Graphical Markov Models

3 Lectures

Giovanni M. Marchetti

2024

1 Conditional independence
Graphs (DAG, UG, BG, RCG)
Markov Properties
Conditional Indep. tests
Examples

2 Undirected graphs.
Markov properties
Separation, decomposable graphs
Fitting by ML, Examples

3 Bidirected graphs
Regression chain graphs
Markov properties
Parametrizations
Fitting graphs to data
Examples

Bidirected graph models

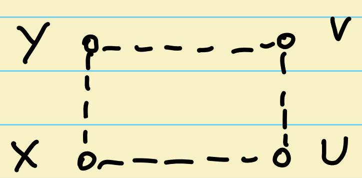
Are graphical models
of marginal independence

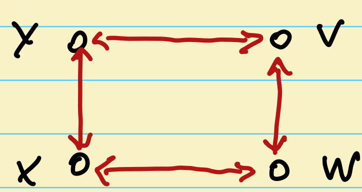
First introduced by Cox and Wermuth (1993)
for Gaussian distributions

Example Study by Kohlmann (1990) 72 students

Strategies to cope with stressful events.

		Y	X	V	U
Y	cognitive avoidance	1	-0.31	0.50	0.23
X	blunting	-0.2	1	0.22	0.51
V	vigilance	0.46	0.00	1	-0.26
U	monitoring	0.01	0.47	-0.15	1





dashed edges

bi-directed edges

Independencies

$Y \perp\!\!\!\perp W \qquad X \perp\!\!\!\perp V$

$\sigma_{YW} = 0 \qquad \sigma_{XV} = 0$

These graphs are called also covariance g.

Fitting Gaussian Bi-directed graphs

```
bg <- 100 * UG(~ Y*X + V*U + X*U + V*Y)
drawGraph(bg)

fitCovGraph(bg, stress, n = 72)

$Shat
           Y          X          V          U
Y  0.9995028 -0.2038079  0.4610985  0.0000000
X -0.2038079  1.0020514  0.0000000  0.4717660
V  0.4610985  0.0000000  1.0015117 -0.1536671
U  0.0000000  0.4717660 -0.1536671  0.9996165

$dev
[1] 0.007675671

$df
[1] 2

$it
[1] 7
```

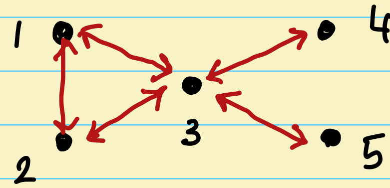
By default, this function gives the ML estimates in the covariance graph model, by iterative conditional fitting (Drton and Richardson, JRSS, 2003). See the help for `fitCovGraph`

Markov properties for Bi-directed graphs

Global MP   Connected set MP

Global MP

If we have a partition $\{A, B, C, D\}$ of V and
A and B are seperated by C
then $A \perp\!\!\!\perp B \mid D$ $D = V \setminus (A \cup B \cup C)$



marginal independencies

1) 1 and 4 separated by $\{2,3,5\}$
$\Rightarrow 1 \perp\!\!\!\perp 4 \mid \emptyset \Rightarrow 1 \perp\!\!\!\perp 4$

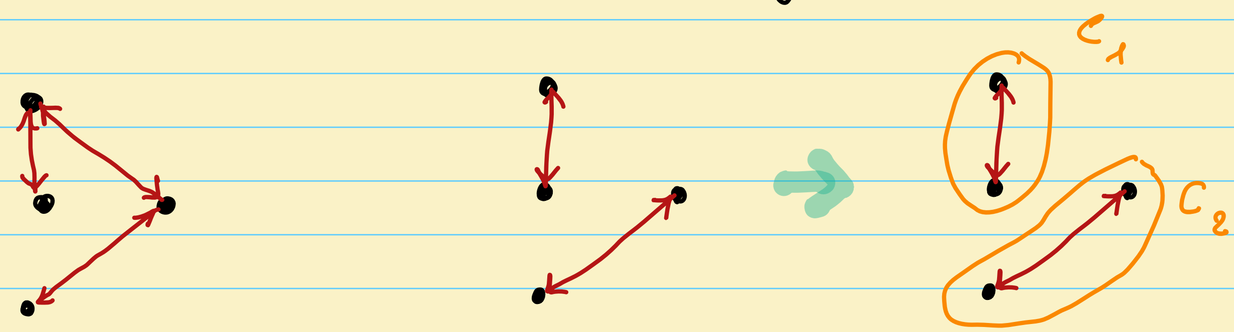
2) 12 and 4 separated by $\{3,5\}$
$\Rightarrow 12 \perp\!\!\!\perp 4 \mid \emptyset \Rightarrow 12 \perp\!\!\!\perp 4.$

but also conditional independencies

3) 1 and 4 separated by 3
$\Rightarrow 1 \perp\!\!\!\perp 4 \mid 2,5$

Connected set MP

A graph is connected if every pair of nodes of V is joined by a path



connected | disconnected | connected components

Definition

The random vector X_V obeys the connected set MP with respect to G

if

for every disconnected subgraph G_A with connected components

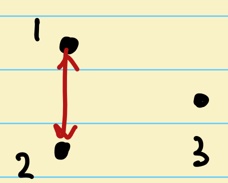
$$C_1, C_2, \dots C_r$$

the joint independence

$$C_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp C_r \quad \text{holds.}$$

Example

G



G_{123}	G_{13}	G_{23}
12 3	1 3	2 3
$12 \perp\!\!\!\perp 3$	$1 \perp\!\!\!\perp 3$	$2 \perp\!\!\!\perp 3$

Relation between the 2 Markov propertie

Global MP $\iff$ Connected set MP

Remarks the connected set MP prevents the violation of the COMPOSITION PROPERTY

Composition property

$$X_1 \perp\!\!\!\perp X_2 \ \& \ X_1 \perp\!\!\!\perp X_3 \Rightarrow X_1 \perp\!\!\!\perp (X_2, X_3)$$

Valid for Gaussian distributions

Not valid for cross-classified Multinomials

Example $X_1 \not\!\perp\!\!\!\perp (X_2, X_3)$

	X_3	0		1		
	X_2	0	1	0	1	sum
X_1	0	0.2	0.05	0.05	0.2	0.5
	1	0.15	0.1	0.1	0.15	0.5
	sum	0.35	0.15	0.15	0.35	1

	$X_2 = 0$	1
$X_1 = 0$	0.25	0.25
1	0.25	0.25

	$X_3 = 0$	1
$X_1 = 0$	0.25	0.25
1	0.25	0.25

$X_1 \perp\!\!\!\perp X_2$ $X_1 \perp\!\!\!\perp X_3$

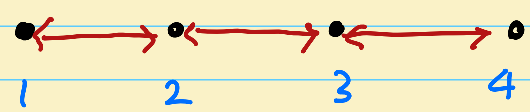
Discrete (categorical) Bi-directed graphical models

Example

Data from Coppen (1966) concerning 4 psychiatric symptoms observed over a sample of 362 patients.

$X1$: Stability (1=extroverted, 2=introverted),
$X2$: Validity (1=psychasthenic, 2=energetic),
$X3$: Depression (yes, no) and
$X4$: Solidity (1=hysteric, 2=rigid).

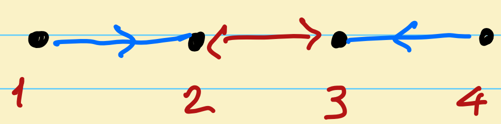
		X_4	1		2	
X_1	X_3	X_2	1	2	1	2
1	y		15	30	9	32
	n		25	22	46	27
2	y		23	22	14	16
	n		14	8	47	12



$X_1 \perp\!\!\!\perp X_3 X_4$ test $G^2_3 = 5.54$ $(p > 0.1)$

$X_1 X_2 \perp\!\!\!\perp X_4$ test $G^2_3 = 3.46$ $(p > 0.3)$

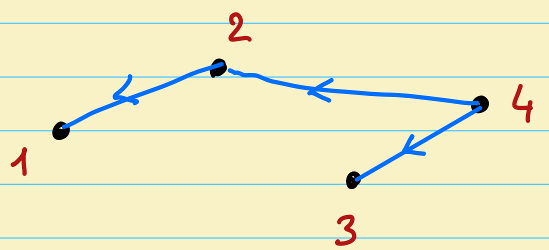
Markov equivalence (see later)



$G^2_5 = 8.6$ $(p > 0.1)$

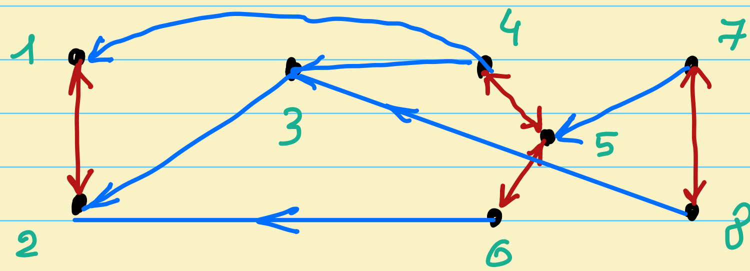
Regression graph models

- univariate regression chain graphs

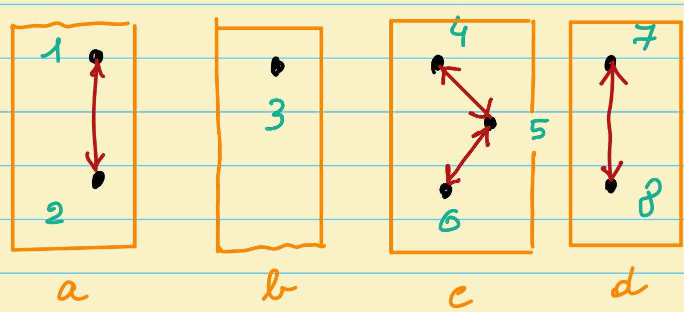


graph = DAG

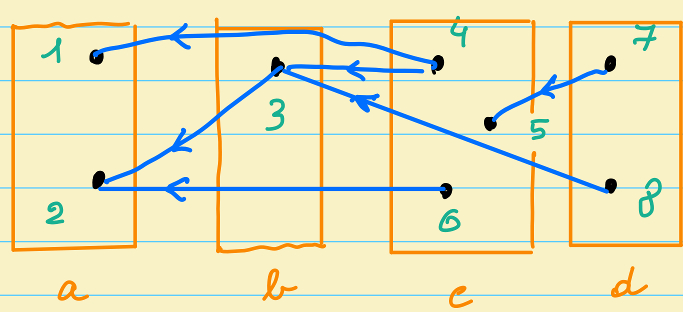
- multivariate regression chain graphs



Both have "blocks" called chain components

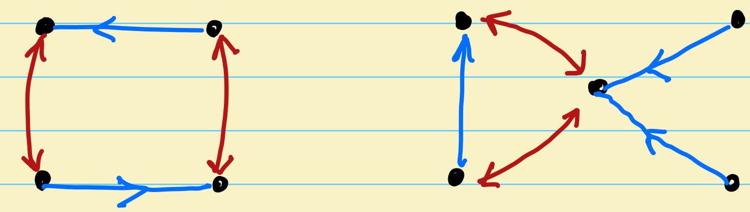


- The subgraphs of the chain components are bidirected



- The directed edges connect nodes in distinct blocks $i \leftarrow j$ with $i < j$.

Examples that are not regression chain graphs



Why the regression chain graphs are useful?

Example Wermuth & Cox (1995, CSDA)

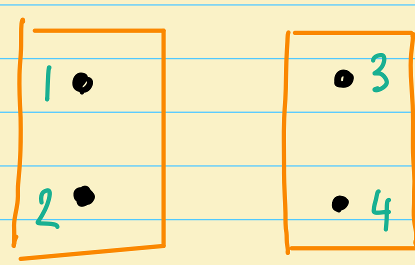
On 44 healthy female patients were collected data on

X_1 log (syst/diast) blood pressure

X_2 log (diast) blood pressure

X_3 weight/(100 height) "BMI"

X_4 age



joint responses

explanatory variables

Regression graph model for Gaussian data

Assume $\underline{X} = (X_1, X_2, X_3, X_4) \sim N(\underline{\mu}, \Sigma)$

call $a = (1, 2)$ $\quad b = (3, 4)$

$$E(\underline{X}) = \underline{\mu} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \qquad cov(\underline{X}) = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \cdot & \Sigma_{bb} \end{pmatrix}$$

We consider the factorization $P_{ab} = P_{a|b} \, P_b$

Model for $P_{a|b}$ consider $X_a \mid X_b$

$$X_a \mid X_b \sim N(\underline{\mu}_{b|a}, \Sigma_{aa|b})$$

equations :
$$X_1 = \beta_{13} x_3 + \beta_{14} x_4 + \varepsilon_1$$
$$X_2 = \beta_{23} x_3 + \beta_{24} x_4 + \varepsilon_2$$

with
$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} \sigma_{11|b} & \sigma_{12|b} \\ \cdot & \sigma_{22|b} \end{pmatrix} \right)$$

Model for p_b Consider $X_b \sim N(\mu_b, \Sigma_{bb})$

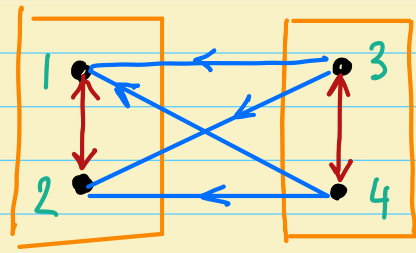
with $$\Sigma_{bb} = \begin{pmatrix} \sigma_{33} & \sigma_{34} \\ \cdot & \sigma_{44} \end{pmatrix}$$

Independence Constraints

$1 \leftarrow 3 = \beta_{13}$

$1 \leftarrow 4 = \beta_{14}$

$1 \leftrightarrow 2 = \sigma_{12|b}$



$3 \leftrightarrow 4 = \sigma_{34}$

$2 \leftarrow 3 = \beta_{23}$

$2 \leftarrow 4 = \beta_{24}$

Interpretation

$\beta_{13} = 0 \iff 1 \perp\!\!\!\perp 3 \mid 4$

$\beta_{14} = 0 \iff 1 \perp\!\!\!\perp 4 \mid 3$

$\beta_{23} = 0 \iff 2 \perp\!\!\!\perp 3 \mid 4$

$\beta_{24} = 0 \iff 2 \perp\!\!\!\perp 4 \mid 3$

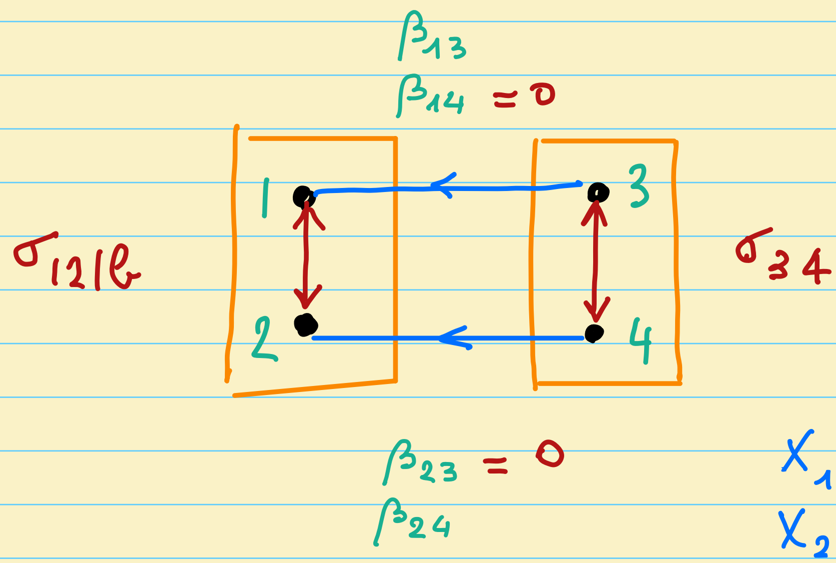
$\sigma_{12|b} = 0 \iff 1 \perp\!\!\!\perp 2 \mid 34$

$\sigma_{34} = 0 \iff 3 \perp\!\!\!\perp 4$

MLE fitting of saturated model

Edge	Estimate	se	est/se
$1 \leftarrow 3$	−0.0042	0.0029	−1.483
$1 \leftarrow 4$	0.0003	0.0016	0.196
$2 \leftarrow 3$	0.0009	0.0037	0.244
$2 \leftarrow 4$	0.0060	0.0021	2.940
$1 \leftrightarrow 2$	−0.0055	0.0017	−3.230

Reduced model



$X_1 \perp\!\!\!\perp X_4 \mid X_3$
$X_2 \perp\!\!\!\perp X_3 \mid X_4$

Seemingly unrelated model

Zellner (1962)

$$X_1 = \beta_{13} X_3 + \varepsilon_1$$
$$X_2 = \beta_{24} X_4 + \varepsilon_2$$
$$X_3 = \varepsilon_3$$
$$X_4 = \varepsilon_4$$

with

$$(\varepsilon_3, \varepsilon_4) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_{33} & \sigma_{34} \\ \cdot & \sigma_{44} \end{pmatrix} \right)$$

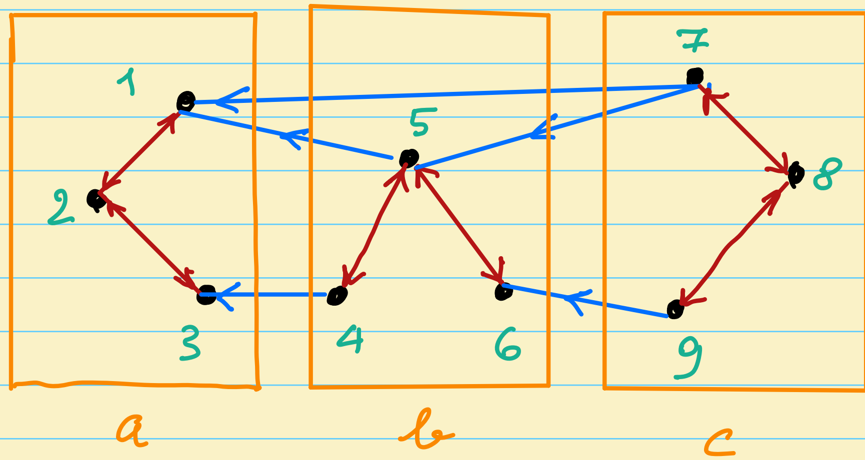
$$(\varepsilon_1, \varepsilon_2) \sim N\left(\begin{pmatrix} \beta_{13} X_3 \\ \beta_{24} X_4 \end{pmatrix}, \begin{pmatrix} \sigma_{11|b} & \sigma_{12|b} \\ \cdot & \sigma_{22|b} \end{pmatrix} \right)$$

The reduced model requires iteration

Use for example R package sem

→ cw_example

Ordered Regression chain
Markov Property



$a = (1,2,3) \qquad b = (4,5,6) \qquad c = (7,8,9)$

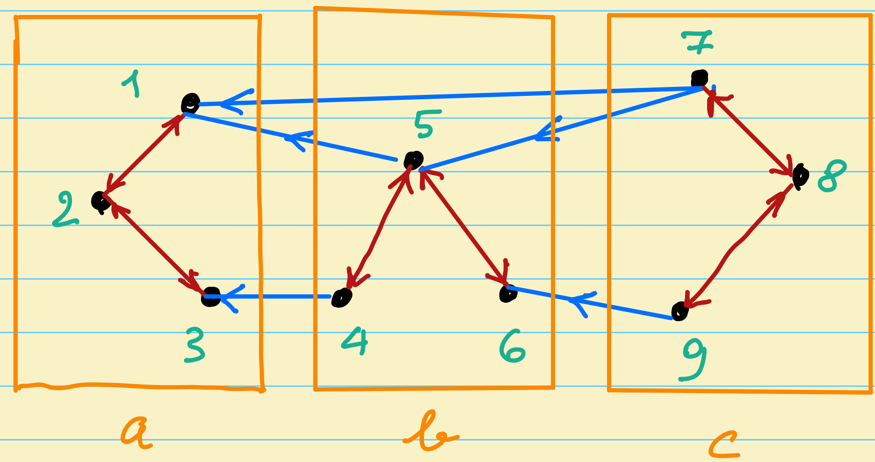
Factorization $\quad P_V = P_{a|bc}\, P_{b|c}\, P_c$

$$= P_{a|457}\, P_{b|79}\, P_c$$

Constraints for disconnected sets in block a

a $\quad P_{13|457} = P_{1|457} \times P_{3|457}$ $\quad$ 13 disconnected

$$X_1 \perp\!\!\!\perp X_3 \mid X_4 X_5 X_7$$



Constraints for connected sets in block a

$$P_{1|457} = P_{1|57} \qquad P_{12|457} = P_{12|57}$$
$$P_{2|457} = P_{2} \qquad P_{23|457} = P_{23|4}$$
$$P_{3|457} = P_{3|4} \qquad P_{123|457} = \text{no constraint}$$

Independencies for block a

$$X_1 \perp\!\!\!\perp X_3 \mid X_4 X_5 X_7$$

$$X_1 \perp\!\!\!\perp X_6 X_8 X_9 \mid X_4 X_5 X_7 \qquad X_1 X_2 \perp\!\!\!\perp X_4 \mid X_5 X_7$$
$$X_2 \perp\!\!\!\perp X_4 X_5 X_7 \qquad X_2 X_3 \perp\!\!\!\perp X_5 X_7 \mid X_4$$
$$X_3 \perp\!\!\!\perp X_5 X_7 \mid X_4$$

Regression chain graph
Global Markov property

Is a generalization of d-separation

A path contains a <u>collider</u> if

$\rightarrow m \leftarrow$ or $\rightarrow m \leftrightarrow$

A path is blocked given C iff

- the path contains a chain $\rightarrow m \rightarrow$
  or a fork $\leftarrow m \rightarrow$
  such that m is in C

- or the path contains a collider (see above)
  such that m is not in C and
  no descendants of m are in C

Two subsets of nodes A an B are said m-separated given C if every path between A and B is blocked by C

Check with 'msep' in ggm
$\rightarrow$ msep-example

MLE of a regression graph for Gaussian variables

The log-likelihood is

$$\ell(\Sigma) = -\frac{N}{2}\log|\Sigma| - \frac{N}{2}\operatorname{tr}(\Sigma^{-1}S)$$

The regression chain graph model is the family of normal distributions with Σ positive def. matrices that satisfy the constraints defined by the graph.

MLE

Existence

the condition $N \geq d$ is sufficient

Uniqueness

can have multiple local maxima

General estimation methods

See

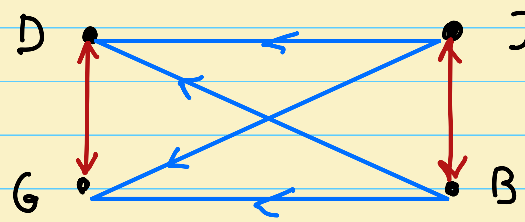
- SEM (structural equation models)
- Iterative Conditional Fitting

→ Drton & Richardson (2004, UAI)

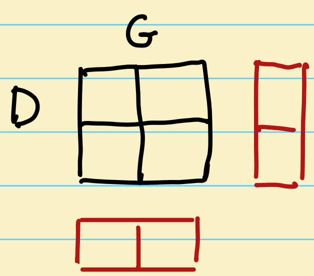
MLE of discrete Regression chain graph models

Example 4 variables from the U.S. General Social Survey (1972 – 2006)

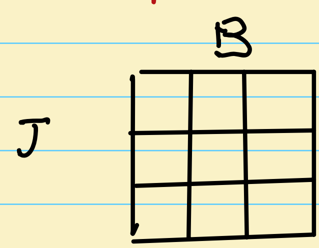
D Death penalty (1 = favor, 2 = oppose)
G Require permit to buy a gun (1, 2)
J Job satisfaction (1 = +, 2 = •, 3 = −)
B confidence in Banks (1 = +, 2 = •, 3 = −)



Responses D, G



Explanatory J, B



Parameters (multivariate logistic)

$$\eta_D = \text{logit}_D \sim J + B$$

$$\eta_G = \text{logit}_G \sim J + B$$

$$\eta_{DG} = \log OR_{DG} \sim 1$$ (i.e. it's constant)

Results

	D			*G*			*D* ∗ *G*		
Model (2)	β	se	p	β	se	p	β	se	p
(const.)	−1.040	-	-	−1.190	-	-	−0.371	0.05	0.000
J_2	0.099	0.04	0.013	0.041	0.04	0.305			
J_3	0.251	0.06	0.000	−0.010	0.06	0.868			
B_2	0.004	0.04	0.920	−0.005	0.05	0.920			
B_3	0.080	0.06	0.182	0.141	0.06	0.019			
Model (3)	β	se	p	β	se	p	β	se	p
(const.)	−1.028	-	-	−1.178	-	-	−0.371	0.05	0.000
J_2	0.103	0.04	0.01						
J_3	0.256	0.06	0.00						
B_2				−0.003	0.05	0.952			
B_3				0.147	0.06	0.014			

model (2) $G^2_{16} = 16.1$

model (3) $G^2_{20} = 19.2$