

**Asymmetric Causal Relationship Through Cumulative Explained Variability
with Continuous Piecewise Affine Function**

by

Siyu Zhang

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2022

Date of final oral examination: 07/15/2022

The dissertation is approved by the following members of the Final Oral Committee:

Zhengjun Zhang, Professor, Statistics

Lu Mao, Assistant Professor, Biostatistics and Medical Informatics

Kris Sankaran, Assistant Professor, Statistics

Vikas Singh, Professor, Biostatistics and Medical Informatics

Chunming Zhang, Professor, Statistics

© Copyright by Siyu Zhang 2022

All Rights Reserved

To my dear parents

Acknowledgments

The five years' Ph.D. study is one of the most precious experience in my whole life. I cannot imagine how lucky I am to know and be supervised by Professor Zhengjun Zhang, from the very first time me visiting the campus. He is the best role model of being a great supervisor and statistician for me, with curiosity, enthusiasm, and desire for exploration in research, as well as consideration, patience, and continuous support for guiding me through my Ph.D. study. His encouragement and suggestions are always the strongest motivation for me, especially during tough times of research. It is my honor to be his Ph.D. student.

Then, I would like to thank my committee members: Professor Lu Mao, Professor Kris Sankaran, Professor Vikas Singh, and Professor Chunming Zhang. They have provided very helpful guidance and suggestions for improving this thesis work. I really appreciate their valuable time and support.

Moreover, I want to thank our former and current group members, Jili Wang, Yuqing Xu, Jingyu Ji, Haoyang Teng, and Yongjoon Kim. They have provided me many suggestions and help during the five years in Madison for my research, course study, and even job hunting. My true and loyal friends, Andrea Gao, Shan Lu, Sheng Wang, Yuqin Mu, also deserve to be acknowledged here. Sharing the joy and pain with them comforted me a lot, not to mention their encouragement led me through many difficulties.

Special thanks to my parents and Mr. You'an Chen, who is a friend of our family and like a godfather to me. They strongly supported my Ph.D. study and provided me the best they have all the time and never asked me for any. This accomplishment would not be possible without their endless love.

Contents

Abstract	xvi
1 Introduction	1
2 WeCon Causality Based on Cumulative Explained Variability	9
2.1 Generalized measures of correlation: A brief review	10
2.2 WeCon causality	12
2.3 Cumulative explained variability with continuous piecewise affine function .	13
2.3.1 Piecewise affine function	14
2.3.2 CEV-PWA estimators	17
3 High-dimension Predictors	20
3.1 Dimension reduction: SEVIS screening	21
3.1.1 Sure explained variability	21
3.1.2 Sure explained variability and independence screening	22
3.2 Cumulatively explained variability with continuous piecewise affine function estimators	23
3.3 Parameter Estimation	25
3.3.1 Estimation via Genetic Algorithms	25
3.3.2 CEV-PWA algorithm	28
3.4 Simulation	29
3.5 Real data application	34

3.5.1	EEG Application: Visual Categorization	34
3.5.2	EEG Application: Parkinson's Disease	40
3.5.3	fMRI Application: Major Depressive Disorder	50
4	Alternative Approach	64
4.1	Principal component approximation	65
4.2	Partial least squares approximation	68
4.3	Simulation	71
4.3.1	Simulation Example 1	72
4.3.2	Simulation Example 2	79
4.4	Real data application	84
4.4.1	fMRI Application: Major Depressive Disorder	84
5	Discussion	91
5.1	Comparison	92
5.2	Conclusion	98
5.3	Future Perspectives	101
A	Appendix	105
A.1	Figures for EEG application: Visual Categorization	107
A.2	Figures for EEG application: Parkinson's Disease	111
A.3	Figures for fMRI application: Major Depressive Disorder	132
	Bibliography	164

List of Tables

3.1	Simulation Statistics.	29
3.2	Simulation Results for Example 1.	31
3.3	Simulation Results for Example 2.	33
3.4	CEV results for top 2 selected highest asymmetric nonlinear correlation channels.	38
3.5	CEV results for top 2 selected highest asymmetric nonlinear correlation channels. (Continued)	39
3.6	The frequency of each channel being selected as top 2 highest asymmetric nonlinear correlation channels.	40
3.7	Harvard Oxford Atlas: IDs and corresponding ROI names	52
4.1	CEV estimators by principal component analysis for both all predictors and partial predictors scenarios.	68
4.2	CEV estimators by partial least squares regression for both all predictors and partial predictors scenarios.	70
4.3	Simulation statistics for alternative approaches.	71
4.4	Simulation results for CEV-PCA using partial predictors for Example 1.	74
4.5	Simulation results for CEV-PCA using all predictors for Example 1.	75
4.6	Simulation results for CEV-PLS using partial predictors for Example 1.	77
4.7	Simulation results for CEV-PLS using all predictors for Example 1.	78
4.8	Simulation results for CEV-PCA using partial predictors for Example 2.	80

4.9	Simulation results for CEV-PCA using all predictors for Example 2.	81
4.10	Simulation results for CEV-PLS using partial predictors for Example 2. . .	82
4.11	Simulation results for CEV-PLS using all predictors for Example 2.	83

List of Figures

3.1	Genetic Algorithm Flow Chart from Singh and Sharma (2014)	26
3.2	(a) shows the distribution of channels in human brain; (b) shows the connectivity network of the channels identified by our model.	35
3.3	Average CEV values for each channel in visual categorization. Channels C4, F3, F4, F7, F8, FP2, O2, P4r, PZ, T4, T5r, T6r are included.	36
3.4	Average CEV values for each channel except $Y X$ and $X Y$ in visual categorization. Channels C4, F3, F4, F7, F8, FP2, O2, P4r, PZ, T4, T5r, T6r are included.	37
3.5	Average CEV values for each channel in the Parkinson's patients group (PD). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.	43
3.6	Average CEV values for each channel in the healthy control group (HC). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.	44
3.7	Average CEV values for each channel except $Y X$ and $X Y$ in the Parkinson's patients group (PD). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.	46
3.8	Average CEV values for each channel except $Y X$ and $X Y$ in the healthy control group (HC). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.	47

3.9	(a) shows the directed connectivity network of channels in Parkinson's patients group (PD); (b) shows the directed connectivity network of channels in healthy control group (HC).	48
3.10		49
3.11	Harvard Oxford Cortical atlas.	51
3.12	Average CEV values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.	54
3.13	Average CEV values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.	55
3.14	Average CEV values for each region except $Y X$ and $X Y$ in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.	57
3.15	Average CEV values for each region except $Y X$ and $X Y$ in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.	58
3.16	(a) shows the heatmap of 48 regions in the major depressive disorder group (MDD); (b) shows the heatmap of 48 regions in the healthy control group (HC).	59
3.18	(a) shows the directed connectivity network of regions in the major depressive disorder group (MDD); (b) shows the directed connectivity network of regions in the healthy control group (HC).	61
4.1	Average CEV-PCA values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.	85
4.2	Average CEV-PCA values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.	86
4.3	Average CEV-PLS values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.	88

4.4	Average CEV-PLS values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.	89
5.1	Boxplot of CEV Values for Simulation Example 1	93
5.2	Boxplot of CEV Values for Simulation Example 2	94
5.3	Comparison of average CEV values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.	96
5.4	Comparison of average CEV values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included. . .	97
A.1	Average CEV values for each channel in visual categorization from CEV- PWA (Part I). Channels C3, C4, CB1, CB1r, CB2, CB2r, F3, F4, F7, F8, FP1, FP2, FZ, I, O1 are included.	107
A.2	Average CEV values for each channel in visual categorization from CEV- PWA (Part II). Channels O1r, O2, O2r, OZ, P3, P4, P4r, PZ, PZr, T3, T4, T5, T5r, T6, T6r are included.	108
A.3	Average CEV values for each channel except $Y X$ and $X Y$ in visual cate- gorization from CEV-PWA (Part I). Channels C3, C4, CB1, CB1r, CB2, CB2r, F3, F4, F7, F8, FP1, FP2, FZ, I, O1 are included.	109
A.4	Average CEV values for each channel except $Y X$ and $X Y$ in visual cate- gorization from CEV-PWA (Part II). Channels O1r, O2, O2r, OZ, P3, P4, P4r, PZ, PZr, T3, T4, T5, T5r, T6, T6r are included.	110
A.5	Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.	111
A.6	Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.	112

A.7	Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.	113
A.8	Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.	114
A.9	Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.	115
A.10	Average CEV values for each channel in the healthy control group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.	116
A.11	Average CEV values for each channel in the healthy control group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.	117
A.12	Average CEV values for each channel in the healthy control group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.	118
A.13	Average CEV values for each channel in the healthy control group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.	119
A.14	Average CEV values for each channel in the healthy control group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.	120
A.15	Average CEV values for each channel except $Y X$ and $X Y$ in the Parkinson's patients group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.	121

A.16 Average CEV values for each channel except $Y X$ and $X Y$ in the Parkinson's patients group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.	122
A.17 Average CEV values for each channel except $Y X$ and $X Y$ in the Parkinson's patients group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.	123
A.18 Average CEV values for each channel except $Y X$ and $X Y$ in the Parkinson's patients group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.	124
A.19 Average CEV values for each channel except $Y X$ and $X Y$ in the Parkinson's patients group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.	125
A.20 Average CEV values for each channel except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.	126
A.21 Average CEV values for each channel except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.	127
A.22 Average CEV values for each channel except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.	128
A.23 Average CEV values for each channel except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.	129
A.24 Average CEV values for each channel except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.	130

A.25 Average CEV values for each region. in the major depressive disorder group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	132
A.26 Average CEV values for each region in the major depressive disorder group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	133
A.27 Average CEV values for each region in the major depressive disorder group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	134
A.28 Average CEV values for each region in the major depressive disorder group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	135
A.29 Average CEV values for each region in the healthy control group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	136
A.30 Average CEV values for each region in the healthy control group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	137
A.31 Average CEV values for each region in the healthy control group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	138
A.32 Average CEV values for each region in the healthy control group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	139
A.33 Average CEV values for each region except $Y X$ and $X Y$ in the major depressive disorder group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	140

A.34 Average CEV values for each region except $Y X$ and $X Y$ in the major depressive disorder group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	141
A.35 Average CEV values for each region except $Y X$ and $X Y$ in the major depressive disorder group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	142
A.36 Average CEV values for each region except $Y X$ and $X Y$ in the major depressive disorder group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	143
A.37 Average CEV values for each region except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	144
A.38 Average CEV values for each region except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	145
A.39 Average CEV values for each region except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	146
A.40 Average CEV values for each region except $Y X$ and $X Y$ in the healthy control group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	147
A.41 Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	148
A.42 Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	149

A.43 Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	150
A.44 Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	151
A.45 Average CEV values for each region in the healthy control group from CEV-PCA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	152
A.46 Average CEV values for each region in the healthy control group from CEV-PCA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	153
A.47 Average CEV values for each region in the healthy control group from CEV-PCA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	154
A.48 Average CEV values for each region in the healthy control group from CEV-PCA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	155
A.49 Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	156
A.50 Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	157
A.51 Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	158

A.52 Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	159
A.53 Average CEV values for each region in the healthy control group from CEV-PLS (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.	160
A.54 Average CEV values for each region in the healthy control group from CEV-PLS (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.	161
A.55 Average CEV values for each region in the healthy control group from CEV-PLS (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.	162
A.56 Average CEV values for each region in the healthy control group from CEV-PLS (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.	163

Abstract

Causal relationships among brain regions have played an essential role in scientific brain functionality studies. Numerous research results have been published, and significant findings in brain functions have enhanced the understanding of brains and the quality of life of humans. Still, there are many unknown issues arising in brain studies. There have also been existing significant uncertainties from published brain research results involving the study designs, experiments, and statistical methods used. Considering the asymmetric and nonlinear functional connectivity between brain regions, we introduce a new type of asymmetric and nonlinear conditional causal relationship, what we call **weighted conditional network causality**, 'WeCon cause,' by cumulative explained variability (CEV). Furthermore, a new estimation approach for CEV is proposed, using continuous piecewise affine function (CEV-PWA) properties. Besides, two alternative methods for estimating CEV have included: the combination of leave-one-out strategy and principal component analysis (CEV-PCA)/ partial least squares regression (CEV-PLS). The superior performance of CEV-PWA approach, as well as asymmetric causal relationship analysis, is illustrated in various simulations. The comparison between CEV-PWA and two alternative approaches, CEV-PCA and CEV-PLS, has proven that the CEV-PWA method achieves the best result through simulation and real data studies. The application of EEG data concerning visual categorization and Parkinson's disease (PD) leads to identifying several influential brain channels through causal analysis. In addition, the research on major depressive disorder (MDD) based on resting-state fMRI data shows the differences in brain activity in several regions between MDD patients and healthy control subjects by causal connectivity obtained from CEV-PWA, CEV-PCA, and

CEV-PLS methods.

Chapter 1

Introduction

The study of the brain has been drawing major attention worldwide; see Grillner et al. (2016) for details of brain study initiatives in the world. Numerous research results have been published, and significant findings in brain functions have enhanced the understanding of brains and the quality of life of humans. Still, there are many unknown issues arising in brain studies. There have also been existing significant uncertainties from published brain research results involving the study designs, experiments, and statistical methods used; see Botvinik-Nezer et al. (2020) for a comprehensive comparison study which has raised great concerns about the repeatability of the study results.

Among many research directions, causal inference of brain functionality has been a contemporary research topic. The need to understand how the interactions and coordination among brain regions contribute to brain functions has led to increasing attention to the investigation of brain connectivity. In addition to anatomical connectivity, two other types of connectivity are defined to show how the dynamic activity of different brain regions is correlated to each other. The first type is functional connectivity, which refers to the statistical dependence between the activity of the regions, while effective connectivity represents the causal influence one neural system exerts over another, Friston et al. (2011). Recent researches include that: Jiang et al. (2019) studied the Granger causality and generalized linear model (GLM) induced causality; López-Madróna et al. (2019) studied correlations associated with causal interactions in brain signals using autoregressive models; Zhang et al. (2016) conducted statistical learning of neuronal functional connectivity; Bielczyk et al. (2019) reviewed current approaches in brain causal inference; Vink et al. (2020) focused on EEG causal and predictions.

To investigate the causal relationship among brain regions, many different approaches developed outside neuroscience are used to study causality. In particular, Granger (1969) proposed Granger causality to infer causality from financial and economic time-series, and later the Granger causality has been widely applied in neuroscience as a model-free approach to study causal relationships among brain regions by neuroscientists and statisticians. We

briefly discuss the inference idea here. Assuming that all relevant information regarding X_t belongs to a vector set $\{Y_t^D\}$ consisting of only two series X_t, Y_t , Granger used $\sigma^2(X | \bar{X})$ to denote the minimum predictor error variance using the past of X data only, and used $\sigma^2(X | \bar{X}, \bar{Y})$ for measuring the minimum predictor error variance using past of both series, X, Y . If the cause Y_t precedes its effect X_t in time, and $\sigma^2(X | \bar{X}) > \sigma^2(X | \bar{X}, \bar{Y})$, then the definition gives $Y_t \Rightarrow X_t$, or we say Y_t (Granger)-causes X_t . In the case that there is a third series Z which causes both X, Y , then spurious causality could arise. For a comprehensive review and discussion of various other causal inference methods and models, we refer to Richardson and Robins (2013) and therein references in which a simple graphical theory unifying causal directed acyclic graphs is presented.

Truly understanding brain functions mainly relies on identifying how the component neurons (circuits, voxels, regions) interact with each other and their relationship. The linear relationship is the most fundamental relationship that has been measured in numerous studies daily. It does offer insights into the variable association of central tendency. Given the vast complexity of a brain, the linear relationship is certainly not enough for exploring and defining nonlinear relationships in a vast array of unknown neurons. The generalized measures of correlation (GMC) is an asymmetric and nonlinear dependence measure utterly different from those measures and model-based approaches studied in the literature. The paired GMCs can be used in identifying the nonlinear relationship between any two components, e.g., pixels/voxels, neurons, or circuits. Zheng et al. (2012) demonstrated that GMC could lead to a more refined version of the concept of Granger causality. Also, Vinod (2017) reminded the reader that Granger also defined "instantaneous causality," which omits the time precedence condition and references to the past values $\bar{X}, \bar{Y}$. From Granger (1969), Y_t instantaneously causes X_t if "the current value X_t is better 'predicted' if the present value Y_t is included in the 'prediction' than if it is not." In a linear Gaussian regression of Y on X , the R^2 is exactly the same as when regressing X on Y . Based on this, Vinod (2017) implemented the instantaneous causality, called kernel causality, in his paper, using

GMC proposed by Zheng et al. (2012). He stated that GMC could be interpreted in terms of kernel causality. $\text{GMC}(X_I|X_J)$, is the coefficient of determination, R^2 , of the Nadaraya Watson nonparametric Kernel regression:

$$X_I = g(X_J) + \epsilon = E(X_I | X_J) + \epsilon,$$

where $g(X_J)$ is a nonparametric, unspecified (nonlinear) function. Interchanging X_J and X_I , we obtained the other $\text{GMC}(X_J|X_I)$ defined as the R^2 of the Kernel regression:

$$X_J = g'(X_I) + \epsilon' = E(X_J | X_I) + \epsilon'.$$

Vinod (2017) defined $\delta = \text{GMC}(X_J|X_I) - \text{GMC}(X_I|X_J)$ as the difference of two population R^2 values. When $\delta < 0$, we know that X_J better predicts X_I than vice-versa. Hence, we define X_J kernel causes X_I provided the true unknown $\delta < 0$. Its estimate $\hat{\delta}$ can be readily computed by means of a regression. Later, Allen and Hooper (2018) used the metric to analyze causal relations between the VIX, S&P500, and the realized volatility (RV) of the S&P500 sampled at 5-min intervals. Allen and McAleer (2020) commented on the GMC metric partly extended the concept of Granger causality and explored previous work related to this measure, including some publications over 100 years earlier by Yule in the Proceedings of the Royal Society, including Yule (1897a), Yule (1897b), and Yule (1900), and by Kendall et al. (1946), as well as research on dependency by Rényi (1959) and Doksum and Samarov (1995), together with the Yule–Simpson paradox by Yule (1903) and Pearson et al. (1899).

There are some empirical drawbacks with GMC. The most important one is that they only take the correlation between two variables into consideration, and work only on pairwise relationships may neglect the joint information of a set of multivariate variables and the joint relationship between X_I and the set of variables. Thus, GMC may not be able to identify covariates that are jointly important with other variables to X_I but marginally insignificant.

To overcome the aforementioned shortcomings, Miao et al. (2018) advanced GMC to a more general form as $\text{CEV}(X_I|X_{J1}, X_{J2}, \dots, X_{JD})$, where CEV stands for cumulative explained variability in X_I by $X_{J1}, X_{J2}, \dots, X_{JD}$. CEV can capture the joint information from the variables, and CEV is the same as GMC when $D = 1$.

Similar to the causal relationship stated by Zheng et al. (2012) and Vinod (2017), we utilize CEV to define a more generalized asymmetric nonlinear causal relationship. Without loss of generality, we fix the present time so as to simplify the notation, and only consider instantaneous causality for now. Since the procedure we use here is a **weighted conditional network**, we call this as **WeCon Causality**. Define

$$\lambda = \text{CEV}(Y|C(X)) - \text{CEV}(C(X)|Y). \quad (1.1)$$

where $C(X)$ can be either an individual variable X_J or the joint of multiple variables $g(X_{J1}, X_{J2}, \dots, X_{JD})$.

- If $\lambda > 0$, $\text{CEV}(Y|C(X)) > \text{CEV}(C(X)|Y)$, then WeCon cause = $C(X)$.
- If $\lambda < 0$, $\text{CEV}(Y|C(X)) < \text{CEV}(C(X)|Y)$, then WeCon cause = Y .

In this dissertation, we introduce a new estimation approach concerning CEV, using continuous piecewise affine function (PWA) properties. Hahn et al. (2017); Cui et al. (2018) stated that statistics regressions with the mean functions could be taken as PWA functions. For this procedure, we first applied CEV to calculate the asymmetric nonlinear correlation between X_I and each covariate, then ranked the covariates based on this. We only choose the top ones with the highest correlations to construct the important covariates set. Next, given the theory from Hempel et al. (2013) that every continuous PWA function can be written as the difference between two convex PWA functions, we use this PWA method to divide the observations into different groups to make the estimation more precise. In order to estimate the parameters in the PWA method, we implement the genetic algorithm and update the searching range iteratively by applying ridge regressions within each group. At the end of this process, we would achieve the highest cumulative explained variability by

selecting the most appropriate combination of important variables for each group. In addition, we propose several CEV estimators so that we not only study the pairwise WeCon causality but also explore more about the causality between X_I and the combination of $X_{J1}, X_{J2}, \dots, X_{JD}$ constructed by PWA functions.

Besides PWA functions, we utilize principal component analysis (PCA) and partial least squares regression (PLS) to construct the network of multiple variables. Using leave-one-out strategy and random search, we would be able to achieve the highest cumulative explained variability by selecting the most appropriate combination of important variables. Similar to CEV-PWA method, these two approaches can also be summarized as CEV-PCA and CEV-PLS algorithms. Several CEV estimators are proposed to show the differences when we consider more components. Then, we compare three algorithms based on the simulation study and real data application.

Two main types of data, electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), are studied about brain functionality. Specifically, we focus on rest and 3 stimuli auditory oddball EEG data for Parkinson’s disease (PD) and resting-state with closed eyes fMRI data for major depressive disorder (MDD).

Parkinson’s disease (PD) is a common progressive neurodegenerative disorder of the central nervous system as stated by Valls-Solé and Valldeoriola (2002), and it’s affecting about 2 – 3% of the elderly population. Savitt et al. (2006) claimed that the clinical motor symptoms of PD, such as tremors, slowness of movement, rigidity, and postural instability, are resulted from a dopaminergic deficiency in the basal ganglia. What’s more, Cooper et al. (1991) concluded that PD could cause non-motor impairments such as cognitive dysfunction in the early stage of PD. However, the diagnosis of PD via clinical symptoms could be very difficult, especially in the early stage when there are no significant motor symptoms and obvious cognitive dysfunction. EEG signals measure brain activities in various regions

and could be used for the detection of PD. Current research includes applying deep learning methods, such as convolutional neural network (CNN), for the diagnosis of PD (Oh et al. (2020), Khare et al. (2021), Pereira et al. (2016), Gunduz (2019)). PCA and PLS are also utilized in the research of Parkinson’s disease. Bhurane et al. (2019) proposed to use a support vector machines classifier with a third-degree polynomial kernel to detect the PD signals automatically and apply feature ranking and PCA for progressive feature addition analysis. This approach can achieve maximum accuracy of $99.1 \pm 0.1\%$, and the presented computer-aided diagnosis system can act as an assistive tool to confirm the finding of PD by the clinicians. In addition, Pyatigorskaya et al. (2021) used partial least squares path modeling (PLS-PM) for estimating a network of causal relationships between blocks of variables and further testing whether brain lesions progress from the medulla oblongata to the cortex or first occurs in the cortex. They finally found that the neurodegeneration pattern of patients with PD with REM sleep behavior disorder (RBD) was consistent with the Braak brainstem-to-cortex model, whereas that of patients without RBD followed the cortex-to-brainstem model.

Major depressive disorder (MDD) is one of the most commonly diagnosed mental disorders in most first-world countries, and it can be featured by feelings of persistent sadness or diminished interest in daily activities. Functional magnetic resonance imaging (fMRI) has been essential for learning the pathogenesis of MDD. Liang et al. (2020) focused on fMRI-data-driven identification and validation of potential DMN-pattern-based (default mode network) MDD subtypes to parse heterogeneity of the disorder. PCA was used in their study to define patient subgroups for statistical comparison. Bessette et al. (2018) also conducted PCA for significant clusters to explore the connectivity components and further study whether there exist time-invariant DMN connectivity differences early in the lifespan course of depression. Sen and Parhi (2019) proposed using PLS for predicting males and females from task-fMRI brain connectivity. Other related researches include Yang et al. (2021), Dong et al. (2019), and Shi et al. (2019). They implemented Granger causal analysis

(GCA) for investigating effective connectivity between regions, while they mostly concentrated on several main regions and considered pairwise connectivity.

In our research for alternative approaches, we combine PCA and PLS with cumulative explained variability (CEV) and conduct CEV screening for high dimensional settings. We not only study the pairwise connectivity among brain regions but also consider the effect of networks jointed by multiple regions. PCA and PLS were utilized for constructing the network of various brain regions, rather than the simple screening and prediction. Furthermore, we analyze all the brain regions and channels from fMRI and EEG data so as to better understand the mechanism between brain activity and diseases, including MDD and PD, while many other researchers only focus on the main regions.

The rest of the dissertation is organized as follows. Chapter 2 reviews generalized measures of correlation and cumulative explained variability and proposes the WeCon causality based on CEV. Chapter 3 introduces cumulative explained variability with continuous piecewise affine function and proposes four CEV estimators. Chapter 3 also details the procedure under a high-dimensional setting and summarizes the proposed algorithm. In addition, simulation study and real data applications for EEG and fMRI data are included in Chapter 3, and we compare several different CEV estimators and further discuss the causality suggested by the simulation and real data results. Chapter 4 provides the alternative approaches utilizing principal component analysis (CEV-PCA) and partial least squares regression (CEV-PLS) instead of piecewise affine function (CEV-PWA) to construct the network. In Chapter 5, the comparison among three approaches proposed in the thesis is shown with regard to simulation study and fMRI application in major depressive disorder (MDD). In addition, Chapter 5 contains the conclusion and discusses future work.

Chapter 2

WeCon Causality Based on Cumulative Explained Variability

The brain's interactions among the circuits (neurons, voxels, regions) vary from one pair to another. Within one pair, the directed effects on each other are asymmetric. As such, the linear correlation has its limitation. Therefore, it becomes an important step to study the asymmetric dependence in the brain to improve the scientific knowledge and understanding of the brain and provide better clinical practice guidance. In this chapter, we introduce our WeCon causality measures based on the generalized measures of correlation and their extensions.

In Section 2.1, we briefly review the generalized measures of correlation, and then introduce our WeCon causality measures in Section 2.2. We introduce continuous piecewise affine functions to reduce dimensionality to address the high-dimension conditional expectation estimation challenge in WeCon causality measures.

2.1 Generalized measures of correlation: A brief review

Consider two neurons (circuits, voxels) i_1 and i_2 and their observations X_{i_1,t_j} , $j = 1, \dots, n$ and X_{i_2,t_j} , $j = 1, \dots, n$, the generalized measures of correlation in Zheng et al. (2012); Chen et al. (2017); Allen and McAleer (2020) for any two paired random variables are defined as:

$$\begin{aligned} \text{GMC}(X_{i_1,t_j}|X_{i_2,t_{j-k}}) &= \frac{\text{var}\{E(X_{i_1,t_j}|X_{i_2,t_{j-k}})\}}{\text{var}(X_{i_1,t_j})} = 1 - \frac{E(\text{var}(X_{i_1,t_j}|X_{i_2,t_{j-k}}))}{\text{var}(X_{i_1,t_j})} \\ &= 1 - \frac{E[\{X_{i_1,t_j} - E(X_{i_1,t_j}|X_{i_2,t_{j-k}})\}^2]}{\text{var}(X_{i_1,t_j})}. \end{aligned} \quad (2.1)$$

Assume that both sequences X_{i_1,t_j} , $j = 1, \dots, n$ and X_{i_2,t_j} , $j = 1, \dots, n$ are stationary. Then $\text{GMC}(X_{i_1,t_j}|X_{i_2,t_{j-k}})$ is the same for all j and a fixed k . $\text{GMC}(X_{i_1,t_j}|X_{i_2,t_{j-k}})$ is interpreted as the variability in X_{i_1,t_j} at time t_j being explained by $X_{i_2,t_{j-k}}$ at prior times t_{j-k} for $k = 0, 1, 2, \dots$. For notation convenience, we set $k = 0$ throughout the dissertation. There are three different nonparametric estimation procedures developed in Zheng et al. (2012); Chen et al. (2017); Miao et al. (2018) for $\text{GMC}(X_{i_1,t_j}|X_{i_2,t_j})$. The proposed research is to use the estimated $\widehat{\text{GMC}}(X_{i_1,t_j}|X_{i_2,t_j})$ from Miao et al. (2018) to form circuits, and from

circuits to form interconnected brain networks. One challenge is the curse of dimensionality. Some recent applications of GMC in asymmetric dependence measures and brain functional correlations can be found in Tang and Zhao (2017); Zhao and Tang (2018); Młyńczak and Krysztofiak (2018, 2019); Badwan et al. (2019).

GMCs are computed pairwise, i.e., they don't contain additional information from other brain components. Miao et al. (2018) advanced Eqn. (2.1) to a more general formula as:

$$\begin{aligned}
& \text{CEV}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) \\
&= \frac{\text{var}(\mathbb{E}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}))}{\text{var}(X_{i_1,t_j})} \\
&= \frac{\mathbb{E}\{\mathbb{E}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d})\} - (E(X_{i_1,t_j}))^2}{\text{var}(X_{i_1,t_j})},
\end{aligned} \tag{2.2}$$

where CEV stands for cumulative explained variability in X_{i_1,t_j} by $X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}$ with $k_2, k_3, \dots, k_d$ being the delayed time units. When $d = 1$, CEV is the same as GMC. The following facts hold:

1. $0 \leq \text{CEV}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) \leq 1$.
2. If $\text{CEV}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) = 0$, then $\text{cor}(X_{i_1,t_j}, X_{i_l,t_j-k_l}) = 0$, for $l = 1, \dots, d$.
3. If at least one of $\text{cor}(X_{i_1,t_j}, X_{i_l,t_j-k_l})$, for $l = 1, \dots, d$ equals to ± 1 , then $\text{CEV}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) = 1$.
4. If $X_{i_1,t_j} = g(X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d})$, almost surely, then

$$\text{CEV}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) = 1.$$

5. If $X_{i_1,t_j} = g(X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) + \epsilon$, and $g(\cdot)$ is independent of ϵ , then
$$\text{CEV}(X_{i_1,t_j} | X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) = \frac{\text{var}(g(X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}))}{\text{var}(Y)}.$$

Facts 1-3 offer the relationships between the cumulative explained variance and Pearson's correlation coefficient. Facts 4 and 5 tell the validity of the definition of the cumulative explained variance of X_{i_1,t_j} by $X_{i_2,t_j-k_2}, X_{i_3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}$. Clearly, CEV shows its advantage of measuring both linear and nonlinear correlations.

2.2 WeCon causality

Zheng et al. (2012) stated that GMC could lead to a more refined version of the concept of Granger causality, and proposed Granger causality GMC (GcGMC) with the following definition.

Definition 2.2.1. *Suppose that $\{X_t, Y_t\}, t > 0$ is a bivariate stationary time series. Define GcGMC as:*

$$GcGMC(X_t | \mathcal{F}_{t-1}) = 1 - \frac{E \left[\{X_t - E(X_t | X_{t-1}, X_{t-2}, \dots, Y_{t-1}, Y_{t-2}, \dots)\}^2 \right]}{E(\text{var}(X_t | X_{t-1}, X_{t-2}, \dots))},$$

$$GcGMC(Y_t | \mathcal{F}_{t-1}) = 1 - \frac{E \left[\{Y_t - E(Y_t | Y_{t-1}, Y_{t-2}, \dots, X_{t-1}, X_{t-2}, \dots)\}^2 \right]}{E(\text{var}(Y_t | Y_{t-1}, Y_{t-2}, \dots))},$$

where $\mathcal{F}_{t-1} = \sigma(X_{t-1}, X_{t-2}, \dots, Y_{t-1}, Y_{t-2}, \dots)$.

Based on the definition, if $GcGMC(X_t | \mathcal{F}_{t-1}) > 0$, it's said Y Granger causes X , while we say X Granger causes Y if $GcGMC(Y_t | \mathcal{F}_{t-1}) > 0$. Also, it has been shown that GcGMC is bounded by 0 and 1. The following proposition relates GcGMC to Grainger's causality.

Proposition 2.2.1. *If $E(\text{var}(X_t | X_{t-1})) = E(\text{var}(Y_t | Y_{t-1}))$ and the strength of that X_t Granger causes Y_t is stronger than the strength of that Y_t Granger causes X_t , that is,*

$$\begin{aligned} & E \left[\{E(Y_t | Y_{t-1}) - E(Y_t | Y_{t-1}, X_{t-1})\}^2 \right] \\ & > E \left[\{E(X_t | X_{t-1}) - E(X_t | X_{t-1}, Y_{t-1})\}^2 \right] \end{aligned}$$

then

$$\text{GcGMC}(Y_t | Y_{t-1}, X_{t-1}) > \text{GcGMC}(X_t | X_{t-1}, Y_{t-1}).$$

As CEV is a more general form of GMC (Miao et al. (2018)), we extend the Granger causality GMC from Zheng et al. (2012) to this newly proposed CEV-based causal relationship. Recalling $\lambda = \text{CEV}(Y|C(X)) - \text{CEV}(C(X)|Y)$ defined in Chapter 1, we formally identify the "WeCon cause" as:

- If $\lambda > 0$, $\text{CEV}(Y|C(X)) > \text{CEV}(C(X)|Y)$, then WeCon cause = $C(X)$.
- If $\lambda < 0$, $\text{CEV}(Y|C(X)) < \text{CEV}(C(X)|Y)$, then WeCon cause = Y .

where $C(X)$ can be either an individual variable X_J or the joint of multiple variables $g(X_{J1}, X_{J2}, \dots, X_{JD})$. This identification should be used only when we already know from theory and/or experiments that Y and $C(X)$ are causally linked.

It is clear in the definition of WeCon causality, especially in brain functional connectivity studies that the dimension of the set of covariates is ultra-high. As a result, it is not practically feasible to compute WeCon causality to apply such a framework. We will explore three types of dimension reductions to overcome this difficulty. We next introduce the continuous piecewise affine functions. Then, in Chapter 4, we study the widely used principal component analysis and partial least squares method.

2.3 Cumulative explained variability with continuous piecewise affine function

A set of random variables $X_{i_1, t_j}, X_{i_2, t_j - k_2}, X_{i_3, t_j - k_3}, \dots, X_{i_d, t_j - k_d}$ can always be connected by

$$X_{i_1, t_j} = g(X_{i_2, t_j - k_2}, X_{i_3, t_j - k_3}, \dots, X_{i_d, t_j - k_d}) + \epsilon_{i_1, t_j},$$

where the function $g(\cdot)$ is either specified or unspecified, and ϵ_{i_1, t_j} is the error term. $g(X_{i_2, t_j - k_2}, X_{i_3, t_j - k_3}, \dots, X_{i_d, t_j - k_d})$ and ϵ_{i_1, t_j} can be independent or dependent depending

on the form of $g(\cdot)$ function and its argument variables. In this dissertation, we choose $g(\cdot)$ to have parametric forms.

2.3.1 Piecewise affine function

Traditionally, a statistical model is often represented by a linear combination of the input variables, leading to linear regression or classification modeling. However, recent researches including Hannah and Dunson (2011), Hannah and Dunson (2013), and Mazumder et al. (2019), pay more attention to the use of convex piecewise affine (PWA) functions, and there are also other researchers, such as Hahn et al. (2017), studied the usage of non-convex PWA estimation functions.

As Hahn et al. (2017) and Scholtes (2012) stated that a parametric statistical model could be expressed as

$$Y = g_\theta(X) + \epsilon \quad (2.3)$$

where Y is the response variable, $X \in \mathcal{X} \subset \mathbb{R}^d (d \geq 1)$, and $g_\theta : \mathbb{R}^d \rightarrow \mathbb{R}$ is continuous k -piecewise affine (k -PWA; $k \geq 2$), which means that there exists a partition of $\mathcal{X}$ into polyhedral sets $\{C_i\}_{i=1}^k$ (i.e., $C_i \subset \mathcal{X}, C_i \cap C_j = \emptyset$ for all $i \neq j$) such that

$$g_\theta(x) = a_i^\top x + b_i \quad \text{if } x \in C_i,$$

and that g is continuous on $\mathcal{X}$. We assume that $k \in \mathbb{N}$ is given and denote the unknown parameter as $\theta = (a_1, \dots, a_k, b_1, \dots, b_k) \in \mathbb{R}^{k(d+1)}$. The error term ϵ is assumed to have zero mean and finite variance.

Given i.i.d. data $\{(X_i, Y_i)\}_{i=1}^n$ from this model, the next goal is to estimate the unknown parameter θ . A naive method for solving this parametric regression problem is to utilize

the least squares estimator:

$$\tilde{\theta} = \underset{\theta \in \mathbb{R}^{k(d+1)}}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - g_{\theta}(X_i))^2,$$

where the minimization is over all continuous PWA g_{θ} .

Consider that the regression model (2.3) where g_{θ} is a continuous piecewise affine (PWA) function. The following lemma states that any continuous PWA function can be expressed as the difference of two convex PWA functions, and it will be pivotal for the rest of the research.

Lemma 2.3.1. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a convex polyhedral region (i.e., a set formed by the intersection of finitely many hyperplanes). Every continuous PWA function $g : \mathcal{X} \rightarrow \mathbb{R}$ defined over a convex polyhedral partition of $\mathcal{X}$ with full dimensional elements $C_i, i = 1, \dots, k$, can be written as the difference of two convex PWA functions, i.e.,*

$$g(x) = g_1(x) - g_2(x), \quad \text{for all } x \in \mathcal{X},$$

where $g_i : \mathcal{X} \rightarrow \mathbb{R}, i = 1, 2$, are convex and PWA functions.

The above result is summarized by Hahn et al. (2017) and proved in Kripfganz and Schulze (1987); Hempel et al. (2013). Therefore, in order to seek for a smooth approximation of a PWA (continuous) g we first consider the case when g is PWA and convex as well. The benefit from using a convex PWA function is that it can be expressed as a maximum of affine functions, i.e.,

$$g(x) = \max_{i=1, \dots, k} \{a_i^{\top} x + b_i\}$$

for some $k \geq 1$ where $a_1, \dots, a_k \in \mathbb{R}^d$ and $b_1, \dots, b_k \in \mathbb{R}$.

Cui et al. (2018) studied statistics regressions with the mean functions being PWA

functions, and we adopt the same setting as their research. Suppose $g(\cdot)$ is a PWA function. Then it can be expressed as:

$$g(x; \mathbf{a}, \mathbf{b}) = \max_{1 \leq i \leq k_1} (a_i^{(1)}x + b_i^{(1)}) - \max_{1 \leq i \leq k_2} (a_i^{(2)}x + b_i^{(2)}) \quad (2.4)$$

where k_1 and k_2 are two nonnegative integer numbers to be determined, $a_i^{(j)}$, $b_i^{(j)}$, $j = 1, 2$; $i = 1, \dots, k_j$, are real valued row vectors of d dimension, and $\mathbf{a} = (a_i^{(j)}, i = 1, \dots, k_j; j = 1, 2)$, $\mathbf{b} = (b_i^{(j)}, i = 1, \dots, k_j; j = 1, 2)$.

This model includes the 1-layer neural network by the rectified linear unit (ReLU) activation function (Nair and Hinton (2010)) that has a simple form $g(x; \mathbf{a}, \mathbf{b}) = \max(a_i x + b_i, 0)$. Furthermore, it has been shown in Cui et al. (2018) that a multi-layer neural network with the ReLU activation function can be written as the difference-of-max form, for example, $g(x; \theta) = \max_{1 \leq j \leq k_1} g_1(x; \theta) - \max_{1 \leq j \leq k_2} g_2(x; \theta)$ for a 2-layer model, where k_1 and k_2 are some positive integers, and $g_i(x; \theta)$ are all once but not twice continuously differentiable piecewise linear-quadratic.

Different from PWA functions though shared similar max-linear structures, using extreme value analysis, Cui and Zhang (2018) introduced a competing copula, and Cui et al. (2021) introduced a max-linear regression model that outperformed random forest model, graphical-lasso models in some general model settings. In a logistic regression framework, Zhang (2021a, 2022a) introduced maxi-linear competing factor logistic models to study COVID-19 classifications, and Zhang (2021b, 2022b) studied lung cancer and breast cancers, with nearly perfect classification rates and the discovery of new disease subtypes. These applications show that PWA has a great potential in improving statistical inference in many areas.

2.3.2 CEV-PWA estimators

Rather than simply studying the pairwise causal relationship between two variables X_{i1} and X_{i2} , we have more interests in applying PWA function $g(\cdot)$ to capture the joint information from multiple variables, and further analyze the causality among different components, including the individual variables and the network of various variables constructed by PWA. When building the network for variable X_{i1} , we have two strategies, either take the information of X_{i2} into consideration or not. This will help us to identify the role of X_{i2} by comparing CEV estimators utilizing different information. The proposed CEV estimators include the following:

$$\begin{aligned} \widehat{\text{CEV}}_1(X_{i1,t_j} | X_{i2,t_j-k_2}, X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) \\ = \max_{a,b} \widehat{\text{CEV}}(X_{i1,t_j} | g((X_{i2,t_j-k_2}, X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}); a, b)), \end{aligned} \quad (2.5)$$

$$\begin{aligned} \widehat{\text{CEV}}_2(X_{i1,t_j} | X_{i2,t_j-k_2}, X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) \\ = \frac{\max_{a,b} \{\widehat{\text{var}}(g((X_{i2,t_j-k_2}, X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}); a, b))\}}{\widehat{\text{var}}(X_{i1,t_j})}, \end{aligned} \quad (2.6)$$

$$s.t. |\widehat{\text{cov}}(g(X_{i2,t_j-k_2}, X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}), \epsilon)| \leq \delta,$$

$$\begin{aligned} \widehat{\text{CEV}}_3(X_{i1,t_j} | X_{i2,t_j-k_2}, X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) \\ = \max_{a,b} \widehat{\text{CEV}}(X_{i1,t_j} | X_{i2,t_j-k_2}, g((X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}); a, b)), \end{aligned} \quad (2.7)$$

$$\begin{aligned} \widehat{\text{CEV}}_4(X_{i1,t_j} | X_{i2,t_j-k_2}, X_{i3,t_j-k_3}, \dots, X_{i_d,t_j-k_d}) \\ = \max_{a,b} \widehat{\text{CEV}}(X_{i1,t_j} | X_{i2,t_j-k_2}, g((X_{i2,t_j-k_2}, \dots, X_{i_d,t_j-k_d}); a, b)). \end{aligned} \quad (2.8)$$

In equation (2.6), we assume that X_{i1,t_j} could be approximated by $g(\cdot)$ function as $X_{i1,t_j} = g(X) + \epsilon$, then we can denote the residual as $\hat{\epsilon}$, the estimated variance and covariance of $g(X)$ as $\widehat{\text{var}}$ and $\widehat{\text{cov}}$, respectively. Comparing with the nonparametric estimator $\widehat{\text{CEV}}$, $\widehat{\text{CEV}}_1$ is semi-parametric and uses $\widehat{\text{GMC}}$ estimation approach; $\widehat{\text{CEV}}_2$ is a parametric estimator; $\widehat{\text{CEV}}_3$ and $\widehat{\text{CEV}}_4$ are semi-parametric and uses $\widehat{\text{CEV}}$ estimation approach. These three new semi-parametric estimators greatly reduce the dimensionality in nonparametric estimation.

We have the following population relationships

$$\text{GMC}(X_{i_1,t_j}|X_{i_2,t_{j-k_2}}) \leq \text{CEV}_2(X_{i_1,t_j}|X_{i_2,t_{j-k_2}}, X_{i_3,t_{j-k_3}}, \dots, X_{i_d,t_{j-k_d}}),$$

$$\begin{aligned} \text{GMC}(X_{i_1,t_j}|X_{i_2,t_{j-k_2}}) &\leq \text{CEV}_1(X_{i_1,t_j}|X_{i_2,t_{j-k_2}}, X_{i_3,t_{j-k_3}}, \dots, X_{i_d,t_{j-k_d}}) \\ &\leq \text{CEV}_3(X_{i_1,t_j}|X_{i_2,t_{j-k_2}}, X_{i_3,t_{j-k_3}}, \dots, X_{i_d,t_{j-k_d}}) \\ &\leq \text{CEV}_4(X_{i_1,t_j}|X_{i_2,t_{j-k_2}}, X_{i_3,t_{j-k_3}}, \dots, X_{i_d,t_{j-k_d}}) \end{aligned}$$

The nonparametric estimation of CEV is mainly built on locally weighted regression, which is proposed by Miao et al. (2018). We can estimate the multivariate exploratory CEV by

$$\widehat{\text{CEV}}(X_{i_1,t_j}|X_{i_2,t_{j-k_2}}, X_{i_3,t_{j-k_3}}, \dots, X_{i_d,t_{j-k_d}}) = S_Y^{-2} \left(\frac{1}{n} \sum_{i=1}^n \hat{m}^2(\mathbf{X}; H) - Y^2 \right) \quad (2.9)$$

where H is a $d \times d$ symmetric positive definite matrix depending on n ; K is a d -variate kernel such that $\int K(u)du = 1$ and $K_H(u) = |H|^{-1/2}K(H^{-1/2}u)$. We will call $H^{-1/2}$ the bandwidth matrix since it is the multivariate extension of the usual bandwidth parameter.

Simple algebra yields the local regression based estimator of $m(\mathbf{x})$ is

$$\hat{m}(\mathbf{x}; H) = e_1^T (\mathbf{X}_x^T \mathbf{W} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W} Y$$

where

$$\mathbf{X}_x = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \end{pmatrix}$$

and $\mathbf{W} = \text{diag}\{K_H(\mathbf{X}_1 - \mathbf{x}), \dots, K_H(\mathbf{X}_n - \mathbf{x})\}$; e_1 is the $(d+1) \times 1$ vector with 1 in the first entry and all other entries 0. The estimator (2.9) was first introduced as a regression estimator by Stone (1977); Ruppert and Wand (1994) studied its asymptotic properties in

a nonparametric regression situation.

Chapter 3

High-dimension Predictors

The data dimensions in brain image and functional connectivity studies are generally ultra-high. The WeCon causality introduced in Chapter 2 involves high dimensional conditional expectations that can further involve conditional density estimations. However, high dimensional conditional density estimation is a challenging task in theory and computation. It is almost impossible to computationally find an efficient algorithm that can handle high dimensional conditional density estimation. As such, dimension reduction is needed. In this dissertation, we adopt an efficient dimension reduction method that is closely related to our newly introduced WeCon causality.

Subsequently, we first review the dimension reduction method, and then introduce the definition of WeCon causality in Section 3.1. To further reduce the computational complexity, we introduce a class of piecewise affine functions and integrate it with our WeCon causality inference in Section 3.2. The parameter estimation is very challenging due to the form of CEV, so we design an alternative approach combining genetic algorithms and ridge regression for solving the optimization problem in Section 3.3. Section 3.4 provides simulation studies for both linear and non-linear settings to monitor the performance of our proposed estimators. In Section 3.5, three real data analysis including EEG application in visual categorization and Parkinson’s disease, and resting-state fMRI application in major depressive disorder were conducted to assess WeCon causality among brain regions/channels.

3.1 Dimension reduction: SEVIS screening

3.1.1 Sure explained variability

Following the definition of the generalized measures of correlation (GMC) by Zheng et al. (2012), another measure of sure explained variability (SEV) is defined by Chen et al. (2017) as follows.

Suppose Y is a response variable and $\{X_i\}_{i=1}^D$ are exploratory variables, a natural ex-

tension of GMC is defined as

$$\text{SEV}(Y | X) = \frac{\text{var}(E(Y | X))}{\text{var}(Y)} = \frac{E[E(Y | X)]^2 - E(Y^2)}{\text{var}(Y)} \quad (3.1)$$

Note that we actually have $\text{SEV}(Y|X) = \text{GMC}(Y|X) = \text{CEV}(Y|X)$ based on the definition. However, $\text{GMC}(Y|X)$ focuses more on studying both directional asymmetrically explained variances, which means that Y and X play the same role and we don't differentiate them by response variable and exploratory variable. On the contrary, Y is a response variable, and X is an exploratory variable in the definition of SEV.

Since CEV is a more general form of SEV, we could still use locally weighted regression as nonparametric estimator of $\text{SEV}(Y|X)$.

3.1.2 Sure explained variability and independence screening

Let $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p)$ and $\mathbf{X}_k = (X_{1k}, \dots, X_{nk})^T$, where n is the sample size. We assume that the dimension p is much larger than the sample size n , i.e. $n \ll p$. We define the index set of active predictors by

$$\mathcal{M} = \{1 \leq k \leq p : \text{the predictor } X_k \text{ has contributed to the response } Y\} \quad (3.2)$$

We call each $X_k, k \in \mathcal{M}$, an active predictor. In order to distinguish active X_k 's among all predictors, we use $\text{SEV}(Y | X_k)$ as a measure of the relations between X_k and Y . Denoted by $\omega_k = \text{SEV}(Y | X_k), k = 1, \dots, p$. Naturally, we select an active predictors set

$$\mathcal{M}^* = \{1 \leq k \leq p : \omega_k > 0\} \quad (3.3)$$

which contains all predictors having contributed to the response Y in terms of sure explained variabilities. Denoting $\hat{\omega}_k$ as the estimator of ω_k , we define a truncated active predictors

set by

$$\widehat{\mathcal{M}}^* = \{1 \leq k \leq p : \widehat{\omega}_k \geq cn^{-\tau}\} \quad (3.4)$$

The constants $c > 0$ and $\tau \geq 0$ have to be prespecified and satisfy certain conditions. Although it is not easy to decide suitable magnitudes of c and τ such that $\widehat{\mathcal{M}}^* \subseteq \widehat{\mathcal{M}}^*$. For ease of implementation, we first rank $\{\widehat{\omega}_k, k = 1, \dots, p\}$ in a decreasing order, and then choose the first d largest ones to get a workable active index set:

$$\widehat{\mathcal{M}}_d^* = \{1 \leq k \leq p : \widehat{\omega}_k \text{ is among the first } d \text{ largest estimates in } \widehat{\mathcal{M}}^*\} \quad (3.5)$$

In numerical studies, we use $\widehat{\mathcal{M}}_d^*$ instead of $\widehat{\mathcal{M}}^*$ and pre-decide d to be a function of n in a general way.

With the established notations, the sure explained variability and independence screening (SEVIS) algorithm is as follows.

SEVIS algorithm:

1. Normalize $X_k, k = 1, \dots, p$, such that $\sum_{i=1}^n X_{ik} = 0$ and $(n-1)^{-1} \sum_{i=1}^n X_{ik}^2 = 1$;
2. Calculate $\widehat{\omega}_k = \text{SEV}(Y | X_k)$ between Y and candidate predictors $X_k, k = 1, \dots, p$;
3. Rank $\widehat{\omega}_k$ in a decreasing order, that is $\widehat{\omega}_{k_1} > \widehat{\omega}_{k_2} > \dots > \widehat{\omega}_{k_p}$;
4. Choose the first $d, X_{k_1}, \dots, X_{k_d}$, as the active predictors. d is a given integer and has the order of $O([n/\log(n)])$.

3.2 Cumulatively explained variability with continuous piece-wise affine function estimators

After the SEVIS screening, we selected a set of important variables to the specific variable with relatively large $\widehat{\omega}_k$. Then we could explore more about the cumulatively explained variability with continuous piece-wise affine function for each given variable. To simplify

the notation of equation (2.5) - (2.8) without loss of generality, the three CEV estimators we have interests in are:

$$\begin{aligned}\widehat{\text{CEV}}_1 &= \max_{\mathbf{a}, \mathbf{b}} \widehat{\text{CEV}}(X_I | g(X_{[-I]}; \mathbf{a}, \mathbf{b})) \\ \widehat{\text{CEV}}_3 &= \max_{\mathbf{a}, \mathbf{b}} \widehat{\text{CEV}}(X_I | X_J, g(X_{[-I, -J]}; \mathbf{a}, \mathbf{b})) \\ \widehat{\text{CEV}}_4 &= \max_{\mathbf{a}, \mathbf{b}} \widehat{\text{CEV}}(X_I | X_J, g(X_{[-I]}; \mathbf{a}, \mathbf{b}))\end{aligned}$$

where $0 < I, J < D$, $I \neq J$, $X_{[-I]}$ denotes for all other variables except X_I , and $X_{[-I, -J]}$ denotes for all other variables except X_I and X_J .

Considering observations may have different characteristics, the relationship between X_I and all other variables and the relationship among these variables could be different for different groups of observations. Using the fact that $g(\cdot)$ function (2.4) contains several groups of parameters, we could divide the population into different groups based on the set of parameters used in the $g(\cdot)$ function. Then we explore the relationships within each group so that more precise results come through. Our proposed cumulatively explained variability (CEV) with continuous piece-wise affine function procedure is based on this structure. Before introducing detailed procedures, we first utilize a simple example to illustrate the unique structure of equation (2.4).

Example. For equation (2.4), suppose $k_1 = k_2 = 2$, then we have formula (3.6) as follows.

$$\begin{aligned}g(x; \mathbf{a}, \mathbf{b}) &= \max_{1 \leq i \leq 2} (a_i^{(1)}x + b_i^{(1)}) - \max_{1 \leq i \leq 2} (a_i^{(2)}x + b_i^{(2)}) \\ &= \begin{cases} (a_1^{(1)} - a_1^{(2)})x + (b_1^{(1)} - b_1^{(2)}) & \text{Case 1,} \\ (a_1^{(1)} - a_2^{(2)})x + (b_1^{(1)} - b_2^{(2)}) & \text{Case 2,} \\ (a_2^{(1)} - a_1^{(2)})x + (b_2^{(1)} - b_1^{(2)}) & \text{Case 3,} \\ (a_2^{(1)} - a_2^{(2)})x + (b_2^{(1)} - b_2^{(2)}) & \text{Case 4.} \end{cases} \quad (3.6)\end{aligned}$$

From this example, we notice that parameter set $\{a_1^{(1)}, b_1^{(1)}\}$ is competing with $\{a_2^{(1)}, b_2^{(1)}\}$ in the first $\max(\cdot)$ function, and it runs the same way in the second $\max(\cdot)$ function. Finally, the winners of both $\max(\cdot)$ functions could be combined together, which means that we can obtain different combinations of parameter sets in different scenarios.

When we apply equation (2.4) to each observation, we could know the exact winning parameter combination. Suppose we have M different winning combinations after checking all observations, then we can use the corresponding winning sets as labels and divide the whole population into M groups, where observations within each group are expected to share similar characteristics. This process helps us to capture the structure of the data.

3.3 Parameter Estimation

3.3.1 Estimation via Genetic Algorithms

Because of the form of CEV, the estimation is difficult, so we design an alternative approach to find a feasible solution for solving the optimization problem. We first decide the lower and upper bound of parameters in the $g(\cdot)$ function, then we search in the determined range to achieve optimal CEV estimators.

Considering to use $g(\cdot)$ function (2.4) for approximating variable X_I , then the following relationship holds

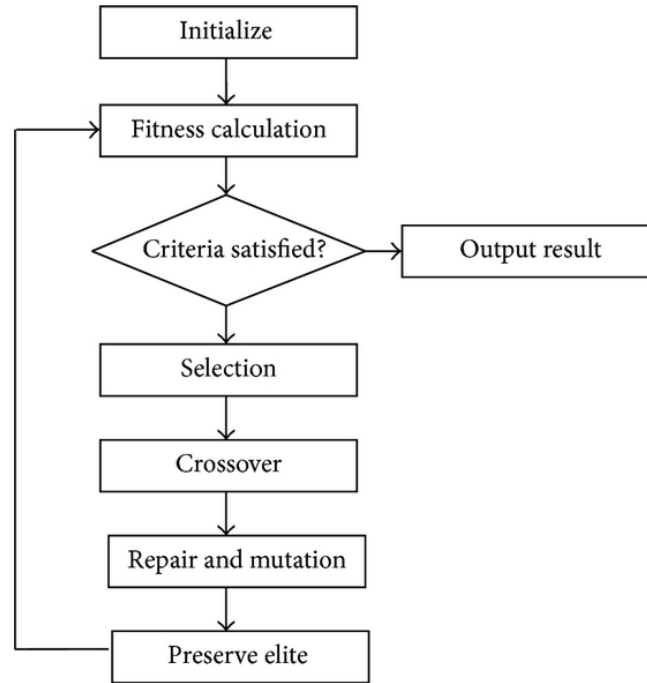
$$X_I = g(X; \mathbf{a}, \mathbf{b}) + \epsilon \quad (3.7)$$

where the random error ϵ follows a standard normal distribution. Then the model parameters are $\boldsymbol{\beta} = \{(a_i^{(j)}, b_i^{(j)}) : j = 1, 2; i = 1, \dots, k_j\}$, and the optimum parameter set could be defined as such that minimizes the sum of mean squared loss

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum (g(\widehat{X}; \boldsymbol{\beta}) - X_I)^2 \quad (3.8)$$

To achieve the optimal solution, we used the combination of genetic algorithm and ridge regression.

Figure 3.1: Genetic Algorithm Flow Chart from Singh and Sharma (2014)



The genetic algorithm (GA), developed by John Holland and his collaborators in the 1970s (John H. Holland (1976)), is a model or abstraction of biological evolution based on Darwin's theory of natural selection. Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as mutation, crossover and selection (Figure 3.1). The essence of GA involves the encoding of an optimization function as arrays of bits or character strings to represent chromosomes, the manipulation operations of strings by genetic operators, and the selection according to their fitness, with the aim to find a good (even optimal) solution to the problem concerned.

This algorithm is implemented by R package GA, and there are two types of inputs that we mostly concentrate on. The first one is the fitness function $F(X)$, which is defined as

the sum of mean squared loss, equation (3.8), so we have

$$F(X) = \sum (g(\widehat{X}; \boldsymbol{\beta}) - X_I)^2. \quad (3.9)$$

In order to achieve the optimal performance of $F(X)$, we need to search the parameters in a certain range. Suppose the initial lower bound of parameter set $\boldsymbol{\beta}$ is specified as $\boldsymbol{\beta}_L = \{a_L^{(1)}, b_L^{(1)}, a_L^{(2)}, b_L^{(2)}\}$, and the upper bound is represented as $\boldsymbol{\beta}_U = \{a_U^{(1)}, b_U^{(1)}, a_U^{(2)}, b_U^{(2)}\}$.

Remember that we have obtained an active variable set $\widehat{\mathcal{M}}_d^*$ after SEVIS screening, we then initialized the lower and upper bound of parameters and conducted genetic algorithm for optimization. The output of GA would be the estimated parameters $\widehat{\boldsymbol{\beta}}_{GA}$. Next, we applied the PWA function $g(X; \widehat{\boldsymbol{\beta}}_{GA})$ so that we could know the label of each observation by checking the winning parameter sets, and further, divide the whole population into M groups, where M is the total number of unique labels. Now, within each group, we conduct ridge regression with 5-fold cross validation, and the relationship between estimated parameters of group m and equation (3.6) from the previous example in Chapter 3.2 is

$$\widehat{a}_{(i_m, GA)}^{(1)} - \widehat{a}_{(i_m, GA)}^{(2)} = \widehat{\boldsymbol{\beta}}_{\text{ridge}, m}$$

If we fix the value of $\widehat{a}_{(i_m, GA)}^{(2)}$, then $\widehat{a}_{(i_m, GA)}^{(1)}$ could be expressed as

$$\widehat{a}_{(i_m, GA, \text{ridge})}^{(1)} = \widehat{a}_{(i_m, GA)}^{(2)} + \widehat{\boldsymbol{\beta}}_{\text{ridge}, m}. \quad (3.10)$$

Finally, the lower bound and upper bound of parameters could be updated as

$$\boldsymbol{\beta}_{L, \text{update}} = \min\{\widehat{\boldsymbol{\beta}}_{(GA)}, \widehat{\boldsymbol{\beta}}_{L, (GA, \text{ridge})}\} \quad (3.11)$$

$$\boldsymbol{\beta}_{U, \text{update}} = \max\{\widehat{\boldsymbol{\beta}}_{(GA)}, \widehat{\boldsymbol{\beta}}_{U, (GA, \text{ridge})}\} \quad (3.12)$$

where $\widehat{\boldsymbol{\beta}}_{(GA, \text{ridge})} = \{\widehat{\boldsymbol{\beta}}_{(i_m, GA, \text{ridge})} : 1 \leq m \leq M\}$.

We run this process iteratively until the optimal solution is achieved or the maximum number of iterations is reached. Then we treat the most up-to-date lower and upper bound of parameters in $g(\cdot)$ as initial values, and search for the best parameter set within this range to achieve the optimal CEV estimators.

3.3.2 CEV-PWA algorithm

The algorithm of cumulatively explained variability with piecewise affine function could be summarized as follows

Algorithm 1 CEV-PWA Algorithm

1. Implement CEV algorithm for screening, so as to obtain an active variable set $\widehat{\mathcal{M}}_d^*$ for variable X_I .
 2. Initialize the lower and upper bound of parameters as $\beta_L = \{a_L^{(1)}, b_L^{(1)}, a_L^{(2)}, b_L^{(2)}\}$ and $\beta_U = \{a_U^{(1)}, b_U^{(1)}, a_U^{(2)}, b_U^{(2)}\}$.
 3. Conduct genetic algorithm with fitness function $F(X)$ and store the estimated parameters $\widehat{\beta}_{GA}$.
 4. Apply the PWA function $g(X; \widehat{\beta}_{GA})$ to label each observation, and divide the population into M groups according to the label.
 5. Within each group, conduct ridge regression with 5-fold cross validation, and obtain the estimated parameters $\widehat{\beta}_{\text{ridge}}$.
 6. Update the lower bound and upper bound of parameters $\beta_{L,\text{update}}$ and $\beta_{U,\text{update}}$ using formula (3.11) and (3.12).
 7. Repeat Step 3-6 until the optimal solution is achieved or the maximum number of iterations is reached and store the determined lower and upper bound of parameters as $\beta_{L,\text{determined}}$ and $\beta_{U,\text{determined}}$.
 8. Utilize $\beta_{L,\text{determined}}$ and $\beta_{U,\text{determined}}$ as the range, search for the best parameter set β_{optimal} to achieve optimal CEV estimators including $\widehat{\text{CEV}}_1$, $\widehat{\text{CEV}}_3$, and $\widehat{\text{CEV}}_4$.
-

3.4 Simulation

In this section, we study the finite sample performance of CEV-PWA through several Monte Carlo simulations. We assess the performance of our proposed method by comparing the CEV results for several different settings. Through out the simulation, the results contain top 10 variables with the highest $CEV(Y|X_j)$ values for each model. Table 3.1 provides the types of CEV estimators that are evaluated in our study.

Table 3.1: Simulation Statistics.

Cases	Statistics	Explanation
Individual	$Y X_j$	The CEV in Y is explained by X_j .
	$X_j Y$	The CEV in X_j is explained by Y .
$g(\cdot)$ without X_j	$Y g(X_{[-j]})$	$\widehat{CEV}_1$. The CEV in Y is explained by $g(X_{[-j]})$.
	$g(X_{[-j]}) Y$	The CEV in $g(X_{[-j]})$ is explained by Y .
	$Y X_j, g(X_{[-j]})$	$\widehat{CEV}_3$. The CEV in Y is explained by X_j and $g(X_{[-j]})$.
$g(\cdot)$ with X_j	$Y g(X)$	The CEV in Y is explained by $g(X)$.
	$g(X) Y$	The CEV in $g(X)$ is explained by Y .
	$Y X_j, g(X)$	$\widehat{CEV}_4$. The CEV in Y is explained by X_j and $g(X)$.

Example 1. In this example, we investigate the performances of CEV-PWA through several linear models. The ordinary linear model (1.a) is similar to Example 1 in Fan and Lv (2008). The linear model with indicator function (1.b), the linear model with interaction (1.c), and the heteroscedastic linear model (1.d) are similar to the examples in Li et al. (2012).

$$(1.a) \ Y = \beta_1 X_1 + \beta_1 X_2 + \beta_1 X_{20} + \beta_1 X_{30} + \varepsilon,$$

$$(1.b) \ Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 I(X_{20} < 0) + \beta_4 X_{30} + \varepsilon,$$

$$(1.c) \ Y = \beta_1 X_1 X_2 + \beta_3 I(X_{20} < 0) + \beta_4 X_{30} + \varepsilon,$$

$$(1.d) \ Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 I(X_{20} < 0) + \beta_5 \exp(|X_{30}|) + \varepsilon.$$

The variables $X = (X_1, \dots, X_p)$ are generated from a normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$, where $\sigma_{ij} = \sigma^{|i-j|}$. The random error ε follows a standard normal distribution. We set the vector $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (5, 2, 7, 5, 2)$. The sample size $n = 500$, the dimension $p = 1000$, $\sigma = 0.5$.

Table 3.2 presents the simulation results for Example 1, and the largest CEV values are marked in bold. One can see that $\text{CEV}(Y|X_j, g(X))$ achieves the highest value almost all the time for model (1.a), and the cumulative explained variability is significantly improved comparing with $\text{CEV}(Y|X_j)$. For model (1.b), we notice that sometimes $\text{CEV}(Y|X_j, g(X_{[-j]}))$ obtains higher value than $\text{CEV}(Y|X_j, g(X))$, but they are still very close to each other, and are both much larger than $\text{CEV}(Y|X_j)$. For model (1.c), $\text{CEV}(Y|X_j, g(X))$ is conspicuously larger than other CEV estimators, especially when the value of $\text{CEV}(Y|X_j)$ is lower than 0.1. For model (1.d), $\text{CEV}(g(X)|Y)$ beats $\text{CEV}(Y|X_j, g(X_{[-j]}))$ many times, although their results are very close to each other. Also, there are two times that $\text{CEV}(g(X_{[-j]}|Y))$ achieve the highest. The results here indicate that in this case Y better 'predicts' the variability of $g(X)$ and $g(X_{[-j]})$. Overall, we observe that $\text{CEV}(Y|g(X_{[-j]})) \leq \text{CEV}(Y|X_j, g(X_{[-j]})) \leq \text{CEV}(Y|X_j, g(X))$, i.e., $\widehat{\text{CEV}}_1 \leq \widehat{\text{CEV}}_3 \leq \widehat{\text{CEV}}_4$. And the result from Example 1 shows that the PWA function $g(\cdot)$ performs well in capturing the structure of Y , and further explains the cumulative variability of Y more than a single variable X_j .

For all four models in Example 1, we use X_1, X_2, X_{20} , and X_{30} to construct the variable Y , so they can be taken as **truly active variables**. In model (1.a), one can notice from Table 3.2 that $X_4, X_{19}, X_{21}, X_{29}$, and X_{31} all have very small $\text{CEV}(Y|X_j)$ values (lower than 0.1), while the values of $\text{CEV}(Y|g(X_{[-j]}))$ and $\text{CEV}(Y|X_j, g(X_{[-j]}))$ are greater than 0.98. The joint information from other variables combined by $g(\cdot)$ function dominates the explained variability of Y , which indicates these 5 individual variables are not active from the simulation result.

Table 3.2: Simulation Results for Example 1.

Model	p = 1000, $\sigma = 0.5$								
	X_j	$Y X_j$	$X_j Y$	$Y g(X_{[-j]})$	$g(X_{[-j]}) Y$	$Y X_j, g(X_{[-j]})$	$Y X_j, g(X)$	$Y g(X)$	$g(X) Y$
1.a	1	0.46945	0.46887	0.84731	0.84610	0.96688	0.99265	0.99265	0.99265
	2	0.45881	0.45840	0.89002	0.88993	0.95462	0.99267	0.99265	0.99265
	3	0.13983	0.13985	0.99262	0.99261	0.99276	0.99274	0.99265	0.99265
	4	0.04247	0.04158	0.99271	0.99270	0.99272	0.99267	0.99265	0.99265
	19	0.03703	0.03909	0.99273	0.99273	0.99274	0.99267	0.99265	0.99265
	20	0.18563	0.18535	0.88877	0.88854	0.94783	0.99265	0.99265	0.99265
	21	0.06151	0.05446	0.99269	0.99268	0.99280	0.99268	0.99265	0.99265
	29	0.06547	0.06639	0.98998	0.98999	0.99007	0.99269	0.99265	0.99265
	30	0.22421	0.22490	0.88193	0.88211	0.94581	0.99265	0.99265	0.99265
	31	0.04857	0.04931	0.99274	0.99275	0.99284	0.99270	0.99265	0.99265
1.b	1	0.41709	0.41791	0.72345	0.72463	0.87966	0.93501	0.93423	0.93411
	2	0.22686	0.22939	0.90929	0.90882	0.92539	0.93448	0.93423	0.93411
	3	0.03689	0.03006	0.93444	0.93421	0.93463	0.93445	0.93423	0.93411
	20	0.09391	0.07323	0.83183	0.83191	0.92448	0.93464	0.93423	0.93411
	28	0.02903	0.02593	0.93468	0.93479	0.93540	0.93490	0.93423	0.93411
	29	0.13862	0.13784	0.93297	0.93315	0.93497	0.93477	0.93423	0.93411
	30	0.33692	0.33445	0.72229	0.72011	0.85740	0.93463	0.93423	0.93411
	31	0.07108	0.06608	0.93442	0.93456	0.93451	0.93443	0.93423	0.93411
	268	0.02892	0.01110	0.93211	0.93192	0.93252	0.93443	0.93423	0.93411
	828	0.02834	0.02256	0.93129	0.93138	0.93153	0.93432	0.93423	0.93411
1.c	1	0.16716	0.02254	0.41396	0.41345	0.45667	0.88037	0.87907	0.87539
	2	0.12724	0.01232	0.64817	0.65404	0.66326	0.87980	0.87907	0.87539
	19	0.03237	0.05943	0.87668	0.87242	0.87827	0.88146	0.87907	0.87539
	20	0.13075	0.14105	0.78455	0.77431	0.84451	0.88332	0.87907	0.87539
	28	0.04981	0.07560	0.85815	0.85150	0.85965	0.88086	0.87907	0.87539
	29	0.14069	0.16793	0.87718	0.87089	0.87929	0.88205	0.87907	0.87539
	30	0.38302	0.40668	0.65835	0.65173	0.79802	0.88102	0.87907	0.87539
	31	0.04658	0.05151	0.87441	0.86955	0.87522	0.87959	0.87907	0.87539
	331	0.02800	0.02463	0.86725	0.86015	0.86759	0.87985	0.87907	0.87539
	986	0.02700	0.00092	0.88151	0.87737	0.88220	0.88001	0.87907	0.87539
1.d	1	0.22419	0.33315	0.74065	0.71500	0.85913	0.84419	0.83513	0.84198
	2	0.12034	0.16186	0.85866	0.86369	0.86696	0.83683	0.83513	0.84198
	3	0.03620	0.04955	0.86582	0.86775	0.86661	0.83691	0.83513	0.84198
	19	0.02460	0.04348	0.85328	0.85467	0.85611	0.83721	0.83513	0.84198
	20	0.05434	0.10171	0.84159	0.83246	0.86964	0.83912	0.83513	0.84198
	29	0.02737	0.00739	0.83425	0.85140	0.83546	0.83656	0.83513	0.84198
	30	0.57178	0.00387	0.34925	0.47340	0.42695	0.83692	0.83513	0.84198
	31	0.04372	0.00297	0.84623	0.86427	0.84644	0.83546	0.83513	0.84198
	479	0.02586	0.01204	0.87057	0.87014	0.87092	0.84078	0.83513	0.84198
	886	0.02578	0.01772	0.86646	0.87081	0.86661	0.83880	0.83513	0.84198

After checking the rest of the variables X_1, X_2, X_3, X_{20} , and X_{30} in Example 1, they can be determined as active variables due to relatively high $\text{CEV}(Y|X_j)$ values and reasonable differences comparing $\text{CEV}(Y|X_j)$ with $\text{CEV}(Y|g(X_{[-j]}))$ and $\text{CEV}(Y|X_j, g(X_{[-j]}))$. Using the same strategy, for model (1.b) and (1.c), X_1, X_2, X_{20}, X_{29} , and X_{30} can be selected as active variables. For model (1.d), X_1, X_2, X_{20} , and X_{30} are taken as active. Overall, the truly active variables X_1, X_2, X_{20} , and X_{30} are selected by all models. By comparing the performances of different CEV estimators, one can find the dominating components for explaining the variability of Y and determine active variables as well.

Example 2. This example is designed to examine the performance of CEV-PWA for nonlinear models. We generate $X = (X_1, \dots, X_p)$ from a normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$, where $\sigma_{ii} = 1, i = 1, \dots, p$ and $\sigma_{ij} = \sigma, i \neq j$. Model (2.a) and (2.b) are non-parametric additive models with polynomial terms. The random error ε follows a standard normal distribution. We set $n = 500$ and $p = 1000$, $\sigma = 0$.

$$(2.a) \ Y = 5X_1 + 3X_2^2 + X_3^3 + \varepsilon$$

$$(2.b) \ Y = 5\sqrt{|X_1|} + 3X_2^2 + X_3^3 + \varepsilon$$

All simulation results for Example 2 are summarised in Table 3.3. For model (2.a), $\text{CEV}(Y|X_j, g(X))$ is most likely to achieve the highest value, while sometimes $\text{CEV}(Y|X_j, g(X_{[-j]}))$ is the largest one. But still, the CEV values of these two estimators are close to each other. This indicates that either the PWA function $g(\cdot)$ contains X_j or not, it still performs good in capturing the joint information from the covariates. As for model (1.b), $\text{CEV}(Y|X_j, g(X))$ always achieves the highest value comparing with other CEV estimators, which exactly fits the expectation from theoretical analysis. For both models, one can see that the relationship of $\widehat{\text{CEV}}_1 \leq \widehat{\text{CEV}}_3 \leq \widehat{\text{CEV}}_4$ holds as observed from Example 1.

Table 3.3: Simulation Results for Example 2.

Model	X_j	$Y X_j$	$X_j Y$	$Y g(X_{[-j]})$	$g(X_{[-j]} Y$	$Y X_j, g(X_{[-j]})$	$Y X_j, g(X)$	$Y g(X)$	$g(X) Y$
2.a	1	0.60206	0.62223	0.35465	0.34947	0.82961	0.86971	0.86357	0.87396
	2	0.10199	0.00964	0.73499	0.74675	0.77045	0.86868	0.86357	0.87396
	3	0.20600	0.18909	0.74853	0.76940	0.86858	0.86853	0.86357	0.87396
	125	0.03557	0.02024	0.87136	0.88218	0.87221	0.88398	0.86357	0.87396
	279	0.02967	0.03939	0.86737	0.87806	0.86853	0.86370	0.86357	0.87396
	306	0.03034	0.02835	0.86787	0.86862	0.87806	0.86589	0.86357	0.87396
	571	0.02975	0.01825	0.87019	0.88064	0.87075	0.86453	0.86357	0.87396
	580	0.02736	0.01121	0.87263	0.88386	0.87301	0.88392	0.86357	0.87396
	755	0.03088	0.01423	0.86955	0.86113	0.87055	0.87422	0.86357	0.87396
	806	0.03223	0.01954	0.87256	0.85316	0.87297	0.87436	0.86357	0.87396
2.b	1	0.12519	0.01415	0.68057	0.62396	0.72719	0.71862	0.66537	0.61494
	2	0.25705	0.01165	0.53913	0.51367	0.65445	0.66736	0.66537	0.61494
	3	0.47073	0.38545	0.31817	0.36828	0.65936	0.69198	0.66537	0.61494
	74	0.02616	0.01401	0.66965	0.63555	0.66781	0.66090	0.66537	0.61494
	181	0.02806	0.02671	0.65008	0.62033	0.65733	0.66547	0.66537	0.61494
	369	0.02409	0.02147	0.65289	0.61836	0.66760	0.67197	0.66537	0.61494
	515	0.02379	0.02342	0.66378	0.62702	0.66521	0.66354	0.66537	0.61494
	635	0.02831	0.00990	0.65635	0.61194	0.66052	0.66345	0.66537	0.61494
	984	0.02334	0.00789	0.65703	0.61330	0.66028	0.66968	0.66537	0.61494
	997	0.02998	0.00254	0.64577	0.61022	0.64969	0.66757	0.66537	0.61494

Based on the true models, X_1 , X_2 , and X_3 are used to construct variable Y and can be regarded as **truly active variables**. By checking the performance of $CEV(Y|X_j)$ in Table 5.3, we find that X_1 , X_2 , and X_3 hold relatively high values for both model (2.a) and (2.b). When comparing $CEV(Y|X_j)$ with $CEV(Y|g(X_{[-j]}))$, we notice that X_1 in model (2.a) has a greater $CEV(Y|X_1)$ value, 0.5025, than $CEV(Y|g(X_{[-1]}))$, 0.31875, while X_2 and X_3 have higher $CEV(Y|g(X_{[-j]}))$ values. One possible reason is that X_1 achieves the highest CEV value when it's evaluated as an individual variable, and it dominates the explained variability of Y . Hence, the joint information from other variables $g(X_{[-j]})$ can't compete with the individual X_1 . The similar condition happens to variable X_3 in model (2.b), which also achieves the highest $CEV(Y|X_j)$ value. If we add the performance of $CEV(Y|X_j, g(X_{[-j]}))$ into comparison, we can conclude that X_1 , X_2 , and X_3 should be taken as active variables. Due to the fact that both models in Example 2 contain polynomial

terms, the results here are consistent with the theoretical analysis that CEVs can capture nonlinear and asymmetric relationships efficiently.

3.5 Real data application

3.5.1 EEG Application: Visual Categorization

The functional connectivity studies the statistical interdependencies between the activities of different areas and is commonly measured with indexes such as correlation and coherence. There is a definition for connectivity which concerns with functional interactions but in a directed manner, named, effective connectivity. This connectivity considers the effect of one node over another, establishing causality in the interaction. Techniques such as the Local Field Potential (LFP) and the electroencephalography (EEG), measure the postsynaptic potential of hundreds of neurons, reflecting the average of their electrical activity.

The EEG data used here were drawn from OpenNeuro (Delorme (2020)) of ultra-rapid visual categorization of novel natural scenes. In this study, participants seated in a dimly lit room at 110 cm from a computer screen piloted from a PC computer. There are two tasks alternated, a categorization task and a recognition task, and target images and non-target images were equally likely presented in both tasks. Participants were tested in two recording phases. The first day contains 13 series, and the second day of 12 series, with 100 images per series. There are 14 participants in total, and we concentrate on the performance of the 1st series in the the first day. For each participant, the signals of 31 channels are measured, and the primary goal is to study the cumulative explained variability regarding to channels.

Figure 3.2(a) shows the relative positions of the 31 channels in human brain, and Figure 3.2(b) shows the connectivity network of these channels by our model. In Figure 3.2(b), the side bar represents the value of $CEV(X_I|X_J)$, and the thicker and darker of the lines between channels, the higher value of the cumulative explained variability is. The size and color of the channel circles also reveal the influence of these channels to other channels. Note that not all the 31 channels are plotted here, and only the most influential ones are

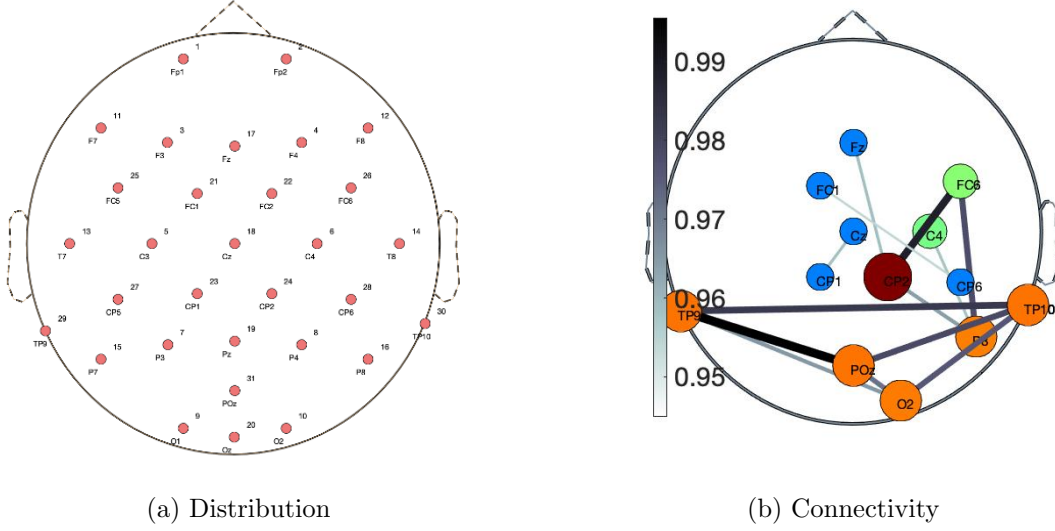


Figure 3.2: (a) shows the distribution of channels in human brain; (b) shows the connectivity network of the channels identified by our model.

shown. One can observe from this figure that some channels are highly correlated with each other and could be formed to different groups based on the level of connectivity. In addition to the pairwise connectivity of the channels shown in this undirected graph, we also want to investigate the causal relationship among these channels as well as the cumulative explained variability from the group of channels joint by the PWA function $g(\cdot)$.

To further study the performance of the proposed CEV-PWA algorithm and the causal relationship among the channels, we summarized the results of 12 channels in Figure 3.3 and Figure 3.4. These 12 channels draw our attention since they are characterized by special trends. The results for all 31 channels are summarized in the Appendix. Using the same statistics provided in Table 3.1, we applied different colors and line types to denote these CEVs. Note that the Y in the legend represents each of the 12 channels shown in the title of the subplot, and X includes the rest 30 channels except Y . Figure 3.3 shows the average CEV values including the pairwise $CEV(Y|X_j)$ (orange line) and $CEV(X_j|Y)$ (yellow dash line). For all of the 12 channels, the lines of $Y|X_j$ and $X_j|Y$ are significantly much lower than the others, which suggests that our proposed CEV-PWA algorithm works

well in explaining more variability of a single variable using the $g(\cdot)$ function. Also, we find that for channels C4, F7, F8, and P4r, the value of $CEV(X_j|Y)$ is greater than $CEV(Y|X_j)$ for almost all the other channels. This means that channels C4, F7, F8, and P4r WeCon cause the other individual channels.

Figure 3.3: Average CEV values for each channel in visual categorization. Channels C4, F3, F4, F7, F8, FP2, O2, P4r, PZ, T4, T5r, T6r are included.



Figure 3.4: Average CEV values for each channel except $Y|X$ and $X|Y$ in visual categorization. Channels C4, F3, F4, F7, F8, FP2, O2, P4r, PZ, T4, T5r, T6r are included.



Then Figure 3.4 provides a clearer view of the other CEVs with the $g(\cdot)$ function. One can see that the pink line is always on the top of other lines or at least very close to the top line for all 12 channels, which indicates that $\{X_j, g(X)\}$ performs good in explaining the variability of Y , and the result here is consistent with the simulation. Also, the blue dot line ($\text{CEV}(g(X)|Y)$) in channels C4, F3, FP2, and T4 is on top of other lines for almost all other 30 channels (X_j), and this suggests that Y WeCon causes $g(X)$, i.e., C4, F3, FP2, and T4 WeCon cause other channels joint by $g(\cdot)$ function. Considering the C4 performance in Figure 3.2, we could identify C4 as an important channel not only WeCon causes other individual channel, but also causes the joint of other channels by $g(\cdot)$. Then for channel O2, the orange line $\text{CEV}(Y|g(X_{[-j]}))$ is higher than the yellow line $\text{CEV}(g(X_{[-j]}|Y))$ all the time, and this suggests $g(X_{[-j]})$ WeCon causes Y , the channel O2.

Table 3.4: CEV results for top 2 selected highest asymmetric nonlinear correlation channels.

Y	X_j	$Y X_j$	$X_j Y$	$Y g(X_{[-j]})$	$g(X_{[-j]} Y)$	$Y X_j, g(X_{[-j]})$	$Y X_j, g(X)$	$Y g(X)$	$g(X) Y$
P4	I	0.89488	0.89645	0.97121	0.96926	0.97442	0.9804	0.97829	0.97814
P4	P3	0.81273	0.84741	0.96971	0.96906	0.97334	0.97804	0.97649	0.97592
P4r	P3	0.74086	0.77406	0.83632	0.83708	0.89516	0.91882	0.90808	0.90771
P4r	F3	0.66634	0.70059	0.90864	0.91023	0.92481	0.90346	0.88971	0.88436
PZ	OZ	0.83247	0.8287	0.93934	0.9436	0.94637	0.94783	0.94263	0.94674
PZ	P4	0.7837	0.79867	0.92894	0.93117	0.94223	0.95107	0.94673	0.94917
PZr	O2r	0.93885	0.94344	0.98236	0.98182	0.98628	0.98519	0.98399	0.98424
PZr	T5	0.92849	0.92369	0.98612	0.9862	0.98927	0.99105	0.98916	0.98904
T3	T5	0.87619	0.88348	0.94402	0.9462	0.95287	0.96069	0.95907	0.95953
T3	I	0.82513	0.85246	0.95181	0.95393	0.95405	0.95522	0.95426	0.95604
T4	CB1r	0.71515	0.72527	0.90506	0.9116	0.90695	0.90398	0.89976	0.90893
T4	C4	0.68498	0.65841	0.9034	0.90925	0.91439	0.91337	0.90766	0.91474
T5	O2	0.94918	0.95484	0.98161	0.98082	0.98231	0.98473	0.98337	0.98205
T5	CB1r	0.94658	0.95269	0.98231	0.98107	0.98324	0.98125	0.98034	0.97893
T5r	F3	0.77631	0.76094	0.93222	0.93418	0.93655	0.9406	0.93671	0.94096
T5r	O1	0.73063	0.74657	0.93304	0.93694	0.93844	0.94162	0.94071	0.94308
T6	O2	0.98116	0.9817	0.98311	0.98295	0.98389	0.98637	0.98429	0.98402
T6	CB2r	0.97644	0.97728	0.98361	0.98373	0.98526	0.99383	0.98419	0.98402
T6r	O1	0.87165	0.86459	0.93747	0.94014	0.94654	0.93808	0.92234	0.92934
T6r	T5r	0.80377	0.79101	0.93055	0.93414	0.94361	0.93732	0.9302	0.93784

Table 3.5: CEV results for top 2 selected highest asymmetric nonlinear correlation channels.
(Continued)

Y	X	$Y X_j$	$X_j Y$	$Y g(X_{[-j]})$	$g(X_{[-j]}) Y$	$Y X_j, g(X_{[-j]})$	$Y X_j, g(X)$	$Y g(X)$	$g(X) Y$
C3	T4	0.84631	0.84453	0.86294	0.85940	0.90375	0.87889	0.87148	0.87213
C3	CB2	0.77338	0.77028	0.89856	0.89788	0.89766	0.89426	0.89550	0.89239
C4	CB1r	0.70682	0.78341	0.94442	0.94865	0.94596	0.95022	0.94826	0.95094
C4	T6	0.68191	0.67610	0.94734	0.95650	0.95235	0.94899	0.94388	0.94973
CB1	PZr	0.72358	0.69435	0.87737	0.86955	0.88985	0.90876	0.90437	0.90643
CB1	FZ	0.70739	0.68712	0.90124	0.89982	0.91284	0.91062	0.90126	0.90246
CB1r	C4	0.82083	0.81964	0.96901	0.95995	0.97220	0.97611	0.97438	0.97646
CB1r	O2r	0.80893	0.82281	0.96779	0.96852	0.97463	0.97556	0.97131	0.97213
CB2	O2r	0.94837	0.92684	0.98215	0.98180	0.98561	0.98979	0.98981	0.98985
CB2	CB2r	0.89642	0.87386	0.98772	0.98824	0.98777	0.99037	0.99035	0.99046
CB2r	I	0.89374	0.91900	0.92842	0.92984	0.92996	0.93132	0.93060	0.93229
CB2r	OZ	0.87484	0.86127	0.93091	0.93043	0.97529	0.94344	0.93052	0.93141
F3	CB1r	0.84822	0.82206	0.93329	0.94391	0.95384	0.95833	0.95347	0.95735
F3	O1r	0.84316	0.85201	0.93942	0.94551	0.94622	0.95180	0.94743	0.95468
F4	C4	0.87871	0.88494	0.94759	0.95062	0.95688	0.96126	0.95506	0.95871
F4	FZ	0.83418	0.80831	0.93162	0.93571	0.94851	0.95663	0.95585	0.96028
F7	P3r	0.78444	0.80708	0.95498	0.95699	0.95455	0.95404	0.95318	0.95412
F7	I	0.77988	0.85976	0.95734	0.95921	0.95944	0.95257	0.95043	0.95218
F8	FP2	0.45311	0.40148	0.77915	0.78931	0.79463	0.81889	0.81407	0.81451
F8	T6r	0.42584	0.53947	0.79600	0.80774	0.80744	0.82691	0.82005	0.82537
FP1	FP2	0.64232	0.65160	0.85541	0.86471	0.91690	0.96282	0.96119	0.96235
FP1	F7	0.55218	0.53642	0.94955	0.95110	0.96139	0.96309	0.95955	0.96026
FP2	FP1	0.69426	0.63272	0.81889	0.83587	0.86161	0.89414	0.87891	0.90097
FP2	C3	0.54597	0.55767	0.84016	0.85044	0.85064	0.88873	0.87997	0.89137
FZ	F3	0.84655	0.86539	0.92129	0.92183	0.95600	0.94730	0.94295	0.94326
FZ	CB2r	0.80845	0.69302	0.94390	0.94416	0.94691	0.94406	0.93950	0.94032
I	P3r	0.97704	0.96644	0.99167	0.99142	0.99270	0.99267	0.99243	0.99195
I	PZ	0.96817	0.96676	0.99216	0.99184	0.99304	0.99380	0.99316	0.99277
O1	P4	0.80531	0.76080	0.91391	0.91792	0.92210	0.94413	0.94289	0.94477
O1	I	0.80340	0.81206	0.91270	0.91000	0.91794	0.93101	0.92513	0.92536
O1r	CB1r	0.96208	0.96632	0.98906	0.98981	0.99063	0.99223	0.99193	0.99223
O1r	O2	0.95053	0.94959	0.99144	0.99145	0.99149	0.98755	0.98716	0.98882
O2	T6	0.83432	0.80572	0.94250	0.93412	0.95046	0.95029	0.94683	0.93713
O2	T5	0.82343	0.82491	0.94412	0.93424	0.94494	0.94875	0.94831	0.93917
O2r	P3r	0.90601	0.91193	0.96805	0.96822	0.96879	0.97497	0.97419	0.97438
O2r	T5	0.89652	0.88798	0.96191	0.96368	0.96605	0.97587	0.97492	0.97478
OZ	P3r	0.93991	0.96597	0.99199	0.99218	0.99469	0.99710	0.99372	0.99379
OZ	C4	0.89506	0.90491	0.99309	0.99306	0.99379	0.99396	0.99363	0.99377
P3	P3r	0.90641	0.85513	0.98138	0.98104	0.98326	0.98348	0.98225	0.98254
P3	O1r	0.89771	0.87297	0.97964	0.97935	0.98335	0.98447	0.98136	0.98241

We select the top 2 highest asymmetric nonlinear correlation channels relating to each of the 31 channels, and compute the average CEV values containing all the statistics from Table 3.1. The summarized results are in Tables 3.4 and 3.5, and the highest value is marked in bold. In most cases, $CEV(Y|X_j, g(X))$ achieves the largest value, although there are times $CEV(Y|X_j, g(X_{[-j]}))$ being greater. But still, the values of these two statistics are close to each other. We can still observe $\widehat{CEV}_1 \leq \widehat{CEV}_3 \leq \widehat{CEV}_4$ holds. We can then conclude that the PWA function $g(\cdot)$ performs well in capturing the structure of Y , and further explains the cumulative variability of Y more than an individual variable X_j .

Table 3.6: The frequency of each channel being selected as top 2 highest asymmetric nonlinear correlation channels.

Channel	C3	C4	CB1	CB1r	CB2	CB2r	F3	F4	F7	F8	FP1	FP2	FZ	I	O1	O1r
Frequency	1	4	0	5	1	3	3	0	1	0	1	2	2	5	2	2
Channel	O2	O2r	OZ	P3	P3r	P4	P4r	PZ	PZr	T3	T4	T5	T5r	T6	T6r	
Frequency	3	3	2	2	5	2	0	1	1	0	1	4	1	2	1	

After counting the number of times each channel is selected as the top 2 important channels shown in Table 3.4 and 3.5, we organize the frequency with the corresponding channel to Table 3.6. As we noticed in Figure 3.3 and Figure 3.4, the frequencies of channel C4, F3, and O2 are all relatively high, which also partially explain the fact that their performance is more distinguishable from the other channels.

3.5.2 EEG Application: Parkinson's Disease

Parkinson's disease (PD) is a common progressive neurodegenerative disorder of the central nervous system as stated by Valls-Solé and Valdeoriola (2002), and it's affecting about 2 – 3% of the elderly population. Savitt et al. (2006) claimed that the motor clinical symptoms of PD, such as tremor, slowness of movement, rigidity and postural instability result from dopaminergic deficiency in the basal ganglia. What's more, Cooper et al. (1991) concluded that PD could cause non-motor impairments such as cognitive dysfunction, in the

early stage of PD. However, the diagnosis of PD via clinical symptoms could be very difficult, especially in the early stage when there is no significant motor symptoms and obvious cognitive dysfunction. Still, the characterization of PD from the neuroscience point of view provides us with an alternative way to explore and quantify the corresponding brain functionality and help to improve diagnostic certainty.

Electroencephalogram (EEG) signals represent changes of brain activities in different regions, but it is hard to manually analyze the signals because of their nature of non-linear, non-stationary, and complex. Using traditional machine learning methods for the detection of PD requires several manual steps including decomposition, feature extraction, and classification. Considering these limitations, many recent researches suggest using deep learning methods, such as convolutional neural network (CNN), for the diagnosis of PD (Oh et al. (2020), Khare et al. (2021), Pereira et al. (2016), Gunduz (2019)). In addition, effective connectivity on resting-state EEG signals is studied to understand cortical-cortical interactions, so that the neurophysiological mechanisms underlying the PD symptoms could be explored (Maggioni et al. (2021), Asher et al. (2021), Formaggio et al. (2021), Arasteh et al. (2021)). In our research, we not only study the pairwise connectivity between EEG channels, but also have interests in the causal relationship between individual channels and the network joint by other channels. Both PD patients and healthy control subjects are considered here, hence, we are able to measure the performances of these two groups and further analyse the differences by comparison.

The data we are using here comes from James F Cavanagh (2021), and can be downloaded from OpenNEURO, which was collected in Cognitive Rhythms and Computation Lab at University of New Mexico. It contains rest and 3 stimulus auditory oddball data with 22 Parkinson patients (PD) and 19 matched controls (HC). PD came in twice separated by a week, either ON or OFF medication, while HC only came in once. The task contains the instructions for eyes open and eyes closed, and triggers are included for instructed one

minute spans for open or closed (OC or CO) before the task. Subjects in this experiment have an accelerometer taped to their most tremor affected hand. For each subject, the signals of 63 channels are recorded.

In order to study the performance of the proposed CEV-PWA algorithm and the causal relationship among these channels for Parkinson’s disease, we summarized the results of 12 channels in the following figures. These 12 channels draw our attention since they are characterized by different trends between the PD group and the healthy control group. The results for all 63 channels are summarized in the Appendix. Using the same statistics provided in Table 3.1, we applied different colors and line types to denote these CEVs. Note that the Y in the legend represents each of the 12 channels shown in the title of the subplot, and X includes the rest 62 channels except Y.

Figure 3.5 and Figure 3.6 provide the average CEV values for each of the 12 channels in both Parkinson’s (PD) and healthy control (HC) groups. Notice that for all of the 12 channels, the lines of $Y|X_j$ and $X_j|Y$ are significantly much lower than the others, which suggests that our proposed CEV-PWA algorithm works great in explaining more variability of a single variable using the $g(\cdot)$ function. By comparing Figure 3.5 and Figure 3.6, we find that antero-frontal channels AF7, AF8, and frontal-pole channel FP2 can actively influence other individual channels in the HC since the line $X_j|Y$ is always higher than the line $Y|X_j$, which is not the case in the PD group. Central-parietal channels CP1, CP5, and frontal channel F3 seem to be passively impacted by other individual channels in the HC as we can observe that $Y|X_j$ is greater than $X_j|Y$, but they don’t show the same trend in PD group. Also, frontal-central channels FC3, FZ, and parietal channel P5 can actively influence other individual channels if we compare the lines of $Y|X_j$ and $X_j|Y$, while the results in the HC don’t indicate this. The sub-figures of frontal-temporal channel FT9 indicate that this channel could significantly influence other individual channels in the PD group, which is not the case in the HC. In addition, channel PO7 shows dominating other individual channels in

the PD group, while in the HC group, it seems to be passively influenced by other channels. The different performance in the PD and the HC groups may indicate that the channels we mentioned above might be influenced by Parkinson's disease itself or the disorder of these channels may cause the Parkinson's disease.

Figure 3.5: Average CEV values for each channel in the Parkinson's patients group (PD). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.

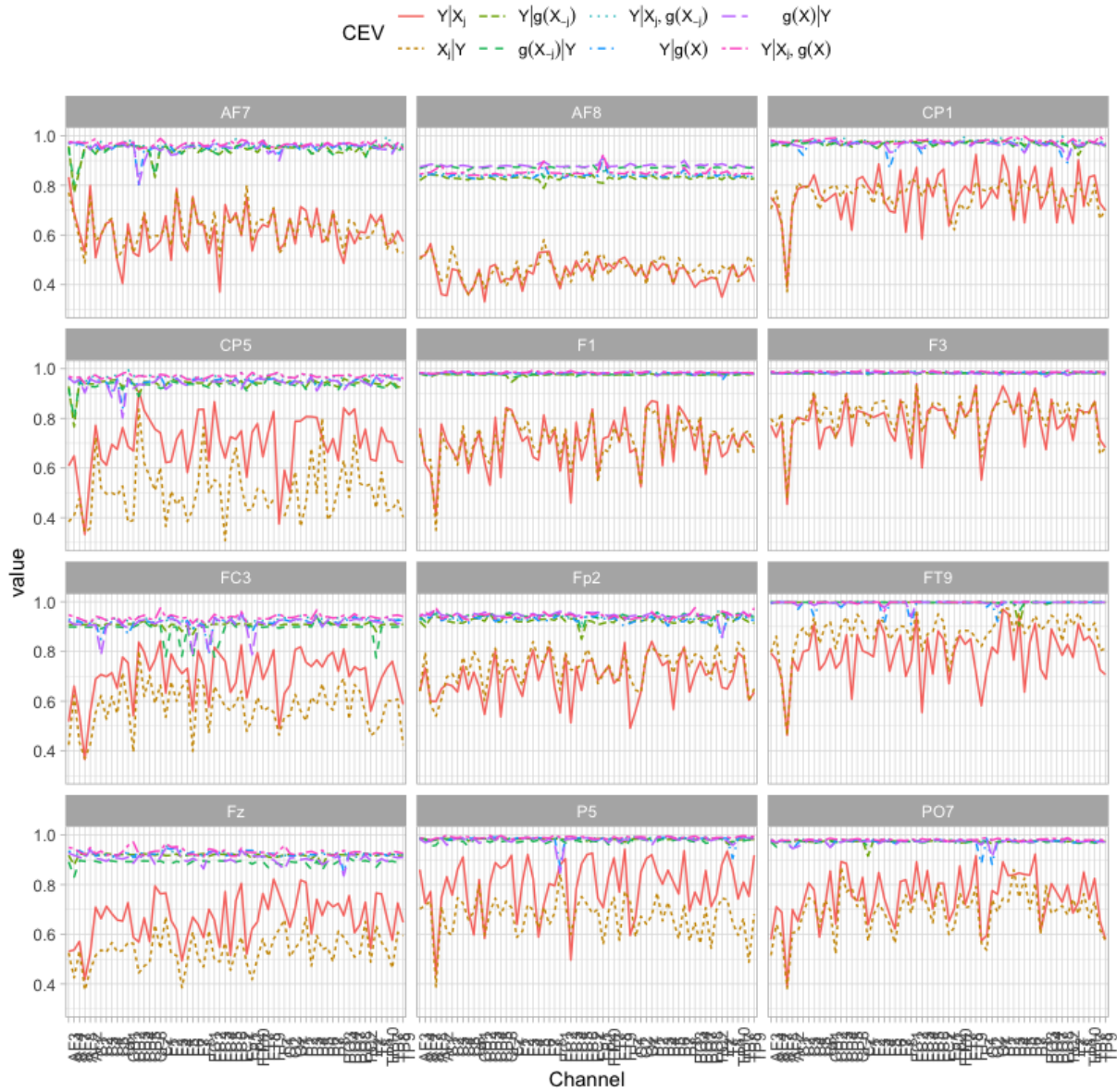
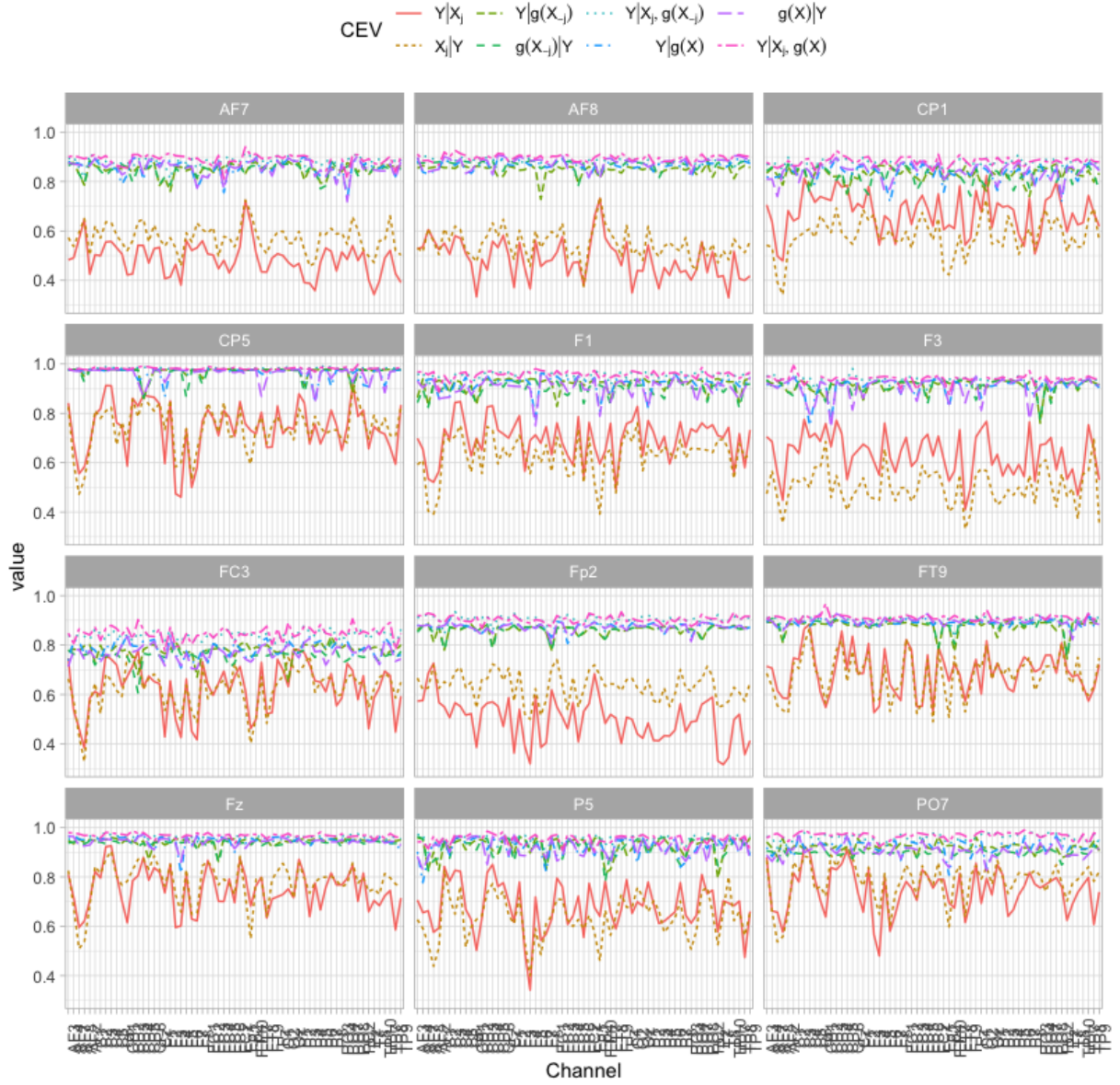


Figure 3.6: Average CEV values for each channel in the healthy control group (HC). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.



When we take a closer look at the results without pairwise CEV values, Figure 3.7 and Figure 3.8 show the average CEV values for each of the 12 channels in both Parkinson's (PD) and healthy control (HC) groups. Antero-frontal channel AF8 shows great impact on the network of the rest channels since the blue dot line $g(X)|Y$ is on top of almost all other lines in the PD group (Figure 3.7), while we can't observe this trend in the HC group (Figure 3.8). Frontal channel F1 seems to be passively influenced by the network of other channels in the HC as the line $Y|g(X_{-j})$ is higher than line $g(X_{-j})|Y$. Frontal-pole Channel FP2 in Figure 3.7 (PD) shows the line $g(X)|Y$ is on top of other lines, meaning that it WeCon causes the network constructed by other channels. While channels FZ and P5 performs in the opposite way. For these two channels, the sub-figures in the PD group display the trend that line $Y|g(X)$ is higher than line $g(X)|Y$, which suggests the network of other channels WeCon causes these two channels. Then, the results of parietal-occipital channel PO7 in the HC group indicate that it could be passively influenced by the network of other channels, while it's not remarkable in the PD group.

Comparing the performances of PD and HC group by both pairwise and network causal relationship, we notice that these channels have significantly different activities in these two groups. Channels FC3, FZ, and P5 in the PD group show that for both pairwise and network relationships, they are always more likely to be passively influenced by other channels. For channels AF8 and FP2, they perform in the opposite way with regard to pairwise and network relationships. AF8 and FP2 dominate other individual channels in the HC, which is not observed in the PD. On the contrary, these two channels can actively influence the network of other channels in the PD, which is not the case for HC. What's more, by analyzing the pairwise relationship of channel PO7, we find that it's passively influenced by other individual channels in the PD group, while it WeCon causes other channels in the HC group.

Figure 3.7: Average CEV values for each channel except $Y|X$ and $X|Y$ in the Parkinson's patients group (PD). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.

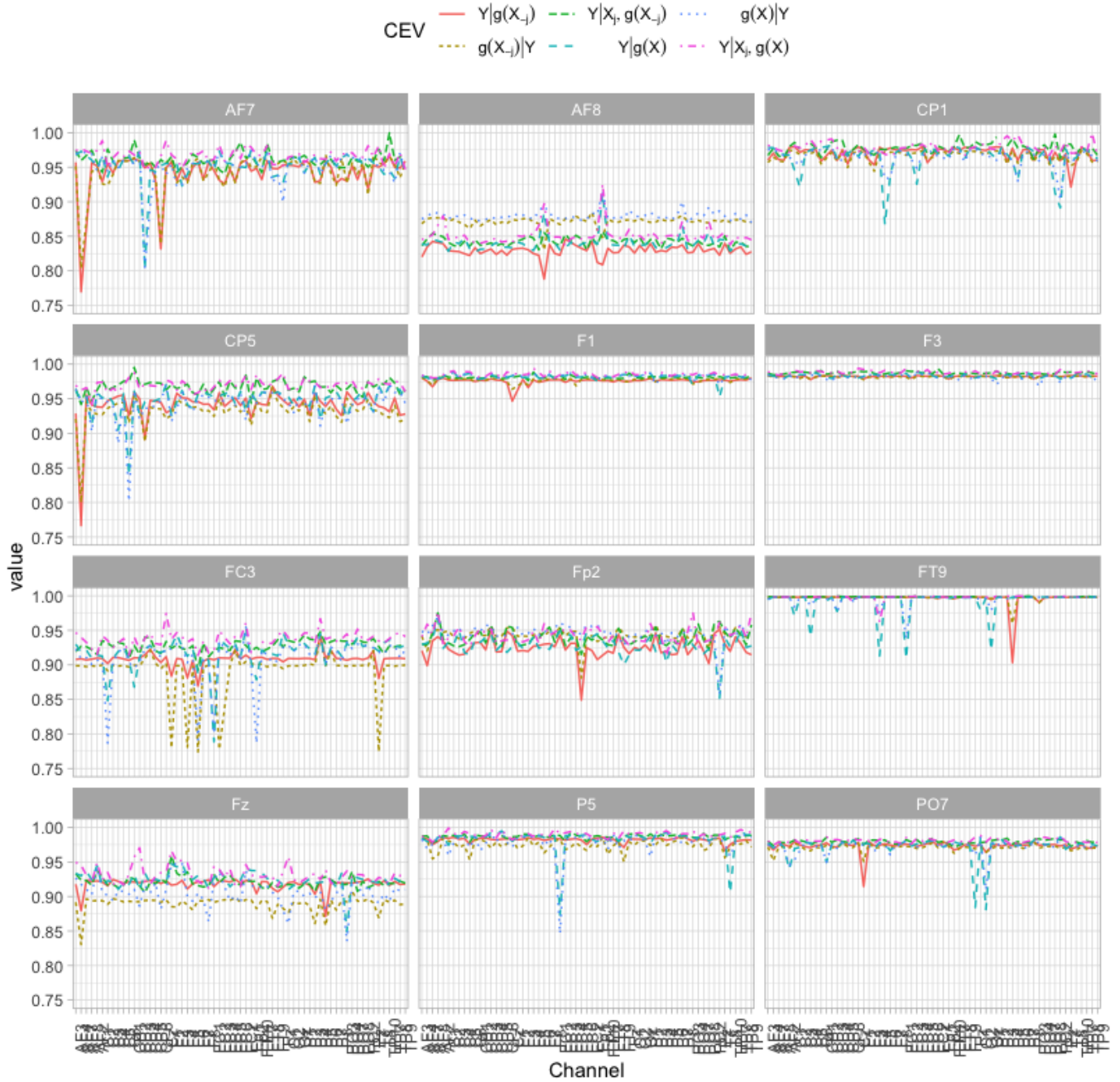
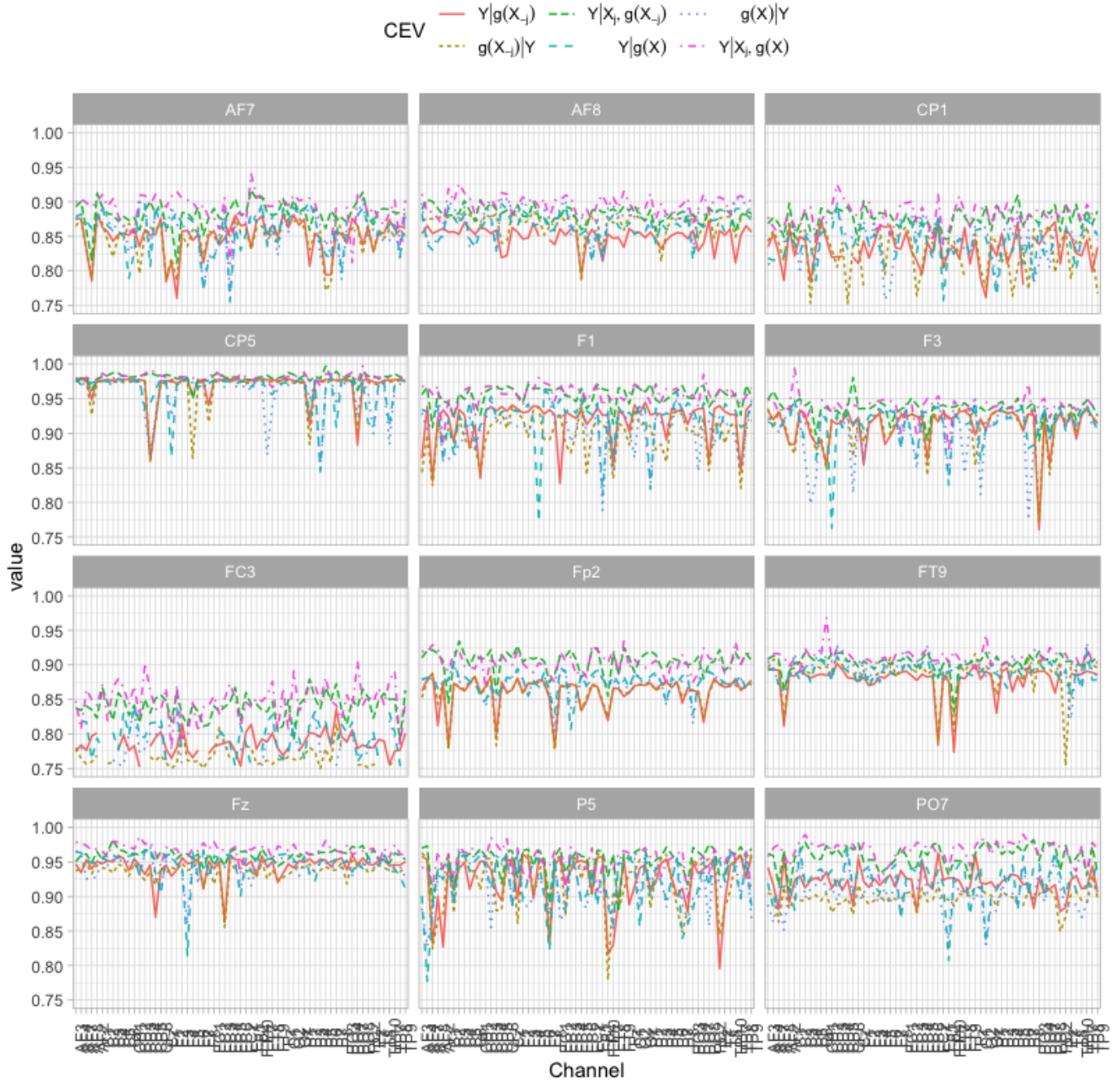


Figure 3.8: Average CEV values for each channel except $Y|X$ and $X|Y$ in the healthy control group (HC). Channels AF7, AF8, CP1, CP5, F1, F3, FC3, Fp2, FT9, Fz, P5, PO7 are included.



To analyze the pairwise causal relationship among these channels, we have Figure 3.9 as the directed network graph with different colors representing different regions in human brain and the width of lines between channels indicating the level of connectivity. The arrows in the network graph represent the causal relationship. Overall, channel FT9 is more actively influencing other channels in the PD group as it's pointing out to others, while it's more likely to be passively influenced by other channels as shown in the HC group. Also, we notice that channels CP5, Fz, and PO7 perform in the opposite way comparing with FT9. It seems that CP5, Fz, and PO7 show more initiative to cause other channels in the HC group, however, they are more likely to be the receiver channels in the PD group.

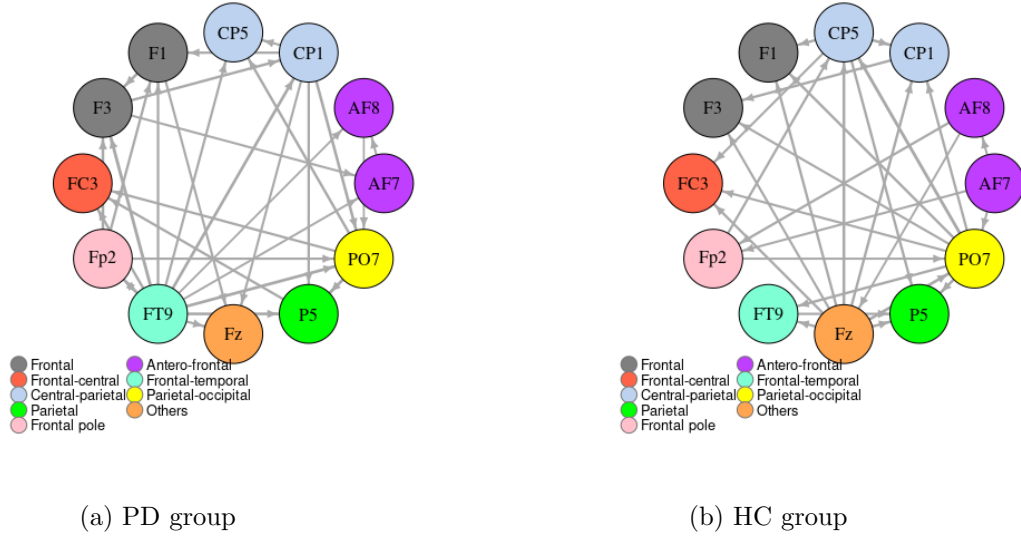
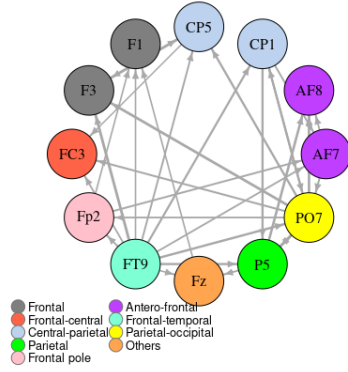


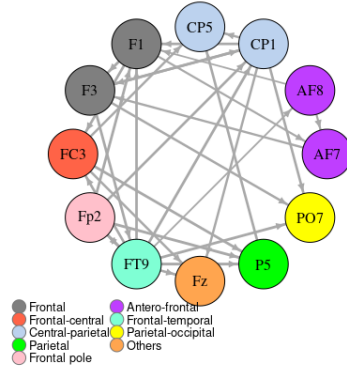
Figure 3.9: (a) shows the directed connectivity network of channels in Parkinson's patients group (PD); (b) shows the directed connectivity network of channels in healthy control group (HC).

Tolosa et al. (2021) stated that age is the most important risk factor for developing Parkinson's disease, hence, we constructed the directed network graph for young ($age < 69$) and old ($age \geq 69$) subjects in both PD and HC group. Figure 3.10 provide the corresponding figures with each sub-figure representing Young and Old sub-groups. First, we notice that the level of connectivity (width of the lines) in the PD-Young group is much

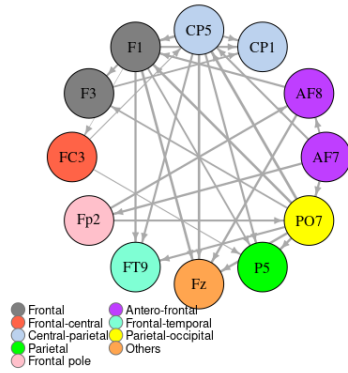
weaker than the PD-Old group, while in the HC group, both Young and Old subgroups show similar level of connectivity.



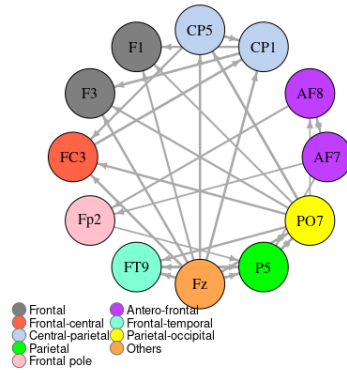
(a) PD - Young group



(b) PD - Old group



(c) HC - Young group



(d) HC - Old group

Figure 3.10: The directed connectivity network of channels in Parkinson's patients group (PD) and healthy control group (HC) for Young ($age < 69$) and Old ($age \geq 69$) sub-groups.

Considering both Young and Old subgroups, it seems that channel CP5, Fz, and PO7 have more connections with other channels in the HC group, however, these channels seem to be less active in the PD group. If we compare only the Young groups for both PD and HC, the overall connections in HC are significantly stronger than those in PD. The weaker

connections are probably caused by Parkinson’s disease for young subjects.

3.5.3 fMRI Application: Major Depressive Disorder

During the last several decades, functional magnetic resonance imaging (fMRI) has played an essential role in studying the pathogenesis of major depressive disorder (MDD). MDD is a mood disorder, which will last for six months or longer. And it can be characterized by feelings of persistent sadness or diminished interest in daily activities. MDD is one of the most commonly diagnosed mental disorder in most first world countries, including Europe, China, and the United States.

Because of the similarities between the symptoms of MDD and other mental disorders including schizophrenia and bipolar disorder, it can be difficult to differentiate the conditions of MDD and often requires highly-trained psychiatrists. However, it’s critical for the patients with MDD to be timely diagnosed and treated. Non-psychiatric disorders, such as central nervous system (CNS) disorders, endocrine disorders, infectious diseases, and sleep disorders might come with depressive-like symptoms, which makes the diagnosis of MDD more complicated in the clinic.

Currently, the diagnosis of MDD depends on behavioral observations and symptoms reported by the patients, along with the Hamilton Scale for Depression and DSM-V criteria as stated in Zhuo et al. (2019). There are no imaging biomarkers widely accepted for the assessment of MDD. Nevertheless, scientists have been utilizing diagnostic imaging techniques to study the underlying mechanism of MDD over the past several decades, mainly concentrating on structural and functional brain alterations.

In order to study the etiology and pathogenesis of psychiatric disorders, such as MDD and schizophrenia, fMRI are widely used for scientific researches. Liang et al. (2020) focused on fMRI-data-driven identification and validation of potential DMN-pattern-based

(default mode network) MDD subtypes to parse heterogeneity of the disorder. Many existing fMRI studies of MDD, such as Friston (2011), focused on undirected synchronizations rather than causal influence (directed connectivity) among the neural populations that lead to the regional fMRI signals. Yang et al. (2021), Dong et al. (2019), Li et al. (2020), and Shi et al. (2019) implemented Granger causal analysis (GCA) to investigate effective connectivity between brain regions, while they mostly concentrated on several main regions and considered pairwise connectivity. Dynamic causal modeling (DCM) is also a popular method utilized for MDD research. Li et al. (2017) measured resting-state effective connectivity (EC) among four regions using spectral DCM with 27 MDD patients and 27 healthy controls. The study showed decreased EC from the left parietal cortex to other regions in the unmedicated patients. Also, Kandilarova et al. (2018) conducted spectral DCM to study EC difference among eight brain regions on a single brain hemisphere and found that depressive patients showed a significantly reduced EC from the anterior insula to the middle frontal gyrus and increased EC from the amygdala to the anterior insula. Although the aforementioned fMRI research were able to feature abnormal EC in the MDD, they are still limited by small number of brain regions (< 10) and considering pairwise causality only. In our research, we not only study the pairwise connectivity among brain regions but also consider the effect of networks jointed by multiple regions. In the meantime, we are able to implement the proposed method for all brain regions (48).

Figure 3.11: Harvard Oxford Cortical atlas.

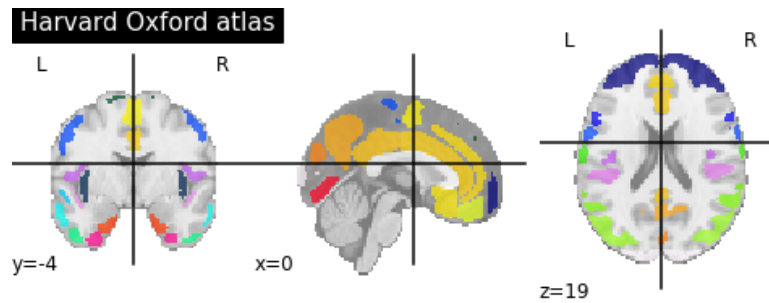


Table 3.7: Harvard Oxford Atlas: IDs and corresponding ROI names

ID	ROI Names	ID	ROI Names
V1	Frontal Pole	V25	Frontal Medial Cortex
V2	Insular Cortex	V26	Juxtapositional Lobule Cortex
V3	Superior Frontal Gyrus	V27	Subcallosal Cortex
V4	Middle Frontal Gyrus	V28	Paracingulate Gyrus
V5	Inferior Frontal Gyrus, pars triangularis	V29	Cingulate Gyrus, anterior division
V6	Inferior Frontal Gyrus, pars opercularis	V30	Cingulate Gyrus, posterior division
V7	Precentral Gyrus	V31	Precuneous Cortex
V8	Temporal Pole	V32	Cuneal Cortex
V9	Superior Temporal Gyrus, anterior division	V33	Frontal Orbital Cortex
V10	Superior Temporal Gyrus, posterior division	V34	Parahippocampal Gyrus, anterior division
V11	Middle Temporal Gyrus, anterior division	V35	Parahippocampal Gyrus, posterior division
V12	Middle Temporal Gyrus, posterior division	V36	Lingual Gyrus
V13	Middle Temporal Gyrus, temporooccipital part	V37	Temporal Fusiform Cortex, anterior division
V14	Inferior Temporal Gyrus, anterior division	V38	Temporal Fusiform Cortex, posterior division
V15	Inferior Temporal Gyrus, posterior division	V39	Temporal Occipital Fusiform Cortex
V16	Inferior Temporal Gyrus, temporooccipital part	V40	Occipital Fusiform Gyrus
V17	Postcentral Gyrus	V41	Frontal Operculum Cortex
V18	Superior Parietal Lobule	V42	Central Opercular Cortex
V19	Supramarginal Gyrus, anterior division	V43	Parietal Operculum Cortex
V20	Supramarginal Gyrus, posterior division	V44	Planum Polare
V21	Angular Gyrus	V45	Heschl's Gyrus (includes H1 and H2)
V22	Lateral Occipital Cortex, superior division	V46	Planum Temporale
lateral occipital cortex (V23)	Lateral Occipital Cortex, inferior division	V47	Supracalcarine Cortex
V24	Intracalcarine Cortex	V48	Occipital Pole

The data we are using here comes from Bezmaternykh et al. (2020), and can be downloaded from OpenNEURO. This fMRI data is used for resting state with closed eyes for 22 patients with depression and 19 healthy participants. Note that the important details about brain activity could be hidden when using whole-brain maps. Also, we would probably find a significant effect of incongruent-congruent. However, the significant effect might be caused by greater incongruent, or a very negative congruent, or the mixture of the two. Region of interest (ROI) analysis is the only way to decide what is driving the effect, and this is especially essential when we need to deal with interactions and more sophisticated study. Normally, researchers use an atlas to create a region for ROI analysis, or a map that

partitions the brain into anatomically distinct regions. Here, we utilize the Harvard-Oxford Cortical atlas shown in Figure 3.11, which divides the whole structure into 48 ROIs. Table 3.7 provides the names of 48 ROIs and the corresponding region ID used here for simpler analysis.

Figure 3.12: Average CEV values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.

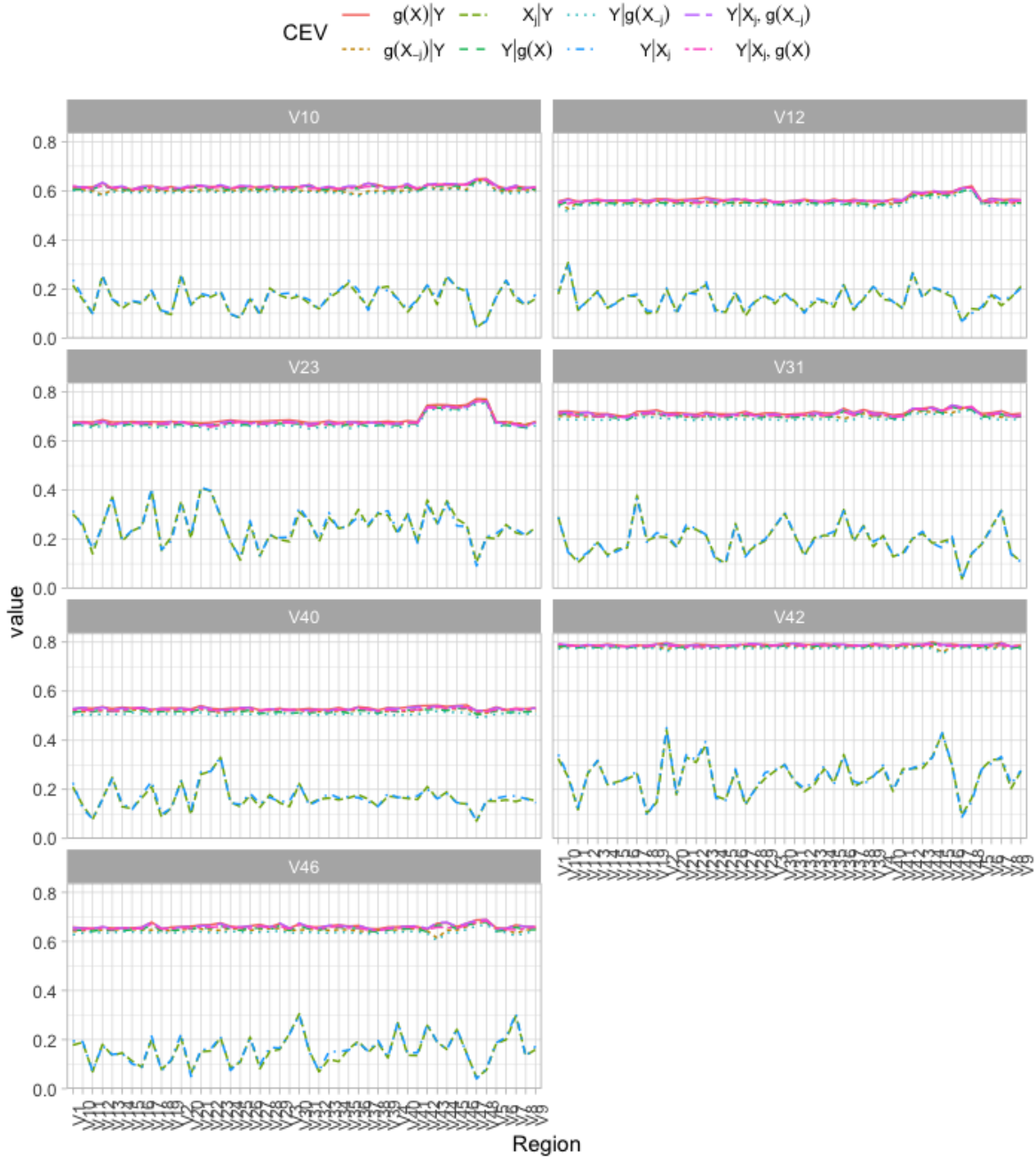
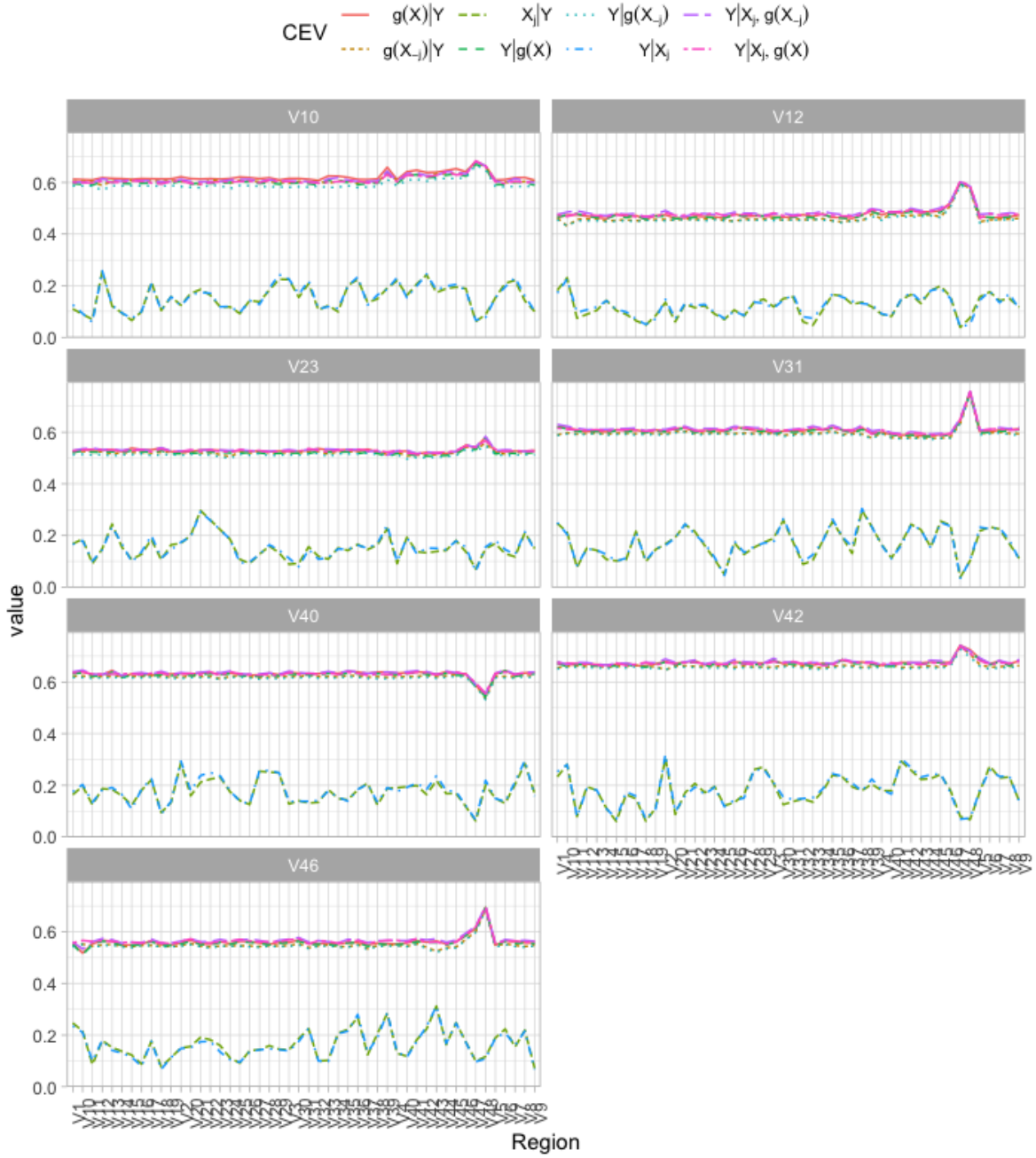


Figure 3.13: Average CEV values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.



In order to study the performance of the proposed CEV-PWA algorithm and the causal relationship among these regions for major depressive disorder, we summarized the results of 7 regions in the following figures. These 7 regions are characterized by different trends between the major depressive disorder group and the healthy control group. The results for all 48 regions are summarized in the Appendix. Using the same statistics provided in Table 3.1, we applied different colors and line types to denote these CEVs. Note that the Y in the legend represents each of the 7 regions shown in the title of the subplot, and X includes the rest 47 regions except Y .

Figure 3.13 and Figure 3.14 provide the average CEV values for each of the 7 ROIs in both major depressive disorder (MDD) and healthy control (HC) groups. Notice that for all of the 7 ROIs, the lines of $Y|X_j$ and $X_j|Y$ are significantly much lower than the others, which suggests that our proposed CEV-PWA algorithm works well in explaining more variability of variable Y using the PWA $g(\cdot)$ function. Also, we notice that the pairwise relationship between region Y and X_j doesn't have much difference for both MDD and HC group, as the line of $Y|X$ and $X|Y$ are very close to each other.

When we take a closer look at the results without pairwise CEV values, Figure 3.14 and Figure 3.15 show the average CEV values for each of the 7 regions in both MDD and HC groups. We observe that middle temporal gyrus (V12), precuneous cortex (V31), occipital fusiform gyrus (V40), and planum temporale (V46) achieve the highest value in $g(X)|Y$ in MDD, which means these four regions can actively influence the network of other regions in the MDD group. Region superior temporal gyrus (V10) performs in the opposite way, and it achieves the highest score in $g(X)|Y$ for the HC group, while this is not the case in the MDD group. Also, we notice that occipital fusiform gyrus (V40) achieves significantly higher values for all CEV estimators in the CT than the MDD group. On the contrary, lateral occipital cortex (V23) and central opercular cortex (V42) has remarkably lower values for all CEV estimators in the HC than the MDD group.

Figure 3.14: Average CEV values for each region except $Y|X$ and $X|Y$ in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.

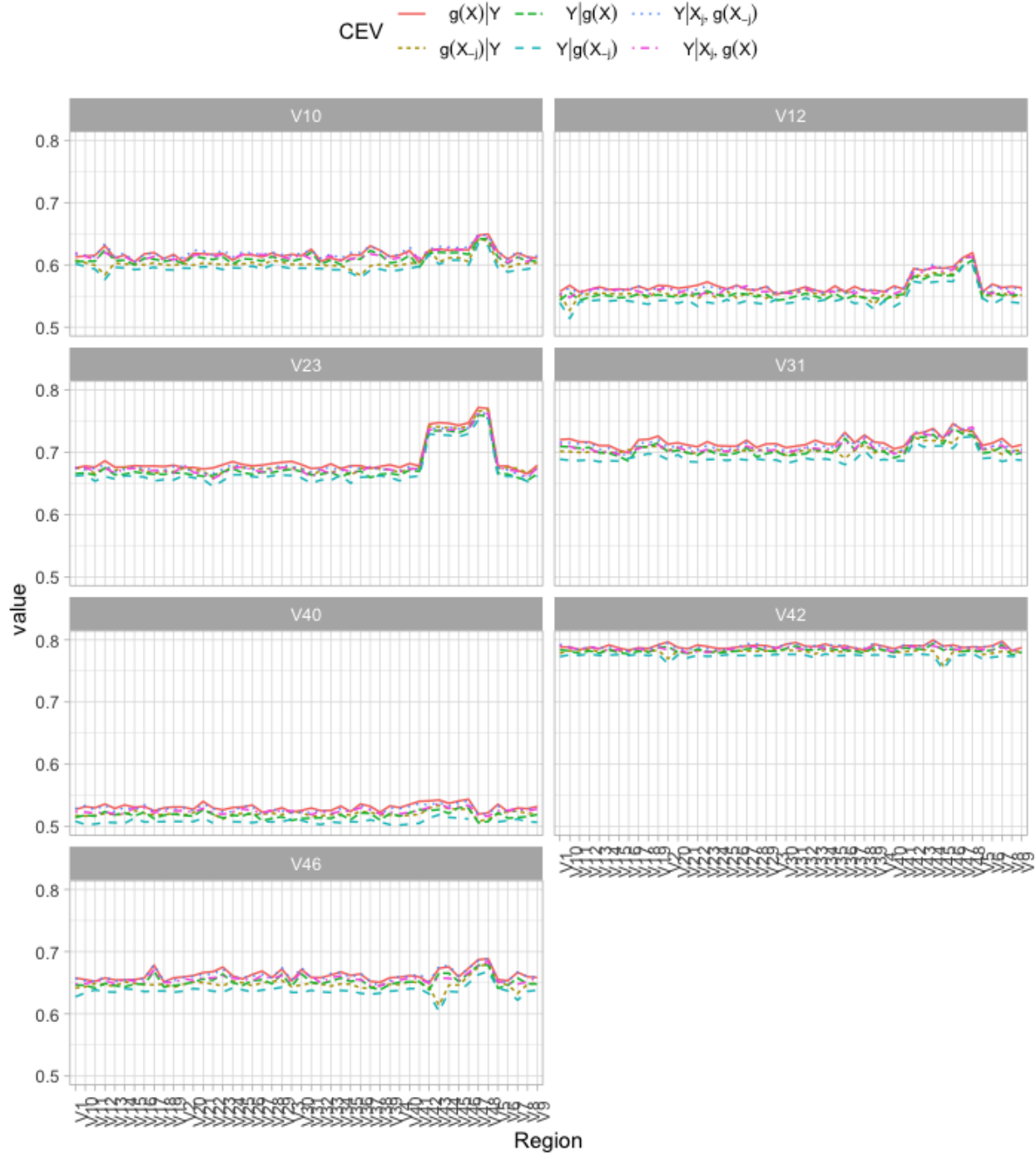
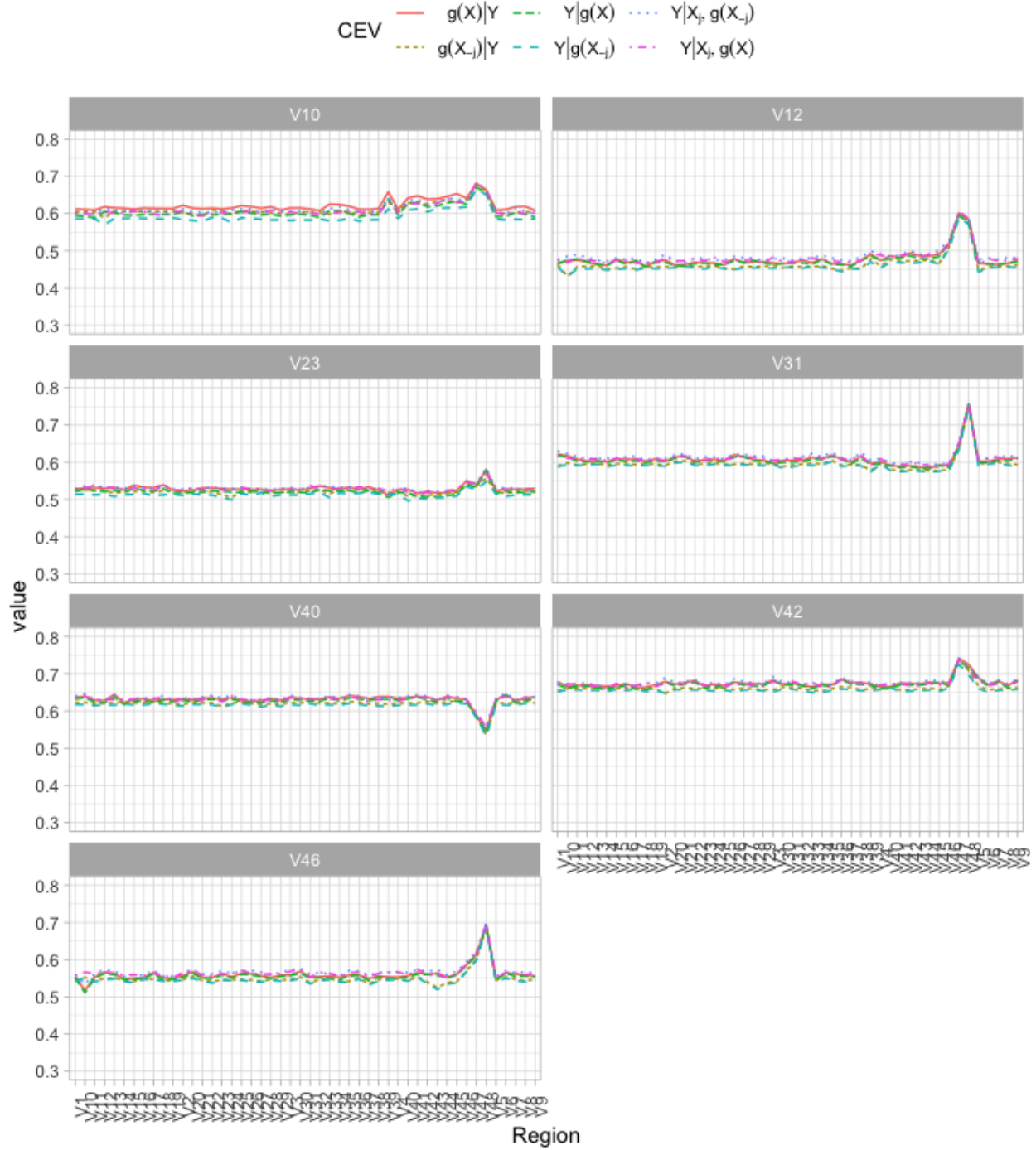
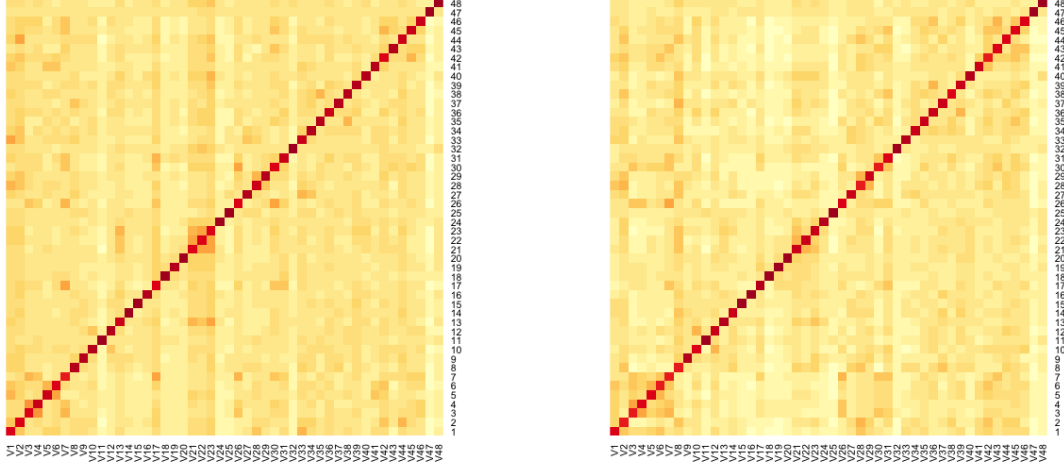


Figure 3.15: Average CEV values for each region except $Y|X$ and $X|Y$ in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.



The heatmaps in Figure 3.16 constructed by pairwise CEV values containing all 48 regions shows the overall interaction for both MDD and HC groups. The heatmap for the MDD group is darker than the one for the HC group, which suggests the pairwise connectivity in the MDD is stronger than the HC overall.

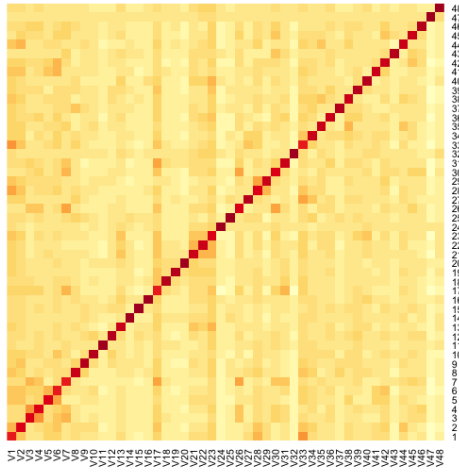


(a) Major depressive disorder group (MDD)

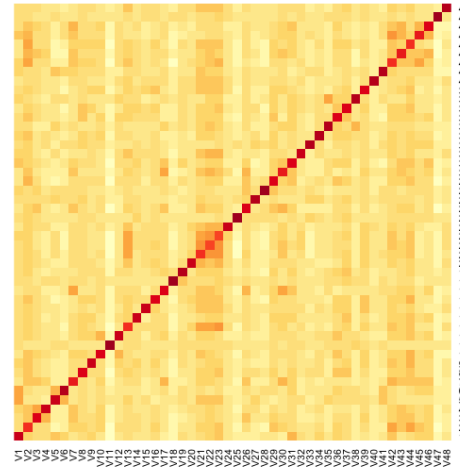
(b) Healthy control group (HC)

Figure 3.16: (a) shows the heatmap of 48 regions in the major depressive disorder group (MDD); (b) shows the heatmap of 48 regions in the healthy control group (HC).

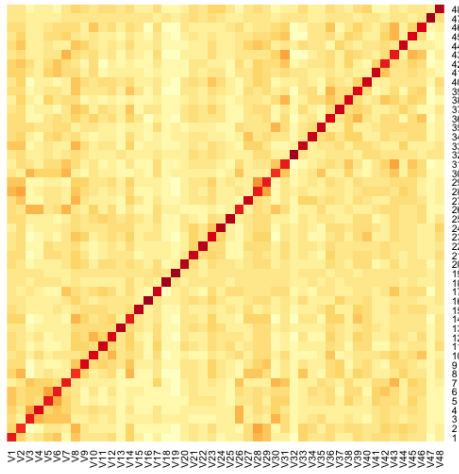
Then, we constructed the heatmap for male and female in both MDD and HC groups. Figure 3.17 provides the corresponding figures with the sub-figure representing each subgroup. It's easy to be noticed that all four heatmaps are asymmetric. When comparing the performance of Female and Male, we observe that Male subgroups for both MDD and HC groups contain stronger connectivity in region angular gyrus (V21), lateral occipital cortex (V22), lateral occipital cortex (V23), and intracalcarine cortex (V24) than Female subgroups, as there exist obvious darker rectangle areas containing the four regions.



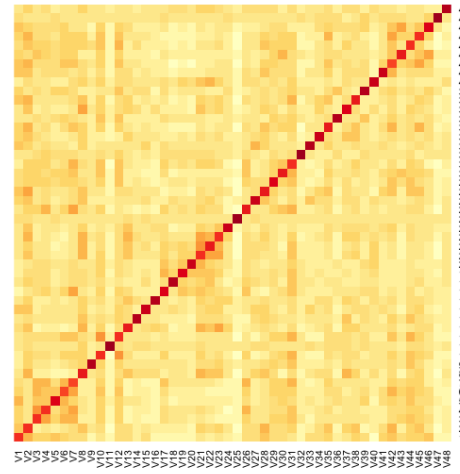
(a) MDD - Female group



(b) MDD - Male group



(c) HC - Female group



(d) HC - Male group

Figure 3.17: The heatmap of 48 regions in the major depressive disorder group (MDD) and the healthy control group (HC) for Male and Female subgroups.

In order to analyze the pairwise causal relationship among these channels, we have Figure 3.18 as the directed network graph with different colors representing different ROIs in human brain and the width of lines between channels indicating the level of connectivity. Central opercular cortex (V42) is very likely to be passively influenced by other regions in the MDD group, while it's less active in the HC. Then, lateral occipital cortex (V23) is more actively affecting other regions in the MDD, comparing with its performance in the HC. But it's noteworthy that V23 always WeCon causes occipital fusiform gyrus (V40) for both MDD and HC groups.

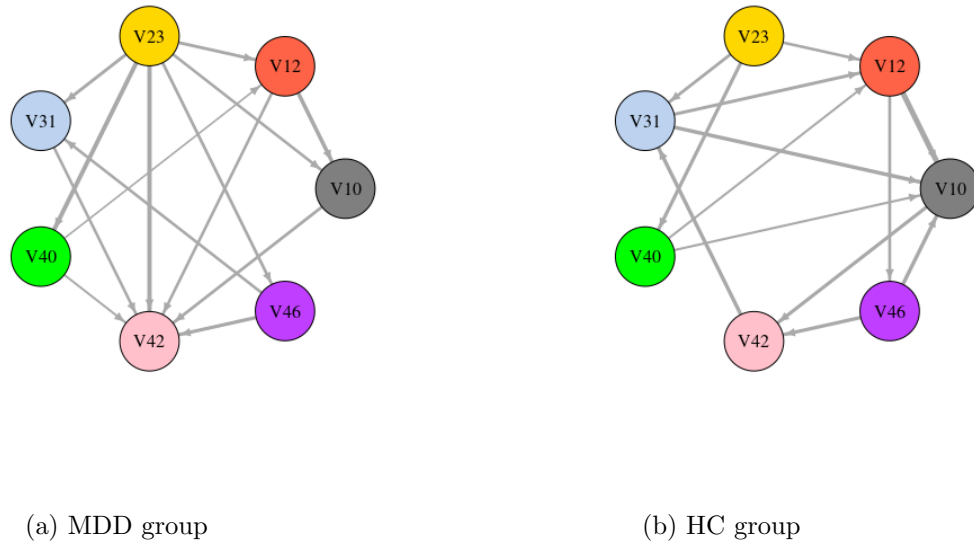
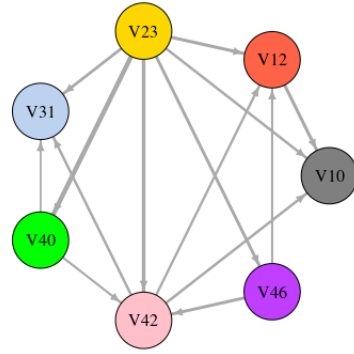
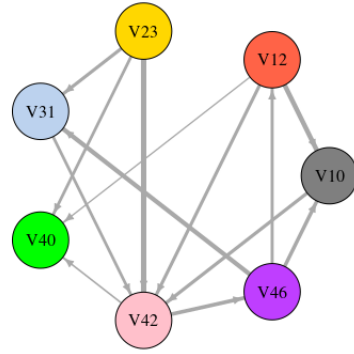


Figure 3.18: (a) shows the directed connectivity network of regions in the major depressive disorder group (MDD); (b) shows the directed connectivity network of regions in the healthy control group (HC).

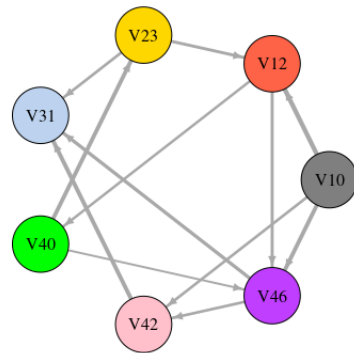
Then, we constructed the directed network graph for male and female in both MDD and HC groups. Figure 3.19 provides the corresponding figures with each sub-figure representing each sub-group. When we compare Female and Male subgroups, we find that middle temporal gyrus (V12) and superior temporal gyrus (V10) show stronger connectivity in Male than Female for both MDD and HC groups. Also, precuneous cortex (V31) seems to be more active in Male than Female for both MDD and HC groups, since the connection lines



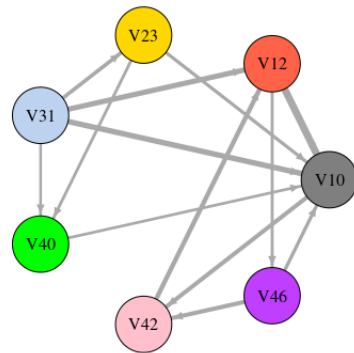
(a) MDD - Female group



(b) MDD - Male group



(c) HC - Female group



(d) HC - Male group

Figure 3.19: The directed connectivity network of regions in the major depressive disorder group (MDD) and the healthy control group (HC) for Female and Male sub-groups.

are thicker. In addition, precuneous cortex (V31) is more likely to be passively affected by other regions in Female than Male as the connectivity arrows are always pointing to this region. When analyzing the MDD and HC groups, we notice that lateral occipital cortex (V23) always WeCon causes region central opercular cortex (V42) for both Female and Male in MDD, which is not obvious in the HC group.

Chapter 4

Alternative Approach

In Chapter 3, we applied cumulative explained variability and piece-wise affine functions to implement dimension reduction and variable selections. In the literature, principal component analysis (PCA) is a powerful approach to realize dimension reduction to a few dimensions. This Chapter 4 is designed to implement PCA in WeCon causality inference. Section 4.1 introduces the integration of PCA and cumulative explained variability, with the combination of leave-one-out strategy and PCA for dimension reduction and parameter optimization. To better explore the effect of different components, several types of estimators are proposed in Section 4.1, and then the corresponding performances are assessed in simulation study from Section 4.3. Section 4.4 provides the real data analysis for resting-state fMRI data in major depressive disorder.

In addition to PCA, partial least squares (PLS) methods have been widely used in regression analysis with high dimensional data. We are going to also implement PLS method in WeCon causality inference. Section 4.2 describes the approach of cumulative explained variability with PLS, where the combination of leave-one-out strategy and PLS is utilized for dimension reduction and parameter optimization. Furthermore, estimators using different PLS components are introduced in Section 4.2, and their performances are measured in simulation study from Section 4.3. Also, Section 4.4 includes the corresponding causal analysis for major depressive disorder based on resting-state fMRI data.

4.1 Principal component approximation

Given that the main goal of this project is to construct networks by $g(\cdot)$ function, where the network is considered as the causal relationship among different features, principal component analysis (PCA) is an alternative choice of $g(\cdot)$ for continuous piece-wise affine function (PWA). principal component analysis was developed and named by Hotelling (1933), and it aims to reduce the dimension of large datasets, increase the interpretability but at the mean time minimize information loss. The structure of continuous PWA function will lead to the grouping of observations, while PCA projects data to a set of orthogonal feature vec-

tors in such a way that the data spread is maximum in the direction of the feature vector. The feature vectors will be evaluated in the decreasing order of explained variances, which are called as first principal component (PC), second principal component, etc. Naturally, the first PC is formed as the linear combination of original features that explains the most variance.

Similar to what was stated in Chapter 3.2, two types of CEV estimator are considered here:

$$\begin{aligned}\widehat{\text{CEV}}_1 &= \max_{\boldsymbol{\beta}} \widehat{\text{CEV}}(X_I | g(PCs; \boldsymbol{\beta})), \\ \widehat{\text{CEV}}_3 &= \max_{\boldsymbol{\beta}} \widehat{\text{CEV}}(X_I | X_J, g(PCs; \boldsymbol{\beta}))\end{aligned}$$

where $0 < I, J < D$, $I \neq J$, and $\boldsymbol{\beta} = \{\beta^{(j)}, j = 1, 2, \dots, D\}$ denotes the model parameters. As for the parameter estimation, we first find the lower and upper bound of parameters of each principal component used in the $g(\cdot)$ function, and then we search in this range to achieve optimal CEV estimators.

Leave-one-out strategy allows the model to implement PCA with a subset of data each time, which will provide us M sets of estimated parameters including $\widehat{\boldsymbol{\beta}}_1, \dots, \widehat{\boldsymbol{\beta}}_M$. Then the lower and upper bound of parameters could be represented as

$$\boldsymbol{\beta}_{L,PC} = \min\{\widehat{\boldsymbol{\beta}}_1, \dots, \widehat{\boldsymbol{\beta}}_M\}, \quad (4.1)$$

$$\boldsymbol{\beta}_{U,PC} = \max\{\widehat{\boldsymbol{\beta}}_1, \dots, \widehat{\boldsymbol{\beta}}_M\} \quad (4.2)$$

Within this specified range, we search for the best parameter set to achieve the optimal CEV estimators.

The algorithm of cumulatively explained variability with PCA could be summarized as

follows

Algorithm 2 CEV-PCA Algorithm

1. Implement CEV algorithm for screening, so as to obtain an active variable set $\widehat{\mathcal{M}}_d^*$ for variable X_I ;
 2. Divide the data into M groups;
 3. Conduct PCA with $M-1$ groups of data each time and store the estimated parameters $\widehat{\beta}_i$.
 7. Run Step 3 iteratively and store the determined lower and upper bound of parameters as $\beta_{L,PC}$ and $\beta_{U,PC}$ using formula (4.1) and (4.2).
 8. Utilize $\beta_{L,PC}$ and $\beta_{U,PC}$ as the range, search for the best parameter set β_{optimal} to achieve optimal CEV estimators including $\widehat{\text{CEV}}_1$ and $\widehat{\text{CEV}}_3$.
-

Principal components (PCs) are used to construct networks that combine the information from various variables. Specifically, when we construct the PCs, we have 2 scenarios listed below.

- **All Predictors:** the first scenario is to consider all predictors in PCA, including the individual variable X_j . Therefore, PCs contain the information from X_j .
- **Partial Predictors:** the second scenario is to exclude the variable X_j and utilize the rest of predictors for PCA. Hence, PCs don't contain the information from X_j .

Comparing these two scenarios, the estimator with all predictors is expected to contain more information than the one with partial predictors, and thus leads to higher CEV values. By evaluating the performance of CEV estimators under different settings, we would be able to better understand the contribution of PCs and the individual variable X_j to explain the variability of Y .

Under high-dimensional condition, sparsity structure is introduced to the proposed model to reduce the dimension. Therefore, we first conduct the CEV screening for dimension reduction, then we implement the traditional principal component analysis and extract the information about PCs. Table 4.1 provides the type of estimators we have interests in, and it suits for both all predictors and partial predictors scenarios.

Table 4.1: CEV estimators by principal component analysis for both all predictors and partial predictors scenarios.

Cases	Estimators
Individual PCs	$CEV(Y PCs)$
Network constructed by multiple PCs	$CEV(Y g(PCs))$
Network constructed by individual variable X_j and PCs	$CEV(Y g(X_j, PCs))$

Note that the $g(\cdot)$ function in Table 4.1 represents the network. For the first case, we evaluate the cumulative explained variability of variable Y given individual PCs. We also have interests in exploring the CEV estimators containing the combination of multiple PCs as the second case stated. Then we consider to combine the individual variable X_j and PCs together, and compute the estimated cumulative explained variability of variable Y . Therefore, not only the performance of individual PCs can be measured, but effect of the mixture of individual variable and various PCs can also be observed.

4.2 Partial least squares approximation

Principal component analysis is an unsupervised learning method, which aims to find hyper-planes between the response and predictors to achieve maximum variance. This means that it doesn't depend on the response variable, and indicating that it loses the corresponding information. In this case, we consider to use partial least squares regression (PLS) as an alternative of continuous piece-wise affine function to construct the network and analyze the casual relationships. Partial least squares regression was introduced by Gerlach et al. (1979) and it generalizes and combines variables from principal component analysis and multiple regression. PLS regression looks for the linear regression model by projecting the predictors and the response variable to a new space, and it tries to find the multidimensional direction in the feature space which explains the maximum multidimensional variance direction in the response space. Similar to principal component, we would be able to obtain partial

least squares component (PL) from this method.

As we stated in Chapter 4.1, leave-one-out strategy allows the model to implement PLS with a subset of data each time, which will provide us M sets of estimated parameters including $\widehat{\beta}_1, \dots, \widehat{\beta}_M$. Then the lower and upper bound of parameters could be represented as

$$\beta_{L,PL} = \min\{\widehat{\beta}_1, \dots, \widehat{\beta}_M\} \quad (4.3)$$

$$\beta_{U,PL} = \max\{\widehat{\beta}_1, \dots, \widehat{\beta}_M\} \quad (4.4)$$

Within this specified range, we search for the best parameter set to achieve the optimal CEV estimators.

The algorithm of cumulatively explained variability with PLS could be summarized as follows

Algorithm 3 CEV-PLS Algorithm

1. Implement CEV algorithm for screening, so as to obtain an active variable set $\widehat{\mathcal{M}}_d^*$ for variable X_I ;
 2. Divide the data into M groups;
 3. Conduct PLS with $M - 1$ groups of data each time and store the estimated parameters $\widehat{\beta}_i$.
 7. Run Step 3 iteratively and store the determined lower and upper bound of parameters as $\beta_{L,PLS}$ and $\beta_{U,PLS}$ using formula (4.3) and (4.4).
 8. Utilize $\beta_{L,PL}$ and $\beta_{U,PL}$ as the range, search for the best parameter set β_{optimal} to achieve optimal CEV estimators including $\widehat{\text{CEV}}_1$ and $\widehat{\text{CEV}}_3$.
-

Partial least squares components (PLs) would be used to construct networks that combine the information from various variables. Also, when we construct the PLs, we have 2 scenarios listed below

- **All Predictors:** the first scenario is to consider all predictors in PLS, including the individual variable X_j . Therefore, PLs contain the information from X_j .

- **Partial Predictors:** the second scenario is to exclude the variable X_j and utilize the rest of predictors for PLS. Hence, PLs don't contain the information from X_j .

When we compare both scenarios, the CEV estimator with all predictors is expected to contain more information than the one with partial predictors, and explain more variability of variable Y , and thus result in higher CEV values. By evaluating the performance of CEV estimators under different conditions, we would be able to better understand the contribution of PLs and the individual variable X_j to explain the variability of Y .

For high dimension problems, we would first conduct CEV screening to reduce the dimension and then implement the partial least squares regression and extract the information about PLs. Table 4.2 provides the type of estimators we have interests in, and it suits for both all predictors and partial predictors scenarios.

Table 4.2: CEV estimators by partial least squares regression for both all predictors and partial predictors scenarios.

Cases	Estimators
Individual PLs	$CEV(Y PLs)$
Network constructed by multiple PLs	$CEV(Y g(PLs))$
Network constructed by individual variable X_j and PLs	$CEV(Y g(X_j, PLs))$

In this table, $g(\cdot)$ represents the network. The first estimator provides the explained variability of Y given individual PLs, while the second estimator aims to evaluate the case when we consider various PLs in the meantime. The third estimator is designed for measuring the effect of the combination of both individual variable X_j and the PLs on explaining the variability of Y . In this setting, we are able to study the causal relationship between variable Y and the network constructed by various combinations of PLs and individual variables.

4.3 Simulation

In this section, we study the finite sample performance of CEV-PCA and CEV-PLS models through several Monte Carlo simulations. We assess the performance of our proposed method by comparing the CEV results for several different settings. Through out the simulation, the results contain top 10 variables with the highest $CEV(Y|X_j)$ values for each model. Table 4.3 provides the types of CEV estimators that are evaluated in our study. In addition, we calculated two types of CEV estimators for both algorithms. For the first type, we constructed the principal components (PCs) and partial least components (PLs) using partial predictors, meaning that we exclude the variable X_j . For the second type, we include the information of X_j to construct PCs and PLs, which means we use all predictors. The results for both partial and all predictors are summarized in this section.

Table 4.3: Simulation statistics for alternative approaches.

Cases	Statistics	Explanation
Individual variable X_j	$Y X_j$	The CEV in Y is explained by X_j .
	$X_j Y$	The CEV in X_j is explained by Y
Individual component C	$Y Ci$	The CEV in Y is explained by Ci , where $i = 1, 2, 3$.
Multiple components Cs	$Y C1, C2$	The CEV in Y is explained by first two components $C1$ and $C2$.
	$Y C1, C2, C3$	The CEV in Y is explained by first three components $C1$, $C2$, and $C3$.
Mixture of variable X_j and components Cs	$Y X_j, C1$	The CEV in Y is explained by X_j and $C1$.
	$Y X_j, C1, C2$	The CEV in Y is explained by X_j and first two components ($C1$ and $C2$).
	$Y X_j, C1, C2, C3$	The CEV in Y is explained by X_j and first three components ($C1$, $C2$, and $C3$).

4.3.1 Simulation Example 1

Example 1. In this example, we investigate the performances of CEV-PCA and CEV-PLS through several linear models. The ordinary linear model (1.a) is similar to Example 1 in Fan and Lv (2008). The linear model with indicator function (1.b), the linear model with interaction (1.c), and the heteroscedastic linear model (1.d) are similar to the examples in Li et al. (2012).

$$(1.a) \ Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{20} + \beta_4 X_{30} + \varepsilon,$$

$$(1.b) \ Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 I(X_{20} < 0) + \beta_4 X_{30} + \varepsilon,$$

$$(1.c) \ Y = \beta_1 X_1 X_2 + \beta_3 I(X_{20} < 0) + \beta_4 X_{30} + \varepsilon,$$

$$(1.d) \ Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 I(X_{20} < 0) + \beta_5 \exp(|X_{30}|) + \varepsilon.$$

The variables $X = (X_1, \dots, X_p)$ are generated from a normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$, where $\sigma_{ij} = \sigma^{|i-j|}$. The random error ε follows a standard normal distribution. We set the vector $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (5, 2, 7, 5, 2)$. The sample size $n = 500$, the dimension $p = 1000$, $\sigma = 0.5$.

Principal component approximation result

For all four models in Example 1, we use X_1 , X_2 , X_{20} , and X_{30} to construct the variable Y , so they can be taken as **truly active variables**. In model (1.a), one can notice from Table 4.4 and Table 4.5 that X_4 , X_{19} , X_{21} , X_{29} , and X_{31} all have very small $\text{CEV}(Y|X_j)$ values (lower than 0.1), while the values of $\text{CEV}(Y|C1)$ and $\text{CEV}(Y|X_j, C1)$ are greater than 0.6. The joint information from other variables combined by PC1 dominates the explained variability of Y , which indicates these 5 individual variables are not active from the simulation result. After checking the rest of the variables X_1 , X_2 , X_3 , X_{20} , and X_{30} , they can be determined as active variables due to relatively high $\text{CEV}(Y|X_j)$ values and reasonable differences comparing $\text{CEV}(Y|X_j)$ with $\text{CEV}(Y|C1)$ and $\text{CEV}(Y|X_j, C1)$. Using

the same strategy, for model (1.b) and (1.c), X_1 , X_2 , X_{20} , X_{29} , and X_{30} can be selected as active variables. For model (1.d), X_1 , X_2 , X_{20} , and X_{30} are taken as active. Overall, the truly active variables X_1 , X_2 , X_{20} , and X_{30} are selected by all models. By comparing the performances of different CEV estimators, one can find the dominating components for explaining the variability of Y and determine active variables as well.

For both partial predictors and all predictors, the estimator $Y|X_j, C1, C2, C3$ achieves the highest score almost all the time or at least very close to the highest value. This fits our expectation to explain the most variability of Y , since $Y|X_j, C1, C2, C3$ contains the most information including both individual variable X_j and first 3 PCs. By comparing the values of CEV estimators for these two types (Table 4.4 and Table 4.5), we find that the values of CEV estimators using all predictors are slightly greater than those using partial predictors.

Table 4.4: Simulation results for CEV-PCA using partial predictors for Example 1.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
1.a	1	0.43615	0.43419	0.5992	0.55226	0.58717	0.70453	0.74111	0.8206	0.86936	0.921
	2	0.45886	0.46234	0.39657	0.59323	0.12045	0.6076	0.6331	0.74881	0.79188	0.79719
	3	0.14234	0.14221	0.69849	0.70188	0.77907	0.89699	0.9513	0.7688	0.90019	0.95154
	4	0.05661	0.05978	0.74523	0.7438	0.69176	0.81896	0.82557	0.7466	0.82497	0.83024
	19	0.07325	0.07555	0.62055	0.2305	0.01411	0.62025	0.77284	0.66445	0.6651	0.77546
	20	0.21389	0.21411	0.6782	0.46371	0.04062	0.68288	0.72306	0.85727	0.87467	0.87933
	21	0.06634	0.06893	0.72176	0.02366	0.69701	0.79149	0.89969	0.76376	0.81423	0.91529
	29	0.04999	0.05633	0.63616	0.80509	0.48674	0.90005	0.94124	0.65652	0.90766	0.94155
	30	0.22561	0.2271	0.71512	0.70529	0.28461	0.78627	0.80027	0.86659	0.88837	0.88893
	31	0.03851	0.03891	0.64509	0.77911	0.10674	0.79584	0.82503	0.65807	0.80272	0.82525
1.b	1	0.45113	0.44993	0.47752	0.16237	0.11385	0.53635	0.55653	0.81228	0.81494	0.81592
	2	0.22396	0.22505	0.53122	0.67269	0.65847	0.74743	0.87115	0.68664	0.79467	0.88503
	3	0.07503	0.07496	0.62778	0.68967	0.77528	0.81202	0.87716	0.64852	0.81543	0.87987
	20	0.13132	0.1009	0.69475	0.58727	0.33711	0.76542	0.77429	0.78659	0.8453	0.84184
	21	0.0333	0.03447	0.67514	0.73949	0.24775	0.7643	0.8655	0.70421	0.79085	0.86445
	29	0.11237	0.10668	0.62898	0.64428	0.07713	0.74084	0.74248	0.6567	0.75169	0.75011
	30	0.38562	0.38624	0.60207	0.37295	0.34074	0.66415	0.68667	0.8707	0.88716	0.89467
	31	0.09967	0.10031	0.62064	0.19361	0.06729	0.69606	0.78391	0.67279	0.72728	0.78882
	441	0.03353	0.01268	0.7943	0.03746	0.64345	0.81565	0.83182	0.79606	0.81729	0.8333
	678	0.03279	0.03279	0.71776	0.05457	0.03671	0.72405	0.81759	0.72252	0.72884	0.82049
1.c	1	0.14645	0.01362	0.36123	0.37742	0.07873	0.39095	0.45286	0.49556	0.5194	0.57347
	2	0.12996	0.00858	0.3951	0.35708	0.27614	0.43565	0.44493	0.52802	0.5662	0.57271
	19	0.04465	0.05427	0.35849	0.32343	0.30355	0.61075	0.69806	0.37281	0.57418	0.62046
	20	0.16555	0.15464	0.3617	0.3235	0.27615	0.65877	0.6935	0.43136	0.69465	0.72531
	21	0.04283	0.05402	0.35071	0.28989	0.30794	0.59991	0.6383	0.37444	0.59088	0.61482
	29	0.09692	0.11867	0.33957	0.3885	0.34392	0.4115	0.7043	0.36542	0.4239	0.68148
	30	0.37985	0.4195	0.20805	0.21842	0.34427	0.35831	0.54481	0.47773	0.62834	0.77771
	31	0.10208	0.10257	0.22128	0.39367	0.30682	0.44635	0.5964	0.27407	0.44848	0.51503
	32	0.03964	0.03682	0.39578	0.40859	0.33604	0.44326	0.73518	0.40432	0.45878	0.70751
	40	0.03797	0.00389	0.41897	0.38125	0.33974	0.4345	0.68062	0.42712	0.44347	0.68313
1.d	1	0.22419	0.33315	0.44435	0.16481	0.1762	0.58371	0.58118	0.55672	0.6309	0.65629
	2	0.12034	0.16186	0.5136	0.21702	0.2651	0.69785	0.66256	0.51706	0.62524	0.64277
	3	0.03619	0.04955	0.49794	0.23905	0.26988	0.68336	0.57174	0.413	0.62005	0.54822
	19	0.0246	0.04348	0.3877	0.3868	0.1724	0.55385	0.67921	0.26187	0.35278	0.51534
	20	0.05434	0.10171	0.40485	0.40222	0.07824	0.48569	0.50735	0.3458	0.46327	0.49815
	29	0.02737	0.00739	0.24997	0.41718	0.46053	0.56661	0.68318	0.28079	0.51586	0.63242
	30	0.57175	0.00387	0.28807	0.12651	0.10383	0.29125	0.36955	0.81571	0.82065	0.79656
	31	0.04372	0.00297	0.23958	0.42944	0.41599	0.62341	0.67362	0.27894	0.62585	0.58443
	479	0.02586	0.01204	0.41923	0.19769	0.2159	0.50089	0.58418	0.30952	0.19804	0.31139
	886	0.02578	0.01772	0.28162	0.2072	0.35212	0.39558	0.42204	0.23724	0.20569	0.32088

Table 4.5: Simulation results for CEV-PCA using all predictors for Example 1.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
1.a	1	0.43615	0.43419	0.7525	0.61903	0.59579	0.85335	0.8718	0.78496	0.88939	0.9016
	2	0.45886	0.46234	0.66806	0.70078	0.65581	0.74411	0.797	0.82035	0.86479	0.87601
	3	0.14234	0.14221	0.66886	0.65943	0.59213	0.79119	0.79225	0.80788	0.86881	0.87414
	4	0.05661	0.05978	0.83517	0.747	0.58172	0.84555	0.848	0.84438	0.85184	0.8557
	19	0.07325	0.07555	0.81159	0.51928	0.59037	0.83295	0.84566	0.81276	0.84785	0.87547
	20	0.21389	0.21411	0.81008	0.77253	0.51543	0.88261	0.89257	0.81568	0.88872	0.8948
	21	0.06634	0.06893	0.81779	0.66673	0.50523	0.8596	0.87662	0.81945	0.86548	0.90835
	29	0.04999	0.05633	0.8965	0.75972	0.56556	0.92057	0.92373	0.91659	0.92338	0.92625
	30	0.22561	0.2271	0.76647	0.58527	0.54042	0.84789	0.85396	0.77442	0.85154	0.85883
	31	0.03851	0.03891	0.74247	0.70047	0.49385	0.85985	0.86407	0.74363	0.86504	0.86803
1.b	1	0.45113	0.44993	0.86492	0.03145	0.03776	0.86844	0.88435	0.86579	0.87587	0.88756
	2	0.22396	0.22505	0.61966	0.03471	0.04083	0.72463	0.73026	0.62004	0.74222	0.74898
	3	0.07503	0.07496	0.75448	0.03048	0.04184	0.7619	0.76641	0.75524	0.77006	0.77402
	20	0.13132	0.1009	0.70805	0.02756	0.03842	0.80212	0.81743	0.74058	0.81126	0.83212
	21	0.0333	0.03447	0.7109	0.03057	0.03358	0.71377	0.77148	0.73328	0.74715	0.7719
	29	0.11237	0.10668	0.80316	0.02565	0.04439	0.82068	0.82513	0.807	0.82093	0.82422
	30	0.38562	0.38624	0.70809	0.02666	0.04011	0.82292	0.90495	0.74564	0.85209	0.90527
	31	0.09967	0.10031	0.71233	0.02135	0.03742	0.75475	0.76054	0.71788	0.82919	0.85897
	441	0.03353	0.01268	0.71594	0.02651	0.04856	0.71929	0.8267	0.74216	0.74337	0.8357
	678	0.03279	0.03279	0.76991	0.03295	0.03831	0.80878	0.81309	0.77087	0.81142	0.81503
1.c	1	0.14645	0.01362	0.40281	0.36392	0.35116	0.4075	0.69858	0.55231	0.56096	0.69632
	2	0.12996	0.00858	0.43667	0.41854	0.33781	0.47437	0.74929	0.57687	0.59881	0.73696
	19	0.04465	0.05427	0.39638	0.36909	0.33099	0.45839	0.7385	0.40157	0.47325	0.71883
	20	0.16555	0.15464	0.43329	0.38588	0.33611	0.46074	0.74433	0.4596	0.4937	0.74883
	21	0.04283	0.05402	0.40087	0.40306	0.32864	0.42775	0.69692	0.41923	0.4406	0.70117
	29	0.09692	0.11867	0.35843	0.37009	0.337	0.39827	0.65315	0.36668	0.39594	0.65655
	30	0.37985	0.4195	0.38498	0.33984	0.3441	0.41157	0.68165	0.45389	0.52126	0.71625
	31	0.10208	0.10257	0.37574	0.34771	0.34487	0.41275	0.71009	0.37909	0.4303	0.70386
	32	0.03964	0.03682	0.34602	0.39082	0.33251	0.3984	0.60428	0.35016	0.40387	0.58214
	40	0.03797	0.00389	0.42972	0.33232	0.33298	0.47757	0.70577	0.43627	0.52979	0.7128
1.d	1	0.22419	0.33315	0.31503	0.22426	0.28181	0.37785	0.4232	0.49318	0.40309	0.4651
	2	0.12034	0.16186	0.3579	0.21533	0.18484	0.45082	0.59797	0.35116	0.28019	0.34524
	3	0.03619	0.04955	0.27156	0.20282	0.20202	0.47074	0.4931	0.27005	0.46355	0.46615
	19	0.0246	0.04348	0.29543	0.21563	0.18995	0.45171	0.46023	0.24216	0.40299	0.43593
	20	0.05434	0.10171	0.31528	0.2172	0.19453	0.36928	0.31745	0.23137	0.35324	0.33542
	29	0.02737	0.00739	0.30514	0.21301	0.18347	0.39016	0.42414	0.33551	0.23591	0.36394
	30	0.57175	0.00387	0.26526	0.20507	0.23409	0.4067	0.57789	0.63617	0.63391	0.83329
	31	0.04372	0.00297	0.2758	0.23311	0.19388	0.48967	0.50373	0.30656	0.50613	0.50729
	479	0.02586	0.01204	0.27791	0.21035	0.23475	0.51035	0.56521	0.2696	0.38778	0.42872
	886	0.02578	0.01772	0.2407	0.22048	0.23282	0.29532	0.26333	0.20639	0.27644	0.27239

Partial least squares approximation result

The performance of CEV-PLS algorithm is similar to CEV-PCA algorithm. Still, in model (1.a), we can notice from Table 4.6 and Table 4.7 that X_4 , X_{19} , X_{21} , X_{29} , and X_{31} all have very small $\text{CEV}(Y|X_j)$ values (lower than 0.1), while the values of $\text{CEV}(Y|C1)$ and $\text{CEV}(Y|X_j, C1)$ are around 0.7 and 0.8. The joint information from other variables combined by PL1 dominates the explained variability of Y , which indicates these 5 individual variables are not active from the simulation result. After checking the rest of the variables X_1 , X_2 , X_3 , X_{20} , and X_{30} , they can be determined as active variables due to relatively high $\text{CEV}(Y|X_j)$ values and reasonable differences comparing $\text{CEV}(Y|X_j)$ with $\text{CEV}(Y|C1)$ and $\text{CEV}(Y|X_j, C1)$. Using the same strategy, for model (1.b) and (1.c), X_1 , X_2 , X_{20} , X_{29} , and X_{30} can be selected as active variables. For model (1.d), X_1 , X_2 , X_{20} , and X_{30} are taken as active. Overall, the truly active variables X_1 , X_2 , X_{20} , and X_{30} are selected by all models. By comparing the performances of different CEV estimators, one can find the dominating components for explaining the variability of Y and determine active variables as well.

For both partial predictors and all predictors, the estimator $Y|X_j, C1, C2, C3$ achieves the highest score almost all the time or at least very close to the highest value. This fits our expectation to explain the most variability of Y , since $Y|X_j, C1, C2, C3$ contains the most information including both individual variable X_j and first 3 PLs. By comparing the values of CEV estimators for these two types (Table 4.6 and Table 4.7), we find that the values of CEV estimators using all predictors are slightly greater than those using partial predictors.

Table 4.6: Simulation results for CEV-PLS using partial predictors for Example 1.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
1.a	1	0.43615	0.43419	0.62849	0.01333	0.01115	0.73737	0.79623	0.83329	0.92571	0.93485
	2	0.45886	0.46234	0.68652	0.01256	0.03918	0.81635	0.86671	0.81514	0.87652	0.97248
	3	0.14238	0.14221	0.84218	0.00608	0.05462	0.92541	0.96514	0.84340	0.92732	0.96683
	4	0.05661	0.05978	0.85392	0.05812	0.07260	0.91530	0.94656	0.85811	0.91609	0.94741
	19	0.07325	0.07555	0.76379	0.00457	0.01069	0.90591	0.95847	0.78353	0.90624	0.95983
	20	0.21389	0.21411	0.69449	0.00364	0.00385	0.82753	0.86530	0.82981	0.88940	0.94797
	21	0.06634	0.06893	0.76375	0.00530	0.01299	0.90841	0.95953	0.78434	0.90897	0.96111
	29	0.04999	0.05633	0.76971	0.00185	0.01434	0.90745	0.95862	0.78240	0.90780	0.96011
	30	0.22561	0.22710	0.69468	0.00524	0.00324	0.82950	0.85270	0.83322	0.89469	0.94475
	31	0.03851	0.03891	0.77515	0.00564	0.00371	0.91278	0.95332	0.79091	0.91295	0.95366
1.b	1	0.45113	0.44993	0.53535	0.00723	0.00093	0.61719	0.67735	0.80933	0.82043	0.89936
	2	0.22396	0.22505	0.66648	0.00502	0.00483	0.82483	0.88027	0.73187	0.83049	0.90713
	3	0.07503	0.07496	0.74543	0.00568	0.01176	0.84127	0.92123	0.74728	0.84766	0.92183
	20	0.13132	0.10090	0.68737	0.01945	0.17185	0.83316	0.83881	0.76718	0.88422	0.90354
	21	0.03330	0.03447	0.71972	0.00387	0.04444	0.87856	0.89025	0.72912	0.88123	0.89106
	29	0.11237	0.10668	0.72709	0.00546	0.02111	0.84813	0.92068	0.73980	0.84901	0.92323
	30	0.38562	0.38624	0.59137	0.00214	0.00258	0.70837	0.72357	0.77485	0.81263	0.84633
	31	0.09967	0.10031	0.71003	0.00404	0.00694	0.86502	0.92101	0.72468	0.86628	0.92231
	441	0.03353	0.01268	0.73399	0.00493	0.00524	0.89087	0.90665	0.73545	0.89176	0.90662
	678	0.03279	0.03279	0.73232	0.00528	0.27486	0.86868	0.91193	0.73647	0.87080	0.91330
1.c	1	0.14645	0.01362	0.36917	0.00416	0.06011	0.44913	0.51161	0.47875	0.56327	0.62291
	2	0.12996	0.00858	0.36770	0.01451	0.00826	0.44059	0.48173	0.48885	0.56293	0.57503
	19	0.04465	0.05427	0.36025	0.00940	0.00333	0.42737	0.49215	0.37246	0.42815	0.50078
	20	0.16555	0.15464	0.32243	0.00683	0.00756	0.38941	0.43179	0.39134	0.41658	0.48059
	21	0.04283	0.05402	0.37236	0.00534	0.00243	0.43652	0.48874	0.37882	0.43743	0.48786
	29	0.09692	0.11867	0.34685	0.00373	0.02103	0.44164	0.49634	0.36720	0.44531	0.49726
	30	0.37985	0.41950	0.23859	0.00295	0.01175	0.26363	0.27856	0.46388	0.47668	0.49143
	31	0.10208	0.10257	0.38777	0.00559	0.00293	0.44019	0.50440	0.39501	0.44446	0.50992
	32	0.03964	0.03682	0.38491	0.00535	0.00252	0.41938	0.49687	0.39071	0.42153	0.50526
	40	0.03797	0.00389	0.36696	0.01182	0.05935	0.43796	0.54063	0.37434	0.44488	0.54610
1.d	1	0.22419	0.33315	0.14222	0.09564	0.12019	0.22044	0.32553	0.28034	0.36512	0.44376
	2	0.12034	0.16186	0.23212	0.13574	0.16379	0.36211	0.48440	0.25837	0.37135	0.48423
	3	0.03620	0.04955	0.27764	0.19759	0.17025	0.47089	0.53529	0.28072	0.46607	0.51153
	19	0.02460	0.04348	0.22017	0.01360	0.10895	0.30057	0.41718	0.23396	0.30409	0.41156
	20	0.05434	0.10171	0.20541	0.01704	0.17063	0.28898	0.44720	0.24941	0.31192	0.45188
	29	0.02737	0.00739	0.24275	0.00233	0.00419	0.29530	0.32243	0.26352	0.32221	0.34744
	30	0.57178	0.00387	0.24034	0.00992	0.00536	0.29161	0.32018	0.68626	0.67434	0.66686
	31	0.04372	0.00297	0.23520	0.00297	0.00774	0.29643	0.32435	0.27402	0.32783	0.35430
	479	0.02586	0.01204	0.24004	0.03109	0.05419	0.31673	0.37870	0.25149	0.33209	0.40232
	886	0.02578	0.01772	0.22769	0.02504	0.04149	0.29036	0.35793	0.24707	0.30314	0.37337

Table 4.7: Simulation results for CEV-PLS using all predictors for Example 1.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
1.a	1	0.43615	0.43419	0.78777	0.00169	0.00838	0.89103	0.95488	0.83027	0.90898	0.96968
	2	0.45886	0.46234	0.79285	0.00138	0.00590	0.89261	0.93712	0.79536	0.91407	0.93771
	3	0.14238	0.14221	0.79247	0.00204	0.00621	0.90327	0.95164	0.85764	0.90415	0.95271
	4	0.05661	0.05978	0.79107	0.00208	0.00627	0.90685	0.95195	0.84917	0.90648	0.95315
	19	0.07325	0.07555	0.80081	0.00240	0.02394	0.89919	0.96256	0.80361	0.91914	0.96302
	20	0.21389	0.21411	0.78495	0.00221	0.00543	0.90784	0.95571	0.80042	0.90828	0.96209
	21	0.06634	0.06893	0.78790	0.00387	0.00557	0.90201	0.95444	0.79012	0.92184	0.95629
	29	0.04999	0.05633	0.79665	0.00184	0.02215	0.89792	0.95083	0.79805	0.90591	0.95103
	30	0.22561	0.22710	0.79091	0.00197	0.00637	0.88671	0.93957	0.81885	0.88690	0.94343
	31	0.03851	0.03891	0.78683	0.00344	0.00381	0.89535	0.95050	0.78790	0.91171	0.95626
1.b	1	0.45113	0.44993	0.71631	0.00530	0.00256	0.87933	0.89569	0.78424	0.87977	0.90126
	2	0.22396	0.22505	0.72130	0.00508	0.05919	0.88358	0.89465	0.72209	0.88675	0.89527
	3	0.07503	0.07496	0.72864	0.01306	0.00653	0.88896	0.91813	0.76953	0.88931	0.91925
	20	0.13132	0.10090	0.73024	0.00428	0.01907	0.87564	0.89414	0.75614	0.88691	0.89898
	21	0.03330	0.03447	0.72322	0.01128	0.08669	0.86356	0.89551	0.72858	0.88645	0.89759
	29	0.11237	0.10668	0.73091	0.00692	0.01918	0.87965	0.90142	0.74437	0.88011	0.90336
	30	0.38562	0.38624	0.72366	0.00362	0.00493	0.87398	0.88794	0.74547	0.89106	0.90151
	31	0.09967	0.10031	0.72640	0.00625	0.00954	0.87049	0.87375	0.73740	0.87935	0.87959
	441	0.03353	0.01268	0.73055	0.00535	0.00744	0.88896	0.90111	0.73773	0.89426	0.90651
1.c	678	0.03279	0.03279	0.72667	0.00575	0.04031	0.85992	0.88521	0.72925	0.86068	0.89130
	1	0.14645	0.01362	0.36637	0.00470	0.06515	0.43333	0.49137	0.47396	0.53331	0.57185
	2	0.12996	0.00858	0.36946	0.02458	0.13177	0.44968	0.55958	0.48669	0.55848	0.61719
	19	0.04465	0.05427	0.36537	0.00801	0.08077	0.45610	0.54660	0.37149	0.46117	0.54711
	20	0.16555	0.15464	0.37103	0.01122	0.06082	0.45051	0.52814	0.38629	0.46450	0.53282
	21	0.04283	0.05402	0.36828	0.00893	0.11536	0.44463	0.51107	0.37934	0.44932	0.51964
	29	0.09692	0.11867	0.36939	0.01280	0.11885	0.44658	0.53700	0.37341	0.48199	0.71019
	30	0.37985	0.41950	0.36937	0.00368	0.04075	0.44369	0.50068	0.44442	0.46452	0.51303
	31	0.10208	0.10257	0.36537	0.01062	0.10122	0.44334	0.52042	0.39316	0.44908	0.52301
	32	0.03964	0.03682	0.36835	0.00463	0.06736	0.45270	0.52802	0.38860	0.45694	0.52990
1.d	40	0.03797	0.00389	0.36627	0.00738	0.06712	0.44708	0.49029	0.37429	0.45148	0.50157
	1	0.22419	0.33315	0.23574	0.02175	0.06177	0.30407	0.38110	0.27306	0.31674	0.41376
	2	0.12034	0.16186	0.23446	0.02155	0.04343	0.30529	0.37791	0.24147	0.31925	0.39893
	3	0.03620	0.04955	0.22948	0.02281	0.05560	0.30598	0.38039	0.28178	0.32948	0.35327
	19	0.02460	0.04348	0.23652	0.02184	0.01849	0.30119	0.34826	0.23889	0.31161	0.36182
	20	0.05434	0.10171	0.24195	0.02139	0.06575	0.30573	0.38539	0.25612	0.31930	0.40469
	29	0.02737	0.00739	0.23222	0.02300	0.03307	0.30701	0.37190	0.25585	0.32812	0.36468
	30	0.57178	0.00387	0.23405	0.01543	0.04606	0.30218	0.37058	0.68094	0.68606	0.64742
	31	0.04372	0.00297	0.23996	0.01798	0.05112	0.30335	0.38383	0.28140	0.32985	0.39489
	479	0.02586	0.01204	0.23591	0.01435	0.06514	0.30076	0.38277	0.25182	0.31967	0.39363
1.e	886	0.02578	0.01772	0.23540	0.01586	0.06004	0.30446	0.37964	0.24664	0.30605	0.37774

For example 1, CEV-PCA and CEV-PLS algorithms show very similar trend with regard to different CEV estimators. After taking PCs and PLs into consideration, we could always obtain higher CEV values, which means more variability of Y has been explained comparing with one individual variable. Also, the results here are consistent with the simulation of CEV-PWA algorithm, and we can always identify the variables that are truly active. Both CEV-PCA and CEV-PLS algorithms work well for capturing linear relationship.

4.3.2 Simulation Example 2

Example 2. This example is designed to examine the performance of CEV-PCA and CEV-PLS for nonlinear models. We generate $X = (X_1, \dots, X_p)$ from a normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$, where $\sigma_{ii} = 1, i = 1, \dots, p$ and $\sigma_{ij} = \sigma, i \neq j$. Model (2.a) and (2.b) are non-parametric additive models with polynomial terms. The random error ε follows a standard normal distribution. We set $n = 500$ and $p = 1000$, $\sigma = 0$.

$$(2.a) \ Y = 5X_1 + 3X_2^2 + X_3^3 + \varepsilon$$

$$(2.b) \ Y = 5\sqrt{|X_1|} + 3X_2^2 + X_3^3 + \varepsilon$$

Principal component approximation result

Table 4.8: Simulation results for CEV-PCA using partial predictors for Example 2.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
2.a	1	0.60206	0.62223	0.12989	0.23792	0.22593	0.24598	0.257	0.65707	0.7375	0.74106
	2	0.102	0.00964	0.64028	0.24911	0.1939	0.64256	0.65516	0.76652	0.76258	0.77114
	3	0.206	0.18909	0.40712	0.20009	0.16356	0.45656	0.54901	0.55727	0.60468	0.65069
	125	0.03557	0.02024	0.52465	0.24348	0.22248	0.55423	0.57429	0.5375	0.56635	0.57947
	279	0.02967	0.03939	0.66102	0.35211	0.21973	0.68679	0.71214	0.6632	0.68109	0.71227
	306	0.03034	0.02835	0.54517	0.33835	0.31885	0.59293	0.6998	0.54866	0.59597	0.69492
	571	0.02975	0.01825	0.55828	0.38306	0.31001	0.66042	0.66243	0.56522	0.66436	0.66356
	580	0.02736	0.01121	0.61821	0.12656	0.19521	0.68744	0.73343	0.62271	0.6838	0.7283
	755	0.03068	0.01423	0.5328	0.30614	0.16165	0.58596	0.6	0.53963	0.58909	0.60142
	806	0.03223	0.01954	0.5093	0.30571	0.2286	0.5144	0.5154	0.51366	0.51718	0.5215
2.b	1	0.12519	0.01415	0.34091	0.24862	0.24893	0.43583	0.453	0.43466	0.45304	0.46101
	2	0.25705	0.01165	0.32889	0.24116	0.11745	0.33826	0.32741	0.55027	0.54495	0.46535
	3	0.47073	0.38545	0.13922	0.12319	0.12078	0.18372	0.218	0.56162	0.63065	0.59564
	74	0.02616	0.01401	0.32537	0.17304	0.29464	0.34055	0.40711	0.31674	0.33515	0.42469
	181	0.02806	0.02671	0.37266	0.24524	0.16275	0.49495	0.54609	0.36275	0.47034	0.5496
	369	0.02409	0.02147	0.31191	0.25271	0.26604	0.37033	0.41312	0.32314	0.37253	0.38421
	515	0.02379	0.02342	0.32183	0.19765	0.23459	0.37737	0.35817	0.32535	0.35907	0.35904
	635	0.02831	0.0099	0.34013	0.21788	0.12944	0.41306	0.44935	0.3425	0.41804	0.38681
	984	0.02334	0.00789	0.34147	0.25369	0.25936	0.38416	0.3371	0.33742	0.34564	0.34735
	997	0.02998	0.00254	0.3784	0.20462	0.16815	0.57802	0.59194	0.3817	0.51987	0.47581

All simulation performances are summarised in Table 4.8 and Table 4.9. For model (2.a), $CEV(Y|X_j, C1, C2, C3)$ is most likely to achieve the highest value, or at least very close to the highest one. This indicates that CEV-PCA algorithm performs good in capturing the joint information from the covariates. As for model (2.b), $CEV(Y|X_j, C1, C2, C3)$ always achieves the highest value comparing with other CEV estimators, which exactly fits the expectation from theoretical analysis. For both models, one can see that the relationship of $CEV(Y|X_j) \leq CEV(Y|C1) \leq CEV(Y|X_j, C1, C2, C3)$ holds as observed from Example 1.

Table 4.9: Simulation results for CEV-PCA using all predictors for Example 2.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
2.a	1	0.60206	0.62223	0.60001	0.24103	0.21397	0.60781	0.63316	0.64542	0.67011	0.70344
	2	0.102	0.00964	0.51019	0.20438	0.20838	0.54666	0.59647	0.69912	0.76493	0.75366
	3	0.206	0.18909	0.51983	0.2273	0.21079	0.60206	0.63634	0.59148	0.66183	0.6759
	125	0.03557	0.02024	0.49245	0.25189	0.22982	0.51258	0.54608	0.50324	0.52142	0.55246
	279	0.02967	0.03939	0.52702	0.22387	0.2261	0.54602	0.5619	0.5378	0.56138	0.57723
	306	0.03034	0.02835	0.60861	0.27363	0.18407	0.6138	0.65002	0.6094	0.61336	0.65109
	571	0.02975	0.01825	0.56793	0.26313	0.20932	0.57451	0.58546	0.57763	0.59177	0.60244
	580	0.02736	0.01121	0.56959	0.19397	0.25775	0.64164	0.64807	0.57592	0.66405	0.66695
	755	0.03068	0.01423	0.54392	0.21654	0.18468	0.56035	0.63118	0.55241	0.5721	0.63908
	806	0.03223	0.01954	0.55636	0.20951	0.20224	0.56118	0.58437	0.56287	0.58547	0.61258
2.b	1	0.12519	0.01415	0.32828	0.23413	0.18541	0.46451	0.51892	0.42456	0.53041	0.58136
	2	0.25705	0.01165	0.32913	0.25011	0.17601	0.4577	0.53871	0.5569	0.59065	0.61374
	3	0.47073	0.38545	0.33283	0.1722	0.18516	0.34559	0.40761	0.49273	0.51227	0.49274
	74	0.02616	0.01401	0.33723	0.2292	0.20135	0.42184	0.40153	0.3472	0.40006	0.4421
	181	0.02806	0.02671	0.30688	0.24708	0.15611	0.45318	0.44953	0.32338	0.44343	0.42958
	369	0.02409	0.02147	0.38939	0.21937	0.17073	0.44178	0.47878	0.37747	0.45401	0.42732
	515	0.02379	0.02342	0.28644	0.27581	0.1707	0.37767	0.54908	0.31757	0.36072	0.46193
	635	0.02831	0.0099	0.32761	0.2301	0.21244	0.40135	0.51004	0.33336	0.39964	0.45685
	984	0.02334	0.00789	0.34303	0.17973	0.19782	0.36207	0.41287	0.32457	0.34983	0.42171
	997	0.02998	0.00254	0.30961	0.22904	0.20784	0.33804	0.33539	0.32669	0.37571	0.37796

Based on the true models, X_1 , X_2 , and X_3 are used to construct variable Y and can be regarded as **truly active variables**. By checking the performance of $CEV(Y|X_j)$ in Table 4.8 and 4.9, we find that X_1 , X_2 , and X_3 hold relatively high values for both model (2.a) and (2.b). When comparing $CEV(Y|X_j)$ with $CEV(Y|C1)$, we notice that X_1 in model (2.a) has a greater $CEV(Y|X_1)$ value, 0.602, than $CEV(Y|C1)$, 0.12989, while X_2 and X_3 have higher $CEV(Y|C1)$ values. One possible reason is that X_1 achieves the highest CEV value when it's evaluated as an individual variable, and it dominates the explained variability of Y . Hence, the joint information from other variables $C1$ can't compete with the individual X_1 . The similar condition happens to variable X_3 in model (2.b), which also achieves the highest $CEV(Y|X_j)$ value. If we add the performance of $CEV(Y|X_j, C1)$ into comparison, we can conclude that X_1 , X_2 , and X_3 should be taken as active variables. Due to the fact

that both models in Example 2 contain polynomial terms, the results here indicate that CEV-PCA can capture nonlinear and asymmetric relationships. By comparing the values of CEV estimators for these two types (Table 4.8 for partial predictors and Table 4.9 for all predictors), we find that the values of CEV estimators using all predictors are slightly greater than those using partial predictors.

Partial least squares approximation result

Table 4.10: Simulation results for CEV-PLS using partial predictors for Example 2.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
2.a	1	0.60206	0.62223	0.24557	0.02854	0.04123	0.26356	0.26592	0.70579	0.73602	0.73726
	2	0.10199	0.00964	0.67024	0.01469	0.01670	0.73508	0.73611	0.77472	0.84888	0.84207
	3	0.20600	0.18909	0.55666	0.02624	0.06741	0.62891	0.63203	0.69210	0.75341	0.75971
	125	0.03557	0.02024	0.65276	0.00373	0.09363	0.72851	0.74181	0.65944	0.73141	0.74198
	279	0.02967	0.03939	0.65838	0.01283	0.03530	0.74359	0.76100	0.66465	0.74764	0.76136
	306	0.03034	0.02835	0.65697	0.00815	0.01133	0.73462	0.74059	0.66161	0.73590	0.74189
	571	0.02975	0.01825	0.64441	0.01077	0.06635	0.73376	0.74823	0.65099	0.73612	0.75052
	580	0.02736	0.01121	0.64381	0.01486	0.06090	0.74034	0.75443	0.64810	0.74124	0.75358
	755	0.03088	0.01423	0.65586	0.00537	0.04821	0.73788	0.74059	0.66140	0.74076	0.74296
	806	0.03223	0.01954	0.66227	0.02944	0.01082	0.74541	0.74854	0.66275	0.74425	0.74868
2.b	1	0.12519	0.01415	0.43730	0.02173	0.07622	0.47363	0.55351	0.51642	0.53675	0.57400
	2	0.25705	0.01165	0.43987	0.00974	0.04626	0.46448	0.46355	0.65318	0.65973	0.62561
	3	0.47073	0.38545	0.09157	0.04925	0.19035	0.13138	0.26078	0.48293	0.48674	0.57703
	74	0.02616	0.01401	0.45319	0.02708	0.07139	0.48512	0.52491	0.45360	0.47894	0.50315
	181	0.02806	0.02671	0.45419	0.01711	0.03210	0.46532	0.48279	0.45386	0.46841	0.47448
	369	0.02409	0.02147	0.44433	0.00809	0.06123	0.46298	0.51299	0.44184	0.44703	0.52070
	515	0.02379	0.02342	0.44247	0.02734	0.15010	0.47434	0.56056	0.44659	0.46966	0.53423
	635	0.02831	0.00990	0.44329	0.02128	0.08188	0.47092	0.50792	0.45215	0.47383	0.51102
	984	0.02334	0.00789	0.43946	0.01481	0.08430	0.47101	0.48680	0.44217	0.47107	0.49930
	997	0.02998	0.00254	0.44583	0.02425	0.06812	0.46470	0.47750	0.44942	0.46560	0.48240

Table 4.11: Simulation results for CEV-PLS using all predictors for Example 2.

Model	X_j	$Y X_j$	$X_j Y$	$Y C1$	$Y C2$	$Y C3$	$Y C1, C2$	$Y C1, C2, C3$	$Y X_j, C1$	$Y X_j, C1, C2$	$Y X_j, C1, C2, C3$
2.a	1	0.60206	0.62223	0.64618	0.01141	0.03610	0.74223	0.75305	0.67355	0.74513	0.75882
	2	0.10199	0.00964	0.65236	0.01745	0.08673	0.74255	0.75730	0.76690	0.85165	0.84407
	3	0.20600	0.18909	0.65693	0.01826	0.07064	0.74052	0.75182	0.69256	0.75622	0.77237
	125	0.03557	0.02024	0.64383	0.01251	0.05914	0.73966	0.74733	0.65928	0.74352	0.75126
	279	0.02967	0.03939	0.64464	0.01828	0.01558	0.73789	0.74709	0.65202	0.74342	0.75081
	306	0.03034	0.02835	0.64764	0.01753	0.05391	0.74322	0.74494	0.66886	0.74443	0.74707
	571	0.02975	0.01825	0.64948	0.01500	0.06525	0.74033	0.74914	0.65746	0.74336	0.74982
	580	0.02736	0.01121	0.65280	0.01145	0.04139	0.74188	0.74925	0.65499	0.74320	0.74870
	755	0.03088	0.01423	0.64515	0.01356	0.03947	0.73864	0.75931	0.64667	0.74272	0.75386
	806	0.03223	0.01954	0.65770	0.01912	0.04919	0.73924	0.74543	0.68068	0.74575	0.75083
2.b	1	0.12519	0.01415	0.44192	0.01762	0.07438	0.46992	0.53026	0.50927	0.53463	0.49192
	2	0.25705	0.01165	0.44148	0.01875	0.07180	0.47298	0.52279	0.65363	0.64626	0.66284
	3	0.47073	0.38545	0.44049	0.01833	0.06427	0.47035	0.49696	0.49794	0.50150	0.51288
	74	0.02616	0.01401	0.43866	0.01441	0.06328	0.47120	0.50721	0.44456	0.48312	0.51280
	181	0.02806	0.02671	0.44169	0.02245	0.05539	0.47679	0.52296	0.44760	0.48278	0.50431
	369	0.02409	0.02147	0.44194	0.01782	0.07241	0.46618	0.50552	0.44454	0.45492	0.51785
	515	0.02379	0.02342	0.43926	0.01515	0.06678	0.46787	0.51778	0.44204	0.47084	0.48895
	635	0.02831	0.00990	0.44302	0.01559	0.08275	0.46460	0.54482	0.44876	0.46901	0.50150
	984	0.02334	0.00789	0.43775	0.01901	0.08369	0.46970	0.50803	0.43855	0.46743	0.53812
	997	0.02998	0.00254	0.44239	0.01965	0.08261	0.47371	0.51962	0.44824	0.47159	0.52130

The overall performance of PLS in Table 4.10 and Table 4.11 is quite similar to the previous CEV-PCA algorithm for Example 2. We can always observe that $\text{CEV}(Y|X_j, C1, C2, C3)$ is most likely to achieve the highest value, or at least very close to the highest one. This suggests that CEV-PLS algorithm performs good in capturing the joint information from the covariates. Also, we find that the values of CEV estimators using all predictors are slightly greater than those using partial predictors. Actually, for both CEV-PCA and CEV-PLS algorithms, the performances are consistent with the simulation of CEV-PWA algorithm. They can always capture more information than individual variable X_j using the network constructed by other variables. And we are able to detect the truly active variables by comparing the results of different CEV estimators. When comparing CEV-PCA and CEV-PLS algorithm, we find that CEV-PLS algorithm performs better than CEV-PCA for most

of the time with regard to both linear and nonlinear simulation examples. We'll show more detailed comparison among three algorithms we proposed in Chapter 5.

4.4 Real data application

4.4.1 fMRI Application: Major Depressive Disorder

In this section, we are using the same data as Chapter 3 to study the alternative approaches in analyzing the fMRI data for major depressive disorder (MDD). In order to study the performance of the proposed CEV-PCA and CEV-PLS algorithms and the causal relationship among these regions for MDD, we summarized the results of 7 regions in the following figures. Using the same statistics provided in Table 4.3, we applied different colors and line types to denote these CEVs. Note that the Y in the legend represents each of the 7 regions shown in the title of the subplot, and X includes the rest 47 regions except Y .

Figure 4.1 and Figure 4.2 provide the average CEV values by CEV-PCA algorithm for each of the 7 regions in both major depressive disorder (MDD) and healthy control (HC) groups. Notice that for all of the 7 regions, the lines of $Y|X_j$ and $X_j|Y$ are significantly much lower than the others, which suggests that our proposed CEV-PCA algorithm works well in explaining more variability of variable Y . Also, we notice that the pairwise relationship between region Y and X_j doesn't have much difference for both MDD and HC groups, as the line of $Y|X$ and $X|Y$ are very close to each other.

From both MDD and HC groups, we notice that the line of $Y|PC1$ is always higher than $Y|PC2$, and $Y|PC2$ achieves greater CEV value than $Y|PC3$, which fits the expectation to PCA. Also, the line of $Y|X_j, PC1, PC2, PC3$ is on top of other lines almost all the time. This indicates that when we consider all information including the individual variable X_j and the first 3 PCs, it explains the most variability of region Y as we expected.

Figure 4.1: Average CEV-PCA values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.

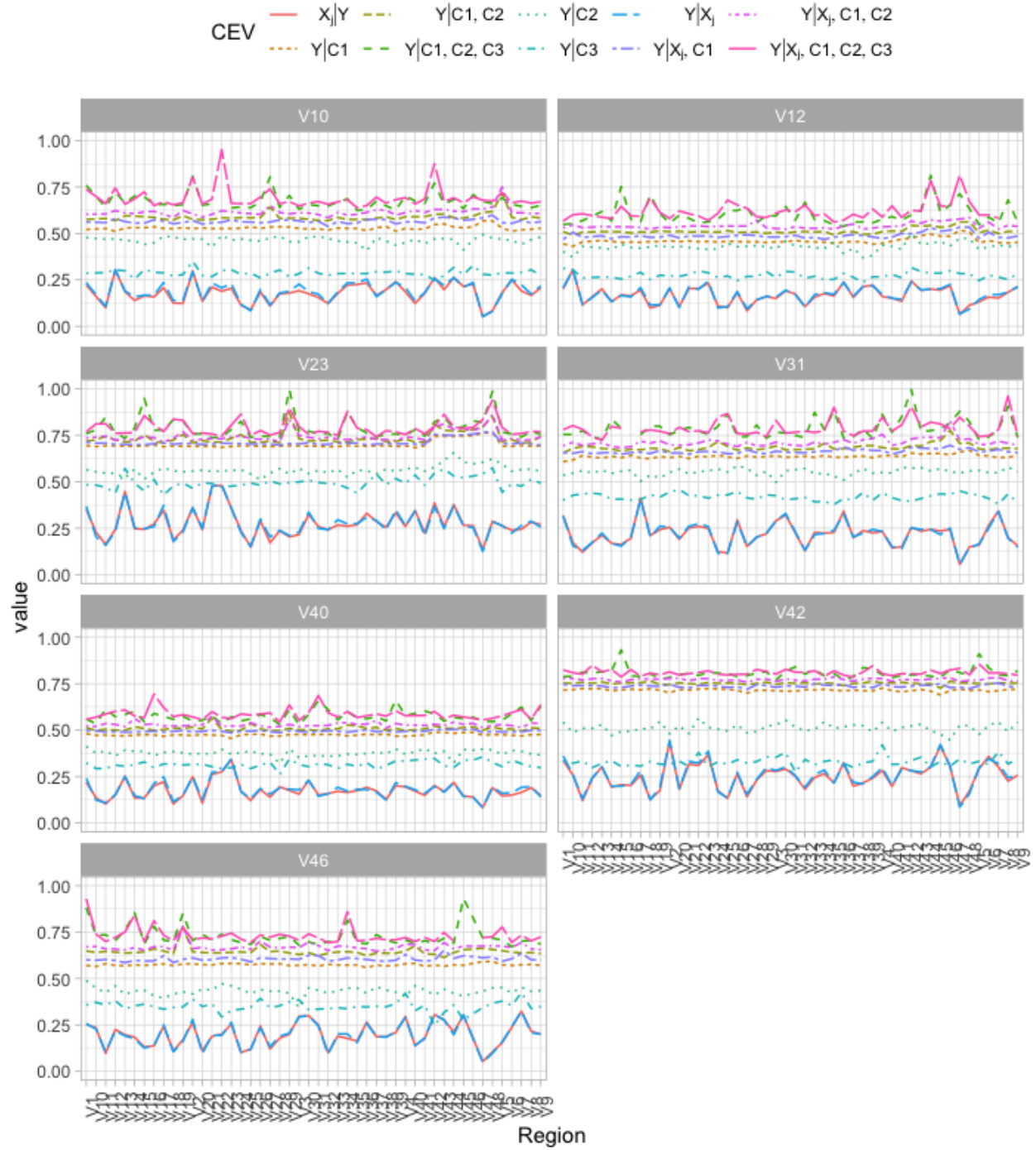
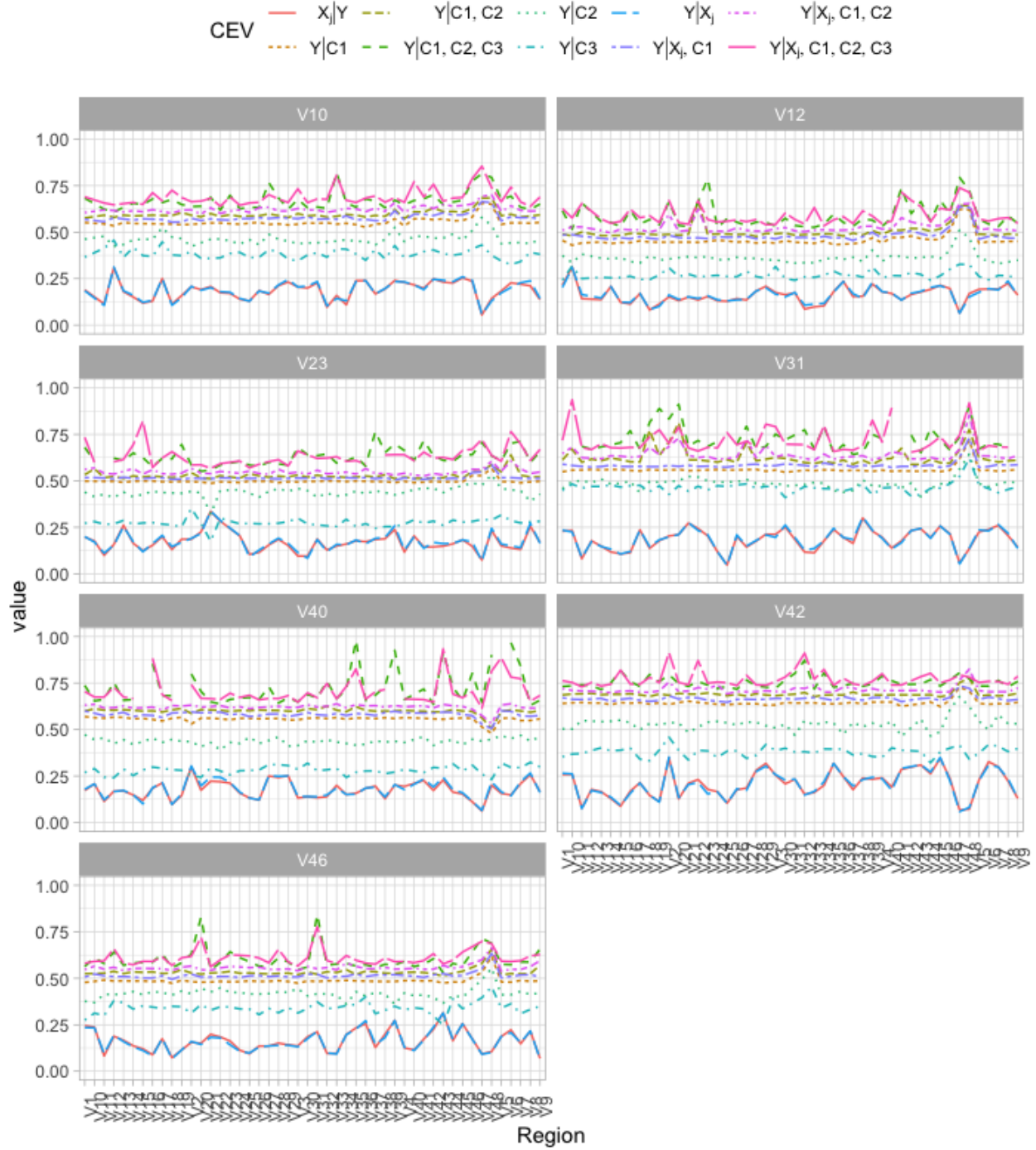


Figure 4.2: Average CEV-PCA values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.



For Figure 4.1 and Figure 4.2, the overall trend and relationship among these regions are quite similar for MDD and HC groups. However, if we compare the CEV values for the same region from these two groups, there are some regions are significantly different. Regions V23, V31, and V46 always achieve greater CEV values in the MDD after we add PCs into consideration than in the HC group, while region V40 works in the opposite way that achieves lower CEV values in the MDD than the HC.

Figure 4.3 and Figure 4.4 provide the average CEV values by CEV-PLS algorithm for each of the 7 regions in both major depressive disorder (MDD) and healthy control (HC) groups. Notice that for all of the 7 regions, the lines of $Y|X_j$ and $X_j|Y$ are significantly much lower than the others, which suggests that our proposed CEV-PLS algorithm works great in explaining more variability of variable Y . Also, we notice that the pairwise relationship between region Y and X_j doesn't have much difference for both MDD and HC groups, as the line of $Y|X$ and $X|Y$ are very close to each other.

Similar to the performance of CEV-PCA algorithm, we find that the line of $Y|PL1$ is always higher than $Y|PL2$ for both MDD and HC groups, and $Y|PL2$ achieves greater CEV value than $Y|PL3$, which fits the expectation to PLS. Also, the line of $Y|X_j, PL1, PL2, PL3$ is on top of other lines almost all the time. This indicates that when we consider all information including the individual variable X_j and the first 3 PLs, it explains the most variability of region Y as we expected. The overall trend and relationship among these regions are quite similar for MDD and HC groups. Then, if we compare the CEV values for the same region from these two groups, there are some regions are significantly different. Regions V23, V31, and V46 always achieve greater CEV values in MDD after we add PLs into consideration than in the HC group, while region V40 works in the opposite way that achieves lower CEV values in the MDD than the HC. The results we observed here are consistent with our previous analysis to CEV-PCA algorithm.

Figure 4.3: Average CEV-PLS values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.

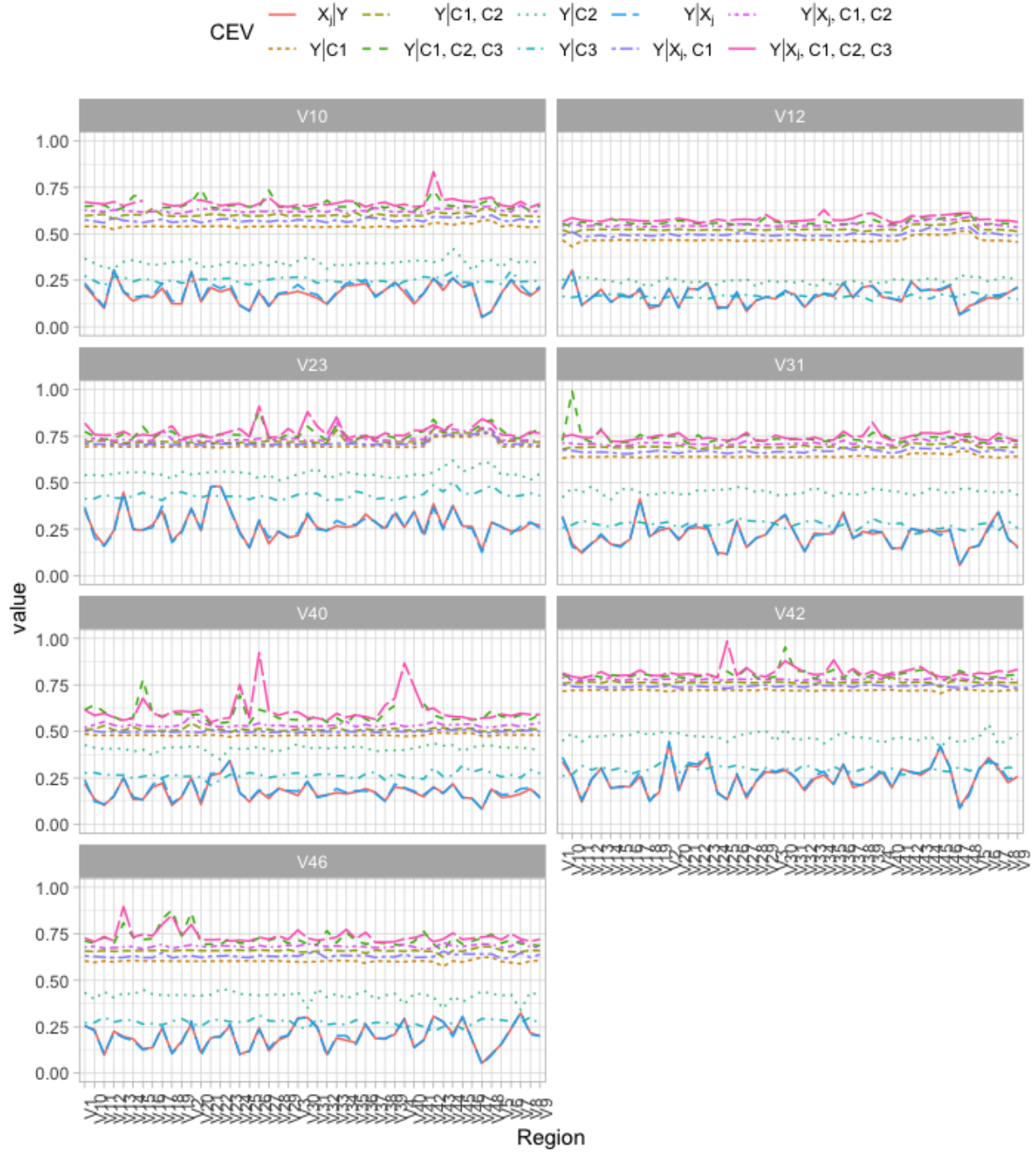
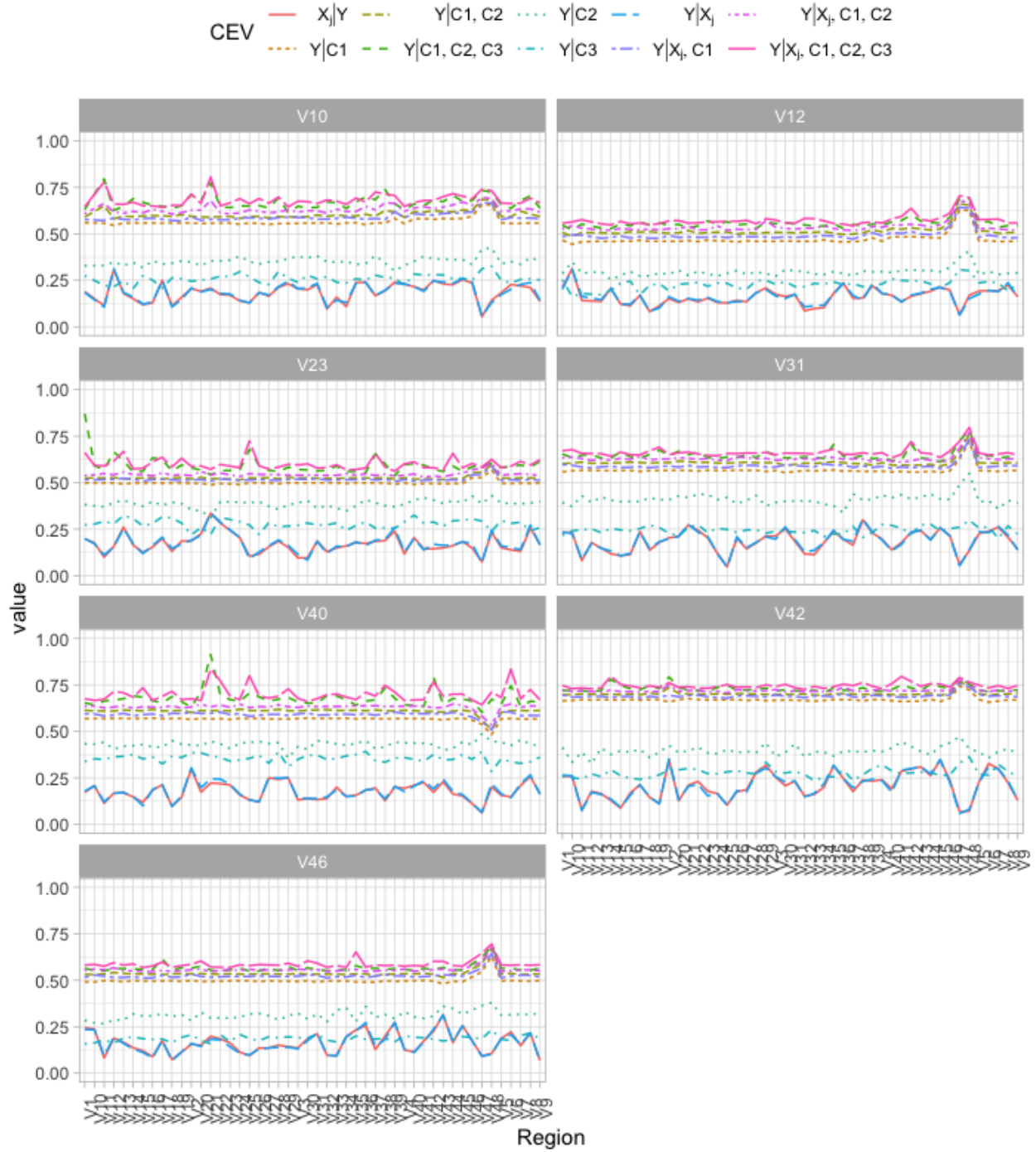


Figure 4.4: Average CEV-PLS values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.



As we compare the performances of CEV-PCA and CEV-PLS, it's remarkable that the overall trend of the CEV estimators are very similar. By checking the CEV values of pairwise connectivity $Y|X_j$ and $X_j|Y$, we find they don't have much difference. When we take PCs or PLs into consideration, the explained variability of Y is increasing as we add more components to the estimator. But still, there are some regions perform differently in MDD and HC groups for both CEV-PCA and CEV-PLS algorithms. Regions V23, V31, and V46 often achieve higher CEV values in MDD group, while region V40 ranks lower in the MDD than in the HC group.

Chapter 5

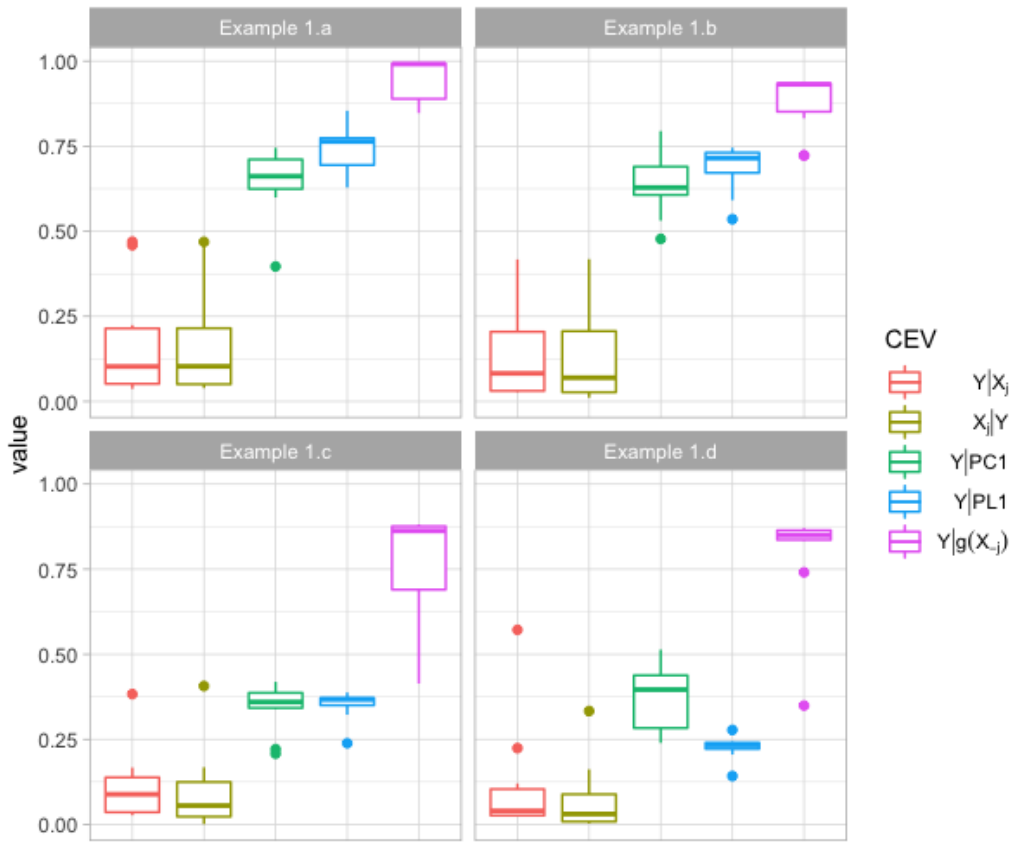
Discussion

A new type of asymmetric and nonlinear conditional causal relationship, WeCon causality, is introduced and three estimation approaches for cumulative explained variability are proposed in the previous chapters. Specifically, Chapter 3 describes the cumulative explained variability and piece-wise affine functions (PWA) for dimension reduction and variable selection, while Chapter 4 provides two alternative approaches, integrating principal component analysis (PCA) and partial least squares regression (PLS) with WeCon causality inference. In this Chapter 5, we will compare these three approaches and then make a conclusion. Section 5.1 provides the comparison among three procedures based on simulation studies and resting-state fMRI data application for major depressive disorder. Then, Section 5.2 discusses the proposed WeCon causality and three estimation approaches and then concludes the dissertation.

5.1 Comparison

Considering the asymmetric and nonlinear functional connectivity between brain regions, we first proposed the CEV-PWA approach to construct the weighted conditional network causality using continuous piecewise affine function (PWA) properties. It has been proved by simulation study and real data analysis that CEV-PWA works well. The PWA function can capture the selected variable's characteristics and increase the explained variability compared with an individual variable. However, as the nature of PWA will lead to the grouping of observations, this is a more complex model than simple linear combination of variables. Hence, a simpler and alternative approach is to use principal component analysis (PCA) and partial least square regression (PLS) for constructing the network of multiple variables.

Figure 5.1: Boxplot of CEV Values for Simulation Example 1



We use Figure 5.1 and Figure 5.2 to compare the performances of these three approaches in the simulation study. Each sub-figure corresponds to the specific simulation example as we stated in Chapter 3 and Chapter 4, and the 5 statistics include pairwise CEV values $Y|X_j$ and $X_j|Y$, the explained variability of Y given the first principal component PC1 ($Y|PC1$) and the first partial least squares component PL1 ($Y|PL1$), and the network constructed by other variables using PWA function ($Y|g(X_j)$). Note that the PC1 and PL1 here don't include the information from the individual variable X_j .

Figure 5.2: Boxplot of CEV Values for Simulation Example 2

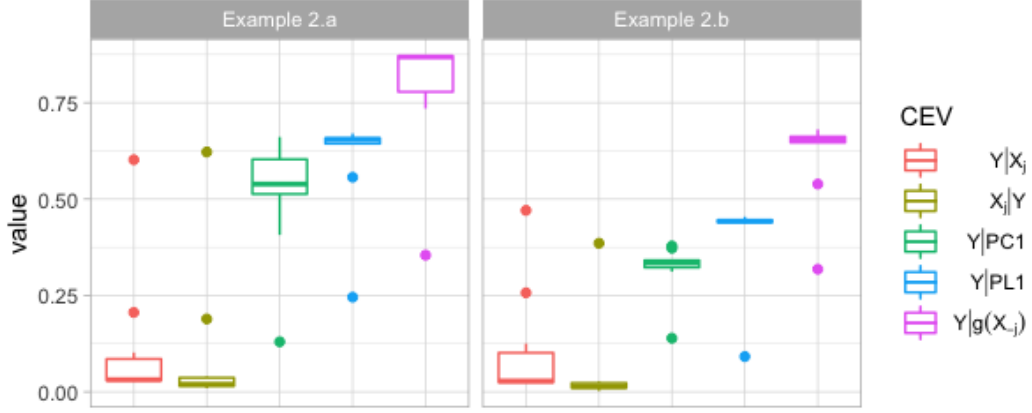


Figure 5.1 provides the comparison for the 4 linear models in simulation example 1. Clearly, the three CEV estimators $Y|PC1$, $Y|PL1$, and $Y|g(X_{-j})$ are all much higher than the pairwise CEV estimators. This means that after we constructed the network using joint information of other variables, the explained variability of Y is always greater than using one single variable. Then, for the first 3 models 1.a, 1.b, and 1.c, we can always rank the three estimators as $CEV(Y|PC1) \leq CEV(Y|PL1) < CEV(Y|g(X_{-j}))$. For the 4th linear model 1.d, it seems that $CEV(Y|PL1) \leq CEV(Y|PC1)$. Overall, the estimator from CEV-PWA approach always achieves the highest value, and the estimator from CEV-PLS works better than the one from CEV-PCA approach.

Simulation example 2 contains two non-linear models, and Figure 5.2 presents the comparison of three proposed methods. We can observe from both models that the three estimators from our proposed methods always perform much better than using one single variable. Of the 3 approaches, CEV-PLS works better than CEV-PCA, and CEV-PWA is more efficient than CEV-PLS.

Both linear and nonlinear simulation examples show the superior performance of the CEV-PWA approach, which means that it can capture the linear and nonlinear relationship

among various variables and significantly improve the explained variability of Y . Also, we find that in most cases, CEV-PLS performs better than CEV-PCA method, and this may be because PLS utilizes the information of Y for implementation while PCA doesn't.

As for real data analysis, we have Figure 5.3 and Figure 5.4 to compare the 3 proposed methods in the application of fMRI data for major depressive disorder (MDD). The lines in different colors represent the 5 estimators as we stated in the simulation example comparison, and each sub-figure corresponds to each of the 7 brain regions we highlighted in Chapter 3 and Chapter 4. Figure 5.3 provides the result for the MDD group, and Figure 5.4 shows the performance for healthy control (HC) group. Obviously, we find that the line of $Y|g(X_{-j})$ is always on top of other lines for almost all brain regions in both MDD and HC groups. This means that CEV-PWA approach performs the best among all estimators. In addition, when we compare CEV-PLS and CEV-PCA algorithms, most of the time they are very close to each other for both MDD and HC groups. But still, for region V46 in the MDD and region V42 in the HC, the line $Y|PL1$ is higher than the line of $Y|PC1$. This indicates that CEV-PLS and CEV-PCA performs similarly for this data, while sometimes CEV-PLS is still more efficient than CEV-PCA approach.

Figure 5.3: Comparison of average CEV values for each region in the major depressive disorder group (MDD). Regions V10, V12, V23, V31, V40, V42, V46 are included.

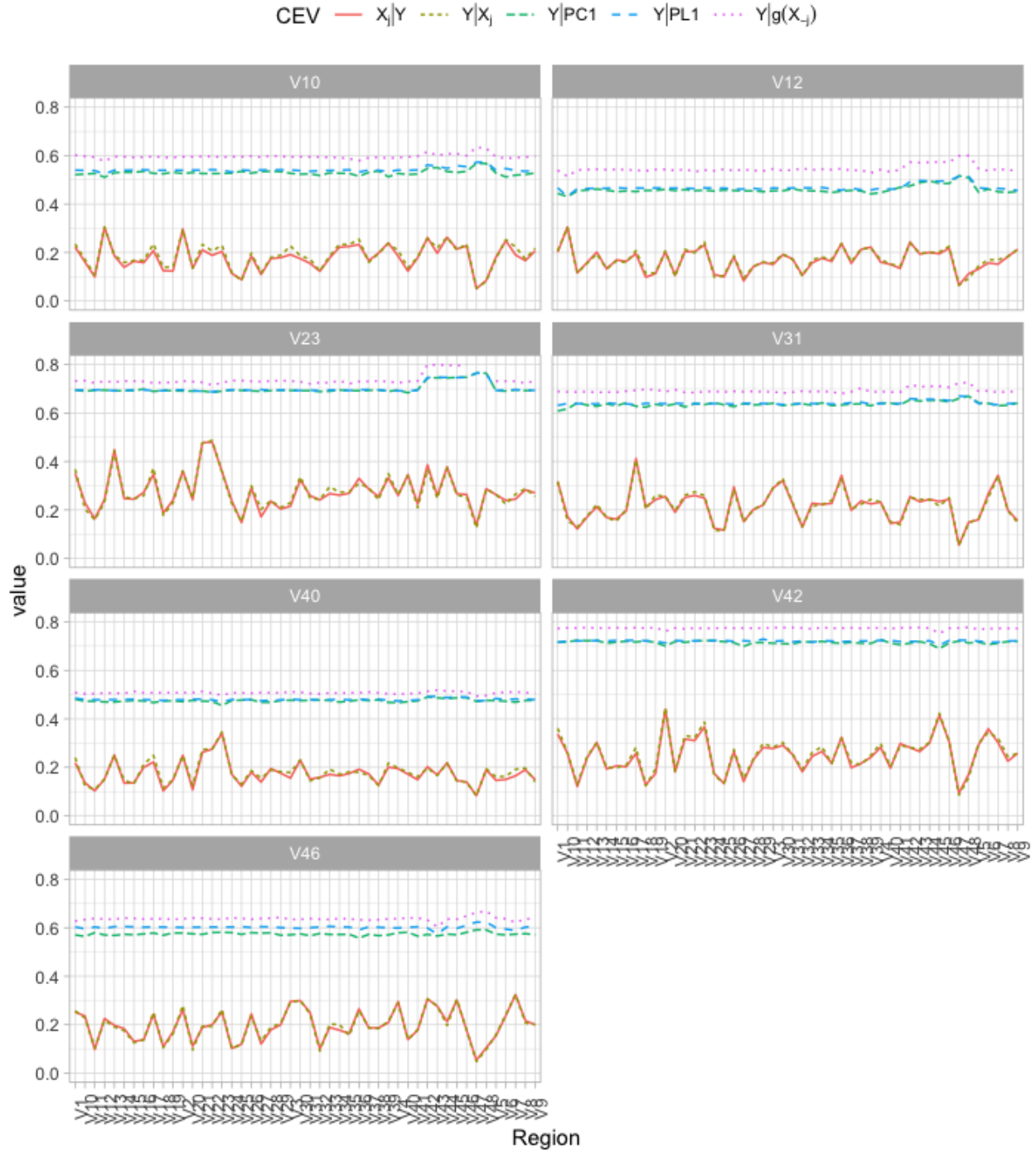
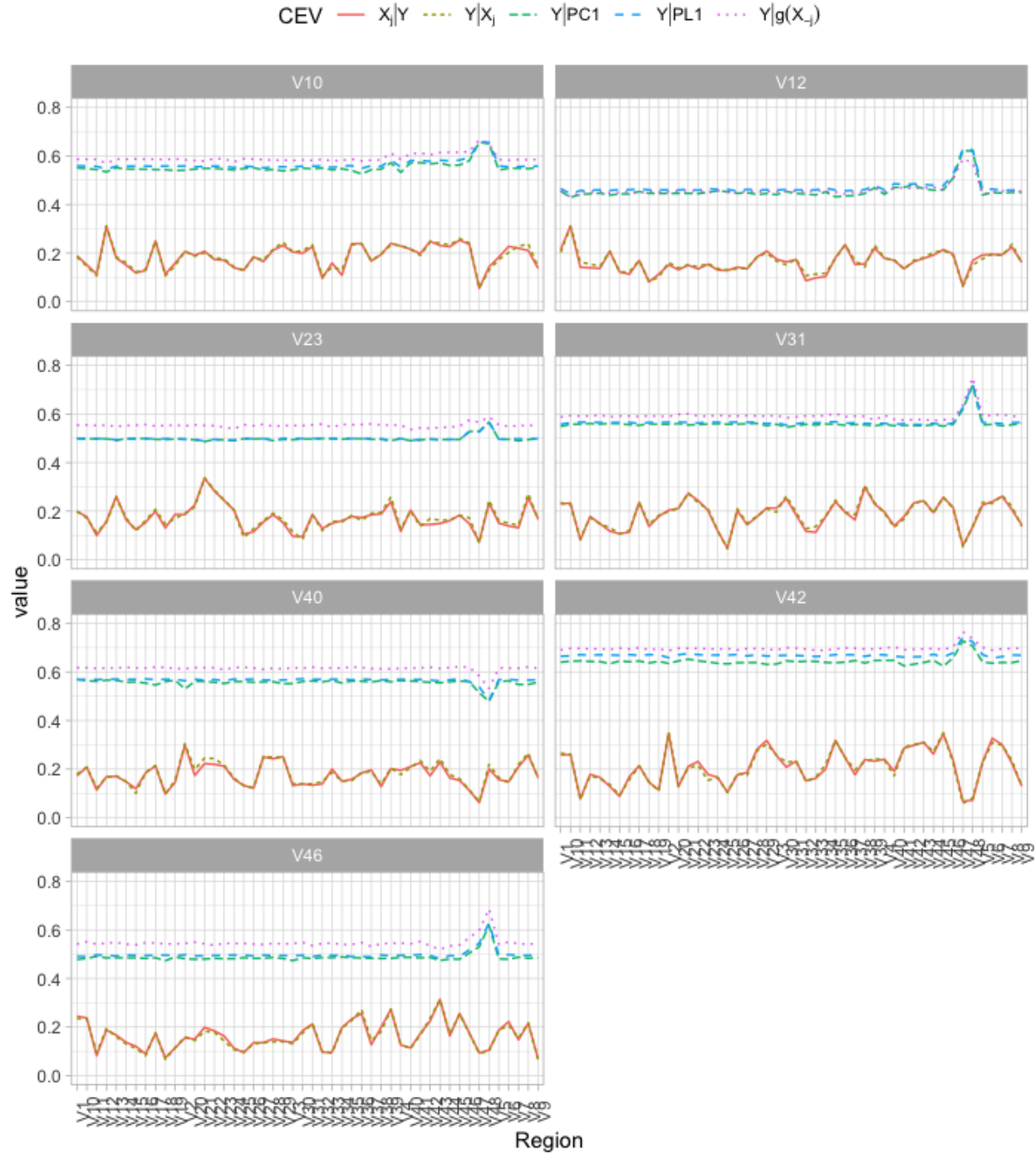


Figure 5.4: Comparison of average CEV values for each region in the healthy control group (HC). Regions V10, V12, V23, V31, V40, V42, V46 are included.



The comparisons from the simulation study and real data application indicate that CEV-PWA approach always achieves the best result, and most of the time, CEV-PLS works better than CEV-PCA, although there are some cases when these two alternative approaches perform similarly.

5.2 Conclusion

A new causal relationship is defined as WeCon causality, which is a more general form of the causal relationship from Zheng et al. (2012) and Vinod (2017). In addition, we introduce a new estimation approach for cumulative explained variability (CEV), using the properties of continuous piecewise affine function (PWA). Then we further proposed the CEV-PWA algorithm. The simulation studies show that the CEV-PWA algorithm performs well, and the PWA function is able to capture the characteristics of the selected variable and increases the explained variability compared with using a single variable. The performance of the algorithm is also verified in three real data analyses, including EEG and fMRI research on visual categorization, Parkinson’s disease (PD), and major depressive disorder (MDD). In the meantime, the proposed method identifies the asymmetric nonlinear causal relationship.

As alternative approximation approaches, principal component analysis and partial least squares regression have been used to construct the network of multiple variables instead of PWA, and the corresponding algorithms are summarized as CEV-PCA and CEV-PLS. Furthermore, simulation and fMRI data application in MDD have been conducted for both alternative approaches, which show that both algorithms perform well and constructed networks could capture the characteristics of the selected variable and achieve a higher explained variability than using an individual variable. Finally, the comparison among the three proposed algorithms shows that CEV-PWA algorithm performs better than CEV-PCA and CEV-PLS algorithms all the time, and CEV-PLS algorithm works better than CEV-PCA algorithm in most cases.

For the research on EEG data for human visual categorization, several influential brain channels are identified by the CEV-PWA approach. We observe that channels C4, F7, F8, and P4r WeCon cause the other individual channels, and channels C4, F3, FP2, and T4 WeCon cause other channels joint by PWA function. Hence, channel C4 is identified as an important channel not only WeCon causes other individual channels but also causes the joint of other channels by PWA function. On the contrary, channel O2 works as a receiver and would be passively influenced by the network constructed by other channels.

The application in EEG data for Parkinson's disease (PD) shows that antero-frontal channels AF7, AF8, and frontal-pole channel FP2 can actively influence other individual channels and central-parietal channels CP1, CP5, and frontal channel F3 seem to be passively impacted by other individual channels in the healthy control group (HC), which is not the case in the PD. In contrast, frontal-central channels FC3, FZ, frontal-temporal channel FT9, and parietal channel P5 WeCon cause other individual channels in PD, while we don't see this trend in the HC. Channel PO7 also draws our attention to the opposite performances in PD and HC groups. The line figure shows that it dominates other individual channels in PD group, while in the HC group, it seems to be passively influenced by other channels. By analyzing both line figures with multiple CEV estimators and the network graph with pairwise connectivity, we find that channels FC3, FZ, and P5 in the PD group show that both pairwise and network relationships are always more likely to be passively influenced by other channels. For channels AF8 and FP2, they perform in the opposite way with regard to pairwise and network relationships. AF8 and FP2 WeCon cause other individual channels in the HC, which is not observed in the PD. On the contrary, these two channels can actively affect the network of other channels in the PD, which is not the case for the HC. As we further divided the subjects in both PD and HC into two subgroups based on age (Young and Old), we noticed from the network figure that the level of connectivity in PD-Young group is much weaker than in the PD-Old group, while in HC group, both

Young and Old subgroups show a similar level of connectivity. What's more, channels CP5, Fz, and PO7 have more connections with other channels in the HC group; however, these channels seem less active in the PD group. When we compare the Young subgroups for both PD and HC, the overall connections in the HC are significantly stronger than those in the PD.

Major depressive disorder (MDD) is studied using resting-state fMRI data by three proposed methods, CEV-PWA, CEV-PCA, and CEV-PLS. The performance of our approaches is accessed by line figures, network connectivity, and heatmaps for pairwise relationships. We notice that the pairwise relationship between brain region Y and X_j doesn't have much difference for both MDD and HC groups, which is different from our observation in EEG brain channels. Middle temporal gyrus (V12), precuneus cortex (V31), occipital fusiform gyrus (V40), and planum temporale (V46) are shown to WeCon cause the network of other regions in the MDD group, while superior temporal gyrus (V10) performs oppositely, and it achieves the highest score in $g(X)|Y$ in the HC group. We also notice that occipital fusiform gyrus (V40) achieves significantly higher values for all CEV estimators in the HC than the MDD group, indicating stronger connectivity. On the contrary, lateral occipital cortex (V23) and central opercular cortex (V42) have remarkably lower values for all CEV estimators in the HC than the MDD group. Then, when comparing the performance of Female and Male for MDD and HC groups via heatmap, we observe that Male subgroups for both groups contain stronger connectivity in region angular gyrus (V21), lateral occipital cortex (V22), lateral occipital cortex (V23), and intracalcarine cortex (V24) than Female subgroups. The heatmap for the MDD group is darker than the one for the HC group, which suggests the pairwise connectivity in MDD is stronger than in the HC overall. The overall directed network graph shows that the central opercular cortex (V42) is very likely to be passively influenced by other regions in the MDD group, while it's less active in the HC. Then, the lateral occipital cortex (V23) is more likely to WeCon cause other regions in the MDD than its HC performance. Also, V23 always WeCon causes occipital fusiform

gyrus (V40) for both MDD and HC groups. Comparing Female and Male subgroups, we find that the middle temporal gyrus (V12) and superior temporal gyrus (V10) show stronger connectivity in Male than Female for both MDD and HC groups, and the precuneus cortex (V31) seems to be more active in Male than Female for both groups as well. Lateral occipital cortex (V23) always WeCon causes region central opercular cortex (V42) for both females and males in the MDD, which is not obvious in the HC group.

In all three real data applications, we observe that the cumulative explained variability of an individual brain channel or region Y would always be significantly improved after considering the network constructed by other channels or regions in our approaches. This suggests that our proposed methods work well for analyzing EEG and fMRI data. Furthermore, by comparing the performance in both simulation and fMRI data of major depressive disorder, the CEV-PWA method is shown to outperform the alternative approaches, CEV-PCA and CEV-PLS, due to its more generalized nature. Overall, it's remarkable that we can always identify influential brain regions or channels for human visual categorization, Parkinson's disease, and major depressive disorder, leading to a better understanding of the specific causal relationship among brain regions in functionality studies. Furthermore, this research could be meaningful in cognitive science and the diagnosis of neurological disorders, including PD and MDD.

5.3 Future Perspectives

This dissertation mainly focused on computational inference. Our proposed CEV-PWA model can be applied to many research areas. Studying theoretical properties of PWA in the CEV-PWA framework can be a future research direction. Finding more efficient algorithm for solving CEV-PWA is very noteworthy. Biologically studying the discovered brain-functional connectivity can be practically important, and image biomarkers may be identified.

As the real data applications to EEG and fMRI data help us to identify some influential brain channels or regions relating to Parkinson's disease and major depressive disorder, we can further compare the key results with other existing researches for a more comprehensive understanding of the findings. Our EEG study on Parkinson's disease suggests that 12 brain channels (F1, F3, Fp2, Fz, AF7, F8, CP1, CP5, FC3, FT9, P5, PO7) can be characterized by different trends between the PD group and the HC group, while some channels were also identified as influential ones by other researchers. Maggioni et al. (2021) utilized the EEG data to investigate the effective connectivity patterns characterizing PD in the resting state and during music listening. By Granger causality analysis, the right central and frontopolar channels (C4 and Fp2) were stated as important channels which showed significantly different inflow and outflow trend between the PD patients and the HC group. The brain activity abnormalities in the early stage of Parkinson's disease was investigated by Han et al. (2013). They applied the AR Burg method and the wavelet packet entropy (WPE) method for analyzing eyes-closed resting state EEG signals recorded from 15 early-stage PD patients and 15 matched healthy controls. After comparing the relative power and WPE of different brain channels, channels Fp2, Fz, and F3 showed remarkably different fluctuations between the PD and the HC groups. In addition, Handojoseno et al. (2015) researched on the brain dynamic changes associated with freezing of gait (FOG) during turning. They focused on determining the most suitable EEG sensor location to detect freezing of gait during turning. Squared generalized partial directed coherence (sGPDC) causality was used to describe the dynamics of the interactions between brain areas and channels F3 and Fz in frontal regions were emphasized as they showed a decrease of information flow with right region of parietal for PD patients with FOG.

Some other related researches including Almalaq et al. (2015), Leodori et al. (2020), and Cakir (2021) featured the channel C4, which is the motor action channel for left-side movement, due to its high causal effect. This is consistent with our results, as the channel

C4 shows strong connectivity with other brain channels for both PD and HC groups. It's noteworthy that our EEG study on visual categorization also suggests that channel C4 is dominant in causal relation and can actively influence other individual channels as well as the network joint by multiple channels.

The CEV-PWA application on resting-state fMRI data provides 7 influential brain regions (V10, V12, V23, V31, V40, V42, V46) with regard to major depressive disorder. Among these regions, the superior temporal gyrus and precuneus cortex (V10 and V31) were also featured by Bessette et al. (2018). They evaluated whether default mode network (DMN) dysconnectivity in remitted depression (rMDD) is reliable over time and symptom-independent, and explored convergent relationships with cognitive features of depression. The superior temporal gyrus and precuneus cortex were identified as influential regions as they have significantly different connectivity between the rMDD and the HC groups. As Utevsy et al. (2014) stated, precuneus (V31) is a functional core of the default-mode network, and this is also supported by many other MDD researches, including Siddiqi et al. (2022), Liang et al. (2020), Jacob et al. (2020), and Loeffler et al. (2018). In addition, both fusiform gyrus and precuneus (V40 and V31) were characterized by the different performances of brain activity between the MDD and the HC groups in Zhong et al. (2016). The authors applied activation likelihood estimation to integrate studies of resting-state fMRI and explore the alterations of abnormal brain activity in first-episode, drug-naive patients with MDD.

Comparing with the work from other researchers, we have some new findings, while our key results are partially consistent with the existing studies. Our proposed CEV-PWA approach goes beyond the fluctuation of individual brain channels or regions and the pairwise connectivity among them, and takes the network information joint by multiple brain regions or channels into consideration, which leads to a more elaborate and comprehensive analysis than the aforementioned researches. Applying the CEV-PWA method to more EEG and

fMRI data sets regarding to Parkinson's disease and major depressive disorder is an important future research goal, so that we can explore the consistency of influential brain channels and regions under different experiment settings and further identify the image biomarker for these diseases.

Appendix A

Appendix

A.1 Figures for EEG application: Visual Categorization

Figure A.1: Average CEV values for each channel in visual categorization from CEV-PWA (Part I). Channels C3, C4, CB1, CB1r, CB2, CB2r, F3, F4, F7, F8, FP1, FP2, FZ, I, O1 are included.



Figure A.2: Average CEV values for each channel in visual categorization from CEV-PWA (Part II). Channels O1r, O2, O2r, OZ, P3, P4, P4r, PZ, PZr, T3, T4, T5, T5r, T6, T6r are included.



Figure A.3: Average CEV values for each channel except $Y|X$ and $X|Y$ in visual categorization from CEV-PWA (Part I). Channels C3, C4, CB1, CB1r, CB2, CB2r, F3, F4, F7, F8, FP1, FP2, FZ, I, O1 are included.

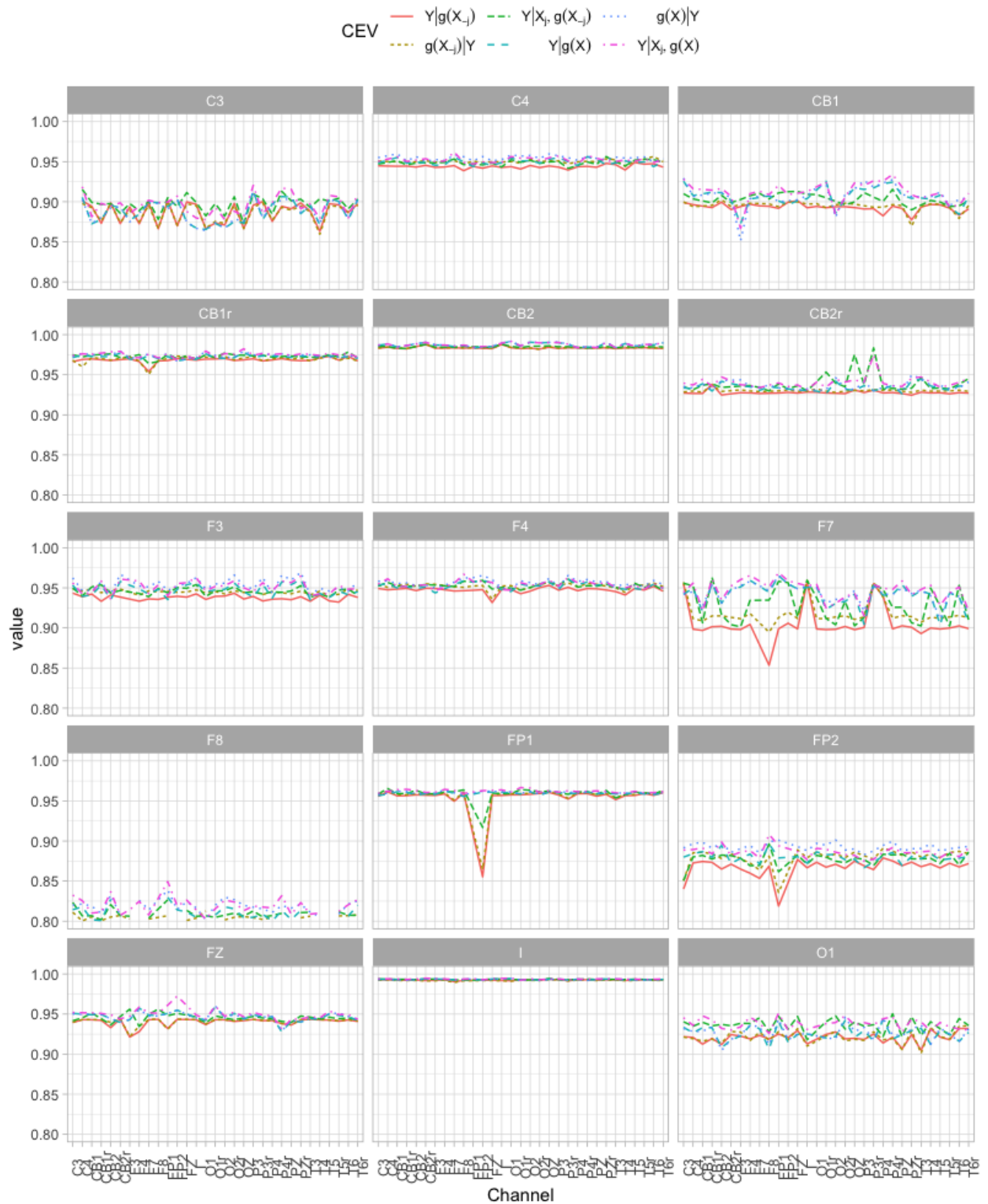
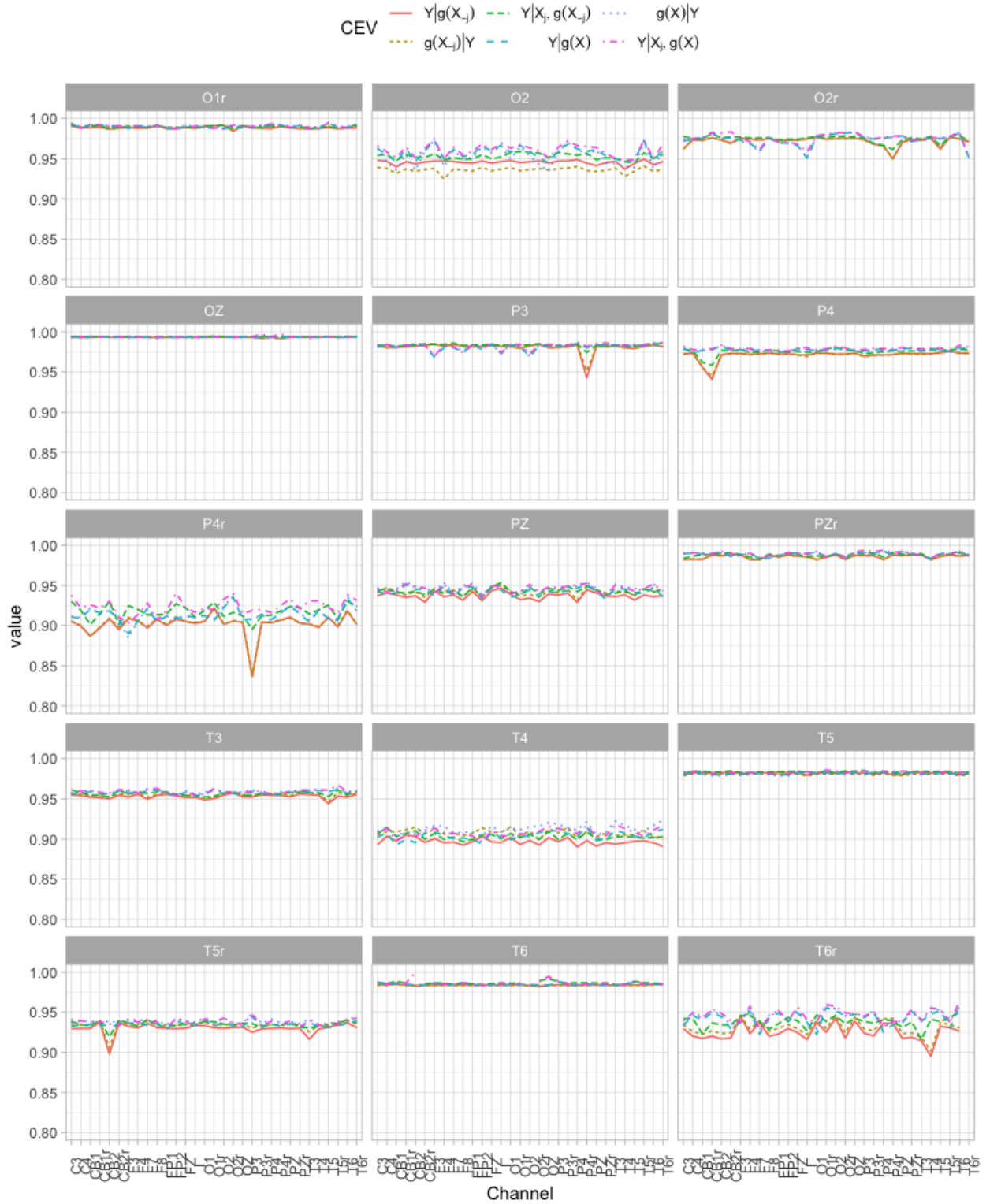


Figure A.4: Average CEV values for each channel except $Y|X$ and $X|Y$ in visual categorization from CEV-PWA (Part II). Channels O1r, O2, O2r, OZ, P3, P4, P4r, PZ, PZr, T3, T4, T5, T5r, T6, T6r are included.



A.2 Figures for EEG application: Parkinson's Disease

Figure A.5: Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.

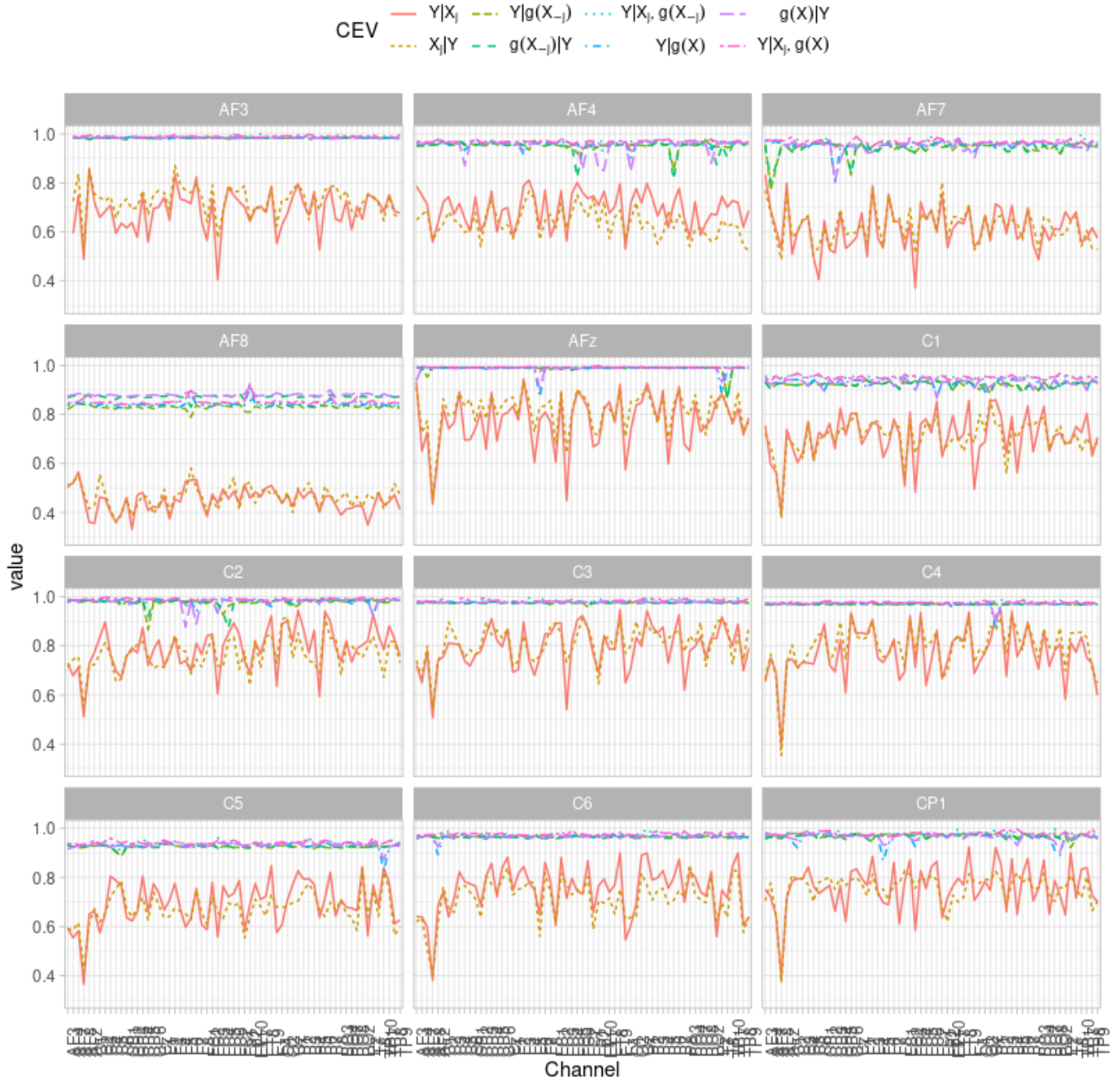


Figure A.6: Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.

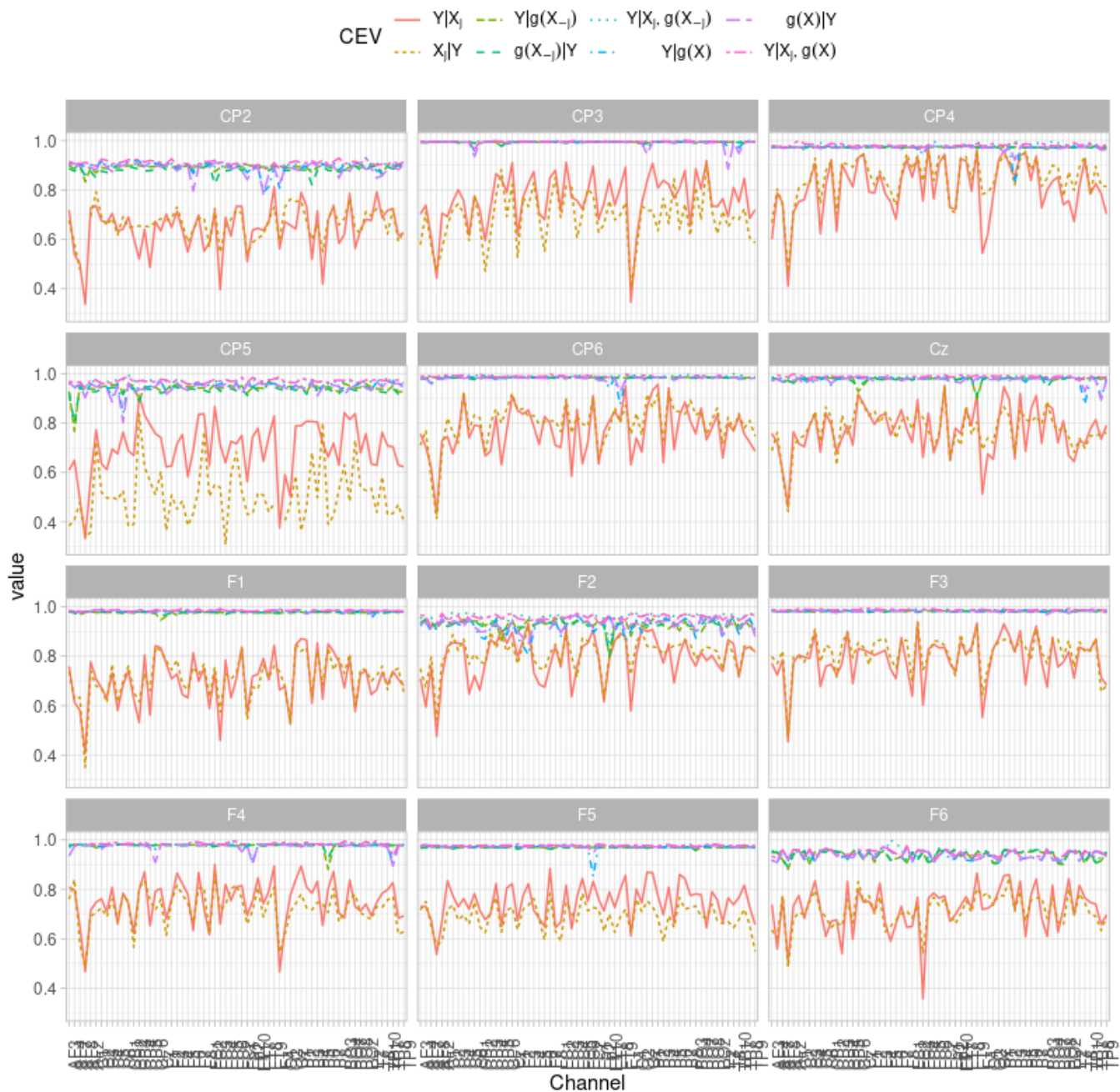


Figure A.7: Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.

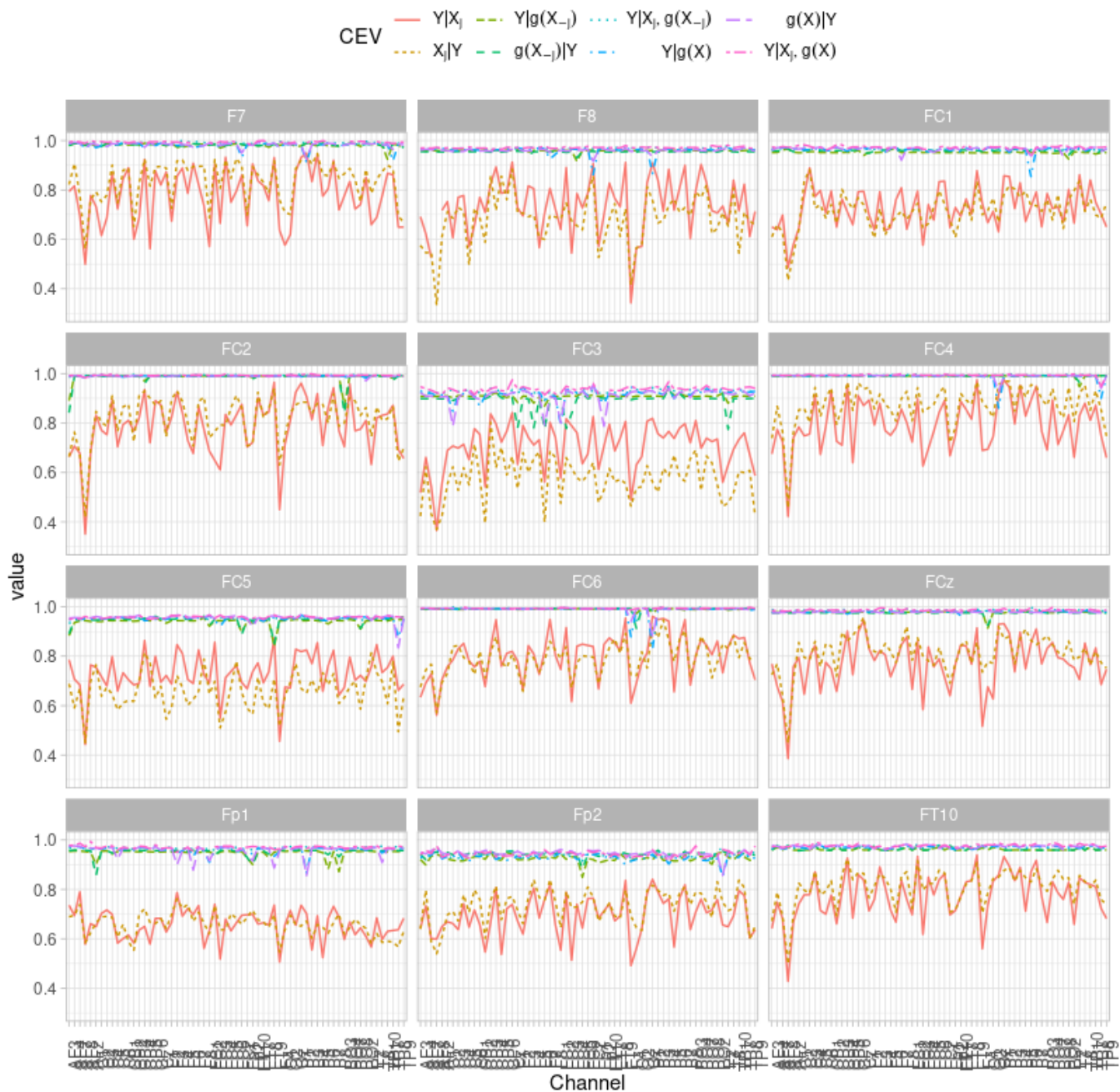


Figure A.8: Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.



Figure A.9: Average CEV values for each channel in the Parkinson's patients group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.

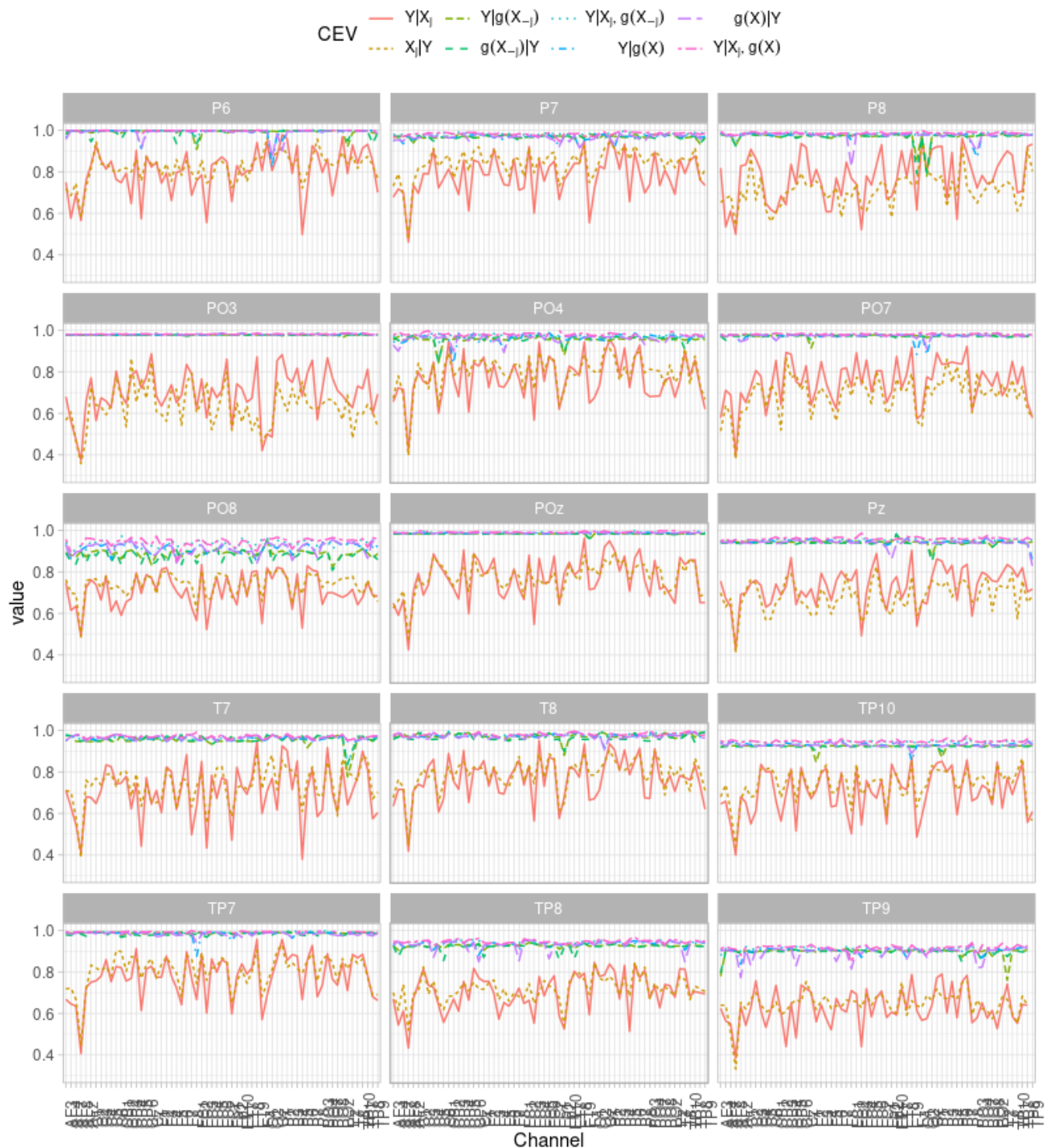


Figure A.10: Average CEV values for each channel in the healthy control group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.



Figure A.11: Average CEV values for each channel in the healthy control group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.

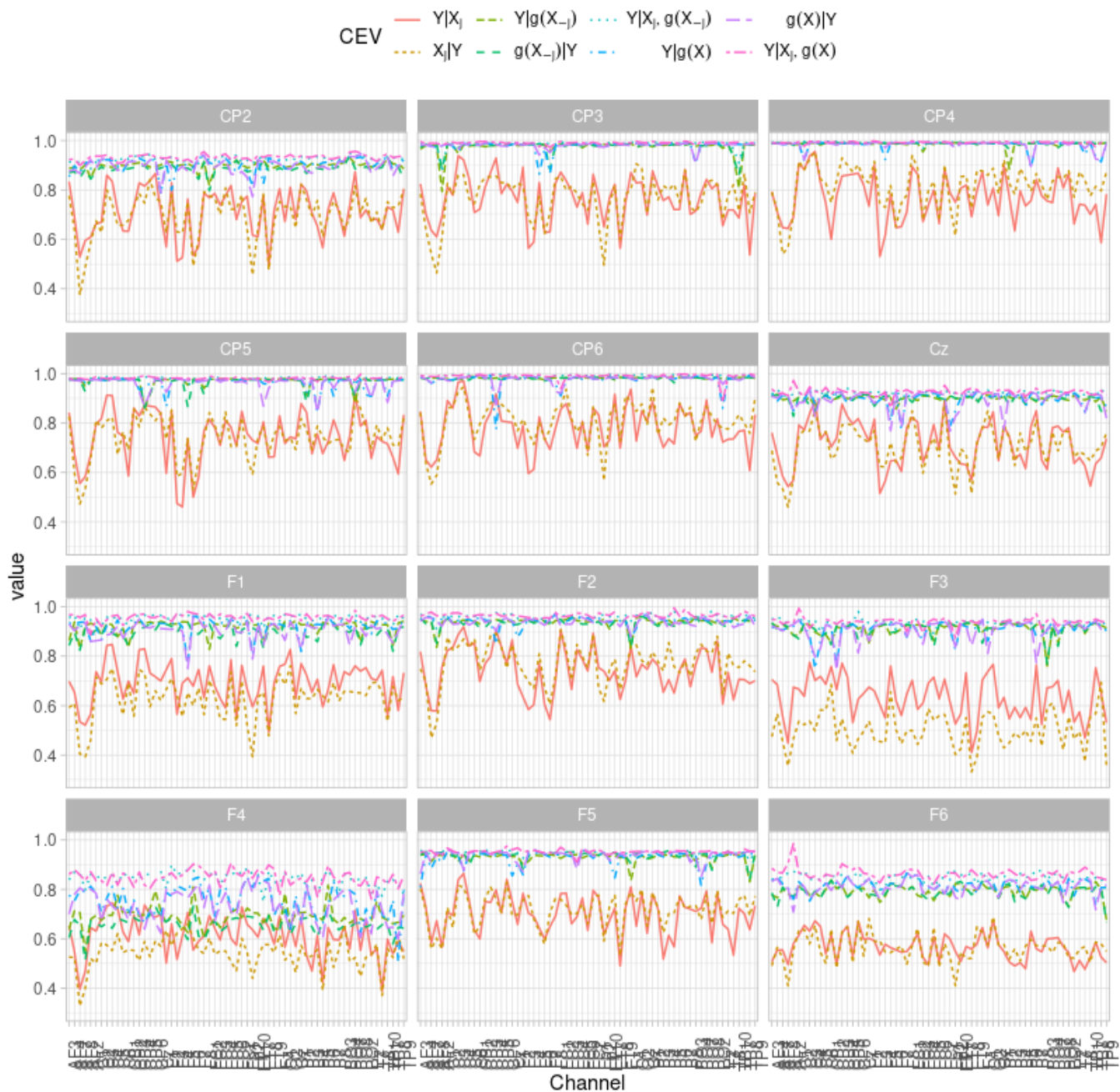


Figure A.12: Average CEV values for each channel in the healthy control group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.

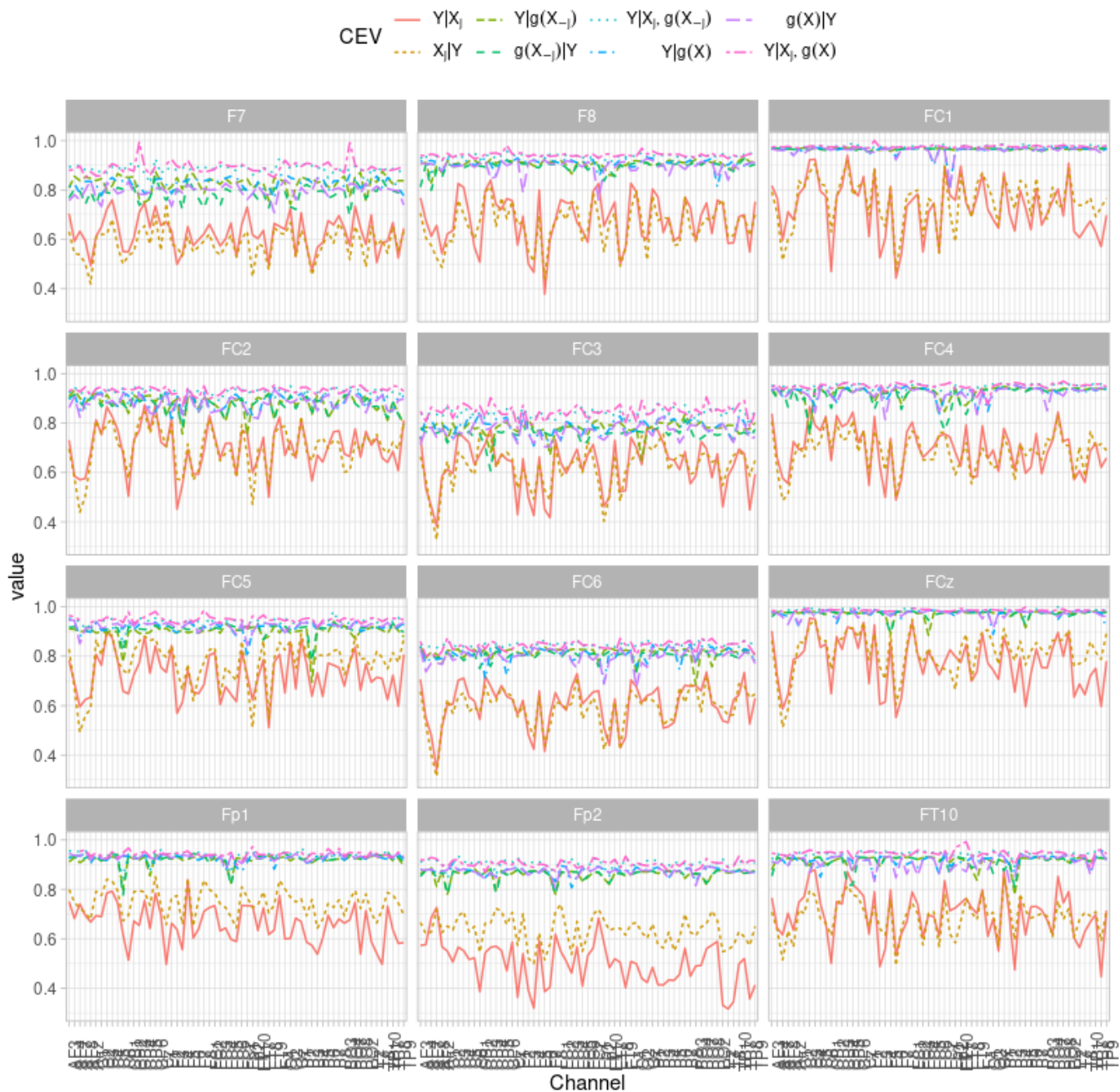


Figure A.13: Average CEV values for each channel in the healthy control group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.

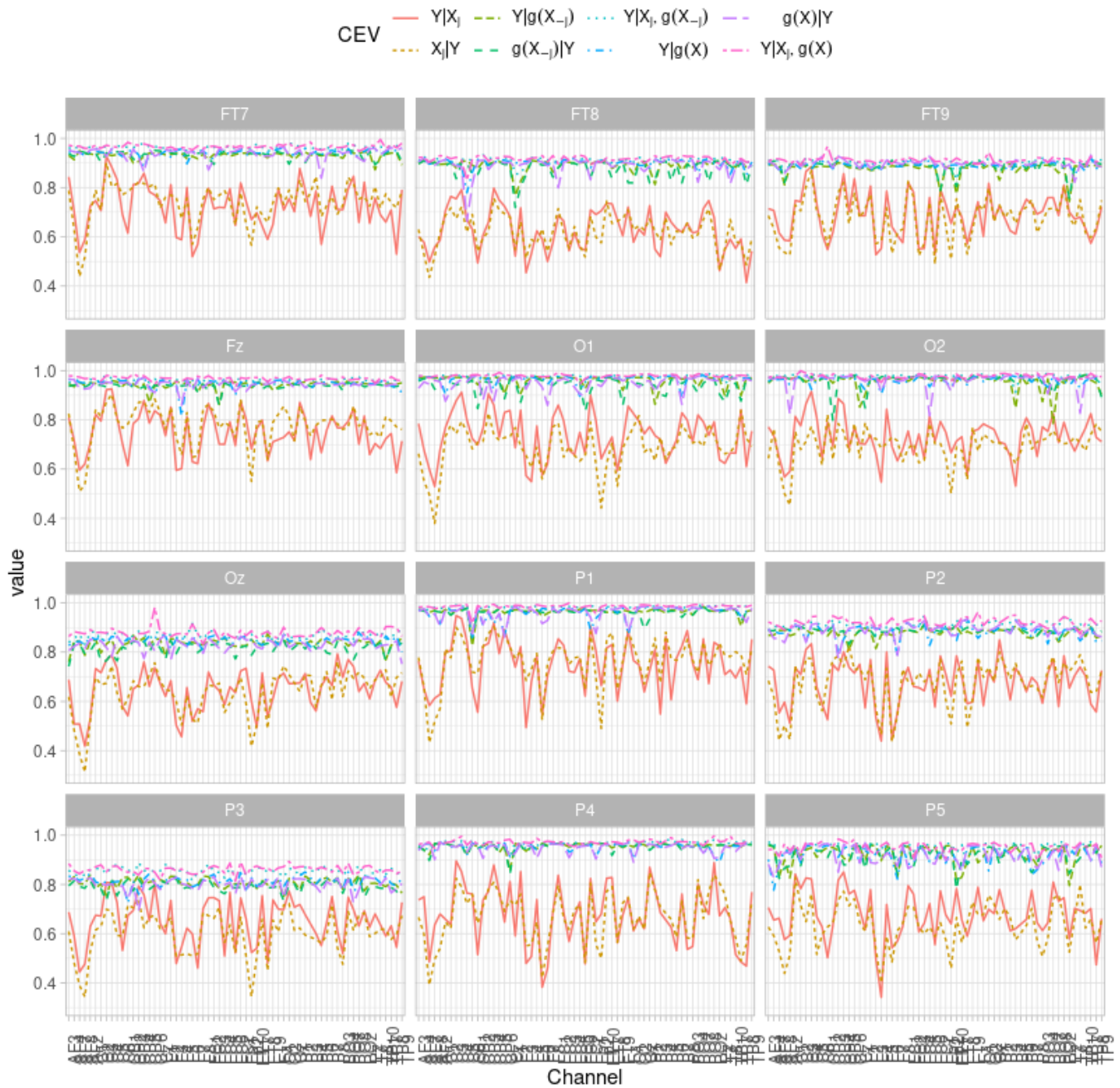


Figure A.14: Average CEV values for each channel in the healthy control group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.



Figure A.15: Average CEV values for each channel except $Y|X$ and $X|Y$ in the Parkinson's patients group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.

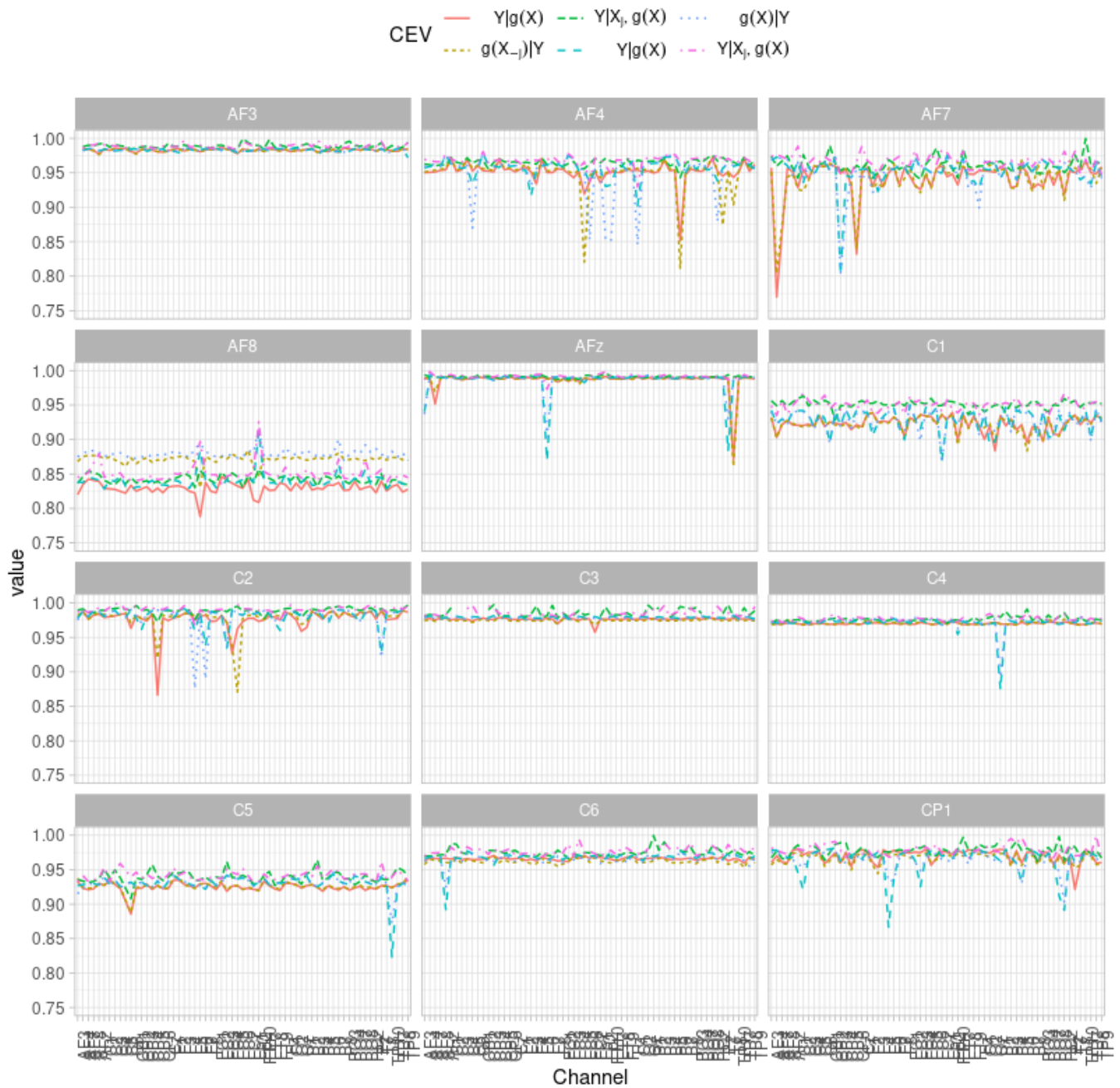


Figure A.16: Average CEV values for each channel except $Y|X$ and $X|Y$ in the Parkinson's patients group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.



Figure A.17: Average CEV values for each channel except $Y|X$ and $X|Y$ in the Parkinson's patients group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.



Figure A.18: Average CEV values for each channel except $Y|X$ and $X|Y$ in the Parkinson's patients group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.



Figure A.19: Average CEV values for each channel except $Y|X$ and $X|Y$ in the Parkinson's patients group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.



Figure A.20: Average CEV values for each channel except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part I). Channels AF3, AF4, AF6, AF8, AFz, C1, C2, C3, C4, C5, C6, CP1 are included.



Figure A.21: Average CEV values for each channel except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part II). Channels CP2, CP3, CP4, CP5, CP6, Cz, F1, F2, F3, F4, F5, F6 are included.

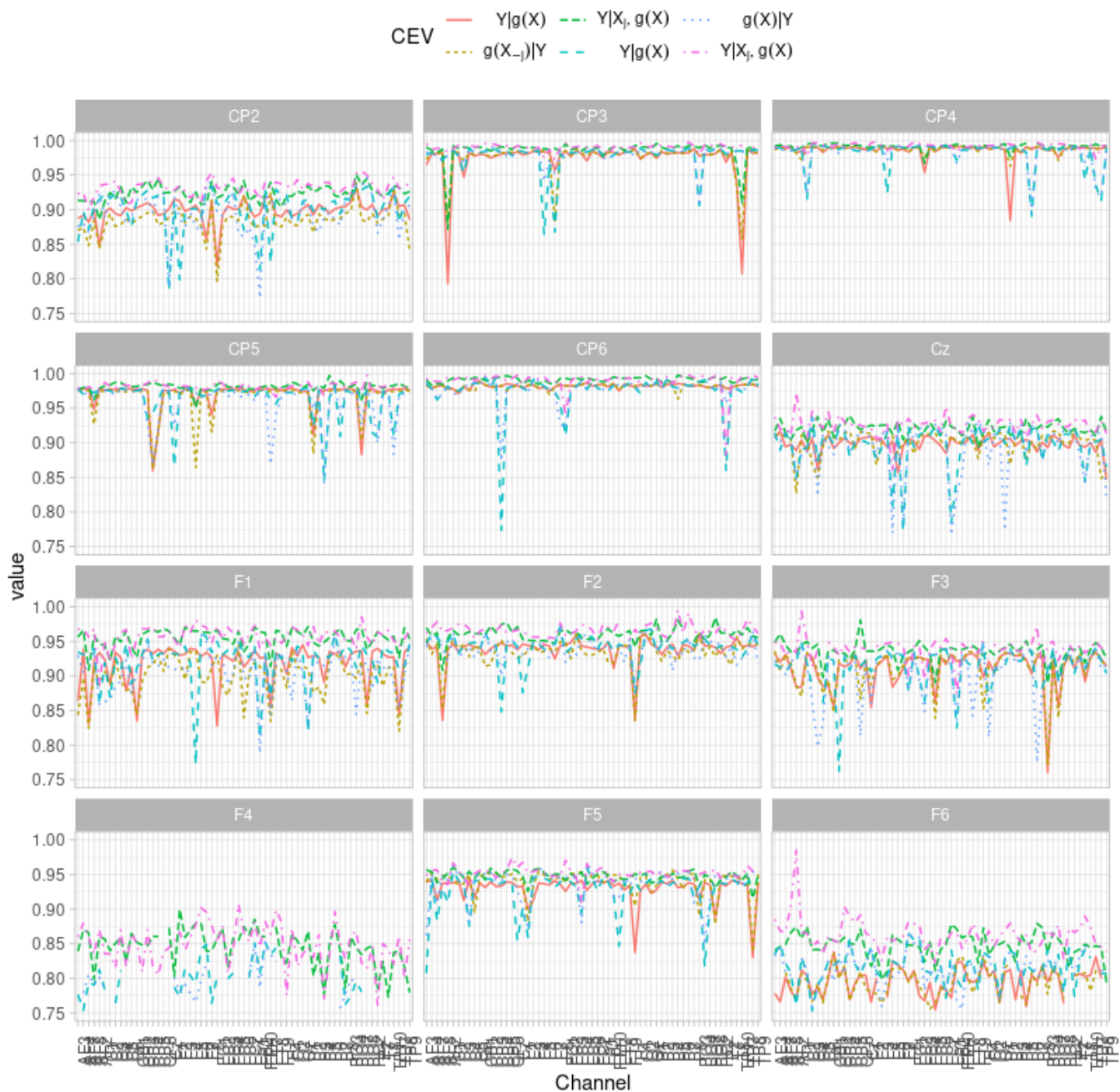


Figure A.22: Average CEV values for each channel except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part III). Channels F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, Fp1, Fp2, FT10 are included.



Figure A.23: Average CEV values for each channel except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part IV). Channels FT7, FT8, FT9, Fz, O1, O2, Oz, P1, P2, P3, P4, P5 are included.



Figure A.24: Average CEV values for each channel except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part V). Channels P6, P7, P8, PO3, PO4, PO7, PO8, POz, Pz, T7, T8, TP10, TP7, TP8, TP9 are included.



A.3 Figures for fMRI application: Major Depressive Disorder

Figure A.25: Average CEV values for each region. in the major depressive disorder group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.



Figure A.26: Average CEV values for each region in the major depressive disorder group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.



Figure A.27: Average CEV values for each region in the major depressive disorder group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.



Figure A.28: Average CEV values for each region in the major depressive disorder group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.



Figure A.29: Average CEV values for each region in the healthy control group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.

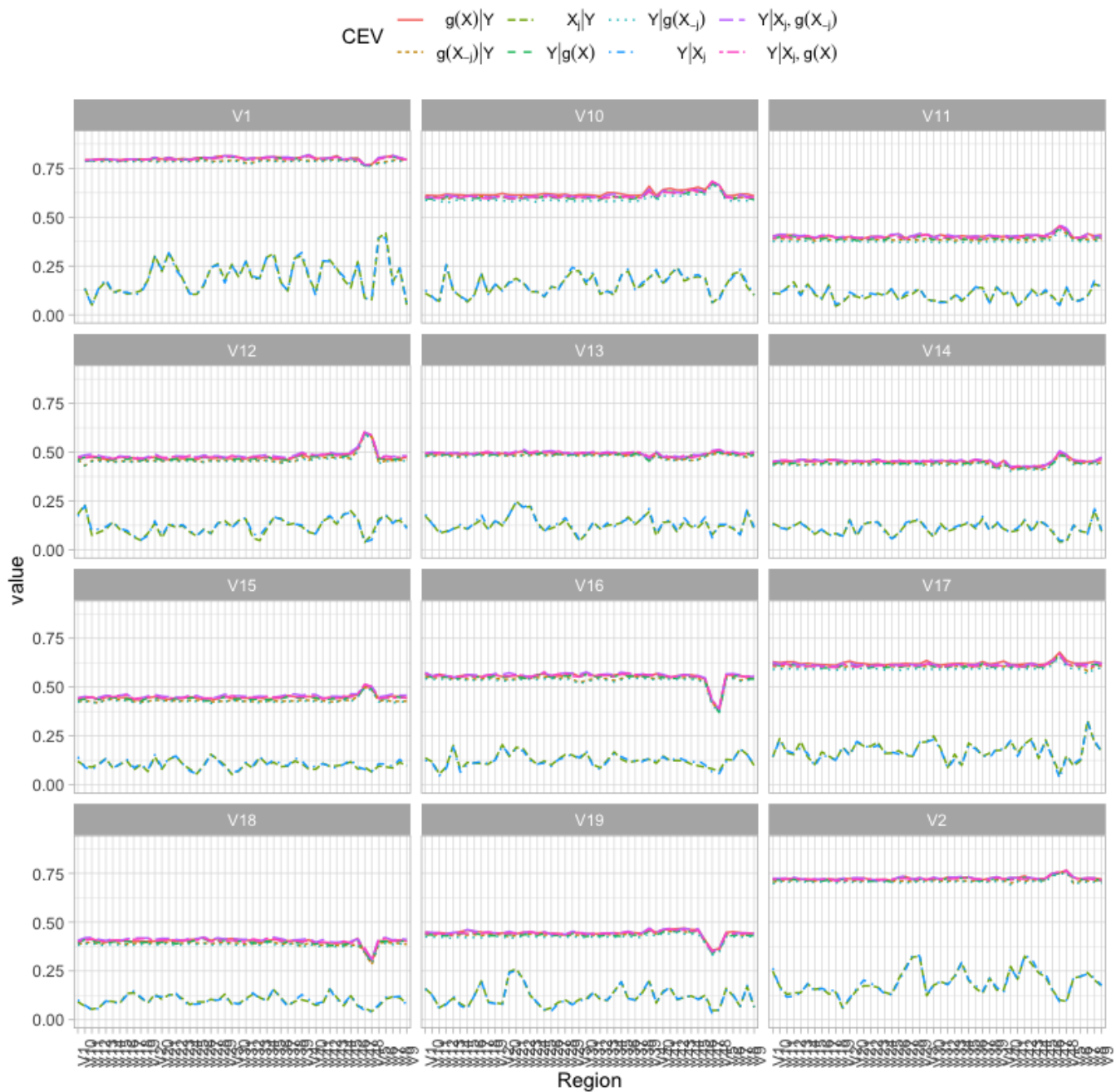


Figure A.30: Average CEV values for each region in the healthy control group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.



Figure A.31: Average CEV values for each region in the healthy control group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.



Figure A.32: Average CEV values for each region in the healthy control group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.



Figure A.33: Average CEV values for each region except $Y|X$ and $X|Y$ in the major depressive disorder group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.

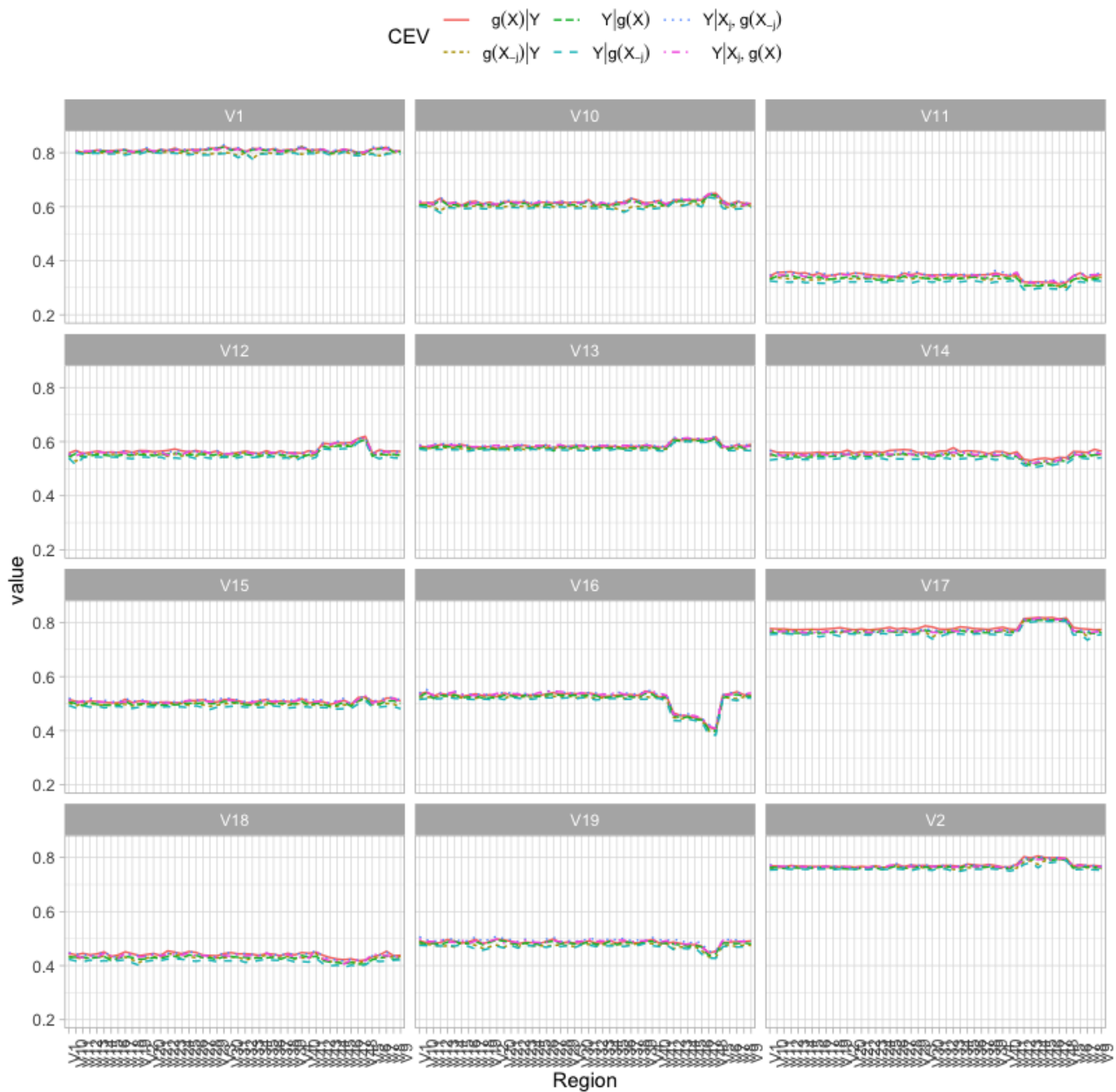


Figure A.34: Average CEV values for each region except $Y|X$ and $X|Y$ in the major depressive disorder group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.

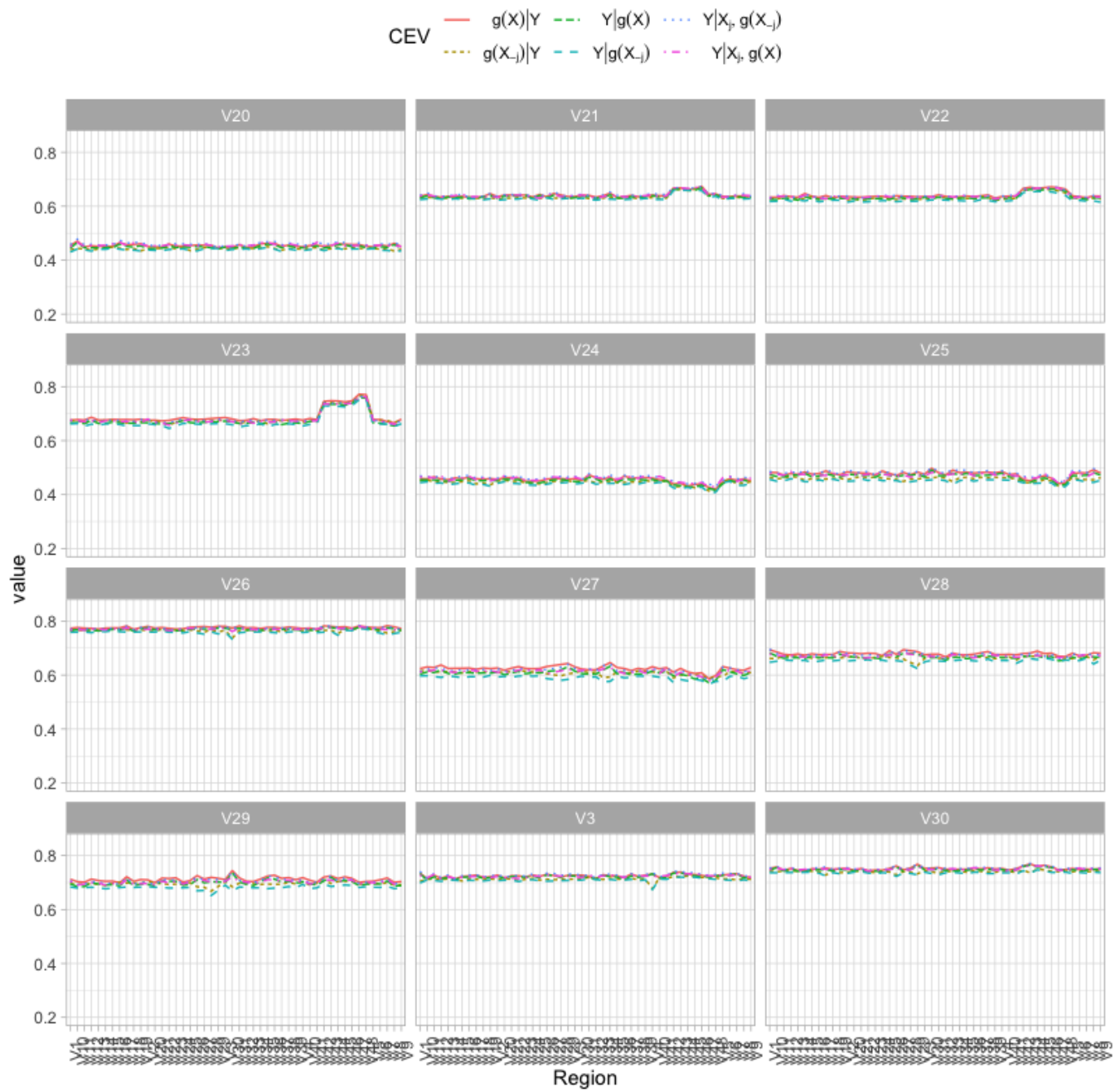


Figure A.35: Average CEV values for each region except $Y|X$ and $X|Y$ in the major depressive disorder group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.

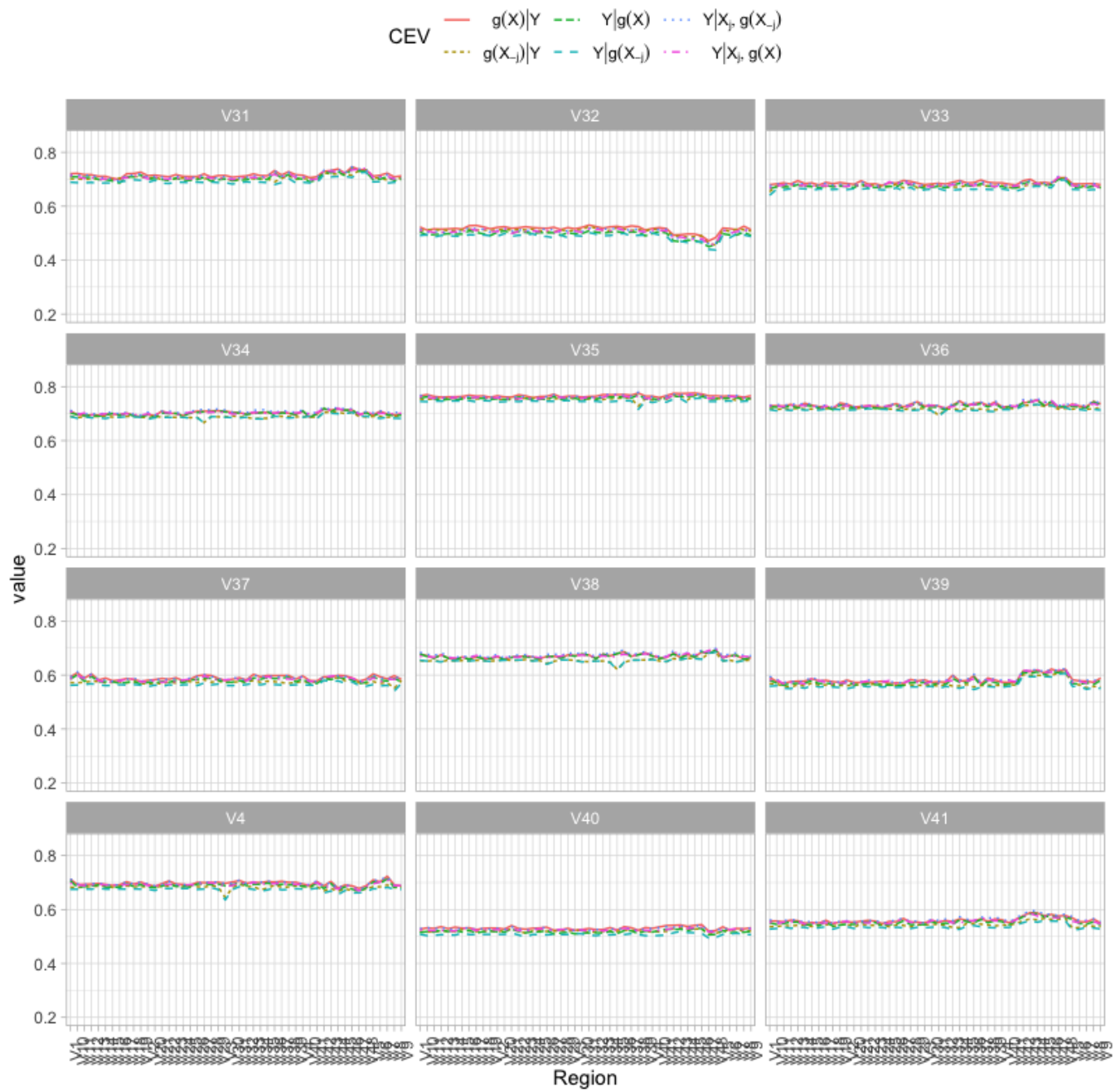


Figure A.36: Average CEV values for each region except $Y|X$ and $X|Y$ in the major depressive disorder group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.

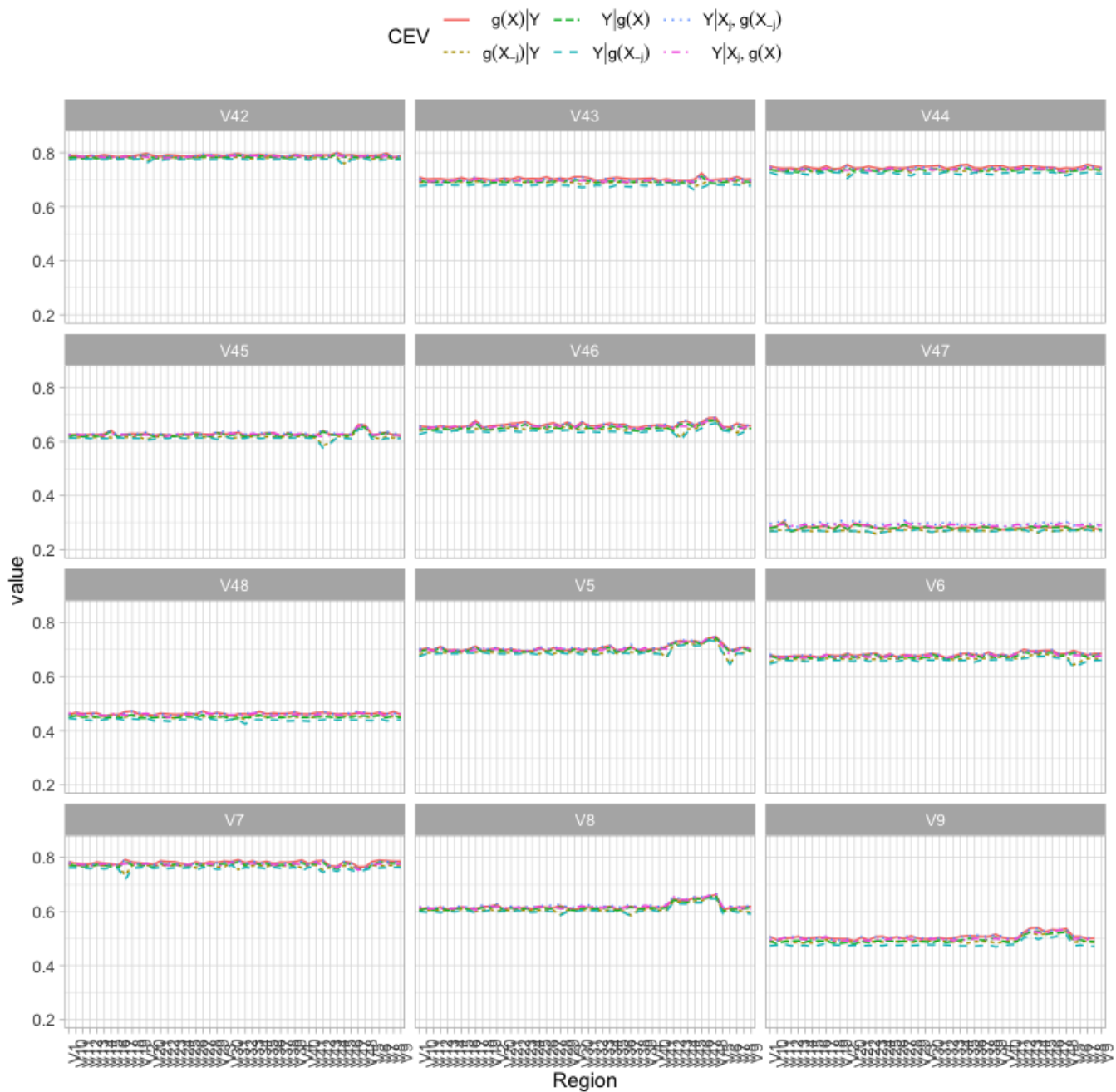


Figure A.37: Average CEV values for each region except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.

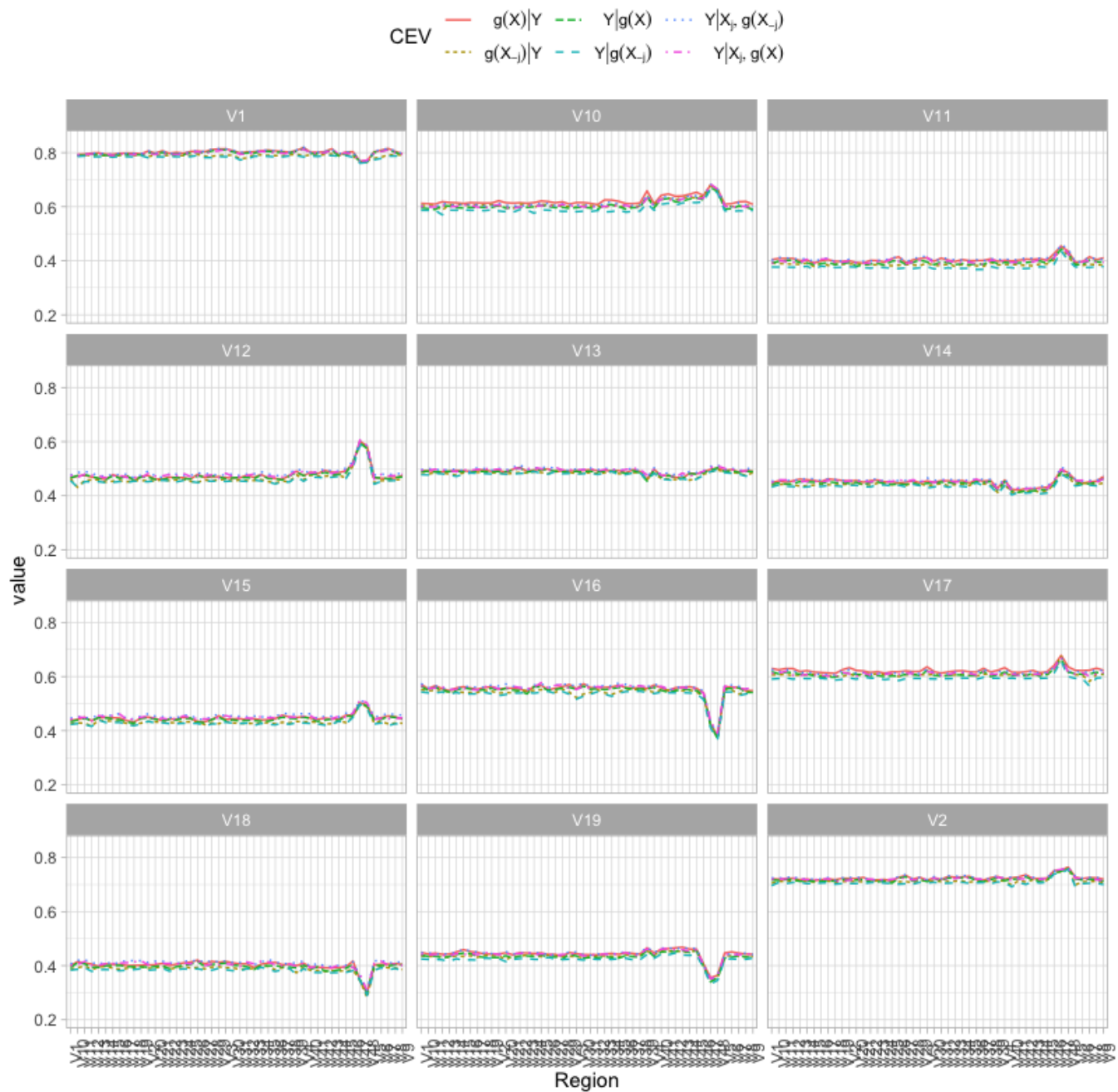


Figure A.38: Average CEV values for each region except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.

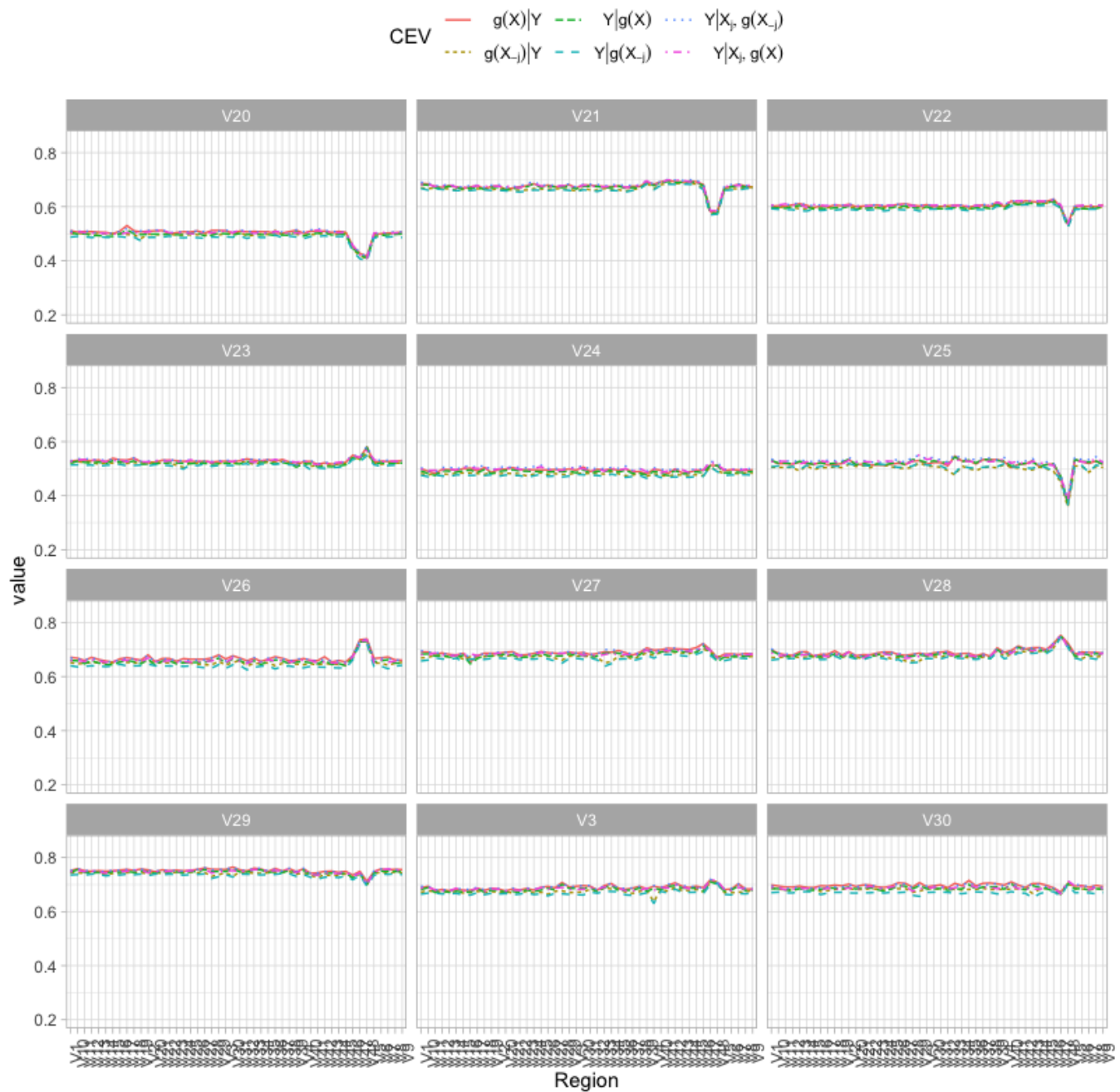


Figure A.39: Average CEV values for each region except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.

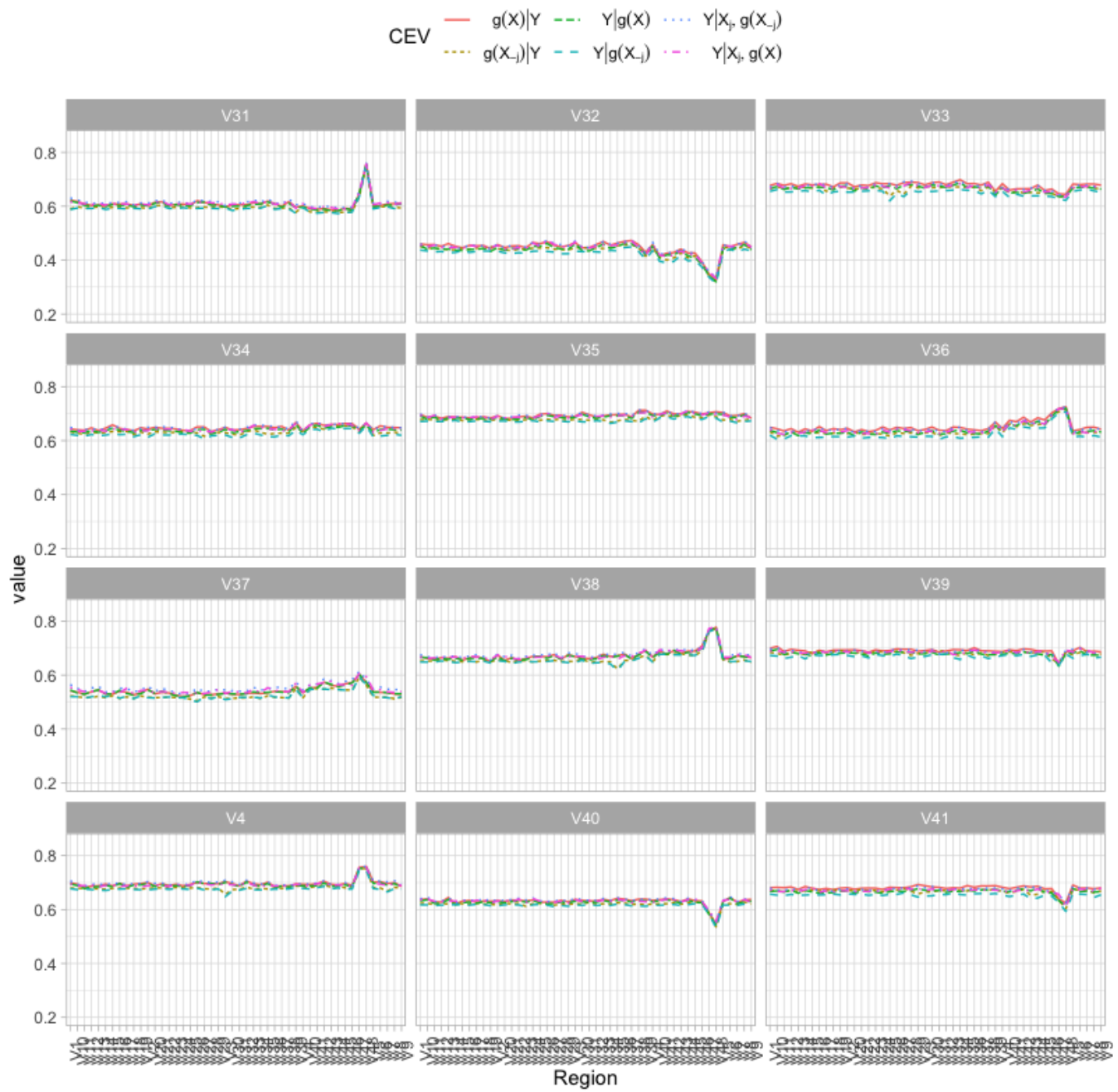


Figure A.40: Average CEV values for each region except $Y|X$ and $X|Y$ in the healthy control group from CEV-PWA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.

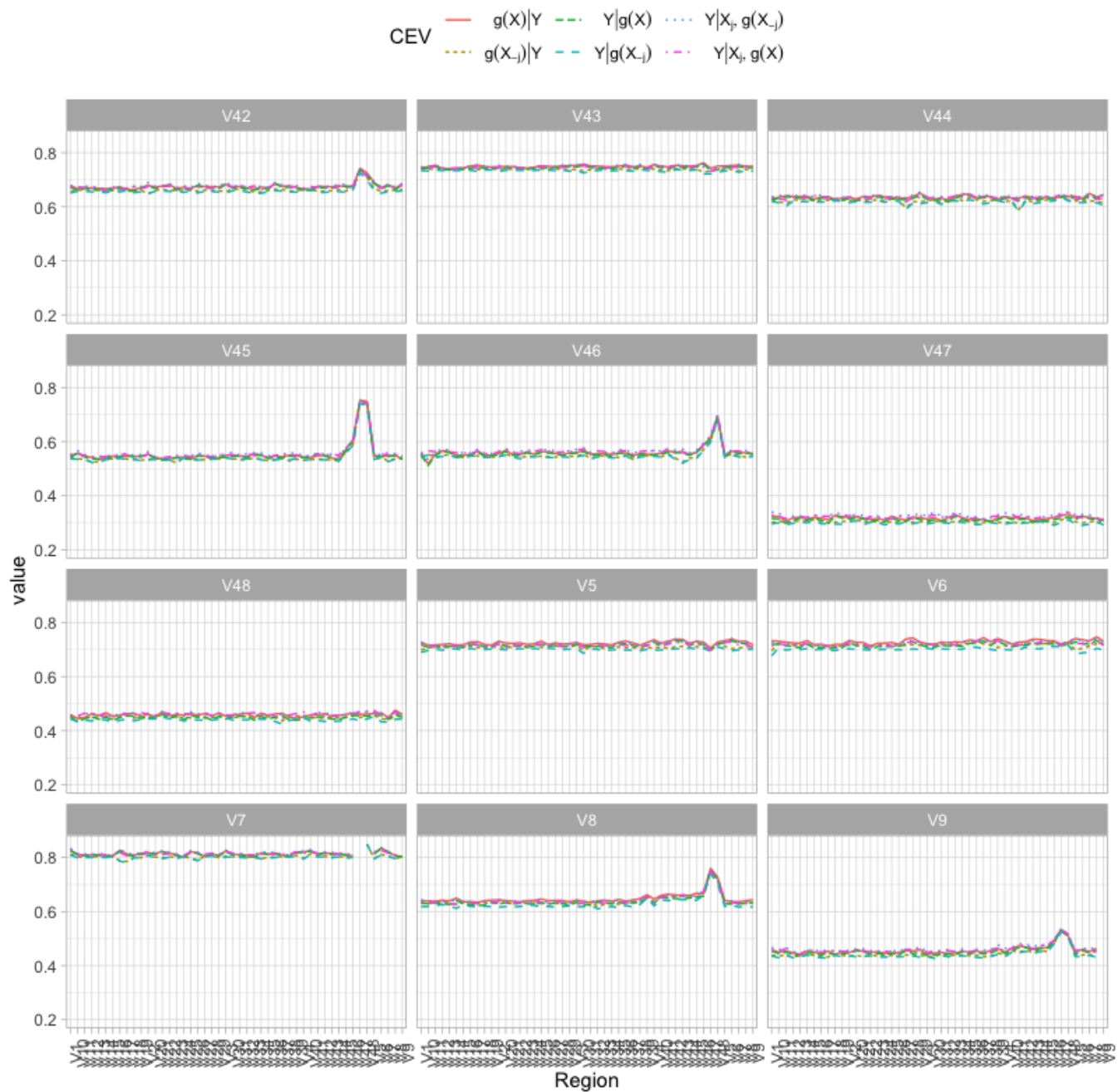


Figure A.41: Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.



Figure A.42: Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.



Figure A.43: Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.



Figure A.44: Average CEV values for each region in the major depressive disorder group from CEV-PCA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.



Figure A.45: Average CEV values for each region in the healthy control group from CEV-PCA (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.



Figure A.46: Average CEV values for each region in the healthy control group from CEV-PCA (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.



Figure A.47: Average CEV values for each region in the healthy control group from CEV-PCA (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.



Figure A.48: Average CEV values for each region in the healthy control group from CEV-PCA (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.



Figure A.49: Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.



Figure A.50: Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.



Figure A.51: Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.



Figure A.52: Average CEV values for each region in the major depressive disorder group from CEV-PLS (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.



Figure A.53: Average CEV values for each region in the healthy control group from CEV-PLS (Part I). Regions V1, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V2 are included.



Figure A.54: Average CEV values for each region in the healthy control group from CEV-PLS (Part II). Regions V20, V21, V22, V23, V24, V25, V26, V27, V28, V29, V3, V30 are included.



Figure A.55: Average CEV values for each region in the healthy control group from CEV-PLS (Part III). Regions V31, V32, V33, V34, V35, V36, V37, V38, V39, V4, V40, V41 are included.



Figure A.56: Average CEV values for each region in the healthy control group from CEV-PLS (Part IV). Regions V42, V43, V44, V45, V46, V47, V48, V5, V6, V7, V8, V9 are included.



Bibliography

- Allen, D. E. and Hooper, V. (2018). Generalized correlation measures of causality and forecasts of the vix using non-linear models. *Sustainability*, 10(8):2695.
- Allen, D. E. and McAleer, M. (2020). “generalized measures of correlation for asymmetry, nonlinearity, and beyond”: Some antecedents on causality. *Journal of the American Statistical Association*, pages 1–11.
- Almalaq, A., Dai, X., Zhang, J., Hanrahan, S., Nedrud, J., and Hebb, A. (2015). Causality graph learning on cortical information flow in parkinson’s disease patients during behaviour tests. In *2015 49th Asilomar Conference on Signals, Systems and Computers*, pages 925–929. IEEE.
- Arasteh, E., Mahdizadeh, A., Mirian, M. S., Lee, S., and McKeown, M. J. (2021). Deep transfer learning for parkinson’s disease monitoring by image-based representation of resting-state eeg using directional connectivity. *Algorithms*, 15(1):5.
- Asher, E. E., Plotnik, M., Günther, M., Moshel, S., Levy, O., Havlin, S., Kantelhardt, J. W., and Bartsch, R. P. (2021). Connectivity of eeg synchronization networks increases for parkinson’s disease patients with freezing of gait. *Communications biology*, 4(1):1–10.
- Badwan, B. A., Creamer, M. S., Zavatone-Veth, J. A., and Clark, D. A. (2019). Dynamic nonlinearities enable direction opponency in drosophila elementary motion detectors. *Nature neuroscience*, 22(8):1318–1326.

- Bessette, K. L., Jenkins, L. M., Skerrett, K. A., Gowins, J. R., DelDonno, S. R., Zubieta, J.-K., McInnis, M. G., Jacobs, R. H., Ajilore, O., and Langenecker, S. A. (2018). Reliability, convergent validity and time invariance of default mode network deviations in early adult major depressive disorder. *Frontiers in psychiatry*, 9:244.
- Bezmaternykh, ME, M., AA, S., and ED, P. (2020). "resting state with closed eyes for patients with depression and healthy participants".
- Bhurane, A. A., Dhok, S., Sharma, M., Yuvaraj, R., Murugappan, M., and Acharya, U. R. (2019). Diagnosis of parkinson's disease from electroencephalography signals using linear and self-similarity features. *Expert Systems*, page e12472.
- Bielczyk, N. Z., Uithol, S., van Mourik, T., Anderson, P., Glennon, J. C., and Buitelaar, J. K. (2019). Disentangling causal webs in the brain using functional magnetic resonance imaging: A review of current approaches. *Network Neuroscience*, 3(2):237–273.
- Botvinik-Nezer, R., Holzmeister, F., and Camerer, C. t. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582:84 – 88.
- Cakir, Y. (2021). Computational neuronal correlation with enhanced synchronized activity in the basal ganglia and the slowing of thalamic theta and alpha rhythms in parkinson's disease. *European Journal of Neuroscience*, 54(3):5203–5223.
- Chen, M., Lian, Y., Chen, Z., and Zhang, Z. (2017). Sure explained variability and independence screening. *Journal of Nonparametric Statistics*, 29(4):849–883.
- Cooper, J. A., Sagar, H. J., Jordan, N., Harvey, N. S., and Sullivan, E. V. (1991). Cognitive impairment in early, untreated parkinson's disease and its relationship to motor disability. *Brain*, 114(5):2095–2122.
- Cui, Q., Xu, Y., Zhang, Z., and Chan, V. (2021). Max-linear regression models with regularization. *Journal of Econometrics*, 222:579–600.

- Cui, Q. and Zhang, Z. (2018). Max-linear competing factor models. *Journal of Business & Economic Statistics*, 36(1):62–74.
- Cui, Y., Pang, J.-S., and Sen, B. (2018). Composite difference-max programs for modern statistical estimation problems. *SIAM Journal on Optimization*, 28(4):3344–3374.
- Delorme, A. (2020). "go-nogo categorization and detection task".
- Doksum, K. and Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics*, pages 1443–1473.
- Dong, M., Xia, L., Lu, M., Li, C., Xu, K., and Zhang, L. (2019). A failed top-down control from the prefrontal cortex to the amygdala in generalized anxiety disorder: evidence from resting-state fmri with granger causality analysis. *Neuroscience Letters*, 707:134314.
- Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911.
- Formaggio, E., Rubega, M., Rupil, J., Antonini, A., Masiero, S., Toffolo, G. M., and Del Felice, A. (2021). Reduced effective connectivity in the motor cortex in parkinson’s disease. *Brain Sciences*, 11(9):1200.
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36.
- Friston, K. J., Li, B., Daunizeau, J., and Stephan, K. E. (2011). Network discovery with dcm. *Neuroimage*, 56(3):1202–1221.
- Gerlach, R. W., Kowalski, B. R., and Wold, H. O. (1979). Partial least-squares path modelling with latent variables. *Analytica Chimica Acta*, 112(4):417–421.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438.

- Grillner, S., Ip, N., Koch, C., Koroshetz, W., Okano, H., Polachek, M., Poo, M. M., and Sejnowski, T. J. (2016). Worldwide initiatives to advance brain research. *Nature neuroscience*, 19(9):1118–1122.
- Gunduz, H. (2019). Deep learning-based parkinson’s disease classification using vocal feature sets. *IEEE Access*, 7:115540–115551.
- Hahn, G., Banerjee, M., and Sen, B. (2017). Parameter estimation and inference in a continuous piecewise linear regression model. *Manuscript, Department of Statistics, Columbia University (December 2016)*, 32(2):407–451.
- Han, C.-X., Wang, J., Yi, G.-S., and Che, Y.-Q. (2013). Investigation of eeg abnormalities in the early stage of parkinson’s disease. *Cognitive neurodynamics*, 7(4):351–359.
- Handojoseno, A. A., Gilat, M., Ly, Q. T., Chamtie, H., Shine, J. M., Nguyen, T. N., Tran, Y., Lewis, S. J., and Nguyen, H. T. (2015). An eeg study of turning freeze in parkinson’s disease patients: the alteration of brain dynamic on the motor and visual cortex. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6618–6621. IEEE.
- Hannah, L. A. and Dunson, D. B. (2011). Bayesian nonparametric multivariate convex regression. *arXiv preprint arXiv:1109.0322*.
- Hannah, L. A. and Dunson, D. B. (2013). Multivariate convex regression with adaptive partitioning. *The Journal of Machine Learning Research*, 14(1):3261–3294.
- Hempel, A. B., Goulart, P. J., and Lygeros, J. (2013). Every continuous piecewise affine function can be obtained by solving a parametric linear program. In *2013 European Control Conference (ECC)*, pages 2657–2662. IEEE.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.

- Jacob, Y., Morris, L. S., Huang, K.-H., Schneider, M., Rutter, S., Verma, G., Murrough, J. W., and Balchandani, P. (2020). Neural correlates of rumination in major depressive disorder: a brain network analysis. *NeuroImage: Clinical*, 25:102142.
- James F Cavanagh, j. (2021). "eeg: 3-stim auditory oddball and rest in parkinson's".
- Jiang, Y., Tian, Y., and Wang, Z. (2019). Causal interactions in human amygdala cortical networks across the lifespan. *Scientific reports*, 9(1):1–11.
- John H. Holland, Sampson, J. R. (1976). Adaptation in natural and artificial systems (john h. holland).
- Kandilarova, S., Stoyanov, D., Kostianev, S., and Specht, K. (2018). Altered resting state effective connectivity of anterior insula in depression. *Frontiers in psychiatry*, 9:83.
- Kendall, M. G. et al. (1946). The advanced theory of statistics. *The advanced theory of statistics.*, (2nd Ed).
- Khare, S. K., Bajaj, V., and Acharya, U. R. (2021). Pdcnnnet: An automatic framework for the detection of parkinson's disease using eeg signals. *IEEE Sensors Journal*, 21(15):17017–17024.
- Kripfganz, A. and Schulze, R. (1987). Piecewise affine functions as a difference of two convex functions. *Optimization*, 18(1):23–29.
- Leodori, G., Belvisi, D., De Bartolo, M. I., Fabbrini, A., Costanzo, M., Vial, F., Conte, A., Hallett, M., and Berardelli, A. (2020). Re-emergent tremor in parkinson's disease: the role of the motor cortex. *Movement Disorders*, 35(6):1002–1011.
- Li, G., Liu, Y., Zheng, Y., Li, D., Liang, X., Chen, Y., Cui, Y., Yap, P.-T., Qiu, S., Zhang, H., et al. (2020). Large-scale dynamic causal modeling of major depressive disorder based on resting-state functional magnetic resonance imaging. *Human brain mapping*, 41(4):865–881.

- Li, L., Li, B., Bai, Y., Liu, W., Wang, H., Leung, H.-C., Tian, P., Zhang, L., Guo, F., Cui, L.-B., et al. (2017). Abnormal resting state effective connectivity within the default mode network in major depressive disorder: a spectral dynamic causal modeling study. *Brain and behavior*, 7(7):e00732.
- Li, R., Zhong, W., and Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499):1129–1139.
- Liang, S., Deng, W., Li, X., Greenshaw, A. J., Wang, Q., Li, M., Ma, X., Bai, T.-J., Bo, Q.-J., Cao, J., et al. (2020). Biotypes of major depressive disorder: Neuroimaging evidence from resting-state default mode network patterns. *NeuroImage: Clinical*, 28:102514.
- Loeffler, L. A., Radke, S., Habel, U., Ciric, R., Satterthwaite, T. D., Schneider, F., and Derntl, B. (2018). The regulation of positive and negative emotions through instructed causal attributions in lifetime depression—a functional magnetic resonance imaging study. *NeuroImage: Clinical*, 20:1233–1245.
- López-Madróna, V. J., Matias, F. S., Mirasso, C. R., Canals, S., and Pereda, E. (2019). Inferring correlations associated to causal interactions in brain signals using autoregressive models. *Scientific reports*, 9(1):1–15.
- Maggioni, E., Arienti, F., Minella, S., Mameli, F., Borellini, L., Nigro, M., Cogiamanian, F., Bianchi, A. M., Cerutti, S., Barbieri, S., et al. (2021). Effective connectivity during rest and music listening: An eeg study on parkinson’s disease. *Frontiers in aging neuroscience*, 13:657221.
- Mazumder, R., Choudhury, A., Iyengar, G., and Sen, B. (2019). A computational framework for multivariate convex regression and its variants. *Journal of the American Statistical Association*, 114(525):318–331.
- Miao, C., Zhang, Z., and Zhou, Y. (2018). Simultaneous screening and cumulatively explained variability screening. *Manuscript*.

- Młyńczak, M. and Krysztofiak, H. (2018). Discovery of causal paths in cardiorespiratory parameters: a time-independent approach in elite athletes. *Frontiers in physiology*, 9:1455.
- Młyńczak, M. and Krysztofiak, H. (2019). Cardiorespiratory temporal causal links and the differences by sport or lack thereof. *Frontiers in physiology*, 10:45.
- Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Icml*.
- Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., and Acharya, U. R. (2020). A deep learning approach for parkinson’s disease diagnosis from eeg signals. *Neural Computing and Applications*, 32(15):10927–10933.
- Pearson, K., Lee, A., and Bramley-Moore, L. (1899). Vi. mathematical contributions to the theory of evolution.—vi. genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, (192):257–330.
- Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., and Papa, J. P. (2016). Deep learning-aided parkinson’s disease diagnosis from handwritten dynamics. In *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 340–346. Ieee.
- Pyatigorskaya, N., Yahia-Cherif, L., Valabregue, R., Gaurav, R., Gargouri, F., Ewencyk, C., Gallea, C., Fernandez-Vidal, S., Arnulf, I., Vidailhet, M., et al. (2021). Parkinson disease propagation using mri biomarkers and partial least squares path modeling. *Neurology*, 96(3):e460–e471.
- Rényi, A. (1959). On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(3-4):441–451.
- Richardson, T. S. and Robins, J. M. (2013). Single world intervention graphs (swigs):a

- unification of the counterfactual and graphical approaches to causality. *Working paper: University of Washington*, 128:1–148.
- Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The annals of statistics*, pages 1346–1370.
- Savitt, J. M., Dawson, V. L., Dawson, T. M., et al. (2006). Diagnosis and treatment of parkinson disease: molecules to medicine. *The Journal of clinical investigation*, 116(7):1744–1754.
- Scholtes, S. (2012). *Introduction to piecewise differentiable equations*. Springer Science & Business Media.
- Sen, B. and Parhi, K. K. (2019). Predicting male vs. female from task-fmri brain connectivity. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4089–4092. IEEE.
- Shi, Y., Liu, W., Liu, R., Zeng, Y., Wu, L., Huang, S., Cai, G., Yang, J., and Wu, W. (2019). Investigation of the emotional network in depression after stroke: A study of multivariate granger causality analysis of fmri data. *Journal of Affective Disorders*, 249:35–44.
- Siddiqi, S. H., Kording, K. P., Parvizi, J., and Fox, M. D. (2022). Causal mapping of human brain function. *Nature reviews neuroscience*, 23(6):361–375.
- Singh, V. K. and Sharma, V. (2014). Elitist genetic algorithm based energy balanced routing strategy to prolong lifetime of wireless sensor networks. *Chin. J. Eng*, 2014:1–6.
- Stone, C. J. (1977). Consistent nonparametric regression. *The annals of statistics*, pages 595–620.
- Tang, B. and Zhao, J. (2017). The hidden relationships between different regions of human brain. *Undergraduate Student Honor Thesis, University of Wisconsin, Supervised by Zhengjun Zhang*, pages 1–13.

- Tolosa, E., Garrido, A., Scholz, S. W., and Poewe, W. (2021). Challenges in the diagnosis of parkinson's disease. *The Lancet Neurology*, 20(5):385–397.
- Utevsky, A. V., Smith, D. V., and Huettel, S. A. (2014). Precuneus is a functional core of the default-mode network. *Journal of Neuroscience*, 34(3):932–940.
- Valls-Solé, J. and Valldeoriola, F. (2002). Neurophysiological correlate of clinical signs in parkinson's disease. *Clinical neurophysiology*, 113(6):792–805.
- Vink, J. J., Klooster, D. C., Ozdemir, R. A., Westover, M. B., Pascual-Leone, A., and Shafi, M. M. (2020). Eeg functional connectivity is a weak predictor of causal brain interactions. *Brain topography*, 33(2):221–237.
- Vinod, H. D. (2017). Generalized correlation and kernel causality with applications in development economics. *Communications in Statistics-Simulation and Computation*, 46(6):4513–4534.
- Yang, L., Wei, A.-H., Ouyang, T.-T., Cao, Z.-Z., Duan, A.-W., and Zhang, H.-H. (2021). Functional plasticity abnormalities over the lifespan of first-episode patients with major depressive disorder: a resting state fmri study. *Annals of Translational Medicine*, 9(4).
- Yule, G. U. (1897a). On the significance of bravais' formulae for regression, &c., in the case of skew correlation. *Proceedings of the Royal Society of London*, 60(359-367):477–489.
- Yule, G. U. (1897b). On the theory of correlation. *Journal of the Royal Statistical Society*, 60(4):812–854.
- Yule, G. U. (1900). Vii. on the association of attributes in statistics: with illustrations from the material of the childhood society, &c. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 194(252-261):257–319.
- Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, 2(2):121–134.

- Zhang, C., Chai, Y., Guo, X., Gao, M., Devilbiss, D., and Zhang, Z. (2016). Statistical learning of neuronal functional connectivity. *Technometrics*, 58(3):350–359.
- Zhang, Z. (2021a). Five critical genes related to seven Covid-19 subtypes: A data science discovery. *Journal of Data Science*, 19(1):142–150.
- Zhang, Z. (2021b). Functional effects of four or fewer critical genes linked to lung cancers and new subtypes detected by a new machine learning classifier. *Journal of Clinical Trials*, 11:S14:100001.
- Zhang, Z. (2022a). The existence of at least three genomic signature patterns and at least seven subtypes of covid-19 and the end of the disease. *Vaccines*, 10(5):761.
- Zhang, Z. (2022b). Lift the veil of breast cancers using four or fewer critical genes. *Cancer Informatics*, 21:1–11.
- Zhao, J. and Tang, B. (2018). The hidden connections and the causes of alzheimer disease. *Undergraduate Student Honor Thesis, University of Wisconsin, Supervised by Zhengjun Zhang*, pages 1–14.
- Zheng, S., Shi, N.-Z., and Zhang, Z. (2012). Generalized measures of correlation for asymmetry, nonlinearity, and beyond. *Journal of the American Statistical Association*, 107(499):1239–1252.
- Zhong, X., Pu, W., and Yao, S. (2016). Functional alterations of fronto-limbic circuit and default mode network systems in first-episode, drug-naïve patients with major depressive disorder: a meta-analysis of resting-state fmri data. *Journal of affective disorders*, 206:280–286.
- Zhuo, C., Li, G., Lin, X., Jiang, D., Xu, Y., Tian, H., Wang, W., and Song, X. (2019). The rise and fall of mri studies in major depressive disorder. *Translational psychiatry*, 9(1):1–14.