Twin data analysis with ACE-decomposed explanatory variables using Stata

German Stata Users Group Meeting, 06/23/2017, Humboldt University Berlin

Volker Lang

Bielefeld University

volker.lang@uni-bielefeld.de

Outline

1. TwinLife

2. Univariate ACE-decompositions: The „classical" twin design

3. **acelong**: **gsem**-wrapper for ACE-decompositions using Stata

4. ACE-β models: Causal analysis based on twin design

ACE-(variance) decomposition: Partitions the variance of an outcome varying within twin pairs into three latent components associated with *additive genetic effects* (**A**), *environmental effects shared* by both twins (**C**) and *environmental effects unique* to each twin (**E**)
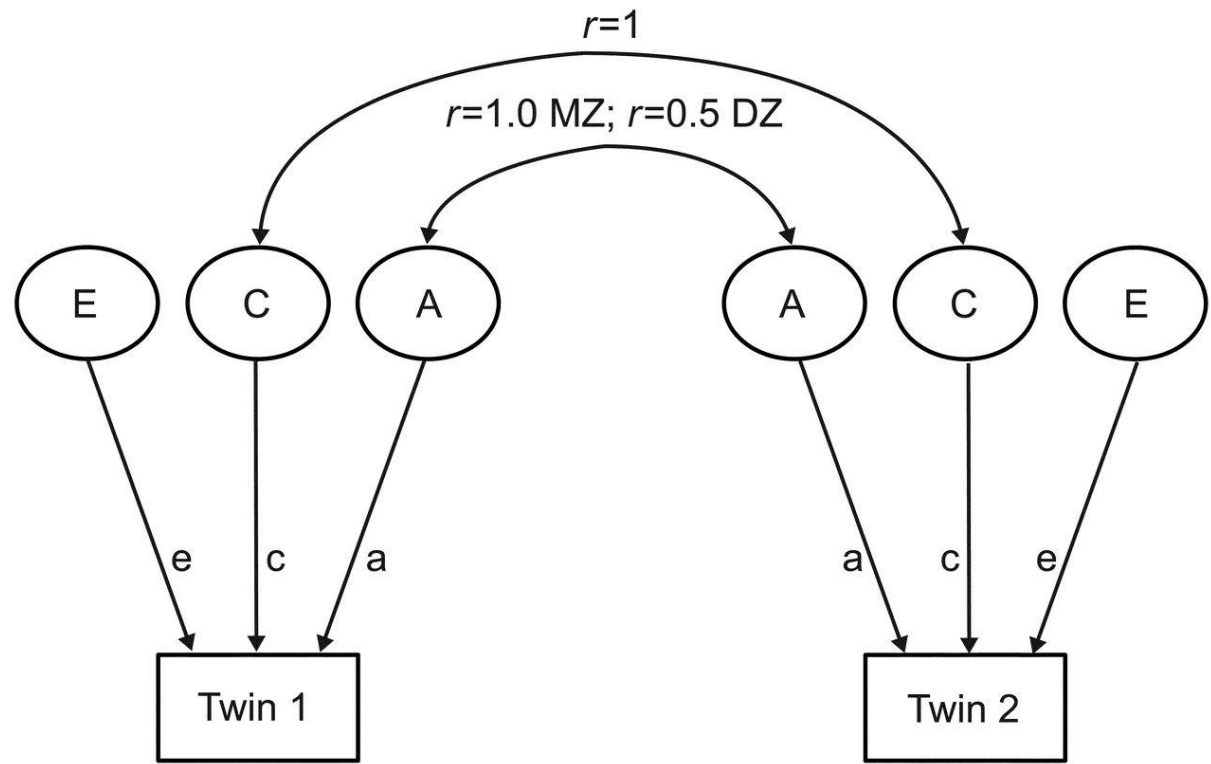
TwinLife

- **German twin family panel** (Diewald et al. 2016)

- Multidimensional social stratified random sample of 4,097 monozygotic (MZ) & same-sex dizygotic (DZ) twin pairs & their families

- Extended twin family design: twins, parents, if applicable sibling & partners

- Comprises four birth cohorts: 1990-93, 1997/98, 2003/04, 2009/10

- Available for the scientific community free of charge at GESIS data catalogue; current release: http://dx.doi.org/10.4232/1.12665

- All examples in this talk use data of the oldest birth cohort (1990-1993)

Univariate ACE-decomposition

- In behavioral genetics typically estimated using structural equation models (SEM) & twin data formatted *one data row per twin pair* ("wide format")

- Additional assumptions:
  - no non-additive genetic effects
  - no assortative mating
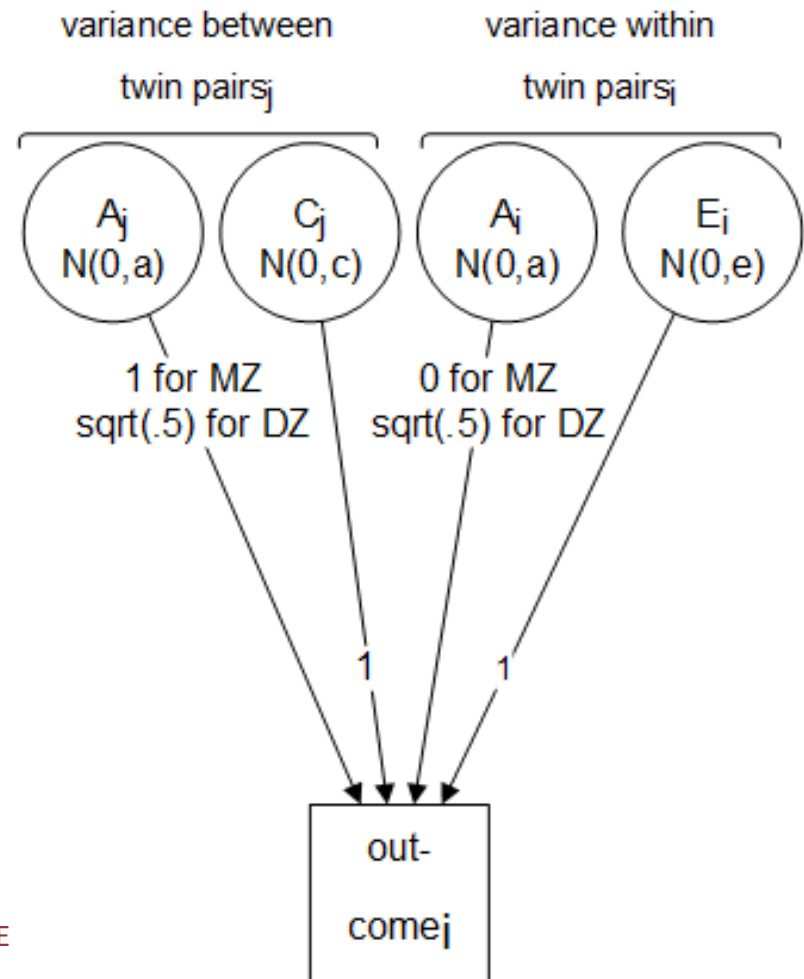 - equal environment effects for MZ & DZ twins (EEA)

(Figure used from Tan et. al (2015))

Multilevel mixed-effect (MME) ACE-decomposition

- Twin data formatted
  one data row per twin
  ("long format", more
  common in social sciences)

- Different implementations:
  - Guo & Wang (2002)
  - McArdle & Prescott (2005)
  - Rabe-Hesketh et al. (2008)

- Same additional assumptions
  like "wide format SEM"

Rabe-Hesketh, Skrondal & Gjessing (2008)-model:

MME ACE-decomposition using Stata: acelong

Necessary information: 1) zygosity of twins: MZ (1) vs. DZ (2) (*zyg*);

2) twin pair identifier (*jid*); 3) twin identifier (*iid*)

Implementation of Rabe-Hesketh et al. (2008)-model using **gsem**:

generate double *aj* = 1

replace *aj* = sqrt(.5) if *zyg* == 2

generate double *ai* = 0

replace *ai* = sqrt(1 - .5) if *zyg* == 2

gsem *outcome* <- C[*jid*]@1 c.*aj*#AJ[*jid*]@1 c.*ai*#AI[*jid*>*iid*]@1, ///

var(AJ[*jid*]@a AI[*jid*>*iid*]@a AJ[*jid*]*AI[*jid*]@0) vce(cluster *jid*)

gsem instead of **meglm** used due to more flexibility in specifying constraints

Alternative implementation using **acelong** (**gsem**-wrapper, Lang 2017):

acelong *outcome zyg jid iid*, vce(cluster *jid*)

MME ACE-decomposition using Stata: acelong

acelong (**gsem**-wrapper, Lang 2017) currently supports:

- Univariate MME ACE, AE & ADE-decompositions

- Different implementations supported: Rabe-Hesketh et al. (2008)-model, Guo & Wang (2002)-model, McArdle & Prescott (2005)-model

- Linear, censored, binary & ordinal outcomes supported

- Absolute & relative ACE-decompositions

- Inclusion of explanatory variables for the mean of the outcome possible

- Flexible specification of DZ twin correlation; e.g., useful for sensitivity tests of no non-additive genetic effects- & no assortative mating-assumptions

- *Currently in beta-testing; available soon!*

Example 1: ACE-decomposition of birth weight

- Birth weight of twins (*bw*) is measured in kg and centered (mean: 2.41 kg)
- Often used as indicator of developmental potential

```
. acelong bw zyg jid iid, vce(cluster jid) nolog

Generalized structural equation model          Number of obs      =       1,494
Response        : bw
Family          : Gaussian
Link            : identity
Log pseudolikelihood = -948.00598
```

bw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mean:						
_cons	-.00003544	.01957597	-0.00	0.999	-.03840363	.03833274
variance:						
A	.01449923	.01196318	1.21	0.226	.00287757	.07305723
C	.23674748	.01587852	14.91	0.000	.20758489	.27000697
E	.07266993	.00639637	11.36	0.000	.06115506	.08635292
A+C+E	.32391663	.01918034	16.89	0.000	.28842331	.36377777
A %	4.4762226	3.6932886	1.21	0.226	.92142848	19.101197
C %	73.089015	4.9020402	14.91	0.000	70.426476	75.594909
E %	22.434763	1.9746959	11.36	0.000	19.575774	25.578501

Example 2: ACE-decomposition of adult height

- Adult height of twins (*ah*) is measured in dm and centered (mean: 17.28 dm)
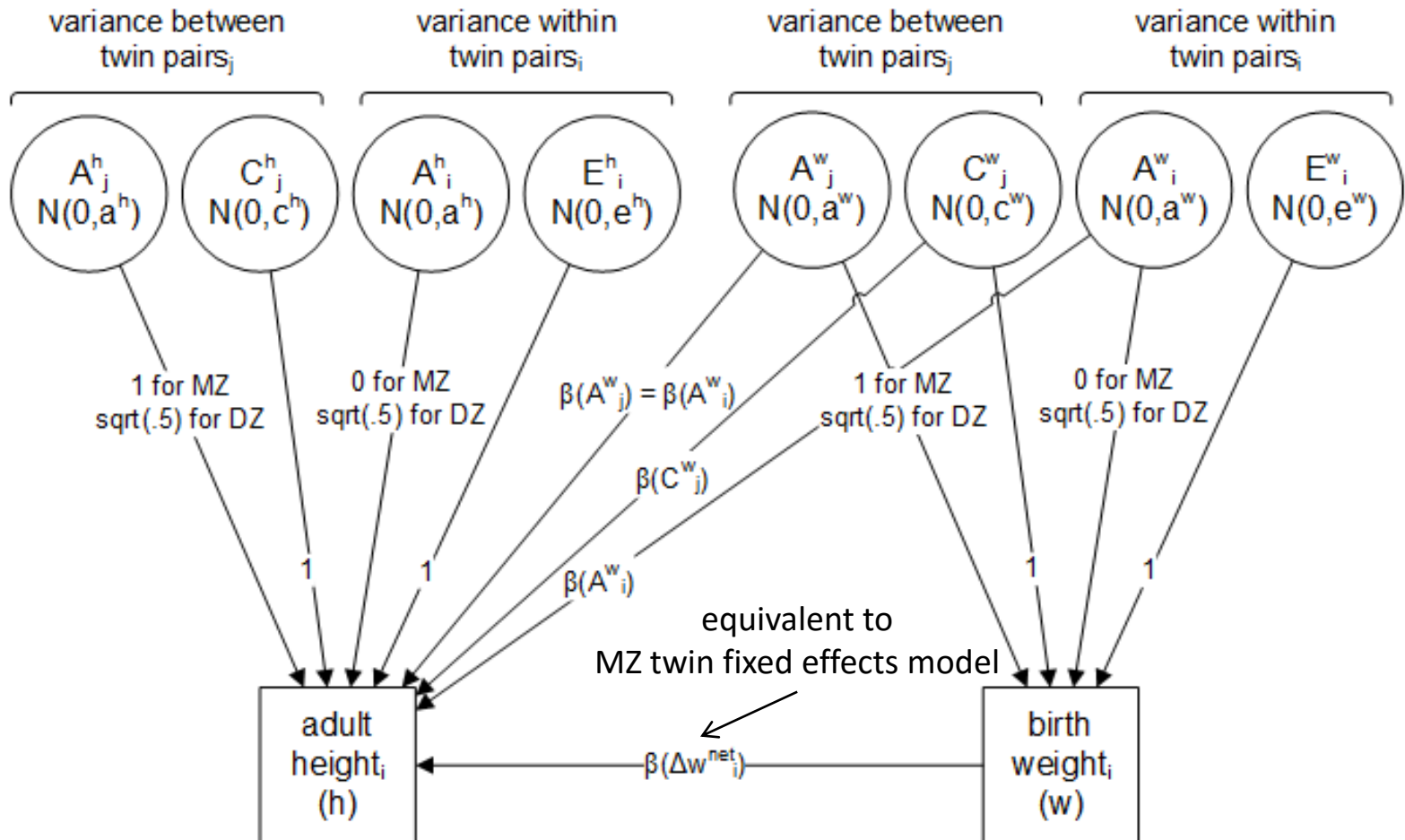
```
. acelong ah zyg jid iid, vce(cluster jid) nolog

Generalized structural equation model          Number of obs     =       1,494
Response        : ah
Family          : Gaussian
Link            : identity
Log pseudolikelihood = -1494.2299
```

ah	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mean:						
_cons	.001461	.03502533	0.04	0.967	-.06718738	.07010938
variance:						
A	.4105211	.04441517	9.24	0.000	.33207956	.50749157
C	.53005424	.05218543	10.16	0.000	.43703576	.64287075
E	.04804335	.00613614	7.83	0.000	.03740391	.06170915
A+C+E	.9886187	.0406705	24.31	0.000	.91203495	1.0716332
A %	41.524716	4.4926494	9.24	0.000	36.485156	46.748032
C %	53.61564	5.2786207	10.16	0.000	48.798086	58.366643
E %	4.8596441	.62067838	7.83	0.000	3.8246119	6.1568496

ACE-β model (MME formulation)

- ACE-β model: Bivariate extension of ACE-decomposition (Kohler et al. 2011)
- Here: MME version of ACE-β model (based on Rabe-Hesketh et al. 2008)

MME ACE-β model using Stata

- Work in progress

- Estimation strategies: a) One-stage maximum likelihood (ML) estimator or

                         b) *Two-stage ML estimator based on plausible values:*

     One-stage estimator is statistically more efficient but

     has more convergence issues (due to large number of random effects) &

     is less flexible regarding extensions (e.g., genXenvironment-interactions)

- Two-stage ML estimator based on plausible values using **acelong**:

     1) Estimate univarite MME ACE-decomposition for the explanatory variable

     2) Generate P plausible values for the A and C components using **predict**

     3) Estimate P univarite MME ACE-decompositions for the outcome including

         the plausible values for the A and C components as explanatory variables

     4) Combine the P results using coefficient & standard error formulas for

         multiple imputed data (Little & Rubin 1989)

Example 3: ACE-β model: Adult height on birth weight

Comparison of different models & estimators:

	ACE-β with PV b / z-value	ACE-β without PV b / z-value	MZ twin fixed effects b / z-value
mean:			
$b(A2)$	2.35 / 2.23**	2.34 / 2.65***	
$b(C2_j)$	0.37 / 3.65***	0.38 / 3.95***	
$b(\Delta w^{net}_i)$	0.29 / 6.81***	0.28 / 6.94***	0.28 / 6.92***
_cons	0.00 / 0.03	0.00 / 0.05	-0.02 / -0.00
variance:			
$A+C_j+E_i$	0.92 / 24.21***	0.92 / 24.40***	
A %	40.12 / 9.56***	40.68 / 9.82***	
C_j %	54.97 / 10.57***	54.70 / 10.91***	
E_i %	4.91 / 7.89***	4.61 / 7.99***	4.31
n(twin pairs)	747	747	408

Concluding remarks

- For the Stata "wish list": **mi** support for **gsem**

  → would make using plausible value estimators easier

- **acelong** is currently in a beta-test cycle;

  if you like to be a beta-tester, please contact me: *vlang@diw.de*

- If you like to use the TwinLife-data for your research,

  please follow instructions on GESIS data catalogue:

  http://dx.doi.org/10.4232/1.12665

Thank you!

Literature

Diewald, M., Riemann, R., Spinath, F. M., Gottschling, J., Hahn, E., Kornadt, A. E.,. . . Peters, A.-L. (2016). TwinLife: GESIS Data Archive. ZA6701 (doi:10.4232/1.12665).

Guo, G., & Wang, J. (2002). The mixed or multilevel model for behavior genetic analysis. Behavior Genetics 32(1), pp37-49.

Kohler, H.-P., Behrman, J. R., & Schnittker, J. (2011). Social science methods for twins data: Integrating causality, endowments, and heritability. Biodemography and Social Biology 57(1), pp88-141.

Lang, V. (2017). The acelong-package: Multilevel mixed-effects ACE, AE and ADE variance decomposition models for "long" formatted twin data using Stata. Working paper (available upon request).

Little, R. J. A., & Rubin, D.B. (1989). The analysis of social science data with missing values. Sociological Methods and Research 18(2-3), pp292–326.

McArdle, J. J., & Prescott, C. A. (2005). Mixed-effects variance components models for biometric family analyses. Behavior Genetics 35(5), pp631–652.

Rabe-Hesketh, S., Skrondal, A., & Gjessing, H. K. (2008). Biometrical modeling of twin and family data using standard mixed model software. Biometrics 64(1), pp280–288.

Tan, Q., Christiansen, L., von Bornemann, J., & Christensen, K. (2015). Twin methodology in epigenetic studies. Journal of Experimental Biology 218, pp134-139.