Social class, financial difficulty, material hardship, family structure.
- Other covariates adjusted in the health submodel: early life health (e.g. BMI at 16); adjusted in partnership submodels (e.g. number of pre-school children, education level, number of previous partners).

# Substantive findings II

Table 1 : Health50 ON childhood SECs

Latent categorical variables	3-step SEM	
	Est.	(SE)
Social class (ref.=High)		
Low	0.41**	(0.188 )
Medium	0.34**	(0.115 )
Financial difficulty (ref.=Low)		
High	0.53**	(0.207 )
Material hardship (ref.=Low)		
Medium	0.34**	(0.106 )
High	0.36**	(0.119 )
Family structure (ref.=Stable)		
Unstable	0.08	(0.133 )

\*\*  $p < 0.05$ , \*  $p < 0.1$

# Substantive findings III

Table 2 : Health on partnership history

Variables	3-step SEM	
	Est.	(SE)
Total number of partners before age 50 (ref.=1)		
0	0.08	(0.307)
2	0.02	(0.127)
3+	0.10	(0.233)
Age at 1 <sup>st</sup> relationship	-0.09**	(0.046)
% time single	1.02**	(0.376)
Random effects parameters		
$\sigma_u^2$	1.15**	(0.132)
$\lambda^{(H)}$	-0.13	(0.123)
$\lambda^{(F)}$	-0.27**	(0.066)
$\lambda^{(D)}$	1	

\*\*  $p < 0.05$ , \*\*  $p < 0.1$

# Substantive findings IV

Table 3 : Partnership history (Formation) ON childhood SECs

Latent categorical variables	3-step SEM	
	Est.	(SE)
Social class (ref.=High)		
Low	0.04	(0.081 )
Medium	0.14**	(0.043 )
Financial difficulty (ref.=Low)		
High	0.23**	(0.098 )
Material hardship (ref.=Low)		
Medium	0.04	(0.041 )
High	0.04	(0.048 )
Family structure (ref.=Stable)		
Unstable	0.10**	(0.052 )

\*\*  $p < 0.05$ , \*\*  $p < 0.1$

# Substantive findings V

Table 4 : Partnership history (Dissolution) ON childhood SECs

Latent categorical variables	3-step SEM	
	Est.	(SE)
Social class (ref.=High)		
Low	−0.02	(0.149 )
Medium	−0.05	(0.084 )
Financial difficulty (ref.=Low)		
High	−0.11	(0.177 )
Material hardship (ref.=Low)		
Medium	−0.13	(0.079 )
High	−0.27**	(0.095 )
Family structure (ref.=Stable)		
Unstable	0.29**	(0.107 )

\*\*  $p < 0.05$ , \*\*  $p < 0.1$



# Summary I

- ① Among the four dimensions of childhood SECs, lower male heads social class, financial difficulty and material hardship have a long-lasting impact on poorer health at age 50, even after controlling for partnership situation during adulthood.
- ② The impact of family instability on later health, however, is fully explained by the impact of the cohort members' own partnership experiences.
  - Unstable family structure → early partnership formation (OR=1.11, 95% CI=[1.00,1.22]), early separation (OR=1.34, 95% CI=[1.08, 1.65]) and increased chance of poor midlife health.
- ③ Among individuals who formed their first partnership at the same age, those who spent more time single have a significantly higher risk of poor health at age 50 (OR=2.77, 95% CI=[1.33,5.80]).

# Summary II

- ④ Estimates of random effect parameters suggest
  - Influences of a common set of unobserved time-invariant characteristics on three processes do exist.
  - A negligible residual association due to time-invariant unobservables between poor midlife health and tendency of dissolution.
    - What is  $u_i$ ? Individual-level characteristics associated with formation, dissolution processes and midlife health. e.g. latent ambition at work.
  - A negative residual association ( $\hat{\lambda}^{(F)} < 0, \hat{\lambda}^{(D)} > 0$ ):  
High values of  $u_i \rightarrow$  delay first partnership union, higher dissolution risk.
  - A positive residual association ( $\hat{\lambda}^{(F)} < 0, \hat{\lambda}^{(H)} < 0$ ):  
High values of  $u_i \rightarrow$  delay first partnership union, better midlife health.

# References

- Asparouhov, T. and Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3):329–341.
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4):450–469.
- Zhu, Y., Steele, F., and Moustaki, I. (2017). A general 3-step maximum likelihood approach to estimate the effects of multiple latent categorical variables on a distal outcome. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(5):643–656.

# Repeated measurements of childhood SECs I

## Measurements of social class

- Based on the occupation of the male head of the cohort member's family. The original coding in the four childhood waves followed the UK official guidance (General Registrar's Social Class) in 1951, 1960, 1966, 1970. Referring to works of Kuh (2003), Chandola et al. (2006) and Case et al. (2011).
- Six ordinal categories: unemployed, unskilled, partially skilled, skilled, managerial and professional.
- Cases with a single mother, the maternal grandfather's social class is used if available; otherwise, occupation is coded as missing.

# Repeated measurements of childhood SECs II

## Measurements of financial difficulty

- Coded in binary using multiple indicators following Bartley et al. (2003).
- At the birth sweep, 1 if a cohort member's father is in a low social class; At age 7, 1 if the father is in the last two social class categories or has requested supplementary benefits or claims to be in financial difficulty. At ages 11 and 16, 1 if there is at least one positive answer to the questions related to financial hardship (being a recipient of free school meals, being a recipient of benefits and father belonging to the last two social class categories). Note that benefits include official supplementary benefits, unemployment support and family income support.

# Repeated measurements of childhood SECs III

## Measurements of material hardship

- Ordinal variable with five categories (from low to high) derived from a summary of yes or no questions related to the following four aspects, following Schoon et al. (2003).
- The existence of overcrowding, no full sole use of household amenities, not owning the property and recipient of support benefits. Answers to these four questions were collected repeatedly at ages 7, 11 and 16 (no question is available at birth sweep).

# Repeated measurements of childhood SECs IV

## Measurements of family structure

- A nominal variable with five categories following Hobcraft et al. (1998) and British Association of Adoption and Fostering.
- The five situations from poor to good are as follows:
  - In care or in foster care or in other similar situations
  - Cared for by other blood relatives
  - Cared for by a single parent (includes individuals cared for by natural parents who are divorced or separated)
  - Cared for by step parents (includes individuals cared for by one natural parent and one step parent)
  - Cared for by joint parents (includes individuals cared for by two natural or adoptive parents)