University of Groningen

The Links in our Mind

Balafas, Spyros

DOI:
[10.33612/diss.202221392](https://doi.org/10.33612/diss.202221392)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Balafas, S. (2022). *The Links in our Mind: Multivariate Statistical Methods for Psychological Data*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.
<https://doi.org/10.33612/diss.202221392>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

university of
 groningen

The Links in our Mind

Multivariate Statistical Methods for Psychological Data

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on
Tuesday 22 February 2022 at 16.15 hours

by

Spyros Balafas

born on 12 July 1985
in
Korinthos, Greece

Supervisors

Prof. E.C. Wit

Prof. M.A. Grzegorzczuk

Assessment Committee

Prof. C.J. Albers

Prof. M.C. Di Serio

Prof. L.J. Waldorp

Submitted thesis material

- **Chapter 2:** Rimvall, M., Wigman, J., Balafas, S., Rosmalen, J., Verhulst, F., van Os, J., Rask, C., & Jeppesen, P. (2020). Functional somatic symptoms and psychotic experiences in adolescence: A network study across two general population cohorts of adolescents. *Journal of Psychosomatic Research*, 133, [110052].
<https://doi.org/10.1016/j.jpsychores.2020.110052>
- **Chapter 2:** Rimvall, M., Wigman, J., Balafas, S., Wit, E. C., Rosmalen, J., Verhulst, F., van Os, J., Rask, C., & Jeppesen, P. (2021). Symptom networks of psychotic experiences and functional somatic symptoms in adolescence: an observational study of two population-based cohorts.
(Under review in *European Child & Adolescent Psychiatry*)
- **Chapter 3:** Balafas, S. E., Booij, S. H., Wigman, J., Wit, E. C., Grzegorzczak, M. A. (2021). Graphical Vector Autoregressive Models in Mental Health Research: A Review with a Tutorial in R
(Under review in *Psychological Methods*)
- **Chapter 4:** van der Tuin, S., Balafas, S. E., Booij, S. H., Wigman, J., Wit, E. C., Oldehinkel, A. J. (2021). Dynamic Symptom Networks Across Different At-risk Stages for Psychosis: An Individual and Transdiagnostic Perspective
(Under review in *Schizophrenia Research*)
- **Chapter 6:** Balafas, S. E., Krijnen, W. P., Post, W. J., Wit, E. C. (2020). mudfold: An R Package for Nonparametric IRT Modelling of Unfolding Processes, *The R Journal*, 12:1, pages 49-75.
<https://doi.org/10.32614/RJ-2020-002>

“It is not easy to deal scientifically with feelings.”

Sigmund Freud

To my parents

Acknowledgements

Throughout my PhD research I received a great deal of support. Foremost, I would like to express my deepest gratitude to my supervisor Prof. Ernst C. Wit for his help, guidance, unconditional support, and for being a source of inspiration, during all these years of working together. The completion of this dissertation (and not only) would not be possible without Ernst. I would like also to give my gratitude to my second supervisor Prof. Marco A. Grzegorzczak for his valuable comments on my research, support, and guidance, and for the nice conversations during my time in the Netherlands.

It is impossible to acknowledge every person who contributed to the completion of this thesis with one way or another. In the remaining of this section, I will try to mention the most important of them. I need to thank Prof. Hanneke Wardenaar–Wigman for giving me the opportunity to work with her, for introducing me into the world of psychiatry, for the support during my PhD, and all the help whenever I needed it. I also want to thank Sanne Booij for the collaboration in chapters of this dissertation, and for the help in the understanding of research procedures, designs, and practice of psychiatry. I am thankful to Prof. Wendy Post for working together from my master studies until today, and for her input on the sixth chapter of this dissertation. I am more than grateful to Wim P. Krijnen for the help across the years, for the insightful comments, and the general collaboration on the sixth chapter of this dissertation.

In addition, I want to thank my colleague Sara van der Tuin for working together on the fourth chapter of this thesis, and for being my paronymph. I need to thank especially my paronymph, colleague, and close friend Giacomo Ceoldo for the support during my PhD. I want to thank also Martin Rimvall for our collaboration on the second chapter of this thesis. I am more than grateful to all my (former) colleagues and teachers: Ineke Schelhaas, Fentaw Abegaz, Reza Mohammadi, Mirko Signorelli, Pariya Behrouzi, Seyed Mahdi Mahmoudi, Mahdi Shafiee Kamalabad, Francisco Richter, Igor Artico, Federica Bianci, Ruta Juozaitiene, Luca Del Core, Lucas Canias. Also, I would like to thank Konstantina Drakou for helping me with the cover of this dissertation, and Evi Kampouri for her comments on the thesis. Finally, I need to thank everyone who supported me in completing this dissertation and I did not mention in this section.

Abstract

Psychopathological processes are characterized by multiple interacting dimensions, that in order to be understood properly have to be studied simultaneously. In this thesis multivariate statistical methods, such as item response theory and graphical models, have been proposed to model psychological observables for explaining complex mental phenomena. After introducing the chronology of multivariate analysis in the behavioural sciences, we describe mixed graphical models for cross-sectional data from two cohort studies, in order to study the relationship between functional somatic symptoms and psychotic experiences in adolescence. Then we focus on the statistical modelling of psychological time-series data and specifically on the class of Graphical Vector Autoregressive (GVAR) models. We implemented GVARs in open-source software that can be used for fitting these models to psychological data with different characteristics. Next, we apply the GVAR to time series data from psychopathology for studying the development of psychosis according to subsequent stages of illness severity. The topic of multivariate time series analysis in psychopathology is concluded by a newly developed hierarchical extension of the traditional GVAR model. This extension can be used for modelling simultaneously time series data from multiple individuals taking into account individual heterogeneity. The thesis is concluded with a software implementation of the nonparametric item response theory model for unfolding processes MUDFOLD, which has been extended to provide uncertainty quantification in parameter estimation using bootstrap, and which handles missing values using multiple imputation by chained equations.

Table of contents

List of figures	xix
List of tables	xxiii
1 Introduction	1
1.1 A brief history of psychometrics	3
1.2 Item response theory models	6
1.2.1 The item response function	7
1.2.2 Parameter estimation in IRT models	9
1.3 Psychological network models	10
1.3.1 Graphical models	11
1.3.2 Undirected graphical models	12
1.3.3 Directed graphical (DAG) models	16
1.3.4 Chain graphical models	17
1.3.5 Graphical vector autoregression	19
1.4 Outline	22
2 Psychotic experiences and functional somatic symptoms in adolescence. A network approach	25
2.1 Introduction	25
2.2 Methods	27
2.2.1 Study populations	27
2.2.2 Measures	28
2.2.3 Statistical analyses	29
2.3 Results	32
2.3.1 Data distribution	32
2.3.2 Symptom-networks across the two cohorts.	32
2.4 Discussion	36
2.4.1 Main findings	36

2.4.2	Methodological considerations	38
2.4.3	Interpretation	38
2.4.4	Explanations for the denser networks in CCC2000	39
3	Exploring the dynamic structure of mental health with graphical vector autore- gression	41
3.1	Introduction	41
3.2	Methods	45
3.2.1	Graphical VAR	45
3.2.2	Lag estimation	46
3.3	Methodological challenges and extensions of the graphical VAR	48
3.3.1	VAR models for high-dimensional symptom time series	49
3.3.2	Non-stationary VAR models	50
3.3.3	Non-Gaussian VAR models	52
3.3.4	VAR models for panel data	54
3.3.5	Missing data in intensive longitudinal studies	57
3.3.6	Unequal spacing in intensive longitudinal data	59
3.4	Estimating graphical VAR models in R	60
3.4.1	Stationary graphical VAR on simulated data	60
3.4.2	Time-varying graphical VAR	62
3.4.3	Stationary inter-individual symptom dynamics	66
3.4.4	MOOVD study: Ordinal intra-individual network dynamics from non-equidistant data	68
3.5	Some final considerations	70
3.6	Conclusions	70
4	Exploring the structure of transdiagnostic psychopathology symptoms	73
4.1	Introduction	73
4.2	Methods	75
4.2.1	Participants and study design	75
4.2.2	Instruments	76
4.2.3	Statistical analyses	78
4.3	Results	79
4.3.1	Study sample characteristics	79
4.3.2	Comparison of contemporaneous symptom networks between sub- groups	80
4.3.3	Comparison of temporal symptom networks between subgroups	81

4.4	Discussion	82
4.4.1	Major findings	82
4.4.2	Strengths and limitations	85
5	Sparse hierarchical vector autoregression	87
5.1	Introduction	87
5.2	Symptom networks for the onset and progression of psychosis	90
5.3	Methods	91
5.3.1	Penalized inference	95
5.3.2	Selecting the optimal regularization parameter	103
5.4	Simulations	104
5.4.1	Fixed effects estimation	105
5.4.2	Random effects covariance estimation	108
5.5	Application	109
5.6	Discussion	113
6	Nonparametric IRT modelling of unfolding processes	115
6.1	Introduction	115
6.2	Methodology	117
6.2.1	Assumptions of the nonparametric unfolding IRT model	118
6.2.2	Errors and Scalability Coefficients	120
6.2.3	Scale construction	123
6.2.4	MUDFOLD diagnostics	127
6.2.5	Uncertainty estimates for MUDFOLD statistics	131
6.2.6	Nonparametric estimation of person ideal points	131
6.2.7	Missing values	133
6.3	The mudfold package	133
6.3.1	Description of the functions <code>mudfold()</code> and <code>as.mudfold()</code>	133
6.3.2	Description of the generic functions	135
6.3.3	The <code>diagnostics()</code> function	137
6.3.4	How to generate data from unfolding processes and description of the function <code>mudfoldsim()</code>	138
6.3.5	Description of the <code>pick()</code> function	140
6.4	Applications	142
6.4.1	Loneliness data	142
6.4.2	Plato's seven works data	146
6.5	Summary	149

References	151
Appendix A Appendix for Chapter 3	175
A.1 Code	175
A.1.1 Example 1	175
A.1.2 Example 2	176
A.1.3 Example 3	179
A.1.4 Example 4	180
Appendix B Appendix for Chapter 4	183
B.1 Testing different imputation strategies	183
B.1.1 Code for testing different imputation strategies	183
B.2 Testing TSCGM assumptions	187
B.3 Additional checks	187
Appendix C Appendix for Chapter 5	189
C.1 Alternative coordinate-wise algorithm for RPQL optimization	189

List of figures

1.1	A) Cattell's data box (Cattell, 1952) where each piece is an observation taken from one individual on one variable at a single time point. Psychological studies often use slices of the data cube. B) they measure sets of variables in multiple individuals, but only once in time (nomothetic approach), and C) they measure sets of variables in multiple occasions, but only from one individual (idiographic approach).	2
1.2	The IRF for two different response processes. The probability of a positive response to an item as a function of the latent trait for a cumulative IRT model (left). The probability of a positive response to an item as a function of the latent trait for an unfolding IRT model (right).	8
1.3	This figure shows two applications of networks in psychology. (left) A figure taken from (Moreno, 1934, p.32), where a network structure was used to describe the mechanism of group formation. (right) An illustration of a theoretical symptom network from psychopathology.	11
1.4	Undirected graph that depicts the joint distribution of five psychological symptoms.	14
1.5	Directed acyclic graph that depicts the joint distribution of five psychological symptoms.	17
1.6	Time series chain graph that describes the first order dynamics of a system consisting of five psychological symptoms.	18
2.1	Left-hand side: Undirected networks without accounting for emotional problems for the TRAILS (upper panel) and CCC2000 (lower panel) cohorts respectively. Right-hand side: Undirected networks after accounting for emotional problems for the TRAILS (upper panel) and CCC2000 (lower panel) cohorts respectively.	34

2.2	Centrality indexes for the networks with and without emotional problems in the model across the two cohorts of adolescents. Centralities for the TRAILS cohort are shown in blue while centralities for the CCC2000 cohort are shown in red.	35
2.3	Bridge centrality indexes for the networks with and without emotional problems in the model across the two cohorts of adolescents. Bridge centralities for the TRAILS cohort are shown in blue while bridge centralities for the CCC2000 cohort are shown in red.	36
2.4	Similarities and differences in the network structure between the two cohorts.	37
3.1	An example of a time series chain graph from a trivariate VAR(1) model. Directed edges (solid links with arrows) representing non-zero entries of the transpose of the matrix $\mathbf{A}^\top$, which correspond to non-zero temporal partial correlations. Undirected edges (dashed links) characterize the entries of Σ_ϵ^{-1} .	47
3.2	An example of a time series chain graph from a trivariate TVVAR(1) model. Directed edges (solid links with arrows) representing non-zero entries of the transpose of the matrix $\mathbf{A}^\top$, which correspond to non-zero temporal partial correlations. Undirected edges (dashed links) characterize the entries of Σ_ϵ^{-1} .	52
3.3	The data generating model parameters (left hand side) versus the estimated model parameters (right hand side) for a system of $p = 10$ symptoms.	61
3.4	The marginal distribution of the symptom 'I prefer stay alone'.	63
3.5	A time-varying VAR(1) on data that has been transformed using a Gaussian Copula. The process has been estimated on 20 estimation points in time and the resulting network in each time instance is shown.	64
3.6	Time variation of selected temporal interactions.	65
3.7	Symptom networks for two groups of participants in the MIRORR study. Green-colored nodes denote positive symptoms and red-colored nodes denote negative symptoms. Gray-colored links denote positive effects while black-colored links denote negative effects. Upper panel: Temporal networks for the first group (left-hand side) and the second group (right-hand side). Lower panel: Contemporaneous networks for the first group (left-hand side) and second group (right-hand side).	67
3.8	Temporal network for a random individual in the MOOVD study with lagged effects between symptoms from the previous time point $t - 1$ to the current time point t	69

- 3.9 Centrality measures for the temporal network of a depressed individual in the MOOVD study. Symptoms are plotted on the x-axis in decreasing order. In the upper panel, the in-degree for each symptom is shown on the left-hand side and the out-degree on the right-hand side. In the two bottom panels, you can see the step one expected influence (left-hand side) and the step two expected influence (right-hand side) for each symptom. 71
- 5.1 The average ROC curve for the fixed effects across 10 simulated datasets using the RPQL (left hand side) and the SPLMM (right hand side) approaches respectively. Each row of the plot corresponds to a different scenario which is compared with the baseline. The ROC from the baseline scenario is plotted in red color, while in blue are the ROCs from all the alternative scenarios. . . 106
- 5.2 The average ROC curve for the random effects across 10 simulated datasets using the RPQL (left hand side) and the SPLMM (right hand side) approaches respectively. Each row of the plot corresponds to a different scenario which is compared with the baseline. The ROC from the baseline scenario is plotted in red color, while in blue are the ROCs from all the alternative scenarios. . . 107
- 5.3 The estimated temporal (left-hand side) and contemporaneous (right-hand side) graphs for each group of participants in the MIRORR study. The groups of participants are ordered according to their risk for psychosis. The graphs from the first group (control) with the lower level of risk for psychosis are shown in the two upper panels, while the graphs from the fourth group with the highest risk for psychosis shown in the two bottom panels. 111
- 5.4 Three different standardized centrality indices for the symptoms in the temporal and contemporaneous networks of each subgroup in the MIRORR study. Centralities for the temporal networks are shown in the left-hand side while the centralities for the contemporaneous networks are shown on the right-hand side. The red lines correspond to the *degree* centrality, the green lines correspond to the *strength* centrality, and the blue lines correspond to the *expected influence* centrality of step two. 112
- 6.1 Left hand side: red squares in the lower triangular part of the matrix represent pairs of conditionally dependent items. Right hand side: red squares represent the position of the observed maxima in the CAM rows for the *Loneliness* unfolding scale. 145
- 6.2 The estimated item response function for each item in the *Loneliness* unfolding scale. 145

6.3	MUDFOLD's empirical estimates of the IRFs for the <i>Loneliness</i> unfolding scale.	146
6.4	The distribution of the estimated person ranks for the <i>Loneliness</i> unfolding scale.	146
6.5	MUDFOLD's empirical estimates of the IRFs for the items in the <i>Plato7</i> unfolding scale.	149
6.6	MUDFOLD's estimates of the IRFs for the items in the <i>Plato7</i> unfolding scale.	149

List of tables

2.1	Distribution of psychotic experiences and functional somatic symptoms in two cohorts of adolescents.	33
3.1	Possible choices for the regularization term $C(T, d, p)$ from (3.5). Each choice of $C(T, d, p)$ corresponds to a different information criterion.	48
3.2	Estimation results of the simulated GVAR(1) model. The recovery of the parameters at the temporal and contemporaneous parts of the model in terms of true positive and true negative rates are presented.	62
4.1	Diary items selected for the analyses.	77
4.2	Demographics and clinical functioning per subgroup	79
4.3	Within-person untransformed median and Inter Quartile Range of scores of the diary items (0 – 100) averaged per subgroup	80
4.4	Descriptive information per item on contemporaneous network density and node strength	80
4.5	Percentage of the times that an item had the highest strength (top 2) in individual symptom networks	81
4.6	Descriptive information per item on temporal network density and node in- and out-strength	82
4.7	Descriptive information per item on temporal network density and node in- and out-strength	83
5.1	The symptoms from MIRORR that are considered in this example. The first column shows how each symptom is coded. The second column provides for each symptom the exact self-report to which a participant needs to provide his response. The third column shows the scale on which is symptom is measured. In the MIRORR example all symptoms are measured on a continuous scale bounded on the range [0, 100].	110