



Robust Graphical Modeling of High-Dimensional Gaussian and Non-Gaussian Models: Theory and Applications

Thesis for the Degree of Doctor of Philosophy (PhD)

Abdaljbbar Babiker Abdaljbbar Dawod

Supervisor: Prof. em. Dr. György Terdik

University of Debrecen

Doctoral Council for Natural Sciences and Engineering

Doctoral School of Informatics

Debrecen, 2026

Hereby I declare that I have prepared this thesis within the Doctoral Council for Natural Sciences and Engineering, Doctoral School of Informatics, University of Debrecen, to obtain a Ph.D. degree in Engineering at Debrecen University.

The results published in the thesis are not reported in any other PhD theses.

Debrecen, June 8, 2026.

signature of the candidate

Hereby, I confirm that Abdaljbbar Babiker Abdaljbbar Dawod candidate, conducted his studies under my supervision within the Applied Information Technology and its Theoretical Background, Doctoral Program of the Doctoral School of Informatics between 2022 and 2026. The candidate's independent studies and research significantly contributed to the results published in the thesis.

I also declare that the results published in the thesis are not reported in any other theses.

I support the acceptance of the thesis.

Debrecen, June 8, 2026.

signature of the supervisor

Robust Graphical Modeling of High-Dimensional Gaussian and Non-Gaussian Models: Theory and Applications

Dissertation submitted in partial fulfillment of the requirements
for the doctoral (PhD) degree in Informatics

Written by **Abdaljbbar Babiker Abdaljbbar Dawod**, B.Sc in
Mathematical Sciences, M.Sc in Statistics.

Prepared in the framework of the Doctoral School of Informatics of
the University of Debrecen
(Applied Information Technology and its Theoretical Background
program)

Dissertation advisor: **Prof. em. Dr. György Terdik**

The official opponents of the dissertation:

Dr.

Dr.

The evaluation committee:

Chairperson: Dr.

Members: Dr.

Dr.

Dr.

Dr.

The date of the dissertation defense: 2026.

ABSTRACT

This dissertation develops robust graphical models for high-dimensional, non-Gaussian data. Sparse precision matrix estimation, typically via the GLASSO, relies heavily on Gaussian assumptions, which limit its use when data exhibit heavy tails, complex dependencies, or irregularities. This research extends sparse graphical modeling beyond the normality assumption while maintaining sparsity, interpretability, and computational speed.

First, clarifying the theoretical link between GLASSO estimation and conditional independence structures. Then, extending the framework to selected elliptical distributions. This improves robustness under non-Gaussian conditions. Comprehensive simulations assess estimation accuracy and structural recovery across diverse scenarios.

The proposed methods are validated through interdisciplinary applications. In statistical process control, we integrate precision matrix estimation into multivariate monitoring to improve the sensitivity of the corresponding control charts. In behavioral science, these models uncover the conditional dependence structures among psychological variables. Finally, we analyze nutrient interactions in baobab leaves to map chemical relationships critical to food system research.

This dissertation strengthens the theoretical foundation of sparse graphical models. It bridges Gaussian and non-Gaussian frameworks and proves their practical utility across high-dimensional, diverse domains.

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to Allah (SWT), the Most Gracious and the Most Merciful, for granting me the strength, health, and perseverance to complete this dissertation.

I would like to express my sincere gratitude to my supervisor, Prof. em. Dr. György Terdik, for his invaluable guidance, continuous support, and insightful feedback throughout my doctoral studies. His expertise, dedication to excellence, and encouragement have been fundamental to the development and completion of this research. I am also thankful to the professors and staff of the Faculty for their knowledge, assistance, and contribution to my academic growth.

My heartfelt appreciation goes to my parents, whose unconditional love, prayers, and encouragement have been the foundation of my achievements. Their values and belief in me have inspired me to pursue excellence. I am equally grateful to my family for their patience and support during this journey, and especially to my wife for her constant encouragement and companionship throughout my studies.

I gratefully acknowledge the financial support provided by the Stipendium Hungaricum Scholarship Programme and the University of Debrecen. Their generous support significantly eased the challenges of my academic journey and allowed me to focus fully on my research.

Abdaljbbar Babiker Abdaljbbar Dawod

Debrecen, Hungary, 2026

Contents

ABSTRACT	i
ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	iv
LIST OF TABLES	v
LIST OF ABBREVIATIONS	vii
1 INTRODUCTION	1
1.1 Overview	1
1.1.1 Gaussian Graphical Models	2
1.1.2 Non-Gaussian and Robust Graphical Models . .	3
1.2 Problem Statement and Research Objectives	4
1.3 Dissertation Structure	5
2 GRAPHICAL MODELS	6
2.1 Introduction	6
2.2 Basic Graphical Concepts	6
2.3 Conditional and Partial Correlations	11
2.3.1 The Inverse Variance	14
2.4 Precision Matrices for Common Multivariate Distributions	15
2.4.1 Elliptically Symmetric Distributions	15
2.4.2 Multivariate Normal distribution	17
2.4.3 Multivariate Student t distribution	17
2.4.4 Multivariate symmetric generalized hyperbolic distribution	18
3 Graphical LASSO	19
3.1 Related Works	19
3.1.1 GLASSO	19
3.2 EM-based Graphical LASSO Algorithms	21
3.2.1 Partial and Conditional Uncorrelations	26
3.2.2 t LASSO	27
3.2.3 Multivariate Symmetric Generalized Hyperbolic .	31

3.2.4	Multivariate Power Exponential Distribution . .	37
3.2.5	Estimation of Nuisance Parameters	42
3.2.6	Theoretical considerations for the estimator of Θ	43
3.2.7	Simulated Results	46
3.2.8	Computational Costs	53
3.3	Thesis 1	56
3.4	Thesis 2	56
4	APPLICATIONS OF GLASSO	57
4.1	Cognitive-Behavioral Networks	57
4.1.1	Post-Traumatic Stress Disorder	58
4.1.2	Emotional monitoring	59
4.2	Robust Quality Inspection via MCV Control Charts . .	64
4.2.1	GLasso-based MCV control charts	65
4.2.2	Wine Quality	67
4.3	Nutrients Interaction Network	69
4.3.1	Materials and Methods	71
4.3.2	Results and Discussion	72
4.3.3	Strengths and Limitations	79
4.4	Thesis 3	81
4.5	Thesis 4	81
4.6	Thesis 5	82
5	CONCLUSION AND FUTURE DIRECTIONS	83
5.1	Overview of Research and Objectives	83
5.2	Summary of Findings	84
5.3	Contributions to Knowledge	85
5.4	Limitations	87
5.5	Future Research Directions	88
	BIBLIOGRAPHY	90

List of Figures

2.1	An undirected graph illustrating maximal cliques.	8
2.2	An ordered chain graph partitioned into blocks.	9
2.3	An undirected decomposable graph with a proper partition.	10
3.1	The subclasses of the GIG distribution	32
3.2	Contour plots of the MPE distribution	39
3.3	ROC curves with MSGH distribution ($d < n$)	49
3.4	ROC curves with MSGH distribution ($d > n$)	51
3.5	ROC curves with MPE distribution ($d < n$)	52
3.6	ROC curves with MPE distribution ($d > n$)	53
4.1	Sparse conditional dependence network of PTSD symp- toms estimated via the GLASSO algorithm	60
4.2	Sparse conditional dependence network of PTSD symp- toms estimated via the EM-GLASSO _P algorithm	60
4.3	Estimated sparse conditional dependence network of stu- dent emotion monitoring variables, generated via GLASSO.	63
4.4	Estimated sparse conditional dependence network of stu- dent emotion monitoring variables, generated via the EM-GLASSO _P	63
4.5	Monitoring the quality of red <i>vinho verde</i> wine using GLASSO framework.	69
4.6	Monitoring the quality of red <i>vinho verde</i> wine using EM- GLASSO _P framework.	70
4.7	Q-Q plot of Baobab nutrients	74
4.8	Heatmap of baobab leaves' correlations	75
4.9	The network analysis of Baobab leaves using SGCGMs	76
4.10	Centrality Measures using SGCGMs	79
4.11	The network analysis of Baobab leaves using EM- GLASSO _H	80
4.12	Centrality Measures using EM-GLASSO _H	80

List of Tables

3.1	Comparison of precision matrix estimation methods for non-Gaussian distributions	25
3.2	Theoretical Computational Complexity Comparison of GLASSO and EM-GLASSO	55
3.3	Computational performance of GLASSO and EM-GLASSO variants ($n = 100, d = 50$, over 1,000 iterations).	55
4.1	Common centrality measures: definitions and interpretation.	73
4.2	Summary statistics of key nutrient components in baobab leaves.	73

LIST OF ABBREVIATIONS

A.GLASSO	Adaptive GLASSO
ARL	Average Run Length
BNs	Bayesian Networks
CDF	Cumulative Distribution Function
CLIME	Constrained ℓ_1 -minimization for inverse matrix estimation
CV	Coefficient of Variation
EM	Expectation-Maximization
FPR	False Positive Rate
GGMs	Gaussian Graphical Models
GIG	Generalized Inverse Gaussian
GLASSO	Graphical Least Absolute Selection and Shrinkage Operator
GSLOPE	Gaussian Sorted- ℓ_1 Penalized Estimators
ICP-OES	Inductively Coupled Plasma-Optical Emission Spectrometry
IG	Inverse Gaussian
IGam	Inverse Gamma
IoT	Internet of Things
LCL	Lower Control Limit
LLSP	Linear Least-Squares Prediction
MCUSUM	Multivariate Cumulative Sum
MEWMA	Multivariate Exponentially Weighted Moving Average

MHWMA	Multivariate Homogeneously Weighted Moving Average
MLE	Maximum Likelihood Estimator
MSGH	Multivariate Symmetric Generalized Hyperbolic
MPE	Multivariate Power Exponential
MSPC	Multivariate Statistical Process Control
PCA	Principal Component Analysis
PC-GLASSO	Partial Correlation GLASSO
PDF	Probability Density Function
pHyp	Positive Hyperbolic
PTSD	Post-Traumatic Stress Disorder
ROC	Overall Edge Recovery Performance
SCAD	Smoothly Clipped Absolute Deviation
SCLIME	Spatial-Signed CLIME
SEM	Structural Equation Model
SGCGMs	Semiparametric Gaussian Copula Graphical Models
StARS	Stability Approach to Regularization Selection
TPR	True Positive Rate
TSLOPE	Student t Sorted- ℓ_1 Penalized Estimators
UCL	Upper Control Limit

1

INTRODUCTION

This chapter summarizes the motivations for this dissertation and provides a brief description of the core concepts used throughout the dissertation. It also outlines the research questions, highlights the contributions made to existing knowledge, and concludes with an overview of the subsequent chapters.

1.1 Overview

Recent advancements in Industry 4.0 and the Internet of Things (IoT) have led to industrial systems that routinely generate high-dimensional data with complex dependence structures. Such data arise in fields including finance, environmental science, image analysis, and genomics, where hundreds of variables are observed simultaneously. Modeling and interpreting the dependence relationships among these variables pose a significant challenge central to statistical inference, prediction, and decision-making.

Graphical models have emerged as a robust and interpretable framework for representing multivariate dependence structures. In particular, undirected graphical models encode conditional independence relationships among variables through the absence of edges in the graph. For continuous data that satisfy Gaussian assumptions, Gaussian graphical models (GGMs) have become a cornerstone of multivariate statistical analysis, in which conditional independencies are characterized by zero entries in the precision (inverse covariance) matrix. Undoubtedly, GGMs provide a strong theoretical foundation and practical utility, enabling dimension reduction, network recovery, and improved inference in high-dimensional settings.

However, GGMs rely heavily on the normality assumption, which

is often violated in practice. Real-world processes frequently exhibit heavy tails, skewness, outliers, and nonlinear dependence structures. In addition, many high-dimensional settings involve a large number of variables that exceed the sample size, rendering conventional estimation approaches infeasible. Hence, to overcome these challenges, several methodological approaches have been developed, including robust statistical techniques and extensions of graphical models to non-Gaussian settings. The use of penalty terms, such as the ℓ_1 -norm, to construct ℓ_1 -penalized likelihoods has enabled consistent estimation of large precision matrices under sparsity assumptions. Moreover, non-Gaussian graphical models—such as those based on elliptical distributions, copulas, and semiparametric formulations—have broadened the applicability of graphical modeling to a wider class of processes.

As a result, this dissertation, **Robust Graphical Modeling of High-Dimensional Gaussian and Non-Gaussian Processes: Theory and Applications**, advances the field by developing and analyzing graphical modeling techniques. By addressing both Gaussian and non-Gaussian processes, the work explores theoretical properties and methodological advancements in high-dimensional settings, with a specific focus on ensuring robustness.

1.1.1 Gaussian Graphical Models

GGMs provide a direct linkage between graph theory and multivariate normal distributions. For a multivariate random variable following a multivariate normal distribution, conditional independence between two variables given the remaining variables is equivalent to a zero entry in their precision matrix[73, 114].

Under the standard settings, involving low-dimensional processes, the maximum likelihood estimator (MLE) of the covariance matrix is straightforward. In contrast, in a high-dimensional setting with more variables than observations, the MLE of the covariance matrix is singular. To address this issue, regularization methods have been introduced to shrink some entries to zero. Most regularizations use the ℓ_1 -norm penalty, for example, Graphical least absolute selection and shrinkage operator (GLASSO) [55], adaptive LASSO (A.GLASSO) [48], and

smoothly clipped absolute deviation (SCAD) [49]. Subsequently, under suitable assumptions, these methods yielded scalable algorithms with strong theoretical guarantees.

The GLASSO algorithm employs an ℓ_1 -norm penalty, defined as the sum of the absolute values of the elements of the precision matrix. Specifically, with penalty parameter ρ , the penalty term is given by

$$\rho \sum_{i=1}^d \sum_{j=1}^d (|\theta_{ij}|) = \rho \|\Theta\|_1,$$

where θ_{ij} denotes the (i, j) -th entry of the precision matrix. Subsequently, Friedman et al. [55] adopted this penalty and introduced a regularized estimator of the precision matrix by solving the following optimization problem.

$$\hat{\Theta}^{GLASSO} = \arg \max_{\Theta} [\log |\Theta| - \text{tr}(\mathbf{S}\Theta) - \rho \|\Theta\|_1] \quad (1.1)$$

The coordinate descent procedure has been used for solving the given objective function. Therefore, the ℓ_1 -penalty shrinks some entries of the precision matrix towards zero as ρ increases, and sets some entries to exactly zero for sufficiently large ρ .

1.1.2 Non-Gaussian and Robust Graphical Models

While GGMs are mathematically convenient, with a closed-form objective function and a clear interpretation of the absence of an edge in the graph, many processes exhibit non-Gaussian features. Several extensions have been proposed to tackle this issue by incorporating other distributions. For instance, preserving the covariance-based dependence structure when having heavier-tailed processes by using an elliptical distribution [52]. Also, semiparametric and copula-based graphical models soften the normality assumptions by modelling the marginal distributions separately from the dependence structure [78].

Therefore, robust graphical models are important in the presence of contaminated data or when the Gaussian assumption is violated. Based on the robust covariance matrix derived from Spearman's ρ or Kendall's τ correlation matrix, the gained graphical model can offer

better interpretable results. Alternatively, using likelihoods derived from heavy-tailed distributions via algorithms such as the Expectation-Maximization (EM) algorithm. These approaches aim to give interpretability of the resulting graphical models while improving their stability and avoiding misspecification.

1.2 Problem Statement and Research Objectives

Despite significant efforts to estimate the structure of graphical models, several critical gaps remain unresolved. First, many existing methods are tailored to Gaussian assumptions, limiting their robustness and applicability to real data when those assumptions are violated. Second, although sparse estimation techniques are well developed, their theoretical interpretation under conditional independence and their empirical performance under non-Gaussian processes are not fully understood to date. Third, offering unified frameworks that accommodate both Gaussian and non-Gaussian processes while maintaining interpretability and computational feasibility is in high demand. Moreover, in many practical cases, such as monitoring an industrial plant and complex system analysis, methods that are statistically sound, implementable and robust to data deficiency are in high demand.

Therefore, this dissertation addresses the problem of effectively modeling, estimating, and applying graphical models to Gaussian and non-Gaussian datasets in high-dimensional settings while preserving their robustness and practical feasibility.

The main objective of this dissertation is to study and develop graphical modeling methodologies for Gaussian and non-Gaussian processes, supported by theoretical analysis and practical applications. To fulfill the main objective, the dissertation pursues the following research sub-objectives:

1. Evaluate the GLASSO framework and its formal connection to conditional independence.
2. Extend GLASSO to selected elliptical distribution families.

3. Model conditional dependence networks in behavioral science applications.
4. Apply these extensions to improve the performance of statistical process control charts.
5. Analyze nutrient interactions in baobab leaves to advance nutritional profiling and food system research.

1.3 Dissertation Structure

This dissertation is structured within a specific paradigm to facilitate a progressive exploration, and it is organized as follows:

- Chapter 2 presents the theoretical foundations of graphical models, including graph theory, conditional independence, and the interpretation of precision matrices.
- Chapter 3 focuses on GLASSO, outlining its theoretical properties and extending it to the Student t distribution. Then, extends graphical modeling to non-Gaussian processes, introducing robust graphical models for elliptical distributions, supported by simulation studies.
- Chapter 4 explores practical applications of GLASSO-based methods across diverse, real-world domains. Specifically, it uses semi-parametric Gaussian copula graphical modeling (SGCGMs) to identify nutrient interaction networks in Baobab leaves and examine their implications for nutrient profiling and nutritional discovery. Additionally, the chapter demonstrates how the GLASSO is applied in the cognitive-behavioral sciences and to enhance statistical process control (SPC) charts
- Chapter 5 concludes the dissertation with a summary of contributions, limitations, and directions for future research.

2

GRAPHICAL MODELS

This chapter introduces graphical models and their interpretation in terms of conditional independence. The chapter is organized as follows: Section 2.1 introduces the topic, while Section 2.2 presents fundamental concepts of graph theory, Markov properties, and their correspondence with conditional independence structures. Section 2.3 describes the use of the precision matrix as a proxy for inferring conditional independence in Gaussian processes. Finally, Section 2.4 presents estimated precision matrices for selected elliptical symmetric distributions.

2.1 Introduction

Graphical modeling combines graph theory and probability theory to visually represent and analyze complex interactions among random variables. It provides a way to show the **dependencies** and **relationships** through intuitive graphical structures and offers a clear, interpretable framework for understanding probabilistic systems. Whittaker [114] and Lauritzen [73] pioneered the adoption of graphical models not only to visualize multivariate relationships, but also to enable efficient inference, prediction, and hypothesis testing. Later, graphical models became invaluable in fields such as epidemiology, social sciences, machine learning, and computational biology [24, 32, 37, 51, 69]. By using graphical models, researchers can uncover hidden patterns, simplify complex systems, and effectively communicate statistical insights.

2.2 Basic Graphical Concepts

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ provides a formal framework for representing the dependencies between random variables, where the vertex set $\mathcal{V}$ corre-

sponds to the variables and the edge set $\mathcal{E}$ encodes the pairwise relationships between them. Graphs are fundamentally categorized by the nature of their edge set $\mathcal{E}$. In **undirected graphs**, edges represent symmetric associations, indicating bidirectional relationships between variables. Conversely, **directed graphs** employ asymmetric edges to denote ordered dependencies, often representing causal pathways. A specialized case is the **marked graph**, defined as a structure in which each vertex has exactly one incoming and one outgoing edge.

In an undirected graph $\mathcal{G}^u = (\mathcal{V}, \mathcal{E})$, an edge between nodes i and j is denoted by $(i - j)$, indicating that i and j are adjacent (or neighbors). A graph is termed **complete** if every pair of distinct vertices is connected by a unique edge, representing a saturated set of pairwise relationships. For any subset $A \subseteq \mathcal{V}$, the **induced subgraph** $\mathcal{G}_A = (A, \mathcal{E}_A)$ is defined by the edge set $\mathcal{E}_A = \mathcal{E} \cap (A \times A)$, containing all original edges between vertices in A . A **path** of length k between vertices i and j is a sequence of $k + 1$ vertices $\{v_0, v_1, \dots, v_k\}$ such that $v_0 = i$, $v_k = j$, and $(v_{m-1}, v_m) \in \mathcal{E}$ for all $m = 1, \dots, k$.

Furthermore, a **clique** is defined as a subset of vertices within an undirected graph where every distinct pair is adjacent, forming a fully connected subgraph (see Example 2.1). A **chain graph** (also known as a dependence chain) is a hybrid graphical model whose vertex set $\mathcal{V}$ is partitioned into an ordered sequence of k disjoint blocks, $\mathcal{V}_1, \dots, \mathcal{V}_k$. Within any single block, edges are strictly undirected, representing symmetric conditional dependencies as illustrated in Example 2.2. Conversely, edges between distinct blocks are directed, capturing asymmetric or causal influences. This dual structure makes chain graphs particularly advantageous for characterizing complex, hybrid dependency architectures in multivariate data.

Example 2.1. Consider an undirected graph $\mathcal{G}^u = (\mathcal{V}, \mathcal{E})$ with the vertex set $\mathcal{V} = \{A, B, C, D, E, F\}$ and edge set $\mathcal{E} = \{(A, B), (A, C), (A, D), (B, C), (B, D), (C, D), (D, E), (E, F)\}$. Within this graph, the subset $\mathcal{C} = \{A, B, C, D\}$ forms a **maximal clique** (i.e., a clique that cannot be extended by adding another vertex). Every distinct pair of vertices within $\mathcal{C}$ is adjacent, and no additional vertex from $\mathcal{V}$ (namely E or F) can be added to $\mathcal{C}$ without violating this fully

connected property. Additionally, the subset $\{E, F\}$ forms a separate maximal clique of size two. See Figure 2.1.

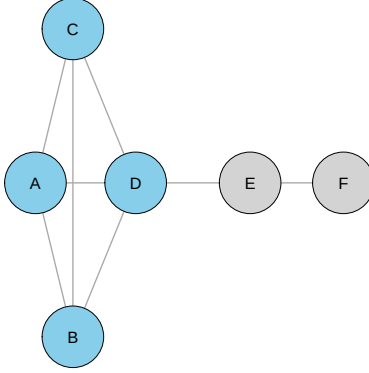


Figure 2.1: An undirected graph illustrating maximal cliques.

Example 2.2. Consider a chain graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the vertex set $\mathcal{V} = \{A, B, C, D, E, F, G, H, I\}$. The edge set $\mathcal{E}$ is partitioned into undirected edges ($\mathcal{E}_U$) and directed edges ($\mathcal{E}_D$), defined as follows:

- $\mathcal{E}_U = \{(A, B), (B, C), (D, E), (F, G), (G, H), (H, I), (I, F)\}$
- $\mathcal{E}_D = \{(C, F), (E, I), (E, F)\}$

The connected components induced by the undirected edges naturally partition the graph into three ordered blocks (or chain components): $\mathcal{V}_1 = \{A, B, C\}$, $\mathcal{V}_2 = \{D, E\}$, and $\mathcal{V}_3 = \{F, G, H, I\}$. This topological structure, illustrating the hybrid dependencies across blocks, is depicted in Figure 2.2.

An undirected graph is defined as **decomposable** if it is either complete or can be recursively decomposed into two decomposable induced subgraphs, $\mathcal{G}_{\mathcal{V}_1 \cup \mathcal{V}_3}$ and $\mathcal{G}_{\mathcal{V}_2 \cup \mathcal{V}_3}$. Here, the triplet $(\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3)$ forms a partition of the vertex set $\mathcal{V}$, where $\mathcal{V}_1$ and $\mathcal{V}_2$ are non-empty and separated by $\mathcal{V}_3$. Furthermore, decomposable graphs are categorized as either *strong* or *weak*. A strongly decomposable graph admits a perfect elimination ordering, which enables recursive factorization into cliques connected by vertex subsets whose removal disconnects the graph, where

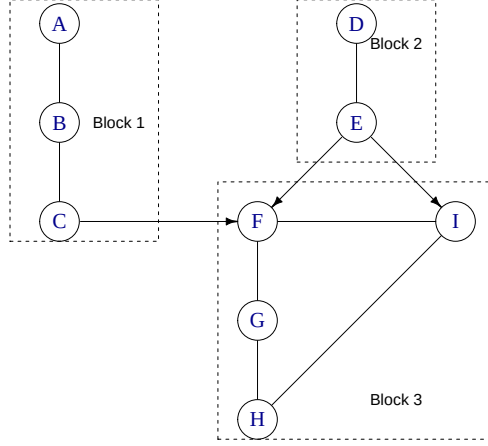


Figure 2.2: An ordered chain graph partitioned into blocks.

each subset forms a fully connected subgraph (**complete separator**). Conversely, while weakly decomposable graphs can also be factored into cliques, their separators are not strictly required to be complete (see Example 2.3).

Example 2.3. Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the vertex set $\mathcal{V} = \{A, B, C, D\}$ and edge set $\mathcal{E} = \{(A, B), (A, C), (B, C), (B, D), (C, D)\}$. This graph admits a proper decomposition where $\mathcal{V}_1 = \{A\}$, $\mathcal{V}_2 = \{B, C\}$ and $\mathcal{V}_3 = \{D\}$, as shown in Figure 2.3.

Markovian properties are closely tied to the graph's conditional independence. Contrary to a common misconception, the marginal independence of a random variable X_1 from variables X_2 and X_3 individually does not necessarily imply independence from the joint pair (X_2, X_3) . Hence, X_1 and X_2 are conditionally independent given X_3 if and only if

$$f_{X_1 X_2 | X_3}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{x}_3) = f_{X_1 | X_3}(\mathbf{x}_1; \mathbf{x}_3) f_{X_2 | X_3}(\mathbf{x}_2; \mathbf{x}_3),$$

for all values of X_3 , where $f_{X_3}(\mathbf{x}_3) > 0$, it is written as $X_1 \perp\!\!\!\perp X_2 | X_3$.

Frydenberg [56] and Andersson [12] studied Markov properties for *directed*, *undirected*, and *chain* graphs, where Andersson [12] highlighted the necessary and sufficient condition for two chain graphs to have the

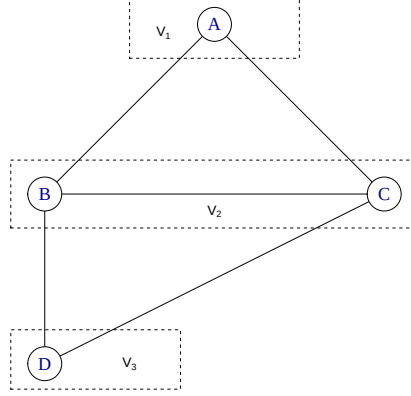


Figure 2.3: An undirected decomposable graph with a proper partition.

same Markov properties. Recently, Cordoba [34] provided a comprehensive review on model inference in directed and undirected Markov models with Gaussian parameterization, focusing on their main characteristics.

An undirected graph $\mathcal{G}^u = (\mathcal{V}, \mathcal{E})$ with disjoint sets $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3 \subseteq \mathcal{V}$, is considered to follow *Global Markov Property* if every path from a node in $\mathcal{V}_1$, to a node in $\mathcal{V}_2$ passes through at least one node in $\mathcal{V}_3$, thus, $\mathcal{V}_1$ and $\mathcal{V}_2$ are said to be separated by $\mathcal{V}_3$. Consequently, it asserts that

$$X_{\mathcal{V}_1} \perp\!\!\!\perp X_{\mathcal{V}_2} | X_{\mathcal{V}_3}$$

It means that variables in $\mathcal{V}_1$ are **conditionally independent** of the variables in $\mathcal{V}_2$ given the variables in $\mathcal{V}_3$.

If two vertices i and j in an undirected graph are not connected by an edge, then

$$X_i \perp\!\!\!\perp X_j | X_{\mathcal{V} \setminus \{i,j\}}$$

This implies that variables X_i and X_j are conditionally independent given all other variables in the graph, which is known as the *Pairwise Markov Property*. This property characterizes conditional independence relationships in the graph.

The *Local Markov Property* states that each node i is conditionally independent of all non-neighboring nodes given its neighbors $ne_{\mathcal{G}}(i)$.

For an undirected graph $\mathcal{G}$ with vertex set $\mathcal{V}$, this can be expressed as:

$$X_i \perp\!\!\!\perp X_{\mathcal{V} \setminus \{i \cup \text{ne}_{\mathcal{G}}(i)\}} | X_{\text{ne}_{\mathcal{G}}(i)}$$

For any arbitrary probability measure $M[\cdot]$ defined on a space $\mathcal{V}_x$, and subsets $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4 \subseteq \mathcal{V}$, the conditional independence relation satisfies the following properties [12, 39, 73]:

Symmetry: $\mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_2 | M[\mathcal{V}_3] \implies \mathcal{V}_2 \perp\!\!\!\perp \mathcal{V}_1 | M[\mathcal{V}_3]$.

Decomposition: $(\mathcal{V}_1 \cup \mathcal{V}_4) \perp\!\!\!\perp \mathcal{V}_2 | \mathcal{V}_3 \implies \mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_2 | \mathcal{V}_3$ and $\mathcal{V}_4 \perp\!\!\!\perp \mathcal{V}_2 | \mathcal{V}_3$.

Reverse Decomposition: $\mathcal{V}_1 \perp\!\!\!\perp (\mathcal{V}_2 \cup \mathcal{V}_3) | M[\mathcal{V}_4] \implies \mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_2 | M[\mathcal{V}_4]$.

Weak Union: $\mathcal{V}_1 \perp\!\!\!\perp (\mathcal{V}_2 \cup \mathcal{V}_3) | M[\mathcal{V}_4] \implies \mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_2 | (\mathcal{V}_3 \cup M[\mathcal{V}_4])$.

Contraction: $\mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_2 | M[\mathcal{V}_4]$ and $\mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_3 | M[\mathcal{V}_2] \implies \mathcal{V}_1 \perp\!\!\!\perp (\mathcal{V}_2 \cup \mathcal{V}_3) | M[\mathcal{V}_4]$.

Intersection: If $M[\cdot]$ has a strictly positive density, then $\mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_2 | (\mathcal{V}_3 \cup M[\mathcal{V}_4])$ and $\mathcal{V}_1 \perp\!\!\!\perp \mathcal{V}_3 | (\mathcal{V}_4 \cup M[\mathcal{V}_2]) \implies \mathcal{V}_1 \perp\!\!\!\perp (\mathcal{V}_2 \cup \mathcal{V}_3) | M[\mathcal{V}_4]$.

If a probability distribution over an undirected graph $\mathcal{G}^u$ satisfies the intersection property (Condition 6), the pairwise, local, and global Markov properties become mathematically equivalent. Thus, the conditional independence is fundamentally characterized by the factorization of the joint Probability Density Function (PDF).

For a partitioned random vector $X = (X_1, X_2, X_3, X_4)$, conditional independence properties allow for the deduction of related independence structures. For instance, by the listed properties, $X_2 \perp\!\!\!\perp (X_3, X_4) | X_1$ implies both $X_2 \perp\!\!\!\perp X_4 | (X_1, X_3)$ and $X_2 \perp\!\!\!\perp X_4 | X_1$.

2.3 Conditional and Partial Correlations

Understanding Markov properties requires characterizing conditional independence structures, a task fundamentally linked to the study of

partial correlations. Although zero partial correlation implies strict conditional independence only for Gaussian distributions—where the structure is explicitly parameterized by the precision matrix—it remains a powerful tool for structural modeling in non-Gaussian settings.

Definition 2.4. Let X_1 and X_2 be scalar random variables, and let $\mathbf{X}_3$ be a d -dimensional random vector. The **conditional covariance** of X_1 and X_2 given $\mathbf{X}_3$ is defined as:

$$\text{Cov}(X_1, X_2 \mid \mathbf{X}_3) = \mathbb{E}[(X_1 - \mathbb{E}[X_1 \mid \mathbf{X}_3])(X_2 - \mathbb{E}[X_2 \mid \mathbf{X}_3]) \mid \mathbf{X}_3].$$

Analogously, the **conditional variance** of X_1 given $\mathbf{X}_3$ naturally follows special case when $X_2 = X_1$. It represents the expected squared deviation of X_1 around its conditional mean:

$$\begin{aligned} \text{Var}(X_1 \mid \mathbf{X}_3) &= \text{Cov}(X_1, X_1 \mid \mathbf{X}_3) \\ &= \mathbb{E}[(X_1 - \mathbb{E}[X_1 \mid \mathbf{X}_3])^2 \mid \mathbf{X}_3] \end{aligned}$$

Definition 2.5. The **partial covariance** between X_1 and X_2 with respect to $\mathbf{X}_3$, denoted as $\text{Cov}(X_1, X_2 \dashv \mathbf{X}_3)$, is defined as the covariance between the residuals of X_1 and X_2 after projecting them linearly onto $\mathbf{X}_3$. The optimal Linear Least Squares Predictor (LLSP) of X_1 based on $\mathbf{X}_3$ is given by:

$$\hat{X}_1 = \mu_{X_1} + \text{Cov}(X_1, \mathbf{X}_3) \text{Cov}(\mathbf{X}_3)^{-1}(\mathbf{X}_3 - \mu_{\mathbf{X}_3}).$$

For a univariate X_3 , the linear projection of X_1 onto X_3 can be parameterized directly by their correlation, $\rho_{X_1 X_3}$, as

$$\hat{X}_1 = \mu_{X_1} + \rho_{X_1 X_3} \frac{\sigma_{X_1}}{\sigma_{X_3}}(X_3 - \mu_{X_3}).$$

Using the predicted values $\hat{X}_1$ and $\hat{X}_2$, the partial covariance is

$$\begin{aligned} \text{Cov}(X_1, X_2 \dashv \mathbf{X}_3) &= \text{Cov}(X_1 - \hat{X}_1, X_2 - \hat{X}_2) \\ &= \text{Cov}(X_1, X_2) - \text{Cov}(X_1, \mathbf{X}_3) \text{Cov}(\mathbf{X}_3)^{-1} \text{Cov}(\mathbf{X}_3, X_2). \end{aligned}$$

Similarly, the **partial variance** of X_1 with respect to $\mathbf{X}_3$ represents the

variance of X_1 residuals after the linear effect of $\mathbf{X}_3$ has been removed:

$$\text{Var}(X_1 \upharpoonright \mathbf{X}_3) = \text{Cov}(X_1, X_1 \upharpoonright \mathbf{X}_3) = \mathbb{E}[(X_1 - \hat{X}_1)^2] = \mathbb{E}[\hat{X}_1 \hat{X}_1^T]$$

Definition 2.6. The **conditional correlation** between two scalar random variables X_1 and X_2 given a random vector $\mathbf{X}_3$, denoted as $\rho_{X_1, X_2 | \mathbf{X}_3}$, is defined as the ratio of their conditional covariance to the product of their conditional standard deviations:

$$\rho_{X_1, X_2 | \mathbf{X}_3} = \frac{\text{Cov}(X_1, X_2 | \mathbf{X}_3)}{\sqrt{\text{Var}(X_1 | \mathbf{X}_3) \text{Var}(X_2 | \mathbf{X}_3)}}.$$

If $\rho_{X_1, X_2 | \mathbf{X}_3} = 0$, X_1 and X_2 are said to be **conditionally uncorrelated** given $\mathbf{X}_3$.

Definition 2.7. The **partial correlation** between scalar random variables X_1 and X_2 controlling for a random vector $\mathbf{X}_3$, denoted by $\rho_{X_1, X_2 \upharpoonright \mathbf{X}_3}$, is defined as the Pearson correlation between their residuals after optimal linear projection onto $\mathbf{X}_3$. From Definition 2.5,

$$\rho_{X_1, X_2 \upharpoonright \mathbf{X}_3} = \frac{\text{Cov}(X_1 - \hat{X}_1, X_2 - \hat{X}_2)}{\sqrt{\text{Var}(X_1 - \hat{X}_1) \text{Var}(X_2 - \hat{X}_2)}}.$$

When the control variable is univariate (i.e., $\mathbf{X}_3 = X_3$), the partial correlation can be expressed directly in terms of the marginal pairwise correlations. As

$$\begin{aligned} \text{Cov}(X_1 - \hat{X}_1, X_2 - \hat{X}_2) &= \sigma_{X_1 X_2} - \frac{\sigma_{X_1 X_3} \sigma_{X_2 X_3}}{\sigma_{X_3}^2} \\ &= \rho_{X_1 X_2} \sigma_{X_1} \sigma_{X_2} - \rho_{X_1 X_3} \rho_{X_2 X_3} \sigma_{X_1} \sigma_{X_2} \\ &= \sigma_{X_1} \sigma_{X_2} (\rho_{X_1 X_2} - \rho_{X_1 X_3} \rho_{X_2 X_3}) \end{aligned}$$

and

$$\begin{aligned} \text{Var}(\hat{X}_1) &= \sigma_{X_1}^2 - \frac{(\rho_{X_1 X_3} \sigma_{X_1} \sigma_{X_3})^2}{\sigma_{X_3}^2} \\ &= \sigma_{X_1}^2 - \rho_{X_1 X_3}^2 \sigma_{X_1}^2 \\ &= \sigma_{X_1}^2 (1 - \rho_{X_1 X_3}^2) \end{aligned}$$

Then, the partial correlation can be formulated as

$$\rho_{X_1, X_2 | X_3} = \frac{\rho_{X_1 X_2} - \rho_{X_1 X_3} \rho_{X_2 X_3}}{\sqrt{(1 - \rho_{X_1 X_3}^2)(1 - \rho_{X_2 X_3}^2)}}.$$

It is well-established that partial and conditional correlations coincide when the conditional expectation is linear, and the conditional correlation remains unaffected by the conditioning variables [11, 16]. Although zero conditional correlation implies conditional independence only under strict Gaussianity [16], this relationship extends partially to elliptical distributions. Specifically, because the best predictor under an elliptical distribution is inherently linear, a zero partial correlation guarantees a zero conditional correlation [101], yielding a slightly more general framework than the Gaussian constraint alone (see Lemma 3.1).

2.3.1 The Inverse Variance

The inverse of the covariance matrix, denoted by $\mathbf{\Omega}$, is commonly referred to as the **precision matrix** or **concentration matrix**. For variables following a multivariate normal distribution, the precision matrix directly encodes the conditional independence structure among the variables [114, 118]. If a d -dimensional vector $\mathbf{X}$ is partitioned into $\mathbf{X}_1 \in \mathbb{R}^{d_1}$, $\mathbf{X}_2 \in \mathbb{R}^{d_2}$, then $\mathbf{\Omega} = \text{Cov}(\mathbf{X}_1, \mathbf{X}_2)^{-1}$ is given by the partitioned matrix as follows

$$\mathbf{\Omega} = \begin{bmatrix} \text{Var}(\mathbf{X}_1) & \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) \\ \text{Cov}(\mathbf{X}_2, \mathbf{X}_1) & \text{Var}(\mathbf{X}_2) \end{bmatrix}^{-1}$$

To derive the inverse of the partitioned joint covariance matrix, let $\mathbf{B}$ denote the coefficient matrix of the LLSP of $\mathbf{X}_2$ based on $\mathbf{X}_1$, defined by

$$\mathbf{B} = \text{Cov}(\mathbf{X}_2, \mathbf{X}_1) \text{Var}(\mathbf{X}_1)^{-1}$$

Applying the block matrix inversion formula via the Schur complement, $\mathbf{\Omega}$ can be expressed as

$$\mathbf{\Omega} = \begin{bmatrix} \text{Var}(\mathbf{X}_1)^{-1} + \mathbf{B}^T \text{Var}(\mathbf{X}_2 | \mathbf{X}_1)^{-1} \mathbf{B} & -\mathbf{B}^T \text{Var}(\mathbf{X}_2 | \mathbf{X}_1)^{-1} \\ -\text{Var}(\mathbf{X}_2 | \mathbf{X}_1)^{-1} \mathbf{B} & \text{Var}(\mathbf{X}_2 | \mathbf{X}_1)^{-1} \end{bmatrix}$$

The matrix-valued multiple correlation coefficient, $\mathbf{R}_{\mathbf{X}_2|\mathbf{X}_1}^2$, measures the proportion of variation in $\mathbf{X}_2$ that is linearly explained by $\mathbf{X}_1$. It is defined and expanded as follows

$$\begin{aligned}\mathbf{R}_{\mathbf{X}_2|\mathbf{X}_1}^2 &= \text{Var}(\mathbf{X}_2)^{-1/2} \text{Cov}(\mathbf{X}_2, \mathbf{X}_1) \text{Var}(\mathbf{X}_1)^{-1} \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) \text{Var}(\mathbf{X}_2)^{-1/2} \\ &= \text{Var}(\mathbf{X}_2)^{-1/2} [\text{Var}(\mathbf{X}_2) - \text{Var}(\mathbf{X}_2 | \mathbf{X}_1)] \text{Var}(\mathbf{X}_2)^{-1/2}\end{aligned}$$

After distributing the terms, $\text{Var}(\mathbf{X}_2)^{-1/2} \text{Var}(\mathbf{X}_2) \text{Var}(\mathbf{X}_2)^{-1/2} = \mathbf{I}$, and $\mathbf{R}_{\mathbf{X}_2|\mathbf{X}_1}^2$ can be simplified to

$$\mathbf{R}_{\mathbf{X}_2|\mathbf{X}_1}^2 = \mathbf{I} - \text{Var}(\mathbf{X}_2)^{-1/2} \text{Var}(\mathbf{X}_2 | \mathbf{X}_1) \text{Var}(\mathbf{X}_2)^{-1/2}$$

Recognizing that $\text{Var}(\mathbf{X}_2 | \mathbf{X}_1)^{-1}$ perfectly corresponds to the bottom-right block of the precision matrix ($\boldsymbol{\Omega}_{22}$), this final result establishes an elegant structural relationship between the matrix-valued $\mathbf{R}^2$ and the precision matrix. More generally, the partial correlation between any two univariate variables X_i and X_j , controlling for the set of all other remaining variables $\mathbf{X}_{\setminus\{i,j\}}$, is obtained by scaling the negative off-diagonal elements of the precision matrix:

$$\rho_{X_i, X_j | \mathbf{X}_{\setminus\{i,j\}}} = -\frac{\boldsymbol{\Omega}_{ij}}{\sqrt{\boldsymbol{\Omega}_{ii}\boldsymbol{\Omega}_{jj}}},$$

where $\boldsymbol{\Omega}_{ij}$ is the (i, j) -th entry of $\boldsymbol{\Omega}$. Therefore, $\boldsymbol{\Omega}_{ij} = 0$ if and only if X_i and X_j are conditionally uncorrelated given all other variables in the system.

The following section presents examples of precision matrices for several important distributions.

2.4 Precision Matrices for Common Multivariate Distributions

2.4.1 Elliptically Symmetric Distributions

A random variable $\mathbf{W}$ is said to have a *spherically symmetric distribution* if its distribution is invariant under rotations of $\mathbb{R}^d$. Such a

distribution admits the stochastic representation

$$\mathbf{W} = R\mathbf{U},$$

where $R \geq 0$ is a radial random variable, $\mathbf{U}$ is uniformly distributed on the unit sphere in $\mathbb{R}^d$, denoting that R and $\mathbf{U}$ are independent [50, 106]. The random variable R is called the generating variate, with the generating distribution F .

An *elliptically symmetric random variable* $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$ is defined by the location-scale extension of $\mathbf{W}$ so that

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{W},$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ is a variance-covariance matrix, $\mathbf{W}$ is spherically distributed, and the function h is called the density generator, and it depends on the Mahalanobis distance. The density function of $\mathbf{X}$ has the form

$$f_{h,\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c_d \boldsymbol{\Sigma}^{-1/2} h\left((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right),$$

where c_d is a constant.

Rewriting the equation of the elliptically symmetric random variable concerning

$$\mathbf{W} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}), R = \|\mathbf{W}\| = (\mathbf{W}^\top \mathbf{W})^{1/2}, \text{ and } \mathbf{U} = \frac{\mathbf{W}}{\|\mathbf{W}\|}.$$

Notice that h is not necessarily dependent on d but on R^2 . Hence, the density of $S = R^2$ is

$$f_{h,S}(s) = c_d \frac{\pi^{d/2} s^{d/2-1}}{\Gamma\left(\frac{d}{2}\right)} h(s); \quad s > 0. \quad (2.1)$$

Based on the generalized precision matrix framework introduced by Torri et al. [107] and Terdik [106], the generalized precision matrix $\boldsymbol{\Omega}$ for any elliptically symmetric distribution is given by:

$$\boldsymbol{\Omega} = \frac{4}{d} \mathbb{E} \left(\frac{h^{(1)2}(S)}{h^2(S)} S \right) \boldsymbol{\Sigma}^{-1}, \quad d \geq 2 \quad (2.2)$$

2.4.2 Multivariate Normal distribution

Let

$$\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{X}})$$

Then, the precision matrix is equal to the negative Hessian matrix of the log-density function $\log f_{\mathbf{X}}(\mathbf{x})$. Specifically,

$$\boldsymbol{\Omega} = -\text{vec}^{-1} D_{\mathbf{x}}^{\otimes 2} \log f_{\mathbf{X}}(\mathbf{x}). \quad (2.3)$$

The PDF of the multivariate normal distribution can be expressed in terms of the scalar quantity

$$h(s) = \exp\left(-\frac{s}{2}\right), \quad s = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

This representation emphasizes the radial structure of the density and explicitly expresses its dependence on s , rather than directly on the vector $\mathbf{X}$. Applying Equation 2.2, the precision matrix is defined as follows:

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} \quad (2.4)$$

2.4.3 Multivariate Student t distribution

Let $\mathbf{X} \in \mathbb{R}^d$ follow a multivariate student's t -distribution with location vector $\boldsymbol{\mu}$, scatter matrix $\boldsymbol{\Sigma}$, and ν degrees of freedom. Its PDF is given by

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \nu^{\frac{d}{2}} \pi^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{1/2}} \left(1 + \nu^{-1} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)^{-\frac{\nu+d}{2}} \quad (2.5)$$

and

$$h(s) = \left(1 + \frac{s}{\nu}\right)^{-\frac{\nu+d}{2}}$$

As the degrees of freedom $\nu \rightarrow 0$, the Student's t -distribution exhibits increasingly heavy tails. Applying Equation 2.2, the generalized precision matrix is given by:

$$\mathbf{\Omega} = \frac{\nu + d}{\nu + d + 2} \mathbf{\Sigma}^{-1} \quad (2.6)$$

2.4.4 Multivariate symmetric generalized hyperbolic distribution

The generalized hyperbolic (GH) family forms a class of normal variance mixtures defined by

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{V} \mathbf{\Sigma}^{1/2} \mathbf{Z},$$

where $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_d)$, $V \sim \mathcal{GIG}(\lambda, \chi, \psi)$ is a generalized inverse Gaussian (GIG) random variable independent of $\mathbf{Z}$, $\boldsymbol{\mu} \in \mathbb{R}^d$, and $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$ is a positive definite matrix. The parameters satisfy $\lambda \in \mathbb{R}$, $\psi > 0$, and $\chi > 0$, see [94, 113]. In the following, appropriate identifiability constraints are assumed during parameter estimation. The density of $\mathbf{X} \sim \mathcal{MSGH}(\lambda, \chi, \psi, \boldsymbol{\mu}, \mathbf{\Sigma})$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = c_{\lambda, \chi, \psi} |\mathbf{\Sigma}_d|^{-1/2} \frac{K_{\lambda-d/2} \left((\psi (\chi + [(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})]))^{1/2} \right)}{(\psi (\chi + [(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})]))^{(d/2-\lambda)/2}}, \quad (2.7)$$

where K_λ denotes a modified Bessel function of the third kind with index λ and the normalizing constant $c_{\lambda, \chi, \psi}$ is given by

$$c_{\lambda, \chi, \psi} = \frac{(\sqrt{\chi\psi})^{-\lambda} \psi^{d/2}}{(2\pi)^{d/2} K_\lambda(\sqrt{\chi\psi})}.$$

Applying Equation 2.2, the generalized precision matrix for the multivariate symmetric GH distribution is

$$\mathbf{\Omega} = \frac{\psi}{d} \mathbb{E} \left[\frac{S}{\chi + S} \left(\frac{K_{\lambda-d/2-1}(\sqrt{(\chi + S)\psi})}{K_{\lambda-d/2}(\sqrt{(\chi + S)\psi})} \right)^2 \right] \mathbf{\Sigma}^{-1},$$

where S denotes the squared Mahalanobis distance.

3

Advancements in Graphical LASSO Algorithms

3.1 Related Works

Over the past two decades, numerous penalized estimation methods have emerged that incorporate regularization terms into model parameters to enhance statistical performance. Friedman et al. [55] proposed the GLASSO framework for precision matrix estimation and graphical network construction. The method estimates sparse precision matrices by shrinking selected entries toward zero, thereby identifying conditional independence relationships among variables. GLASSO-based methods are primarily used for: (1) variable selection and coefficient shrinkage, and (2) inducing sparsity in precision matrix estimation.

This chapter presents the GLASSO framework for estimating Gaussian precision matrices. We then extend this methodology to selected elliptical distributions.

3.1.1 GLASSO

Building on Chapter 2, GGMs model conditional independence in Gaussian processes through undirected graphs [55, 85]. To capture these dependencies, we must estimate the graph's sparse structure and compute a regularized precision matrix.

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be an independent and identically distributed sample from a d -dimensional multivariate normal distribution, $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We define the precision matrix as $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$. The profiled log-likelihood function for $\boldsymbol{\Theta}$ is

$$\ell(\boldsymbol{\Theta}) = -\frac{nd}{2} \log(2\pi) + \frac{n}{2} \log |\boldsymbol{\Theta}| - \frac{n}{2} \text{tr}(\mathbf{S}\boldsymbol{\Theta}),$$

where $\mathbf{S}$ is the sample covariance matrix.

To estimate a sparse precision matrix, GLASSO maximizes a penalized log-likelihood. It imposes an ℓ_1 -norm penalty on Θ . This penalty acts as a convex relaxation of the ℓ_0 -norm and explicitly enforces structural sparsity. Because the objective function is strictly convex over the cone of positive definite matrices, we define the sparse estimator $\hat{\Theta}$ via its Lagrangian formulation:

$$\hat{\Theta} = \arg \max_{\Theta \succ 0} (\log |\Theta| - \text{tr}(\mathbf{S}\Theta) - \rho |\Theta|_1), \quad (3.1)$$

where the tuning parameter $\rho \geq 0$ controls the degree of penalization.

To get the first-order optimality condition, we differentiate $\ell(\mathbf{x})$ with respect to Θ . The matrix Θ is 2-symmetric and contains $n_{d,2} = d(d+1)/2$ distinct elements [105]. Let $\mathbf{v} = \mathbf{Q}_{d,2}^+ \text{vec } \Theta$, where $\mathbf{Q}_{d,2}^+ \in \mathbb{R}^{n_{d,2} \times d^2}$ denotes the 2-way elimination matrix and $\text{vec } \Theta = \mathbf{Q}_{d,2} \mathbf{v}$, $\mathbf{Q}_{d,2} \in \mathbb{R}^{d^2 \times n_{d,2}}$ is the duplication matrix. This yields the derivative:

$$D_{\mathbf{v}}^{\otimes} \ell(\mathbf{x}) = (\tau_{d,2}) \mathbf{Q}_{d,2}^+ \left(\text{vec } \Theta^{-1} - \text{vec } \mathbf{S} - \rho \text{vec sgn}(\Theta) \right), \quad (3.2)$$

where $\text{diag } \tau_{d,2}$ contains the diagonal norm-weights used to scale the distinct values.

To obtain the gradient equation, we replace $\mathbf{W} = \Theta^{-1}$ in the matrix form of Equation 3.2, then we get the following

$$\mathbf{W} - \mathbf{S} - \rho \text{sgn}(\Theta) = \mathbf{0} \quad (3.3)$$

Although ℓ_2 -regularization often yields superior predictive accuracy, the GLASSO framework specifically employs an ℓ_1 -norm penalty to induce structural sparsity in the estimated network. To compute this estimator, the gradient condition in Equation 3.3 is solved using a **block coordinate descent** algorithm. This approach iteratively optimizes over each row, and the corresponding column of the working covariance matrix $\mathbf{W}$ [17]. Upon convergence, this procedure yields the regularized Gaussian precision matrix estimator defined in Equation 1.1. Subsequently, the corresponding estimation procedure for the precision

matrix Θ is outlined in Algorithm 1.

Algorithm 1 GLASSO

```

1: Input:
2:    $\mathbf{X} \in \mathbb{R}^{n \times p}$  ▷ sampled data
3:    $\rho > 0$  ▷ regularization parameter
4:    $\epsilon > 0$  ▷ convergence tolerance
5: Output:
6:    $\Theta \in \mathbb{R}^{p \times p}$  ▷ precision matrix
7: procedure GLASSO( $\mathbf{X}, \rho, \epsilon$ )
8:   Compute empirical covariance:  $\mathbf{S} \leftarrow \frac{1}{n} \mathbf{X}^\top \mathbf{X}$ 
9:   Initialize  $\Theta \leftarrow \text{diag}(\mathbf{S}^{-1})$  or  $\Theta \leftarrow I_p$ 
10:  repeat
11:    for  $j \leftarrow 1$  to  $p$  do
12:      Partition matrices:
          
$$\Theta = \begin{pmatrix} \Theta_{11} & \theta_{12} \\ \theta_{12}^\top & \theta_{22} \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{s}_{12} \\ \mathbf{s}_{12}^\top & s_{22} \end{pmatrix}$$

13:      Solve GLASSO:
          
$$\beta^* \leftarrow \arg \min_{\beta} \left( \frac{1}{2} \beta^\top \Theta_{11} \beta - \beta^\top \mathbf{s}_{12} + \rho \|\beta\|_1 \right)$$

14:      Update  $\Theta$ :
          
$$\begin{aligned} \theta_{12} &\leftarrow -\Theta_{11} \beta^* \\ \theta_{22} &\leftarrow s_{22} + \beta^{*\top} \Theta_{11} \beta^* \end{aligned}$$

15:    end for
16:  until  $\|\Theta^{\text{new}} - \Theta^{\text{old}}\|_F < \epsilon$  ▷  $\Theta^{\text{new}}$  is the updated version of  $\Theta$ 
17:  return  $\Theta$ 
18: end procedure

```

3.2 EM-based Graphical LASSO Algorithms

This section extends the GLASSO framework to estimate sparse concentration matrices for non-Gaussian elliptical models. We focus specifically on elliptically symmetric distributions. Simulations and real-data

applications demonstrate that this EM-based approach is robust beyond the Gaussian setting.

Chapter 2 details the strict link between partial correlation and conditional independence in multivariate normal data. Kelker [68] and Baba et al. [83] explored this relationship across broader elliptically symmetric distributions. They proved that zero partial correlation guarantees conditional independence if and only if the data are multivariate normal. For non-Gaussian elliptical families, a zero partial correlation strictly implies zero conditional correlation.

The elliptically symmetric variates $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$ are defined by the location-scale extension of a spherically distributed variate $\mathbf{W}$ so that $\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{W}$, where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$ is a full rank variance matrix; see [50, 89]. Elliptically symmetric variates provide a broad framework for modeling dependence structures in multivariate data. Within this class, the Multivariate Symmetric Generalized Hyperbolic (MSGH) and Multivariate Power Exponential (MPE) distributions offer greater flexibility in capturing heavy tails, peakedness, and non-Gaussian behavior—making them particularly valuable for realistic modeling and robust inference in high-dimensional settings. These distributions play an essential role in modern statistical modeling, particularly in areas where data exhibit heavy tails, or kurtosis beyond Gaussian assumptions [87, 101, 113]. These flexible distributions extend the normal framework, enabling better modeling of complex real-world phenomena across finance, signal processing, environmental science, and machine learning [18, 96].

MSGH models provide a unifying structure that encompasses several important distributions, such as Student’s t , Laplace, etc., and hyperbolic distributions, offering versatility in capturing tail behavior [25, 41]. Similarly, the MPE family generalizes the Gaussian model by controlling kurtosis via shape parameters, enabling robust estimation under outlier contamination [38, 57].

Building on Chapter 2, Graphical models facilitate inference by representing dependencies through undirected graphs, where edges correspond to non-zero entries in the concentration matrix $\boldsymbol{\Theta}$. In high-dimensional settings, GLASSO is commonly used to induce sparsity in

Θ , enabling efficient recovery of the underlying graph structure. The resulting Θ identifies non-zero off-diagonal entries as conditional dependencies, while zero entries indicate conditionally uncorrelated associations under elliptical symmetric distributions.

Early foundation work by Ravikumar et al. [97] and Janková and van de Geer [66] established consistent, asymptotically normal estimation of sparse concentration matrices in high-dimensional Gaussian and non-Gaussian settings. [54] introduced a mixture of shifted asymmetric Laplace distributions to applied domains such as robust clustering and classification, demonstrating that models capable of accommodating elliptical and heavy-tailed distributions can outperform Gaussian models in capturing complex dependence structures and tail behavior. These findings highlight the limitations of Gaussian assumptions in practice and motivate the development of robust alternatives.

In response, a growing body of literature has focused on estimating sparse concentration matrices for high-dimensional, heavy-tailed, elliptically symmetric distributions. One prominent line of research replaces the Gaussian likelihood with heavy-tailed elliptical likelihoods, such as the multivariate Student- t distribution, to improve robustness to outliers and heavy tails [52] (see section 3.2.2). By explicitly modeling heavy tails, these approaches reduce sensitivity to extreme observations while retaining the interpretability of concentration-based graphical models.

A second mainstream approach adapts the penalty to modify the regularization structure, improving model selection and estimation accuracy. This includes the use of Sorted- ℓ_1 Penalized Estimators (SLOPE) and their graphical extensions, such as GSLOPE for the Gaussian models and (TSLOPE) for Student- t variants [99]. Related approaches incorporate adaptive or data-driven weights into the graphical LASSO penalty (WGLASSO), allowing different shrinkage of entries in the concentration matrix and improving recovery of the network structure under heavy-tailed noise.

A third contribution uses robust and shrinkage approaches, such as Tyler's shape estimator [31] or Maronna-type M-estimators [15], to obtain a robust estimate of the underlying covariance or scatter

matrix. Notable examples include Partial Correlation GLASSO (PC-GLASSO) [119], which regularizes partial correlations rather than covariance entries, thereby improving robustness and interpretability in the presence of heavy tails.

A fourth direction of research reformulates the estimation problem by using alternative objective functions that downweight outliers or contaminated observations. Divergence-based objective function, such as the γ -LASSO [92], replaces the log-likelihood with a robust divergence measure, while the trimmed objective function [116] explicitly remove a fraction of extreme observations during optimization. However, these approaches offer strong robustness guarantees under outliers; they may involve additional tuning parameters or computational complexity.

Finally, rank-based and nonparanormal transformation methods aim to relax distributional assumptions altogether by estimating dependence structures through correlation or copulas. Methods such as CLIME and T-CLIME, along with their spatial-sign extensions (SGLASSO and SCLIME), rely on constrained ℓ_1 minimization and rank-based transformations to achieve robustness against heavy tails [81]. These approaches are computationally attractive and scalable, making them well-suited for extremely high-dimensional settings, though they may sacrifice some efficiency under correctly specified parametric models. A summary of the previously highlighted methods, including their robustness mechanisms, adopted penalty terms, assumptions, and the main related reference, is provided in Table 3.1.

Existing methods struggle with rigid tails, model misspecification, and high-dimensional computational costs. To overcome these limits, we propose the EM-GLASSO framework. It exploits the equivalence between zero partial correlations and conditional correlations under linear prediction to recover sparse concentration matrices for heavy-tailed MSGH and MPE distributions. We formulate an EM-based penalized likelihood to ensure flexible tail modeling and guarantee robust, positive-definite estimates. The algorithm executes closed-form E-steps followed by penalized M-steps.

Table 3.1: Comparison of precision matrix estimation methods for non-Gaussian distributions, with robustness properties, penalties, assumptions, and key references.

Approach	Robustness	Penalty	Assumption	Ref.
tLASSO	Strong	ℓ_1 penalty	Multivariate t	[74]
Tslope	Strong	SLOPE (sorted ℓ_1)	Multivariate t	[98]
WGLASSO	Moderate to strong	Weighted ℓ_1	Weighted Gaussian	[104]
Tyler + shrinkage	Strong	Shrinkage ℓ_1	Elliptical	[31]
Rank-based nonparanormal	Moderate	ℓ_1	Gaussian copula	[115]
PC-GLASSO	Moderate	ℓ_1 on partial correlations	Gaussian	[28]
SCLIME	Low to moderate	spatial-sign variants (SCLIME)	heavy-tailed	[81]
CLIME, ACLIME	Low to moderate	Column-wise ℓ_1	Gaussian / sub-Gaussian	[27]
Model-agnostic	Moderate to strong	ℓ_1 tuning-free	Robust kernel-based	[108]
γ -LASSO	Strong	ℓ_1 + gamma-divergence	Robust divergence	[92]
Trimmed-gLASSO	Strong	ℓ_1 + trimming	Gaussian with trimming	[116]

3.2.1 Partial and Conditional Uncorrelations

In the context of elliptical distributions, a zero partial correlation implies zero conditional correlation, since the best predictor is linear [101]. This result follows from the linearity of the best predictor and is slightly more general than the corresponding Gaussian case. For notational convenience, Let $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}$,

Lemma 3.1. *Assume $\mathbf{Y} = [Y_i, Y_j, \mathbf{Y}_{\setminus ij}]$ is a multivariate process. Considering that the best prediction of Y_i with respect to $\mathbf{Y}_{\setminus i}$ is linear that is,*

$$\mathbb{E}(Y_i | \mathbf{Y}_{\setminus i}) = \hat{Y}_{i|\mathbf{Y}_{\setminus i}}.$$

If $Y_i \perp Y_j \nmid \mathbf{Y}_{\setminus ij}$, then $Y_i \perp Y_j | \mathbf{Y}_{\setminus ij}$. In other words, if Y_i and Y_j are partially uncorrelated, then they are conditionally uncorrelated as well.

Proof. Let us begin by assuming that the conditional expectation of Y_i is independent of Y_j , given the remaining variables $\mathbf{Y}_{\setminus ij}$:

$$\mathbb{E}(Y_i | Y_j, \mathbf{Y}_{\setminus ij}) = \mathbb{E}(Y_i | \mathbf{Y}_{\setminus ij}). \quad (3.4)$$

This assumption implies that Y_i and Y_j are conditionally uncorrelated. By applying the law of iterated expectations, we can expand the product of the conditional means as follows:

$$\begin{aligned} \mathbb{E}(Y_i | \mathbf{Y}_{\setminus ij}) \mathbb{E}(Y_j | \mathbf{Y}_{\setminus ij}) &= \mathbb{E}(Y_j \mathbb{E}(Y_i | \mathbf{Y}_{\setminus ij}) | \mathbf{Y}_{\setminus ij}) \\ &= \mathbb{E}(Y_j \mathbb{E}(Y_i | Y_j, \mathbf{Y}_{\setminus ij}) | \mathbf{Y}_{\setminus ij}) \\ &= \mathbb{E}(\mathbb{E}(Y_i Y_j | Y_j, \mathbf{Y}_{\setminus ij}) | \mathbf{Y}_{\setminus ij}) \\ &= \mathbb{E}(Y_i Y_j | \mathbf{Y}_{\setminus ij}). \end{aligned}$$

Recall that the conditional covariance between Y_i and Y_j is defined by the following relation

$$C_{Y_i, Y_j | \mathbf{Y}_{\setminus ij}} = \mathbb{E}(Y_i Y_j | \mathbf{Y}_{\setminus ij}) - \mathbb{E}(Y_i | \mathbf{Y}_{\setminus ij}) \mathbb{E}(Y_j | \mathbf{Y}_{\setminus ij}),$$

Substituting the result from our previous derivation, it follows di-

rectly from equation (3.4) that:

$$\text{Cov}\left(Y_i, Y_j | \mathbf{Y}_{\setminus ij}\right) = 0,$$

To complete the proof, we must show that the condition $C_{Y_i, Y_j | \mathbf{Y}_{\setminus ij}} = \mathbf{0}$ fulfills the original assumption. We represent the predictor as:

$$\hat{Y}_{i| [Y_j, \mathbf{Y}_{\setminus ij}]} = \hat{Y}_{i| \mathbf{Y}_{\setminus ij}},$$

Using the Additivity Property of the LLSP, we can decompose the predictor of Y_i into two orthogonal components:

$$\hat{Y}_{i| [Y_j, \mathbf{Y}_{\setminus ij}]} = \hat{Y}_{i| \mathbf{Y}_{\setminus ij}} + \hat{Y}_{i| \left(Y_j - \hat{Y}_{j| \mathbf{Y}_{\setminus ij}}\right)},$$

Additivity of LLSP. Therefore, in one way

$$\hat{Y}_{i| [Y_j, \mathbf{Y}_{\setminus ij}]} = \hat{Y}_{i| [Y_j, \mathbf{Y}_{\setminus ij}]} - \hat{Y}_{i| \mathbf{Y}_{\setminus ij}}$$

The second term in this decomposition can be expressed explicitly using the conditional covariance and the inverse variance of the residuals:

$$\hat{Y}_{i| \left(Y_j - \hat{Y}_{j| \mathbf{Y}_{\setminus ij}}\right)} = C_{Y_i, Y_j | \mathbf{Y}_{\setminus ij}} \Sigma_{Y_j | \mathbf{Y}_{\setminus ij}}^{-1} \left(Y_j - \hat{Y}_{j| \mathbf{Y}_{\setminus ij}}\right),$$

Now, If the conditional covariance $C_{Y_i, Y_j | \mathbf{Y}_{\setminus ij}}$ is zero (indicating zero partial correlation), then the adjustment term vanishes:

$$0 = \hat{Y}_{i| [Y_j, \mathbf{Y}_{\setminus ij}]} - \hat{Y}_{i| \mathbf{Y}_{\setminus ij}} = \mathbb{E}\left(Y_i | Y_j, \mathbf{Y}_{\setminus ij}\right) - \mathbb{E}\left(Y_i | \mathbf{Y}_{\setminus ij}\right).$$

Thus, whenever the partial correlation is zero, our initial assumption is satisfied, and we conclude that $Y_i \perp Y_j | \mathbf{Y}_{\setminus ij}$. □

3.2.2 t LASSO

Consider the multivariate Student's t -distribution with the PDF in Equation 2.5, where $\delta_{\mathbf{y}} = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$.

We can represent this distribution via a scale mixture. Let $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ is a multivariate normal vector and $\tau \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ be an

independent univariate gamma variable. The transformed variable $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{X}/\sqrt{\tau}$ exactly follows the multivariate Student's t -distribution defined in Equation (2.5). Its mean and variance are:

$$\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} \quad \text{and} \quad \text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma} \frac{\nu}{\nu - 2}$$

It is known that the conditional distribution of $\mathbf{Y}|\tau \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau)$ which can be represented as

$$f(\mathbf{y}|\tau) = \frac{\tau^{d/2}}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{\tau}{2} \delta_{\mathbf{y}}\right) \quad (3.5)$$

On the other side, the conditional distribution of $\tau|Y$ can be derived using Bayes' Theorem. The result is shown in the following steps

$$\begin{aligned} f(\tau|\mathbf{y}) &= \frac{f(\mathbf{y}|\tau) \times f(\tau)}{f(\mathbf{y})} \\ &= \frac{\left(\frac{\nu}{2}\right)^{(\nu+d)/2}}{\Gamma\left(\frac{\nu+d}{2}\right)} \exp\left(-\frac{\nu + \delta_{\mathbf{y}}}{2} \tau\right) \left(1 + \frac{\delta_{\mathbf{y}}}{\nu}\right)^{(\nu+d)/2} \tau^{(\nu+d)/2-1} \\ &= \frac{\tau^{(\nu+d)/2-1}}{\Gamma\left(\frac{\nu+d}{2}\right)} \exp\left(-\frac{\nu + \delta_{\mathbf{y}}}{2} \tau\right) \left(\frac{\nu + \delta_{\mathbf{y}}}{2}\right)^{(\nu+d)/2} \\ &= \frac{\left(\frac{\nu + \delta_{\mathbf{y}}}{2}\right)^{(\nu+d)/2}}{\Gamma\left(\frac{\nu+d}{2}\right)} \tau^{(\nu+d)/2-1} \exp\left(-\frac{\nu + \delta_{\mathbf{y}}}{2} \tau\right) \end{aligned} \quad (3.6)$$

Therefore, $\tau|Y \sim \Gamma\left(\frac{\nu+d}{2}, \frac{\nu + \delta_{\mathbf{y}}}{2}\right)$.

By conditioning Y on τ , we recover key properties of the Gaussian distribution. Under this conditioning, a zero entry in the precision matrix strictly guarantees zero conditional correlation, rather than merely zero partial correlation. Therefore, if $\theta_{ij} = 0$, the variables y_i and y_j are conditionally uncorrelated given the remaining variables, $Y_{\setminus ij}$. The conditional variance of Y_i given $Y_{\setminus ij}$ is:

$$\text{Var}(y_i|\mathbf{Y}_{\setminus i}) = \mathbb{E}\left(\frac{x_i^2}{\tau}|\mathbf{Y}_{\setminus i}\right) = \mathbb{E}(x_i^2|\mathbf{Y}_{\setminus i}) \times \mathbb{E}\left(\frac{1}{\tau}|\mathbf{Y}_{\setminus i}\right)$$

We know that

$$\begin{aligned}\mathbb{E}\left(x_i^2|\mathbf{Y}_{\setminus i}\right) &= \mathbb{E}\left(x_i^2|\mathbf{X}_{\setminus i}, \tau\right) = \mathbb{E}\left(x_i^2|\mathbf{X}_{\setminus i}\right) \\ &= \theta_{ii}^{-1}\end{aligned}$$

Generally, if $\tau|\mathbf{Y}_{\setminus i} \sim \Gamma(\alpha, \beta)$ then

$$\begin{aligned}\mathbb{E}\left(\frac{1}{\tau}|\mathbf{Y}_{\setminus i}\right) &= \int_0^\infty \frac{1}{\tau} f\left(\tau|\mathbf{Y}_{\setminus i}\right) d\tau \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \tau^{\alpha-2} e^{-\beta\tau} d\tau\end{aligned}$$

From the definition of the gamma function

$$\int_0^\infty \tau^{\alpha-2} e^{-\beta\tau} d\tau = \frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left(\frac{1}{\tau}|\mathbf{Y}_{\setminus i}\right) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}} \\ &= \frac{\beta}{\alpha-1}\end{aligned}$$

For our situation, we directly apply with the value of α and β , and we get

$$\mathbb{E}\left(\frac{1}{\tau}|\mathbf{Y}_{\setminus i}\right) = \frac{\nu + \delta_{\mathbf{Y}_{\setminus i}}}{\nu + d - 2}$$

Therefore,

$$\text{Var}\left(y_i|\mathbf{Y}_{\setminus i}\right) = \theta_{ii}^{-1} \times \frac{\nu + \delta_{\mathbf{Y}_{\setminus i}}}{\nu + d - 2}$$

Having a penalized log-likelihood of (3.5), derived as

$$\begin{aligned}& \ell_t(\boldsymbol{\mu}, \boldsymbol{\Theta}; \mathbf{y}, \tau) \\ &= \sum_{i=1}^n \log f_{\mathbf{y}|\tau}(\mathbf{y}_i|\tau; \boldsymbol{\mu}, \boldsymbol{\Theta}) - \rho \|\boldsymbol{\Theta}\|_1 \\ &= \text{const} + \frac{n}{2} \log(|\boldsymbol{\Theta}|) - \frac{1}{2} \sum_{i=1}^n \text{tr}[\boldsymbol{\Theta} \tau_i (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top] - \rho \|\boldsymbol{\Theta}\|_1\end{aligned}$$

Hence, it can be written as

$$\ell_t(\boldsymbol{\mu}, \boldsymbol{\Theta}; \mathbf{y}, \tau) = \frac{n}{2} \log |\boldsymbol{\Theta}| - \frac{1}{2} \text{tr} \left[\boldsymbol{\Theta} \sum_{i=1}^n \tau_i (\mathbf{y}_i - \boldsymbol{\mu}) (\mathbf{y}_i - \boldsymbol{\mu})^\top \right] - \rho \|\boldsymbol{\Theta}\|_1$$

Since the middle term is linear in τ , the conditional distribution of Y given τ is available in closed form. Exploiting this property, the EM algorithm can be applied iteratively to optimize the objective function and obtain the GLASSO estimate of $\boldsymbol{\Theta}$.

EM algorithm

We start the algorithm by finding the conditional expectation from (3.6)

$$\mathbb{E}(\tau_i | \mathbf{y}_i) = \frac{\alpha}{\beta} = \frac{\nu + d}{\nu + \delta_{\mathbf{y}_i}}$$

which represents the expectation stage in the algorithm at iteration $k + 1$, where the $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{\Theta}^{(k)}$ of the i th sample were carried from the previous iteration. This conditional expectation represents the weight that is going to affect the estimated mean and the covariance as well in iteration $k + 1$, thus, it will be denoted as τ and it is given by

$$\tau_i^{(k+1)} = \mathbb{E}(\tau_i | \mathbf{y}_i, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Theta}^{(k)}) = \frac{\nu + d}{\nu + \delta_{\mathbf{y}_i}}$$

Moreover, the maximization step targets maximizing $\boldsymbol{\mu}^{(k+1)}$ and $\boldsymbol{\Theta}^{(k+1)}$ given the weight $\tau_i^{(k+1)}$. From the formula, it is clear that $\boldsymbol{\mu}^{(k+1)}$ depends only on $\tau_i^{(k+1)}$ and the observed sample $\mathbf{y}_i$

$$\boldsymbol{\mu}^{(k+1)} = \frac{\sum_{i=1}^n \tau_i^{(k+1)} \mathbf{y}_i}{\sum_{i=1}^n \tau_i^{(k+1)}}$$

Let us represent the middle term as S , having $\tau_i^{(k+1)}$ and $\boldsymbol{\mu}^{(k+1)}$ it can be rewrite as

$$\mathbf{S}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \tau_i^{(k+1)} (\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)}) (\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)})^\top$$

As a result, the log-likelihood function subject to ℓ_1 -norm penalty represents our objective function to be maximized, it can be rewritten as

$$\ell_t(\boldsymbol{\mu}, \boldsymbol{\Theta}; \mathbf{y}, \tau) = \frac{n}{2} \log(|\boldsymbol{\Theta}|) - \frac{n}{2} \text{tr} \left[\boldsymbol{\Theta} \mathbf{S}^{(k+1)} \right] - \rho \|\boldsymbol{\Theta}\|_1$$

Successfully, after the EM steps, we ended up with an objective function that has a closed-form equivalent to the objective function of the Gaussian distribution. Therefore, Block-wise coordinate descent can be used to maximize $\boldsymbol{\Theta}^{(k+1)}$.

$$\boldsymbol{\Theta}^{(k+1)} = \arg \max_{\boldsymbol{\Theta}} \left[\log(|\boldsymbol{\Theta}|) - \text{tr} \left[\boldsymbol{\Theta} \mathbf{S}^{(k+1)} \right] - \rho \|\boldsymbol{\Theta}\|_1 \right] \quad (3.7)$$

The algorithm is applied repeatedly until the difference $\boldsymbol{\Theta}^{(k)} - \boldsymbol{\Theta}^{(k+1)}$ is less than a specific threshold. See Algorithm 2, where the weight $\tau_i^{(k+1)} = (\nu + d) / (\nu + \delta_{\mathbf{y}_i})$.

3.2.3 Multivariate Symmetric Generalized Hyperbolic

The MSGH distributions are constructed as a normal variance mixture with Generalized Inverse Gaussian (GIG) weights (see [87], Example 3.8). As a subfamily of elliptically symmetric variates, the MSGH distribution ensures that zeros within the concentration matrix imply conditional uncorrelation between random variables.

The GIG mixing variates encompass several special cases, including the Gamma (G), Inverse Gamma (IGam), Exponential (Exp), Positive Hyperbolic (pHyp), Lévy, Inverse Gaussian (IG), and the Dirac distribution as a limiting case (see [94], section 9.4) as illustrated in Figure 3.1. In this section, we provide an explicit formula for predicting the weight variable, which facilitates the implementation of the LASSO algorithm.

Let us start with the stochastic representation of an MSGH variate $\mathbf{X}$ with parameters $\mathcal{MSGH}(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{V} \boldsymbol{\Sigma}^{1/2} \mathbf{Z},$$

where $\mathbf{Z} \in \mathcal{N}(0, \mathbf{I}_d)$, $V \in \mathcal{GIG}(\lambda, \chi, \psi)$ is a non-negative scalar-valued

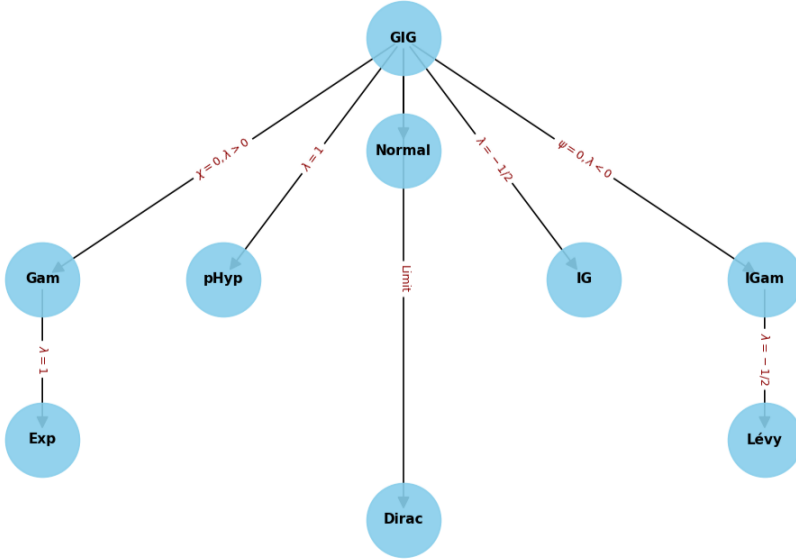


Figure 3.1: The subclasses of the GIG distribution

variate, which is independent on $\mathbf{Z}$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ is a positive definite matrix, $\lambda \in \mathbb{R}$, $\psi > 0$, $\chi > 0$, see [94] for further properties and [113] for different parameterizations. We assume throughout that the standard identifiability constraints are satisfied during parameter estimation. Equation 2.7 defines the PDF for $\mathbf{X} \sim \text{MSGH}(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. A defining property of the MSGH family is its scale-mixture representation. In this framework, a latent mixing variable strictly controls the heavy-tailed behavior. Specifically, let

$$V \sim \text{GIG}(\lambda, \chi, \psi),$$

with density

$$f_V(x) = c_V x^{\lambda-1} \exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right), \quad c_V = \frac{\chi^{-\lambda} (\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})}.$$

To obtain a more convenient formulation for inference, we apply the transformation $V = 1/\tau$, with Jacobian $dv = \tau^{-2}d\tau$. This yields the

equivalent stochastic representation

$$\mathbf{X} = \boldsymbol{\mu} + \frac{1}{\sqrt{\tau}} \boldsymbol{\Sigma}^{1/2} \mathbf{Z}, \quad (3.8)$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and the latent variable τ follows the density

$$f_{\tau}(x) = \frac{c_V}{2} x^{(-\lambda+1)/2-2} \exp\left(-\frac{1}{2}(\chi x + \psi x^{-1})\right). \quad (3.9)$$

This representation reveals an important structural insight: conditional on τ , the random vector $\mathbf{X}$ is Gaussian. Indeed, given τ , the conditional density of $\mathbf{X}$ is

$$f(\mathbf{x}|\tau) = \frac{\tau^{d/2}}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{\tau}{2} \delta(\mathbf{x})\right).$$

This conditional Gaussianity is central to the EM-based estimation framework developed later.

Combining the conditional Gaussian likelihood with the prior density of τ , Bayes' theorem yields the conditional distribution of τ given $\mathbf{X} = \mathbf{x}$:

$$\begin{aligned} f(\tau|\mathbf{x}) &= \frac{f(\mathbf{x}|\tau)f(\tau)}{f(\mathbf{x})} \\ &= c\tau^{d/2} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{\tau}{2} \delta(\mathbf{x})\right) \tau^{-\lambda-1} \exp\left(-\frac{1}{2}(\chi\tau + \psi\tau^{-1})\right) \\ &\quad \times |\boldsymbol{\Sigma}_d|^{1/2} \frac{(\psi(\chi + \delta(\mathbf{x})))^{(d/2-\lambda)/2}}{K_{\lambda-d/2}((\psi(\chi + \delta(\mathbf{x})))^{1/2})} \\ &= c \frac{(\psi(\chi + \delta(\mathbf{x})))^{(d/2-\lambda)/2}}{K_{\lambda-d/2}((\psi(\chi + \delta(\mathbf{x})))^{1/2})} \tau^{d/2-\lambda-1} \exp\left(-\frac{1}{2}((\chi + \delta(\mathbf{x}))\tau + \psi\tau^{-1})\right), \end{aligned}$$

or in short

$$f(\tau | \mathbf{x}) \propto \tau^{d/2-\lambda-1} \exp\left(-\frac{1}{2}((\chi + \delta(\mathbf{x}))\tau + \psi\tau^{-1})\right).$$

Up to a normalizing constant, this density corresponds to a gener-

alized inverse Gaussian distribution. Consequently,

$$\tau | \mathbf{X} \sim \mathcal{GIG}(\lambda_\tau, \chi_\tau, \psi_\tau)$$

with updated parameters

$$\lambda_\tau = \frac{d}{2} - \lambda, \quad \chi_\tau = \psi, \quad \psi_\tau = \chi + \delta(\mathbf{x})$$

The moments of a $\mathcal{GIG}(\lambda, \chi, \psi)$ random variable are available in closed form (see [94], Section 9.4). In particular,

$$\mathbb{E}[V^r] = \eta^r \frac{K_{\lambda+r}(\tilde{\omega})}{K_\lambda(\tilde{\omega})},$$

where $\eta = \sqrt{\chi/\psi}$ and $\tilde{\omega} = \sqrt{\chi\psi}$. Applying this result to the conditional distribution of $\tau | \mathbf{X}$ yields

$$\begin{aligned} \mathbb{E}[\tau | \mathbf{X}] &= \sqrt{\frac{\chi_\tau}{\psi_\tau}} \frac{K_{\lambda_\tau+1}(\sqrt{\chi_\tau \psi_\tau})}{K_{\lambda_\tau}(\sqrt{\chi_\tau \psi_\tau})} \\ &= \sqrt{\frac{\psi}{\chi + \delta(\mathbf{x})}} \frac{K_{d/2-\lambda+1}(\sqrt{\psi(\chi + \delta(\mathbf{x}))})}{K_{d/2-\lambda}(\sqrt{\psi(\chi + \delta(\mathbf{x}))})}. \end{aligned}$$

This closed-form expression plays a crucial role in the E-step of the proposed EM algorithm, enabling efficient and stable estimation in high-dimensional heavy-tailed settings.

Based on the previous result, we get the following Lemma.

Lemma 3.2. *Let $\mathbf{X} \sim \mathcal{MSGH}(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ admit the stochastic representation (3.8) with mixing density given in (3.9). Then, the best predictor of the latent weight variable τ , given $\mathbf{X} = \mathbf{x}$, is*

$$\hat{\tau}(\mathbf{x}) = \sqrt{\frac{\psi}{\chi + \delta(\mathbf{x})}} \frac{K_{\frac{d}{2}-\lambda+1}(\sqrt{\psi(\chi + \delta(\mathbf{x}))})}{K_{\frac{d}{2}-\lambda}(\sqrt{\psi(\chi + \delta(\mathbf{x}))})}. \quad (3.10)$$

Remark 3.3. To show that $\hat{\tau} = \mathbb{E}[\tau | \mathbf{X}]$ of a $\mathcal{GIG}(\lambda = -\nu/2, \chi = \nu, \psi)$ distribution converges to the t -distribution predictor as $\psi \rightarrow 0$. The

predictor is given by the ratio of modified Bessel functions:

$$\hat{\tau} = \sqrt{\frac{\psi}{\nu + \delta(\mathbf{x})}} \frac{K_{d/2+\nu/2+1}(\sqrt{\psi(\nu + \delta(\mathbf{x}))})}{K_{d/2+\nu/2}(\sqrt{\psi(\nu + \delta(\mathbf{x}))})}. \quad (3.11)$$

To evaluate the limit as $\psi \rightarrow 0$, let $z = \sqrt{\psi(\nu + \delta(\mathbf{x}))}$. As $\psi \rightarrow 0$, $z \rightarrow 0$. We apply the asymptotic formula for $K_\nu(z)$ as $z \rightarrow 0$:

$$K_\nu(z) \sim \frac{1}{2} \Gamma(\nu) \left(\frac{1}{2}z\right)^{-\nu}.$$

The ratio of the Bessel functions in the limit becomes:

$$\begin{aligned} \lim_{z \rightarrow 0} \frac{K_{d/2+\nu/2+1}(z)}{K_{d/2+\nu/2}(z)} &= \frac{\frac{1}{2} \Gamma(d/2 + \nu/2 + 1) \left(\frac{z}{2}\right)^{-(d/2+\nu/2+1)}}{\frac{1}{2} \Gamma(d/2 + \nu/2) \left(\frac{z}{2}\right)^{-(d/2+\nu/2)}} \\ &= \frac{d + \nu}{z}. \end{aligned}$$

Substituting $z = \sqrt{\psi(\nu + \delta(\mathbf{x}))}$ back into the full expression for $\hat{\tau}$:

$$\begin{aligned} \lim_{\psi \rightarrow 0} \hat{\tau} &= \lim_{\psi \rightarrow 0} \sqrt{\frac{\psi}{\nu + \delta(\mathbf{x})}} \times \frac{d + \nu}{\sqrt{\psi(\nu + \delta(\mathbf{x}))}} \\ &= \frac{d + \nu}{\nu + \delta(\mathbf{x})}. \end{aligned}$$

This result coincides exactly with the prediction $\hat{\tau}$ for the multivariate t -distribution.

EM-GLASSO_H estimator Consider the model in (3.8). Assume that the distribution of $\mathbf{X}$ and its associated parameters are known (or have been estimated), and that a closed-form expression for the conditional expectation $\hat{\tau} = \mathbb{E}(\tau \mid \mathbf{X})$ is available. In our case, suppose that $\mathbf{X}$ is a d -dimensional multivariate normal random vector, $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In this setting, the GLASSO-regularized estimator $\hat{\boldsymbol{\Theta}}$ is obtained as follows:

$$\hat{\boldsymbol{\Theta}} = \arg \max_{\boldsymbol{\Theta}} \left(\log |\boldsymbol{\Theta}| - \text{tr} [\boldsymbol{\Theta} \hat{\mathbf{S}}_{\mathbf{x}}] - \rho \|\boldsymbol{\Theta}\|_1 \right), \quad (3.12)$$

where $\hat{\mathbf{S}}_{\mathbf{x}} = 1/n \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^\top$, $\hat{\boldsymbol{\mu}} = 1/n \sum_{i=1}^n \mathbf{x}_i$, and $\text{tr}(\cdot)$ denotes the trace of a matrix.

Note that if the weight variable τ in the normal variance mixture model (3.8) is known, the corresponding GLASSO estimator is obtained by maximizing the Gaussian log-likelihood, conditional on τ

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Theta} | \mathbf{x}, \tau) \simeq \frac{1}{2} \log |\boldsymbol{\Theta}| - \tau(\mathbf{x}) \delta(\mathbf{x}),$$

where

$$\delta(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Theta} (\mathbf{x} - \boldsymbol{\mu}) = \text{tr}[\boldsymbol{\Theta} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top],$$

and irrelevant additive constants are omitted. Then, having $\hat{\tau}_i = \hat{\tau}(\mathbf{x}_i)$,

$$\hat{\boldsymbol{\mu}} = \frac{1}{\sum_{i=1}^n \hat{\tau}_i} \sum_{i=1}^n \hat{\tau}_i \mathbf{x}_i \quad \text{and} \quad \hat{\mathbf{S}}_{\mathbf{x}, \tau} = \frac{1}{n} \sum_{i=1}^n \hat{\tau}_i (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^\top.$$

Hence, the GLASSO estimate of $\boldsymbol{\Theta}$ becomes

$$\hat{\boldsymbol{\Theta}} = \arg \max_{\boldsymbol{\Theta}} \left(\frac{n}{2} \log |\boldsymbol{\Theta}| - \frac{n}{2} \text{tr}[\boldsymbol{\Theta} \hat{\mathbf{S}}_{\mathbf{x}, \tau}] - \rho \|\boldsymbol{\Theta}\|_1 \right),$$

When the latent variable τ is unobserved, it is replaced by its conditional expectation, $\hat{\tau} = \mathbb{E}(\tau | \mathbf{X})$, computed from the observed data. This substitution naturally leads to the EM algorithm for parameter estimation.

As an initialization step, the concentration matrix is first estimated by $\hat{\boldsymbol{\Theta}}_0$ using (3.12), corresponding to the standard GLASSO estimator.

E-step. Use the current estimate $\hat{\boldsymbol{\Theta}}_k$ to calculate $\hat{\tau}_i, i = 1 : n$ by (3.10), put

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{\sum_{i=1}^n \hat{\tau}_i} \sum_{i=1}^n \hat{\tau}_i \mathbf{x}_i \quad \text{and} \quad \hat{\mathbf{S}}_{\mathbf{x}, \tau} = \frac{1}{n} \sum_{i=1}^n \hat{\tau}_i (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)^\top$$

M-step. Obtain the next estimate $\hat{\boldsymbol{\Theta}}_{i+1}$ of the concentration matrix

by

$$\hat{\Theta}_{i+1} = \arg \max_{\Theta} \left(\frac{n}{2} \log |\Theta| - \frac{n}{2} \text{tr} [\Theta \hat{\mathbf{S}}_{\mathbf{x}, \tau}] - \rho \|\Theta\|_1 \right).$$

The iteration of these E- and M-steps will be referred to as the EM-GLASSO algorithm. Then, one can follow the algorithm EM-GLASSO_H for estimating the zeros of the concentration matrix. The pseudocode for the algorithm is given in Algorithm 2.

3.2.4 Multivariate Power Exponential Distribution

If $\mathbf{x}_{1:n}$ is a random vector of dimension d , then the multivariate power exponential distribution for $\mathbf{x}$ is a special case of the elliptically contoured distributions with the formula given as follows

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta) = \frac{d\Gamma\left(\frac{d}{2}\right)}{\pi^{d/2} |\boldsymbol{\Sigma}|^{1/2} \Gamma\left(1 + \frac{d}{2\beta}\right) 2^{1+\frac{d}{2\beta}}} \exp\left(-\frac{1}{2} \delta(\mathbf{x})^\beta\right), \quad (3.13)$$

where $\beta > 0$ denotes the shape parameter. This model yields the Gaussian distribution at $\beta = 1$ and the Laplace distribution at $\beta = 1/2$. As $\beta \rightarrow \infty$, it converges to the uniform distribution [57, 77]. Setting $\beta < 1$ produces heavy tails (Figure 3.2). Because this model is a special case of elliptically contoured distributions, Lemma 3.1 strictly applies.

Let us consider having Laplace distribution $\mathcal{MPE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, 1/2)$, whose probability density function is given by

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, 1/2) = \frac{d\Gamma(d/2)}{\pi^{d/2} 2^{d+1} |\boldsymbol{\Sigma}|^{1/2} \Gamma(d+1)} \exp\left(-\frac{1}{2} [\delta(\mathbf{x})]^{1/2}\right),$$

with

$$\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu} \quad \text{and} \quad \text{Var}(\mathbf{X}) = 4(d+1) \boldsymbol{\Sigma}$$

The variable $\mathbf{X}$ can be stochastically represented as

$$\mathbf{X} = \boldsymbol{\mu} + V \boldsymbol{\Sigma}^{1/2} \mathbf{Z},$$

Algorithm 2 Implementation of the EM-GLASSO.

1: **Input:**
 2: Data: $\mathbf{X}_{1:n} \in \mathbb{R}^{n \times d}$
 3: Function of the weight variate $\widehat{\tau}(\mathbf{x}) = \mathbb{E} \tau | \mathbf{x}$ be given
 4: Regularization parameter: $\rho > 0$
 5: Convergence threshold: ϵ
 6: **Output:** $\boldsymbol{\mu}, \boldsymbol{\Theta}$
 7: **Initialization:**
 8: Initialize $\boldsymbol{\mu}^{(0)}, \boldsymbol{\Theta}^{(0)}, t = 0$
 9: **repeat**
 10: **Expectation step:**
 11: **for** $i = 1$ to n **do**
 12: Compute Mahalanobis distance:

$$\delta^{(t)}(\mathbf{x}_i) = (\mathbf{x}_i - \boldsymbol{\mu}^{(t)})^\top \boldsymbol{\Theta}^{(t)} (\mathbf{x}_i - \boldsymbol{\mu}^{(t)})$$

13: Compute Weight: $\widehat{\tau}^{(t+1)}(\mathbf{x}_i)$
 14: **end for**
 15: **Maximization step:**
 16: Update the weighted mean vector:

$$\boldsymbol{\mu}^{(t+1)} = \frac{\sum_{i=1}^n \widehat{\tau}^{(t+1)}(\mathbf{x}_i) \mathbf{x}_i}{\sum_{i=1}^n \widehat{\tau}^{(t+1)}(\mathbf{x}_i)}$$

17: Update the weighted empirical covariance matrix:

$$\mathbf{S}^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \tau^{(t+1)}(\mathbf{x}_i) (\mathbf{x}_i - \boldsymbol{\mu}^{(t+1)}) (\mathbf{x}_i - \boldsymbol{\mu}^{(t+1)})^\top$$

18: Maximize the concentration matrix with ℓ_1 penalty:

$$\boldsymbol{\Theta}^{(t+1)} = \arg \max_{\boldsymbol{\Theta} \succ 0} (\log |\boldsymbol{\Theta}| - \text{tr}(\mathbf{S}^{(t+1)} \boldsymbol{\Theta}) - \rho \|\boldsymbol{\Theta}\|_1)$$

19: **Convergence Check:**
 20: **if** $\|\boldsymbol{\Theta}^{(t+1)} - \boldsymbol{\Theta}^{(t)}\|_F < \epsilon$ **then**
 21: **Break**
 22: **end if**
 23: Increment $t = t + 1$
 24: **until** convergence
 25: **Return:** $\boldsymbol{\mu} = \boldsymbol{\mu}^{(t+1)}, \boldsymbol{\Theta} = \boldsymbol{\Theta}^{(t+1)}$

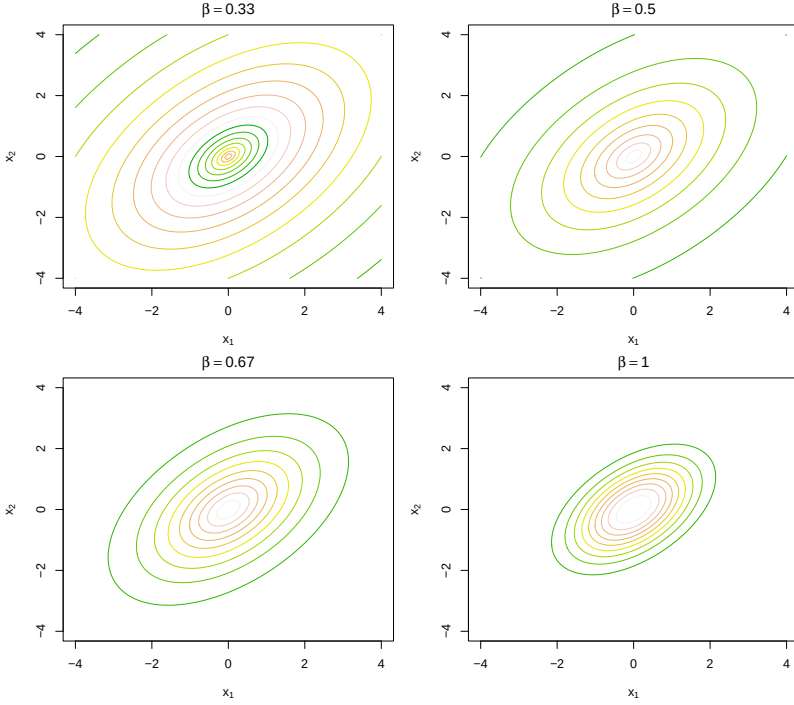


Figure 3.2: Contour plots of the MPE distribution

where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\mathbf{Z} \in \mathcal{N}(0, \mathbf{I}_d)$, $V > 0$, is independent of $\mathbf{Z}$. Here, the variate V is connected to the Lévy distribution [58],

$$f_{V,1/2}(x) = \frac{1}{2^{\frac{3d+1}{2}} \Gamma\left(\frac{d+1}{2}\right)} x^d e^{-x^2/8}, 0 < x < \infty,$$

We apply the transformation $V = 1/\sqrt{\tau}$ with Jacobian $dx = 1/2\tau^{-3/2}d\tau$, and obtain

$$\mathbf{X} = \boldsymbol{\mu} + \frac{1}{\sqrt{\tau}} \boldsymbol{\Sigma}^{1/2} \mathbf{Z},$$

where the τ has the generalized Gamma density

$$f_{\tau,1/2}(\tau) = \frac{1}{2^{\frac{3(d+1)}{2}} \Gamma\left(\frac{d+1}{2}\right)} \tau^{-\frac{d+3}{2}} e^{-\frac{1}{8\tau}},$$

Now, the conditional distribution of $\mathbf{X}$ is $\mathbf{X}|\tau \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau)$ therefore

$$f_{\mathbf{X}|\tau}(\mathbf{x}|\tau) = \frac{\tau^{d/2}}{(2\pi)^{d/2}} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{\tau}{2} \delta(\mathbf{x})\right).$$

To find the conditional expectation $E[\tau|\mathbf{X}]$ we consider

$$\begin{aligned} f_{\tau|\mathbf{X}}(\tau|\mathbf{x}) &= \frac{f_{\mathbf{X}|\tau}(\mathbf{x}|\tau) f_{\tau,1/2}(\tau)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{\Gamma(d) \tau^{d/2} \tau^{-(d+3)/2} e^{-1/8\tau}}{2^{d+1/2} \Gamma(d/2 + 1/2) \Gamma(d/2)} \exp\left(-\frac{\tau}{2} \delta(\mathbf{x}) + \frac{1}{2} \sqrt{\delta(\mathbf{x})}\right) \\ &= \frac{\Gamma(d) \tau^{-3/2} e^{\sqrt{\delta(\mathbf{x})}/2}}{2^{d+1/2} \Gamma(d/2 + 1/2) \Gamma(d/2)} \exp\left(-\frac{1}{2} \left(\tau \delta(\mathbf{x}) + \frac{1}{4\tau}\right)\right) \end{aligned}$$

The conditional density $f_{\tau|\mathbf{X}}$ proves to be inverse Gaussian $\mathcal{IG}(\chi, \psi)$ with parameters $\chi = 1/4$, $\psi = \delta(\mathbf{x})$, see [94], p. 313. Therefore, the conditional expectation follows from the expectation of the $\mathcal{IG}(\chi, \psi)$ variate Y , say, which is $E(Y) = \sqrt{\chi/\psi}$. Applying the actual parameters, we obtain

$$\hat{\tau} = E[\tau|\mathbf{X}] = \sqrt{\frac{\chi}{\psi}} = \frac{1}{2\sqrt{\delta(\mathbf{x})}}. \quad (3.14)$$

We now consider the more general class of MPE distributions with shape parameter $\beta \in (0, 1]$ [30]. In this setting, the distribution admits a variance-mixture representation

$$\mathbf{X} = \boldsymbol{\mu} + V \boldsymbol{\Sigma}^{1/2} \mathbf{Z},$$

is given in general by [58], Theorem 2.1, (2). The density of the mixing variable V cannot, in general, be expressed in closed form using elementary functions, except in the special case $\beta = \frac{1}{2}$ discussed in the previous section. Nevertheless, the scale-mixture-of-normals representation remains valid if and only if $\beta \in (0, 1]$. In this way for $\beta \in (0, 1)$ we have the density of the weight-variate V by

$$f_{V,\beta}(x) = \frac{2^{(1+d/2-d/2\beta)} \Gamma(d/2 + 1)}{\Gamma(d/2\beta + 1)} x^{d-3} S_{\beta}(x^{-2}, 2^{1-1/\beta}), 0 < x < \infty,$$

where $S_\beta(x^{-2}, 2^{1-1/\beta})$ denotes the stable density of variate V . The weight variable V follows a symmetric α -stable distribution, denoted by $S_\beta(u, \sigma)$, which is most conveniently characterized through its Laplace transform,

$$\mathcal{L}_{S_\beta}(u) = e^{-\sigma^\beta u^\beta}$$

Note that the stability parameter of the symmetric stable distribution lies in the interval $(0, 2]$; in the present setting, we restrict attention to the range $(0, 1)$. Applying the transformation $V = 1/\sqrt{\tau}$, with Jacobian $dv = \frac{1}{2}\tau^{-3/2}d\tau$, yields the corresponding density of τ as

$$f_{\tau, \beta}(\tau) = \frac{2^{d/2(1-\beta)}\Gamma(d/2 + 1)}{\Gamma(d/2\beta + 1)}\tau^{-d/2}S_\beta\left(\tau, 2^{1-1/\beta}\right),$$

We then employ the density function $f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$ of $\mathbf{X}$, given in (3.13), together with the conditional density of $\mathbf{X}$

$$f_{\mathbf{X}|\tau}(\mathbf{x}|\tau) = \frac{\tau^{d/2}}{(2\pi)^{d/2}}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{\tau}{2}\delta(\mathbf{x})\right),$$

conditional on τ , to derive the posterior (conditional) density of τ given the observation $\mathbf{X}$ in the usual manner. Specifically, we obtain

$$\begin{aligned} f_{\tau|\mathbf{X}}(\tau|\mathbf{x}) &= \frac{f_{\mathbf{X}|\tau}(\mathbf{x}|\tau)f_{\tau, \beta}(\tau)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= c_{\tau|\mathbf{X}}\exp\left(-\frac{\tau}{2}\delta(\mathbf{x}) + \frac{1}{2}\delta(\mathbf{x})^\beta\right)S_\beta\left(\tau, 2^{1-1/\beta}\right) \\ &= c_{\tau|\mathbf{X}}e^{\delta(\mathbf{x})^\beta/2}\exp\left(-\frac{\tau}{2}\delta(\mathbf{x})\right)S_\beta\left(\tau, 2^{1-1/\beta}\right), \end{aligned}$$

where

$$c_{\tau|\mathbf{X}} = \frac{\frac{\tau^{d/2}}{(2\pi)^{d/2}}|\boldsymbol{\Sigma}|^{-1/2} \times \frac{2^{d/2(1-\beta)}\Gamma(d/2+1)}{\Gamma(d/2\beta+1)}\tau^{-d/2}}{\frac{d\Gamma(d/2)|\boldsymbol{\Sigma}|^{-1/2}}{\pi^{d/2}\Gamma(1+d/2\beta)2^{1+d/2\beta}}}.$$

In this case, the normalizing constant satisfies $c_{\tau|\mathbf{X}} = 1$. The conditional expectation $\mathbb{E}(\tau | \mathbf{X})$ then follows directly from the integral

$$\mathbb{E}[\tau|\mathbf{X}] = e^{\delta(\mathbf{x})^{\beta/2}} \int_0^\infty \exp\left(-\frac{\tau}{2}\delta(\mathbf{x})\right) \tau S_\beta\left(\tau, 2^{1-1/\beta}\right) d\tau.$$

Let $u = \delta(\mathbf{x})/2$. Observe that the resulting integral is directly related to the Laplace transform of the symmetric stable density $S_\beta\left(\tau, 2^{1-1/\beta}\right)$, given by

$$\begin{aligned} \mathcal{L}_{S_\beta}(u) &= \int_0^\infty \exp(-\tau u) S_\beta\left(\tau, 2^{1-1/\beta}\right) d\tau \\ &= e^{-2^{\beta-1}u^\beta}, u > 0. \end{aligned}$$

The expectation with respect to the density $S_\beta\left(\tau, 2^{1-1/\beta}\right)$ can be obtained by differentiating its Laplace transform, namely

$$-\frac{d}{du} \mathcal{L}_{S_\beta}(u) = 2^{\beta-1} \beta u^{\beta-1} e^{-2^{\beta-1}u^\beta},$$

Therefore, to obtain $\mathbb{E}(\tau | \mathbf{X})$, it suffices to substitute $u = \delta(\mathbf{x})/2$ into the derivative, yielding

$$\begin{aligned} \mathbb{E}[\tau|\mathbf{X}] &= e^{\delta(\mathbf{x})^{\beta/2}} 2^{\beta-1} \beta \left(\frac{2}{\delta(\mathbf{x})}\right)^{1-\beta} e^{-2^{\beta-1}(\delta(\mathbf{x})/2)^\beta} \\ &= \beta \delta(\mathbf{x})^{\beta-1}. \end{aligned} \tag{3.15}$$

This result yields the best predictor of the weight variable τ given the observation $\mathbf{x}$. Accordingly, the resulting estimator is termed EM-GLASSO_p and is implemented via Algorithm 2, using the weight specified in (3.15).

3.2.5 Estimation of Nuisance Parameters

In the context of the EM-GLASSO framework, the estimation of the concentration matrix Θ implicitly depends on additional distribution-specific parameters, which we refer to as nuisance distributional parameters. These include the parameters (λ, χ, ψ) for the MSGH family and the shape parameter β for the MPE family. Since these param-

ters directly influence the conditional expectation of the latent weights w_i in (3.10) and (3.15), their treatment is an essential component of the methodology. In this study, we adopted a two-step estimation approach, in which nuisance parameters are estimated separately from the concentration matrix. Specifically:

- **Step 1 (Nuisance parameter estimation):** The distributional parameters η , which represent the parameters (λ, χ, ψ) in the case of MSGH and λ for MPE, are estimated using marginal maximum likelihood based on the observed data, independently of the sparsity constraint on Θ . Denote the resulting estimator by

$$\hat{\eta} = \arg \max_{\eta} \sum_{i=1}^n \log f(\mathbf{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$$

- **Step 2 (Concentration matrix estimation):** Conditional on $\hat{\eta}$, the EM-GLASSO algorithm is applied to estimate Θ , where the latent weights are computed as

$$\hat{\tau}_i^{(k+1)} = \mathbb{E} \left[\tau_i | \mathbf{x}_i; \hat{\eta}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)} \right]$$

This separation ensures computational tractability and avoids the need for high-dimensional joint optimization over both η and Θ .

Therefore, the EM algorithm is applied only to the latent scale variables τ , while the nuisance parameters η are treated as fixed after their initial estimation.

3.2.6 Theoretical considerations for the estimator of Θ

While the proposed EM-GLASSO extends the GLASSO framework to non-Gaussian settings, its theoretical properties require careful discussion due to the presence of the latent variable τ and nuisance distributional parameters η .

Connection to GLASSO: A key observation is that, conditional on the latent weights $\{\tau_i\}_{i=1}^n$, the model reduces to a weighted Gaussian

likelihood. In particular, the M-step of the EM algorithm solves a penalized likelihood problem of the form (3.12), which is equivalent to the classic GLASSO problem with a reweighted empirical covariance matrix $\hat{\mathbf{S}}_{\mathbf{x}, \tau}$. Therefore, conditional on consistent estimation of the weights, the estimator inherits the structural properties of the classical GLASSO estimator, including sparsity and positive definiteness.

Consistency under known nuisance parameters: Consider first the oracle setting in which the nuisance parameters η are known. In this case, the latent weights satisfy

$$\hat{\tau}_i = \mathbb{E} \left[\tau_i | \mathbf{x}_i; \eta, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)} \right],$$

and the estimator reduces to a weighted GLASSO with the convergence rate of the classical GLASSO as

$$\left\| \hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} \right\|_F = O_p \left(\sqrt{\frac{s \log d}{n}} \right)$$

where s denotes the number of nonzero entries in $\boldsymbol{\Theta}$. Similar guarantees can be established for element-wise consistency and support recovery. Check [85, 97] regarding the consistency of the ℓ_1 -penalized estimator.

Effect of nuisance parameter estimation: When estimating the nuisance parameters η from the data, it yields the weights

$$\hat{\tau}_i = \mathbb{E} \left[\tau_i | \mathbf{x}_i; \hat{\eta}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)} \right],$$

In this case, the estimation error can be decomposed as

$$\left\| \hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} \right\|_F \leq \underbrace{\left\| \hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}(\hat{\eta}) \right\|_F}_{\text{GLASSO error}} + \underbrace{\left\| \hat{\boldsymbol{\Theta}}(\hat{\eta}) - \boldsymbol{\Theta}(\eta) \right\|_F}_{\text{nuisance error}}$$

Therefore, if $\hat{\eta}$ is a consistent estimator of η , and the mapping from η to τ is sufficiently smooth, the second term is asymptotically negligible. Consequently, the estimator $\hat{\boldsymbol{\Theta}}$ retains the same convergence rate as in the oracle case.

Evaluation Metrics

To evaluate the efficacy of the GLASSO in recovering sparse precision matrices, we compare the estimated precision matrices against the ground-truth structure using data generated from a known multivariate distribution. Performance is quantified using the following criteria.

Edge Detection Metrics We treat structure recovery as a binary classification problem, where True Positives (TP) and True Negatives (TN) denote the number of correctly identified edges (non-zero entries) and absent edges (zero entries), respectively. False Positives (FP) and False Negatives (FN) represent the corresponding misclassifications. Based on these counts, we compute:

- *Precision*: $TP/(TP + FP)$. This measures the proportion of correctly estimated edges among all identified edges; a higher value indicates a lower false discovery rate.
- *Recall* (Sensitivity): $TP/(TP + FN)$. This quantifies the proportion of true edges that the estimator successfully recovered.
- *F1-score*: $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$. This metric computes the harmonic mean of precision and recall, providing a balanced summary of the estimator's structural accuracy.

Receiver Operating Characteristic (ROC) The ROC curve visually summarizes the performance of the estimator by plotting the true positive rate (Recall) against the false positive rate ($FP/(FP + TN)$) across a sequence of possible classification thresholds (e.g., varying the regularization parameter).

Selection of the Regularization Parameter

Selecting the tuning parameter ρ is a critical challenge in Graphical Models, as it directly controls the sparsity of the estimated graph and the false discovery rate of edges. However, the definition of an "optimal" ρ is not absolute; it depends heavily on the specific analytical objectives, such as prioritizing predictive accuracy versus true structural recovery.

Mazumder and Hastie [85] examined practical considerations and methodological trade-offs for selecting ρ in the GLASSO framework, while Abbruzzo et al. [2] proposed an estimator for the degrees of freedom in ℓ_1 -penalized models to facilitate criterion-based selection. Building on these foundations, several standard strategies are employed to determine ρ in high-dimensional settings:

Cross-Validation (CV) Friedman et al. [55] adapted k -fold CV for the GLASSO framework, and Vujačić et al. [110] subsequently introduced a computationally efficient alternative. Despite its widespread use, CV must be applied cautiously in structural learning. It is known to yield inconsistent graph estimates and is highly sensitive to the random seed, the number of folds, and the specific evaluation metric employed (e.g., log-likelihood, AIC, or BIC) [82].

Information Criteria (AIC/BIC/EBIC) While traditional Akaike and Bayesian Information Criteria (AIC and BIC) are frequently used, they often struggle in high-dimensional spaces. To address this, Foygel and Drton [53] proposed the extended BIC (eBIC), which incorporates an additional penalty on the size of the graph to better balance model fit and sparsity.

Stability Selection To enhance robustness in parameter selection, Meinshausen and Bühlmann [88] developed stability selection, a subsampling-based approach designed to consistently recover the true underlying graphical structure across varying subsamples.

3.2.7 Simulated Results

A preconfigured simulation has been used to evaluate the performance of the proposed method against competing methods across different environments. We benchmarked the proposed EM-GLAASO, GLASSO, γ -LASSO, and SGLASSO using the ROC curve to assess the accuracy of graph structure recovery. The proposed approach is computationally more demanding than GLASSO because GLASSO optimizes at each M-step of the EM algorithm. The simulation results were evaluated using

ROC curves for all comparisons. For SGLASSO, the tuning parameter was set to $\lambda = 10^{-5}$ following Lu and Feng [81], while for γ -LASSO, we used $\gamma = 0.05$ as suggested by Onizuka and Hashimoto [92]. A cutoff threshold of 10^{-3} was applied to define the true graph structure. To ensure stable results, the simulation was repeated 200 times, and the average ROC curve was calculated. We employed the StARS in our simulations, as it determines the tuning parameter by assessing edge stability across subsamples. This method enhances robustness and reproducibility, and alleviates the limitations of likelihood-based criteria by more reliably recovering the underlying graph structure.

Case 1: $MSGH, d = 50$; **sparsity**= 30%; $n = 100$: True sparse concentration matrices of size 50×50 with density 30% were generated using the algorithm described in Algorithm 3. The algorithm enforces strict diagonal dominance to ensure positive definiteness. Under the settings of the MSGH distribution, datasets were simulated with $n = 50$ and $n = 100$ from $MSGH_d(\lambda, \chi = 0.5, \psi = 0.5, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1})$, where $\lambda \in (2, 5, 10, 15)$.

As a result, Figure 3.3 highlights the performance of the GLASSO, EM-GLASSO_H, γ -LASSO, and SGLASSO estimators. Based on the provided ROC curves, the performance of the estimators varies significantly with λ . For $\lambda = 2$, all five estimators—GLASSO, EM-GLASSO_H, γ -LASSO, and SGLASSO—exhibit tightly clustered performance with comparable recovery of the true sparsity pattern. However, as λ increases to 5, 10, and 15, a distinct separation emerges where the EM-GLASSO_H estimator consistently demonstrates superior performance, achieving the highest True Positive Rate (TPR) for a given False Positive Rate (FPR). The SGLASSO estimator generally ranks second, outperforming the remaining methods but falling below EM-GLASSO_H in these higher λ settings. Meanwhile, the GLASSO and γ -LASSO estimators display nearly identical, lower-tier performance profiles across the higher λ values, clustering together at the bottom of the ROC space.

Case 2: $MSGH, d = 100$; **sparsity**= 10%; $n = 50$: In the high-dimensional settings $d > n$ under the MSGH distribution, the EM-

Algorithm 3 Simulation of a precision matrix

```

1: Input:
2:  $\mathbf{Z} \in \{0, 1\}^{d \times d}$   $\triangleright d \times d$  binary matrix marking zero off-diagonal
   entries.
3: Output:  $\Omega \succ 0$   $\triangleright$  Symmetric positive definite precision matrix
4: procedure
5:   Initialize  $\Omega \leftarrow \mathbf{0}_{d \times d}$   $\triangleright$  Zero matrix
6:   for  $i \leftarrow 1$  to  $d - 1$  do  $\triangleright$  Construct off-diagonal elements
7:     for  $j \leftarrow i + 1$  to  $d$  do
8:       if  $Z_{ij} = 0$  then  $\triangleright$  Non-zero entry allowed
9:          $\Omega_{ji} \leftarrow \mathcal{U}(0.1, 1.0)$   $\triangleright$  Sample from uniform
           distribution, maintain symmetry
10:      end if
11:    end for
12:  end for
13:  for  $i \leftarrow 1$  to  $d$  do
14:     $\Omega_{ii} \leftarrow \sum_{k=1}^d |\Omega_{ik}| + \mathcal{U}(0.5, 1.0)$   $\triangleright$  Ensure diagonal dominance
15:  end for
16:   $\lambda_{\min} \leftarrow \min(\text{eig}(\Omega))$   $\triangleright$  Guarantee positive definiteness
17:  while  $\lambda_{\min} \leq 0$  do
18:     $\Omega \leftarrow \Omega + 10^{-5} \cdot I_d$   $\triangleright$  Add scaled identity matrix
19:     $\lambda_{\min} \leftarrow \min(\text{eig}(\Omega))$ 
20:  end while
21:  return  $\Omega$ 
22: end procedure

```

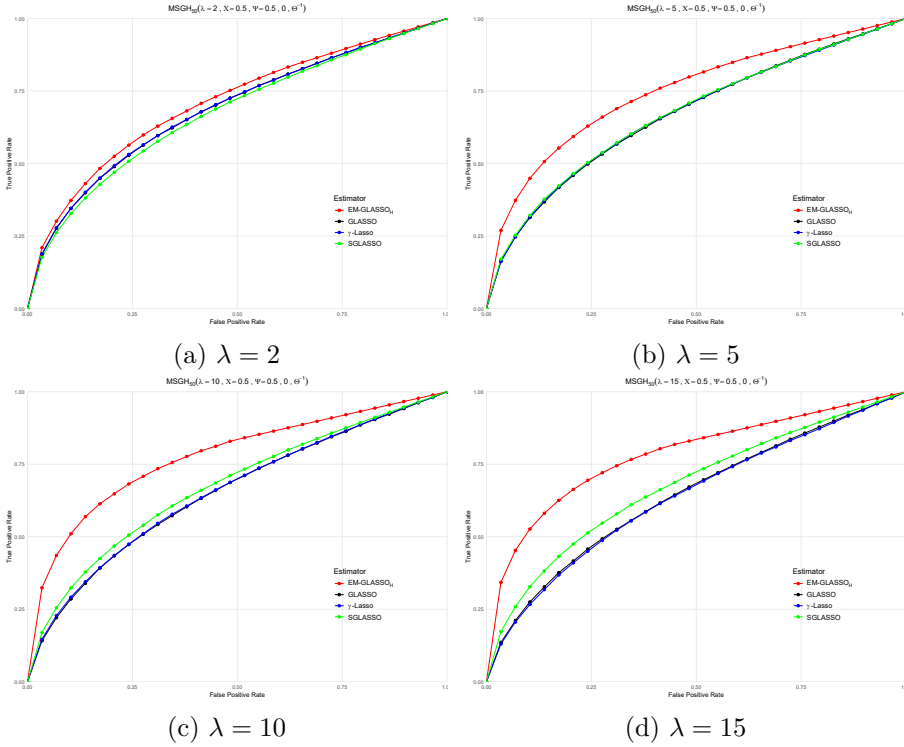


Figure 3.3: ROC curves of the GLASSO, EM-GLASSO_H, γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $MSGH_{50}(\lambda, \chi = 0.5, \psi = 0.5, \mu = \mathbf{0}, \Sigma = \Theta^{-1})$. Where $\lambda \in (2, 5, 10, 15)$

GLASSO_H estimator consistently exhibits superior performance in recovering the true sparsity pattern. While all estimators display comparable efficacy at the lowest λ parameter ($\lambda = 2$), a clear performance hierarchy emerges as λ increases to 5, 10, and 15. In these heavier-tailed scenarios, EM-GLASSO_H dominates the ROC space, achieving the highest True Positive Rate for any given False Positive Rate. The SGLASSO estimator consistently ranks second, outperforming the remaining methods. Meanwhile, the GLASSO and γ -LASSO estimators cluster together with notably lower sensitivity, indicating that they are less robust to the heavy tails introduced by the higher λ values in this regime, as highlighted in Figure 3.4.

Case 3: $\mathcal{MPE}, d = 50$; **sparsity**= 30%; $n = 100$: In this case, datasets were simulated from $\mathcal{MPE}_d(\mathbf{0}, \mathbf{\Theta}^{-1}, \beta)$ distribution, across different sample sizes ($n = 50$ and 100). To evaluate the impact of the shape parameter on the performance of each method, we examined varying values of $\beta \in \{1/3, 1/2, 2/3, 3/4, 1\}$. After the simulation, Figure 3.5 shows different behaviors of the GLASSO, EM-GLASSO_P, γ -LASSO, and SGLASSO estimators when the real concentration matrix is generated from the $\mathcal{MPE}_{50}(\mathbf{0}, \mathbf{\Theta}^{-1}, \beta)$ distribution. The proposed method outperforms the GLASSO approach when the shape parameter is small; however, its performance deteriorates as β approaches 1 (in the Gaussian case), at which point it performs comparably to GLASSO. For small to moderate values of β , EM-GLASSO_P outperforms the competing methods. The results depicted by Figure 3.5 reveal that the EM-GLASSO_P estimator demonstrates a clear and consistent superiority in recovering the true sparsity pattern across all considered values of the shape parameter β . Regardless of whether the distribution is more heavy-tailed ($\beta = 1/3$) or approaches a lighter-tailed shape ($\beta = 3/4$), EM-GLASSO_P consistently yields the highest TPR for any given FPR, effectively dominating the ROC space. The SGLASSO estimator generally achieves second-best performance, showing a distinct advantage over standard methods but remaining below the EM-GLASSO_P curve in all scenarios. In contrast, the GLASSO and γ -LASSO estimators exhibit the lowest performance, clustering together with significantly

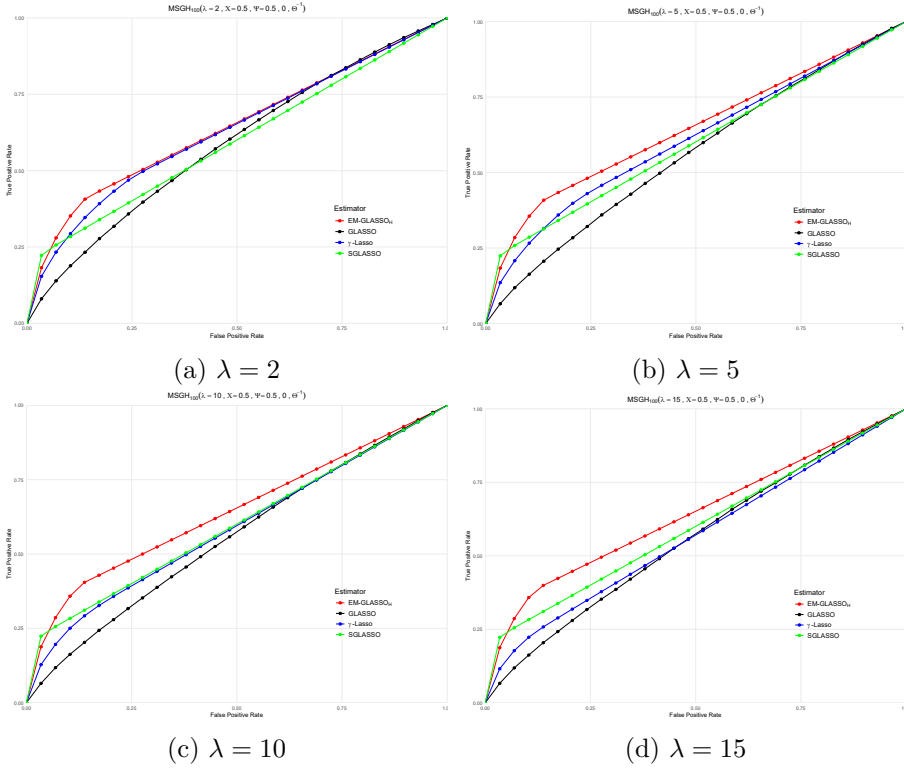


Figure 3.4: ROC curves of GLASSO, EM-GLASSO_H, γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $\text{MSGH}_{100}(\lambda, \chi = 0.5, \psi = 0.5, \mu = \mathbf{0}, \Sigma = \Theta^{-1})$. Where $\lambda \in (2, 5, 10, 15)$ and $n = 50$.

lower sensitivity. This hierarchy remains stable across all variations of β , underscoring the robustness of EM-GLASSO_P to the specific tail behavior of the multivariate power exponential distribution.

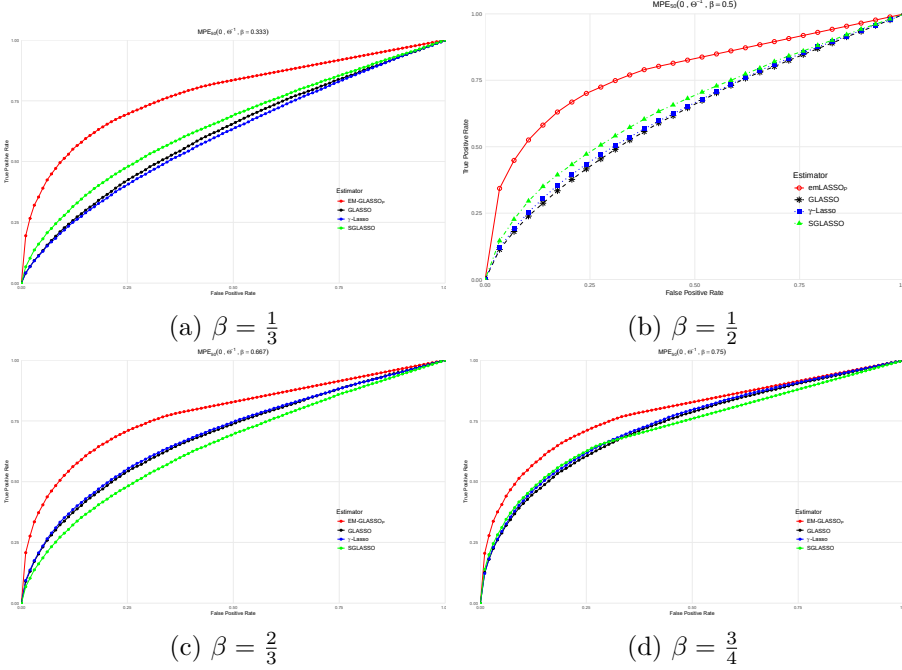


Figure 3.5: ROC curves of the GLASSO, EM-GLASSO_P, γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $\mathcal{MP}\mathcal{E}_{50}(\mathbf{0}, \Theta^{-1}, \beta)$ distribution, and different values of β .

Case 4: $\mathcal{MP}\mathcal{E}, d = 100$; **sparsity**= 10%; $n = 50$: In the case of a high-dimensional context as $d > n$, Figure 3.6 shows the ROC curves with a high dimension $d = 100$ and a limited sample size ($n = 50$). The EM-GLASSO_P estimator demonstrates a commanding performance advantage over competing methods across all considered shape parameters ($\beta \in \{1/3, 1/2, 2/3, 3/4\}$). As illustrated in the ROC curves, EM-GLASSO_P consistently achieves the highest True Positive Rate for low False Positive Rates, exhibiting significant robustness to the

heavy-tailed nature of the MPE distribution. The SGLASSO estimator consistently ranks second, clearly separating itself from the remaining estimators. Conversely, the GLASSO and γ -LASSO estimators show substantially degraded performance in this "large d , small n " regime, clustering together with much lower sensitivity compared to the top-performing EM-GLASSO_P.

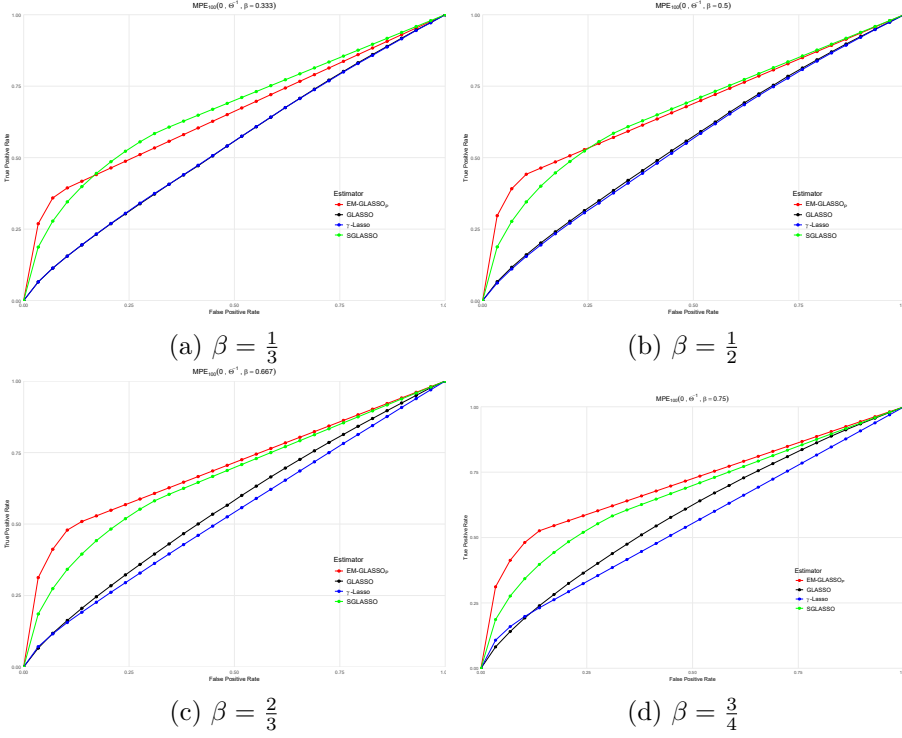


Figure 3.6: ROC curves of the GLASSO, EM-GLASSO_P, γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $MPE_{100}(\mathbf{0}, \Theta^{-1}, \beta)$ distribution, and different values of β and $n = 50$.

3.2.8 Computational Costs

The computational complexity of the proposed EM-GLASSO procedures is theoretically summarized in Table 3.2 and can be compared

with that of the classical GLASSO estimator. The GLASSO algorithm involves a single penalized-likelihood optimization with a computational cost of $\mathcal{O}(d^3)$. In contrast, the proposed methods rely on an EM-type iterative scheme. For the Student- t and MPE models, the E-step is computationally inexpensive, and the dominant cost within each iteration arises from the computation of the weighted sample covariance matrix and the subsequent GLASSO update, yielding a per-iteration complexity of order $\mathcal{O}(nd^2 + d^3)$. The overall computational cost is therefore $\mathcal{O}(T(nd^2 + d^3))$, where T denotes the number of EM iterations. For the MSGH model, the evaluation of modified Bessel functions introduces an additional computational component, resulting in a per-iteration cost of order $\mathcal{O}(nd^2 C_{\text{Bessel}} + d^3)$. As a consequence, this variant is computationally more demanding than the Student- t and MPE cases. Overall, the additional cost of the proposed framework is primarily driven by the EM iterations and the complexity of the underlying distributional model.

The empirical benchmarking results in Table 3.3 indicate that EM-GLASSO_P strikes the most favorable balance between computational speed and memory overhead among the proposed variants. Although GLASSO remains the most efficient (0.61ms), EM-GLASSO_P (4.8ms) remains feasible for high-dimensional applications. In contrast, EM-GLASSO_H exhibits significant memory churn (24.41MB/itr), which is reflected in the high frequency of garbage collection events (gc/sec=8.26).

Therefore, these theoretical findings in Table 3.2 are consistent with the empirical computation times reported in Table 3.3, where the Student- t and MPE variants exhibit moderate overhead relative to GLASSO. At the same time, the MSGH-based approach is comparatively slower.

Choosing between MSGH and MPE

Preliminary descriptive analysis and measures of multivariate kurtosis can guide the choice between the MPE and MSGH models. In particular, Mardia's multivariate kurtosis provides a useful diagnostic for assessing tail behavior relative to the Gaussian benchmark (See [105]).

Table 3.2: Theoretical Computational Complexity Comparison of GLASSO and EM-GLASSO

Method	Model	E-step Cost	Per-Iteration	Total Complexity	Speed
GLASSO	Normal	–	$\mathcal{O}(d^3)$	$\mathcal{O}(d^3)$	Fastest
EM-GLASSO _T	t	$\mathcal{O}(1)$	$\mathcal{O}(nd^2 + d^3)$	$\mathcal{O}(T(nd^2 + d^3))$	Fast
EM-GLASSO _P	MPE	$\mathcal{O}(1)$ (power)	$\mathcal{O}(nd^2 + d^3)$	$\mathcal{O}(T(nd^2 + d^3))$	Fast
EM-GLASSO _H	MSGH	$\mathcal{O}(C_{\text{Bessel}})$	$\mathcal{O}(nd^2 C_{\text{Bessel}} + d^3)$	$\mathcal{O}(T(nd^2 C_{\text{Bessel}} + d^3))$	Slow

Notes: n denotes sample size, d dimension, and T the number of EM iterations.

C_{Bessel} represents the computational cost of evaluating modified Bessel functions.

All EM-GLASSO variants share the same asymptotic complexity but differ in constant factors.

Table 3.3: Computational performance of GLASSO and EM-GLASSO variants ($n = 100, d = 50$, over 1,000 iterations).

Method	Median Time	Relative Speed	Memory Alloc.	Garbage Collection/sec
GLASSO	0.61 ms	1.0×	0.25 MB	6.43
EM-GLASSO _P	4.80 ms	7.8×	2.13 MB	8.28
EM-GLASSO _T	12.90 ms	21.1×	6.45 MB	10.10
EM-GLASSO _H	94.60 ms	154.4×	24.41 MB	8.26

Here are the highlights of their preferences:

- **Tail Flexibility Comparison:** MSGH is exceptionally powerful for **heavy-tailed (leptokurtic)** data. It can model more complex "tail-decay" patterns than MPE. If your data is strictly heavy-tailed with outliers, emLASSO_H will often produce results that are equal to or better than emLASSO_P. MPE is unique because it can also model thin-tailed (**platykurtic**) data ($\beta > 1$). MSGH is generally restricted to the leptokurtic domain. Therefore, if your dataset contains "flat" distributions, MSGH cannot replicate the MPE results.
- **The "Heavy-Tailed" Convergence:** In cases of moderate heavy-tailedness, the two methods often converge on the same sparsity pattern. As seen in the ROC results for MSGH contamination, emLASSO_H and emLASSO_P both significantly outperform standard GLASSO, often showing overlapping performance curves when the contamination is not extreme.

- **Parameter Sensitivity:** MSGH requires tuning multiple nuisance parameters (λ, χ, ψ) . If these are tuned perfectly, MSGH can "mimic" the robust behavior of an MPE model for heavy-tailed data. However, the MPE's single β parameter makes it easier to achieve these results reliably without additional uncertainty.

3.3 Thesis 1

Objective: Extend the GLASSO framework under the MSGH distribution while explicitly preserving interpretable conditional relationships.

Key Findings: The proposed EM-GLASSO_H estimator effectively recovers sparse, interpretable graphical structures under the MSGH distribution, offering robustness to non-normality and improved modeling of heavy-tailed multivariate data. We have open-sourced the robust R implementation of our algorithm, available on GitHub at <https://github.com/AbdoDawod/emLASSO>.

Related publication: Terdik, Gy. and Dawod, ABA., 2026. Extending GLASSO to Non-Gaussian Settings: Sparse Concentration Estimation via EM Algorithms, Computational Statistics.

3.4 Thesis 2

Objective: Extend the GLASSO framework under the MPE distribution while explicitly preserving interpretable conditional relationships.

Key Findings: The proposed EM-GLASSO_P effectively recovers sparse, interpretable graphical structures under the MPE distribution, offering robustness to non-normality and improved modeling of heavy-tailed multivariate data.

Related publication: Terdik, Gy. and Dawod, ABA., 2026. Extending GLASSO to Non-Gaussian Settings: Sparse Concentration Estimation via EM Algorithms, Computational Statistics.

4

APPLICATIONS OF GLASSO

This chapter evaluates the GLASSO, SGCGM, and EM-GLASSO algorithms on real-world data. Specifically, it is broken down into: (i) Cognitive-behavioral networks were estimated using the GLASSO and EM-GLASSO algorithms, and the results were validated against findings reported in the literature. (ii) Railway track quality was monitored using GLASSO and EM-GLASSO with sampled data obtained through zero-crossing sampling. (iii) The nutrient interactions in baobab leaves were mapped to identify the chemical relationships underlying plant physiology. Because the underlying data distribution was unknown, the EM-GLASSO algorithm was compared with the SGCGM method.

4.1 Cognitive-Behavioral Networks

Visualizing networks of thoughts, emotions, and behaviors is central to contemporary models of psychopathology [19, 33, 40]. Mental health disorders are increasingly viewed as emergent properties of complex interactions among symptoms, cognitions, and behaviors [46, 61]. However, accurately recovering the structure of these systems from empirical data remains methodologically challenging [45]. Using the pairwise correlations or factor models may obscure direct relationships among variables due to indirect associations or restrictive conditional independence assumptions [46, 47]. In contrast, GGMs provide a flexible framework in which variables are represented as nodes and conditional dependencies as edges [22, 47].

This section applies GLASSO and EM-GLASSO to estimate cognitive-behavioral networks. Graphical modeling is well established in psychopathology. Therefore, we do not propose a new psychological model. Instead, this case study serves as an empirical benchmark for

the EM-GLASSO algorithm. We evaluate the interpretability, sparsity, and stability of the EM-GLASSO framework. The analysis demonstrates that both methods recover meaningful conditional dependencies from high-dimensional psychological data under the assumption of normality. The next two subsections detail these results.

4.1.1 Post-Traumatic Stress Disorder

Having Trauma triggers Post-Traumatic Stress Disorder (PTSD); patients suffer from flashbacks, severe anxiety, and hypervigilance. This complex clinical structure provides an ideal benchmark. It demonstrates the practical utility and interpretability of the generalized precision matrix framework. We analyzed open-source data from Armour et al. [13] (available on OSF¹). The sample contains $n = 221$ observations, and the networks are estimated across $d = 20$ distinct PTSD symptoms. These variables group into four primary domains:

- **B: Intrusion symptoms:** Intrusive thoughts, nightmares, flashbacks, emotional cue reactivity, and psychological cue reactivity (B1–B5).
- **C: Avoidance symptoms:** Avoidance of thoughts and avoidance of reminders (C1–C2).
- **D: Negative alterations in cognition and mood:** Trauma-related amnesia, negative beliefs, blame of self, negative trauma-related emotions, loss of interest, detachment, and restricted affect (D1–D7).
- **E: Arousal and reactivity symptoms:** Irritability, self-destructive behavior, hypervigilance, exaggerated startle response, difficulty concentrating, and sleep disturbance (E1–E6).

Initially, the regularization parameter (ρ) was estimated using the eBIC, yielding $\rho = 0.4858$. Second, using the parsimonious properties and Mardia’s tests for Kurtosis and Skewness, the shape parameter of the MPE was estimated as $\beta = 0.997$, indicating a nearly Gaussian

¹<https://osf.io/mj5wa/files/osfstorage>

distribution. Then, GLASSO and the EM-GLASSO_P were applied to estimate the graph, where both algorithms identified the same core network. However, the EM-GLASSO_P is stricter, and the primary dependency structure remains stable. For example, GLASSO retained a weak conditional dependency between variables D5 and E1, but the EM-GLASSO_P algorithm forced this connection to exactly zero. Furthermore, the EM-GLASSO_P slightly reduced the magnitude of the surviving partial correlations across the matrix. Consequently, EM-GLASSO_P delivers a cleaner and more conservative structural estimate of the conditional dependence structure for the PTSD symptom network, which is presented in Figures 4.1 and 4.2.

Both GLASSO and EM-GLASSO_P results align with the established clinical literature. Trauma-related amnesia (D1) and avoidance symptoms (C1, C2) remain isolated in the network as they lack conditional dependence on other symptoms, confirming prior findings [21]. Furthermore, the models identify active conditional dependencies among negative cognitive and mood alterations (D5–D7). These alterations link directly to intrusion symptoms [62]. Finally, intrusion symptoms exhibit strong conditional dependencies with negative beliefs, self-blame, and trauma-related emotions (D2–D4). These specific connections persist even when controlling for all other variables [86].

4.1.2 Emotional monitoring

This section analyzes an open-source dataset² of simulated biosensor and behavioral data for emotional monitoring and feedback systems. The data monitors student engagement and emotional states. We estimate the underlying network using the GLASSO and EM-GLASSO frameworks independently. The empirical evaluation relies on a sample of $n = 100$ observations. The dataset contains 14 features:

- **Heart rate (bpm):** High values indicate stress. Low values suggest calm engagement.

²<https://www.kaggle.com/datasets/ziya07/emotional-monitoring-dataset/data>

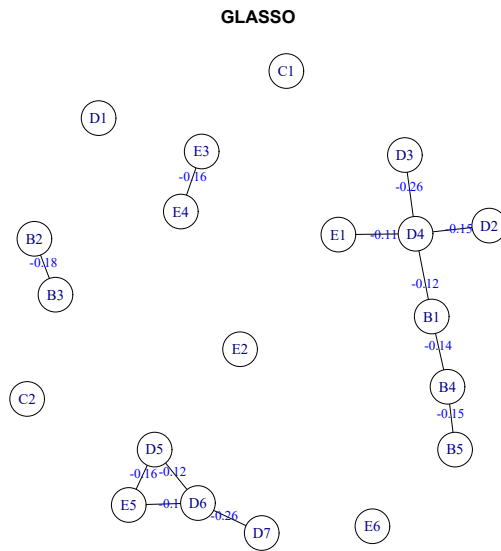


Figure 4.1: Sparse conditional dependence network of PTSD symptoms estimated via the GLASSO algorithm

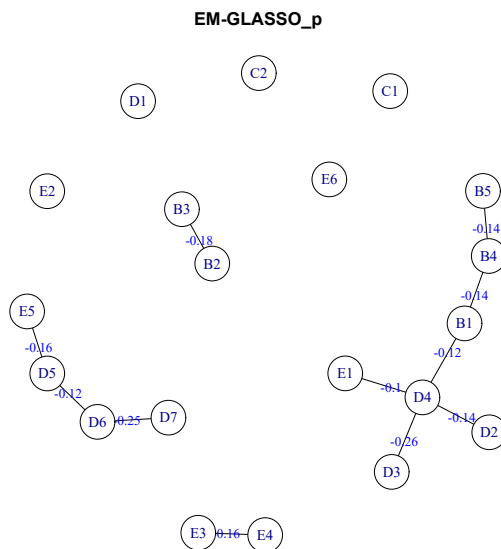


Figure 4.2: Sparse conditional dependence network of PTSD symptoms estimated via the EM-GLASSO_p algorithm

- **Skin conductance:** Measures emotional arousal. High values correlate with stress.
- **Electroencephalography (EEG):** Brain wave frequency. High values indicate focus. Low values suggest distraction.
- **Body temperature (°C):** Fluctuations indicate stress. High values reflect physical tension.
- **Pupil diameter:** Indicates cognitive load. Dilation reflects focus.
- **Smile intensity (0.0–1.0):** High values indicate positive emotion and interest.
- **Frown intensity (0.0–1.0):** High values indicate negative emotion and disengagement.
- **Cortisol level:** High levels indicate stress.
- **Activity level:** Physical movement intensity. High levels correlate with stress or arousal.
- **Ambient noise level:** Background noise disrupting focus.
- **Lighting level:** Environmental illumination affecting attention.
- **Emotional state:** The student’s reported feeling.
- **Cognitive state:** The student’s level of concentration.
- **Engagement level:** The student’s overall engagement.

With a sample size of $n = 100$ and $d = 14$ nodes, there are 91 unique pairwise conditional dependencies. Because the number of parameters approaches the sample size, traditional estimation fails. We optimized the tuning parameter ρ using the eBIC, yielding $\rho = 0.13097$. Estimating the shape parameter of the MPE distribution resulted in $\beta = 1.0043$, which effectively recovers the standard Gaussian case.

Figures 4.3 and 4.4 present the estimated networks. Both GLASSO and EM-GLASSO_P yield sparse structures that align tightly with established psychophysiological literature. We draw two primary structural conclusions. At its center, heart rate (HR), skin conductance, EEG activity, and physical activity form a tight cluster, suggesting that physiological arousal and movement rise and fall together. Emotional state is most strongly linked to HR and temperature, with a smaller positive association with skin conductance, indicating that emotion is reflected primarily through autonomic responses. Activity emerges as another important hub, showing positive associations with HR and cortisol and negative associations with EEG and pupil size. Engagement is positively associated with EEG and skin conductance, suggesting that greater engagement is accompanied by heightened neural and physiological activation. In contrast, environmental factors remain peripheral: lighting is negatively associated with engagement, and ambient conditions are not. Several variables, including smile, frown, and cognitive state, are isolated from the main network, suggesting that, under the EM-GLASSO model, they contribute little unique information once the stronger physiological and behavioral relationships are accounted for. Like a Hemingway story, the network says much with few words: the strongest signals come from the body, while the quieter variables remain at the edges, disconnected from the central action. See [23, 26, 47, 76, 103, 121]

Overall, the results show that EM-GLASSO_P performs comparably to GLASSO when the Gaussian assumption is correctly specified. Both methods recover clear and interpretable dependence structures in high-dimensional cognitive and behavioral data. These findings suggest that EM-GLASSO preserves the strengths of GLASSO while remaining suitable for complex statistical settings commonly encountered in behavioral research.

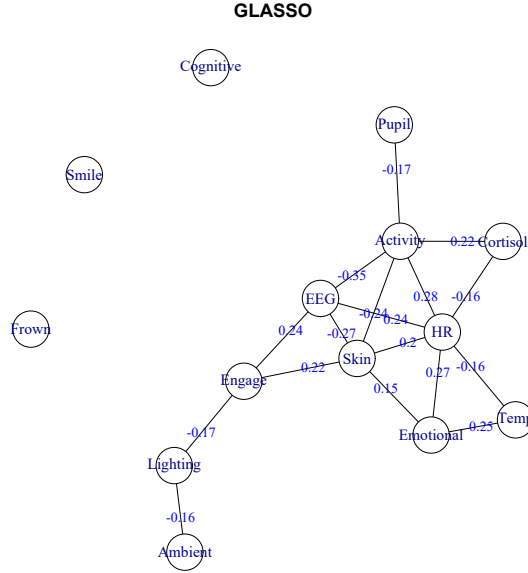


Figure 4.3: Estimated sparse conditional dependence network of student emotion monitoring variables, generated via GLASSO.

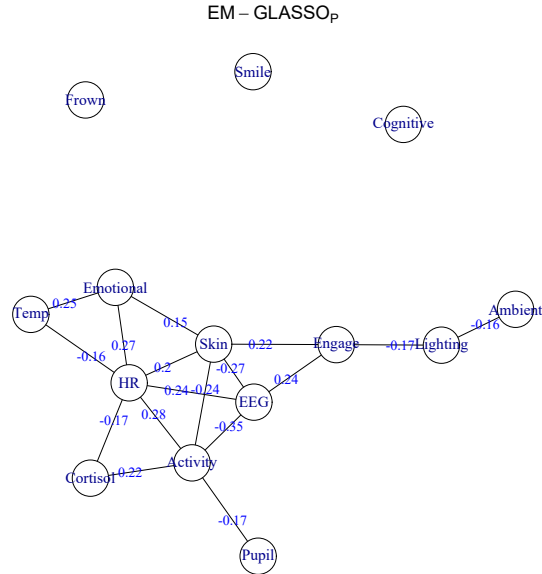


Figure 4.4: Estimated sparse conditional dependence network of student emotion monitoring variables, generated via the EM-GLASSO_P.

4.2 Robust Quality Inspection via MCV Control Charts

SPC has long employed control charts to identify and eliminate sources of transient variations that affect quality. In univariate SPC, Shewhart charts (e.g., $\bar{X}$ and R charts) monitor single quality characteristics. For multivariate SPC (MSPC), Hotelling's T^2 chart was the first extension of Shewhart [63]. It signals when the Mahalanobis distance of a sample mean exceeds control limits, and it is powerful for detecting large shifts in the process mean vector. To improve sensitivity to smaller, gradual shifts, memory-type charts were developed. For example: Multivariate Cumulative SUM (MCUSUM) [36], Multivariate Exponentially Weighted Moving Average (MEWMA) [80], and Multivariate Homogeneously Weighted Moving Average (MHWMA) [6].

These MSPC tools use the full covariance structure among variables, making them more efficient than separate univariate charts when variables are correlated. In practice, SPC typically proceeds in two phases: Phase I (retrospective analysis to estimate in-control parameters from historical data) and Phase II (prospective monitoring). In Phase I, one estimates the mean and covariance from an in-control dataset; in Phase II, the chart monitors for deviations beyond those limits [36, 63].

As manufacturing and sensor technologies advance, modern processes often involve high-dimensional data streams. Conventional MSPC charts face severe challenges; for example, Hotelling's T^2 requires inverting the sample covariance matrix. If $n < d$, the covariance estimate is singular and T^2 is undefined [64]. Even when $d \simeq n$, the covariance estimate can be unstable, leading to inflated false alarms or missed shifts, so the control charts break down as d grows. These limitations have motivated new methods for high-dimensional MSPC that exploit sparsity or other structure. LASSO or GLASSO impose an ℓ_1 penalty to shrink many parameters to zeros or produce a sparse inverse covariance.

More broadly, many recent studies have applied sparse covariance/precision estimation to control charts. For instance, adaptive thresholding LASSO methods have been used for Phase I covariance

monitoring [3], and penalized likelihood charts have been extended to jointly monitor the mean and variance in high-dimensional data [7, 84, 120] and to handle complex data structures by identifying the most relevant quality characteristics. A key advancement was the use of GLASSO for sparse estimation and variable selection. For instance, Zou and Qiu [120] proposed an MSPC chart based on LASSO, and Maboudou-Tchao et al. [84] proposed GLASSO-based charts for monitoring the covariance matrix.

If a process has either a non-constant mean or variance throughout time, monitoring its coefficient of variation (CV), it is the process variation relative to the mean ($CV = \sigma/\mu$) that is itself the quality characteristic of interest [1, 29]. Yeong et al. [117] introduced one-sided Shewhart charts for monitoring shifts in the multivariate CV (MCV). To improve the sensitivity of the MCV chart, Khaw et al. [70] and Nguyen et al. [90] enhanced control charts using Markov chain and synthetic chart approaches.

Adegoke et al. [7] applied GLASSO to establish two one-sided charts—upward and downward—for detecting shifts in the process’s MCV. Their new charts outperformed the classical charts when small subgroup sizes were used for monitoring.

Therefore, this study leverages the GLASSO and EM-GLASSO frameworks to improve MCV control charts for high-dimensional processes. We compare the relative efficiency of both estimators. Ultimately, these methods enhance the chart’s sensitivity to small and moderate structural shifts.

4.2.1 GLasso-based MCV control charts

Assume $\mathbf{X}_{1:n} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a d -variates sample from a Gaussian process. Then, according to Voinov and Nikulin [91], the MCV can be defined by:

$$\varphi = \left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right)^{-1/2} \quad (4.1)$$

Usually, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, but they can be estimated from the sample data using the first and second moments:

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^\top$$

which serve as the sample mean vector and covariance matrix, respectively. Therefore, the resulting MCV based on the sample is:

$$\hat{\varphi} = \left(\bar{\mathbf{X}}^\top \mathbf{S}^{-1} \bar{\mathbf{X}} \right)^{-1/2} \quad (4.2)$$

Yeong et al. [117] showed that $\hat{\varphi}$ follows a non-central F-distribution ($n > d$) with non-centrality parameter ω ; its cumulative distribution function (CDF) is given by:

$$F_{\hat{\varphi}}^{-1}(\alpha|n, d, \omega) = \sqrt{\frac{n(n-d)}{(n-1)d} \left[\frac{1}{F_F^{-1}(1-\alpha|d, n-d, \omega)} \right]} \quad (4.3)$$

where $F_F^{-1}(\cdot|\cdot)$ is the inverse CDF of the non-central F distribution. As the resulting distribution is skewed, using a single two-sided chart to monitor both increases and decreases in MCV will likely yield a biased average run length (ARL) [90]. To overcome this issue, separate one-sided upward and downward charts are proposed. The lower control limit (LCL) of the downward MCV chart is given as:

$$LCL = F_{\hat{\varphi}}^{-1}(\alpha|n, d, \omega), \quad (4.4)$$

where $\omega = n/\varphi_0$ and φ_0 is the initial value of MCV specified in Phase I. Similarly, the upper control limit (UCL) of the upward MCV chart is given by:

$$UCL = F_{\hat{\varphi}}^{-1}(1-\alpha|n, d, \omega) \quad (4.5)$$

Using the GLASSO formula in Equation 1.1 to estimate the precision matrix $\hat{\boldsymbol{\Theta}}^{GL}$. Then, the coefficient of variation φ can be estimated as:

$$\varphi^{GL} = \left(\bar{\mathbf{X}}^\top \hat{\boldsymbol{\Theta}}^{GL} \bar{\mathbf{X}} \right)^{-1/2} \quad (4.6)$$

The GLASSO-based MCV control chart limits are given as:

$$\begin{aligned} LCL &= F_{\hat{\varphi}^{GL}}(\alpha|d, n) \\ UCL &= F_{\hat{\varphi}^{GL}}(1 - \alpha|d, n) \end{aligned}$$

Following the same steps but using EM-GLASSO_P instead of GLASSO to get the EM version of $\hat{\varphi}$. Similarly, we follow the same steps for the EM-GLASSO_P. The control chart triggers an out-of-control signal if either $\hat{\varphi} < LCL$ or $\hat{\varphi} > UCL$.

4.2.2 Wine Quality

This example evaluates production stability using the Wine Quality dataset from the UCI Machine Learning Repository. Paulo Cortez et al. [35] compiled the data from red and white *vinho verde* wines produced in Portugal. The dataset contains 11 physicochemical variables, including fixed acidity, residual sugar, chlorides, pH, and alcohol content, as well as a discrete sensory quality score. We apply the MCV with GLASSO and EM-GLASSO to assess this high-dimensional quality profile. Maintaining chemical consistency is essential in industrial winemaking.

The MCV measures the relative dispersion of the 11 correlated variables. We compute the sample MCV ($\hat{\varphi}$) to monitor the joint relative variability of production batches. A stable $\hat{\varphi}$ indicates controlled fermentation and blending, whereas systematic shifts suggest an out-of-control process with increased chemical variability. This application demonstrates the usefulness of MCV-based monitoring as an empirical benchmark for oenological quality assurance.

To evaluate the robustness of EM-GLASSO under non-Gaussian conditions, we assessed multivariate normality using Mardia's test. The results showed significant departures from normality, with kurtosis = 310.6495 (p -value = 0) and skewness = 91.3705 (p -value = 0). Based on the descriptive statistics and model parsimony, the MPE distribution was selected over the MSGIH distribution for comparison with the Gaussian case. The estimated MPE shape parameter was $\beta = 0.9$.

Phase I: At this stage, the process is assumed to be stable and in control. From the dataset, we selected samples with quality scores greater than 5. The control chart parameters (UCL, CL, LCL, ω , and ρ) were estimated using 57 batches, each of size 15. The tuning parameter ρ was selected using the eBIC criterion, resulting in $\rho = 0.2846$. The remaining parameters were estimated using the precision matrix estimator and are given as

- GLASSO: LCL = 0.00686, CL = 0.06625, UCL = 0.07141.
- EM-GLASSO: LCL = 0.00665, CL = 0.06421, UCL = 0.06921.

Phase II: This stage corresponds to the inspection phase, in which the Phase I control charts are used to monitor the process under real operating conditions, including out-of-control samples. In this phase, 40 batches, each of size 15, were monitored across 11 variables.

The EM-GLASSO framework demonstrates greater sensitivity than the standard GLASSO approach, as illustrated in Figures 4.5 and 4.6. The GLASSO chart detected four out-of-control batches, whereas the EM-GLASSO chart identified ten. This higher detection rate indicates that EM-GLASSO is more effective at detecting subtle structural shifts in the process. In industrial quality monitoring, these control charts assess the multivariate coefficient of variation by tracking the relative dispersion of complex production data, thereby providing early warnings of process instability before critical failures occur. These findings are consistent with the simulation results reported by Adegoke et al. [7] and with the simulation results presented in Chapter 3.

This study employs one-sided control charts to monitor directional shifts in the process MCV. We estimate the underlying precision matrices using both the GLASSO and EM-GLASSO frameworks. The empirical results demonstrate that EM-GLASSO charts consistently outperform their standard GLASSO counterparts

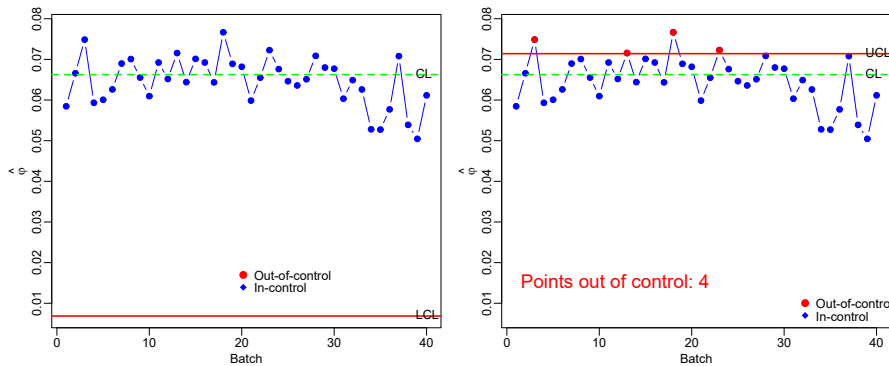


Figure 4.5: Monitoring the quality of red *vinho verde* wine using GLASSO framework.

4.3 Identifying Nutrient Interaction Networks in Baobab Leaves

Baobab leaves (*Adansonia digitata*) are notable for their high levels of minerals, quality protein, and bioactive compounds with strong antioxidant properties [5, 9]. Recent studies report substantial spatial variability in their nutrient composition, highlighting the roles of soil, climate, and genetic diversity [9, 44].

Recently, graphical models have been increasingly applied in nutritional science to reveal cooperative and antagonistic relationships among nutrients and bioactive compounds, improving understanding of bioavailability and compositional complexity [43]. Mixed GGMs extend this framework to accommodate datasets with both continuous and categorical variables, thereby enhancing its applicability to nutritional and behavioral analyses [102]. Bayesian Networks (BNs) provide a probabilistic framework for modeling directional associations and integrating prior biological or dietary knowledge, supporting causal inference in nutritional epidemiology [60]. Dynamic network models capture time-dependent nutrient relationships, enabling researchers to characterize the evolution of dietary patterns and assess the impact of interventions [75]. While hypergraph-based approaches generalize pairwise interactions to multi-nutrient relationships, enabling the modeling of syner-

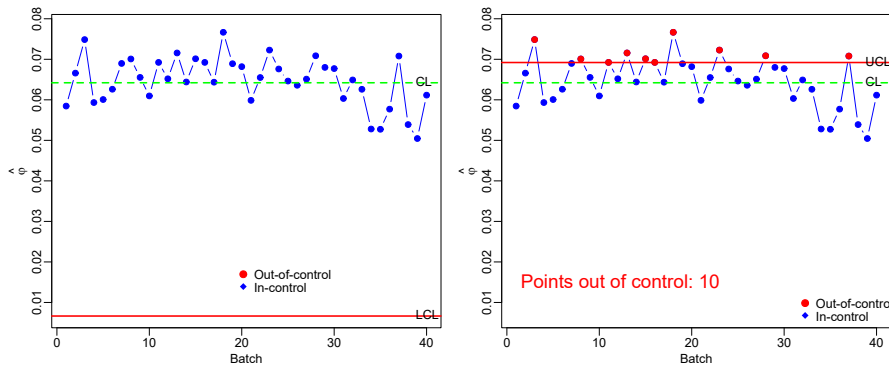


Figure 4.6: Monitoring the quality of red *vinho verde* wine using EM-GLASSO_P framework.

gistic effects among food components [59]. Together, these frameworks offer complementary strategies for disentangling the complex, conditional structure of dietary systems.

Recent studies highlight the effectiveness of GGMs in nutritional network analysis. Aguirre-Quezada and Aranda-Ramírez [8] demonstrated that GGMs can uncover distinct conditional dependency structures linking nutritional, psychological, and metabolic factors in obese populations, revealing patterns not captured by simple correlations. Extending this network perspective, Lailo et al. [71] combined GGMs with survival models to relate dietary networks to cardiovascular risk, illustrating their translational value. Similarly, GLASSO-based approaches have been used to infer high-dimensional nutrient and metabolite interaction networks, emphasizing their scalability and utility for complex dietary and lifestyle data [95].

Applying graphical models to baobab leaf composition enables a systems-level understanding of how nutrients co-vary and influence one another within this highly nutritious leafy vegetable. By discovering these nutrient networks, this study aims to (i) identify key synergistic and antagonistic interactions, (ii) highlight the biochemical components underlying these relationships, and (iii) advance the use of data-driven analytics in food composition research. Such insights have direct implications for dietary formulation, fortification strategies, and food system

modeling, supporting broader efforts toward sustainable nutrition and improved public health.

4.3.1 Materials and Methods

Baobab leaf samples were collected from two ecological zones- the Blue Nile and North Kordofan in Sudan. In each region, 30 trees were randomly sampled over an area of 1,500 km², with two leaf batches collected per tree. The leaves were shade-dried to about 93% dry matter, properly sealed, and subsequently analyzed for nutrient composition. Inductively coupled plasma–optical emission spectrometry (ICP–OES) was used for analysis.

The clear acid extract was used to quantify main elements, including calcium (Ca), magnesium (Mg), potassium (K), sodium (Na), copper (Cu), iron (Fe), manganese (Mn), and zinc (Zn), expressed as parts per million (ppm) dry matter (DM). Calibration of the instrument was achieved using certified standard solutions set for each element to confirm analytical accuracy and precision.

Due to the normality assumptions of GGMs, they are sensitive to non-normal distributions, which can distort results in datasets with significant deviations [10]. A nonparametric extension of GGMs, SGCGMs can be used to analyze skewed data prevalent in food and nutrition studies. Using SGCGMs to assess dietary intake can reveal conditional relationships among foods consumed in a single meal, thereby enhancing our understanding of nutritional patterns and food interactions. SGCGMs relax the stringent normality assumption while effectively reconstructing sparse conditional dependence structures. Rather than directly modeling the joint distribution through multivariate normality, SGCGMs utilize rank-based or nonparanormal transformations to map observed variables into a latent Gaussian space. Hence, employing SGCGMs is beneficial for uncovering complex relationships among variables without the constraints of Gaussian assumptions. This flexibility improves the accuracy of modeling nonlinear interactions and enhances the interpretability of conditional dependencies in high-dimensional data settings [42, 100].

In SGCGMs, the variable $\mathbf{X} = (X_1, \dots, X_d)^\top$ is replaced by the trans-

formed variable $f(\mathbf{X}) = (f_1(X_1), \dots, f_d(X_d))^T$ under the assumption that $f(\mathbf{X})$ follows a multivariate Gaussian distribution and $f_i(X_i)$ are monotone univariate functions. The model depends on two components: the set of univariate transformation functions $f_i(\cdot)$ and a correlation matrix $\Theta \in \mathbb{R}^{d \times d}$ be a positive-definite correlation matrix with $\text{diag}(\Theta) = \mathbf{1}_d$, both of which are estimated from the data. The correlation matrix is typically estimated using Spearman's rho and Kendall's tau, and the resulting matrix is then used as input to GLASSO.

Centrality measures describe the role and importance of variables in a graphical model. They show which nodes are most connected, most influential, or most critical to the flow of information through the network. Measures such as degree, betweenness, closeness, and eigenvector centrality help reveal the structure of dependence among variables in high-dimensional systems. In statistical applications, centrality measures improve the interpretation of graphical models by identifying variables that may drive system behavior, transmit effects, or serve as key points for monitoring and intervention. See Table 4.1.

4.3.2 Results and Discussion

Table 4.2 states the descriptive statistics and Shapiro-Wilk normality test for the proximate composition, pH, and mineral contents of baobab leaves, giving insights into their nutritional value and compositional stability. The data show a high dry matter content (92.38%), indicating low moisture and good stability during storage, which aligns with previous findings by [4]. The ash content (10.96 g/100 g) indicates a high mineral load, demonstrating that baobab leaves are an excellent source of essential inorganic nutrients. The reasonable protein level (15.40 g/100 g) implies a considerable contribution of amino acids to the plant's nutritional value. In contrast, the low-fat content (1.89 g/100 g) is average for leafy vegetables with low lipid resources.

Among minerals, $\text{Ca} = (25219 \text{ ppm})$ and $\text{K} = (14918 \text{ ppm})$ are principal, followed by $\text{Mg} = 4,999 \text{ (ppm)}$ and trace elements like $\text{Fe} = (149 \text{ ppm})$ and $\text{Mn} = (40 \text{ ppm})$. This order reveals the typical nutrient profile of African ILVs, where Ca and K are key in osmoregulation and membrane functions [72, 79]. The slightly acidic pH of 5.9 in this

Table 4.1: Common centrality measures: definitions and interpretation.

Measure	Formulation	Interpretation
Degree	$C_D(i) = \sum_{j \in V} \mathbb{I}(e_{ij})$, where $\mathbb{I}$ is an indicator function and e_{ij} is an edge between i and j .	It represents the number of edges connected to a node. Counts how many edges connect to a node. A higher value means the node has more direct links. It shows simple connectivity.
Strength	$C_S(i) = \sum_{j \in V} w_{ij} $ where $ w_{ij} $ is the absolute weight between i and j .	It is the sum of the absolute weights (e.g., partial correlation coefficients) of the edges connected to a node. A higher value means stronger overall connections. It reflects influence through weighted links.
Closeness	$C_C(i) = \left[\sum_{j \in V} d(i, j) \right]^{-1}$, where $d(i, j)$ is the minimum distance between i and j	It is the inverse of the sum of the (minimum) distance of a node to all the other nodes in the network. It can be normalized as the average length of the weighted shortest paths to all the other nodes. Measures how close a node is to all others. A higher value means shorter paths. Such nodes can quickly spread or receive information.
Betweenness	$C_B(i) = \sum_{j \neq k} g_{jk}(i)$, where $g_{jk}(i)$ is the number of shortest paths between nodes j and k that pass through i	It represents the number of geodesics between any two other nodes that pass through node i . Counts how often a node lies on shortest paths. A higher value means the node acts as a bridge. It connects different parts of the network.

Table 4.2: Summary statistics of key nutrient components in baobab leaves.

	mean	sd	min	max	range	skew	kurt	Shapiro-Wilk
Dry m.	92.38	0.27	91.70	92.91	1.21	-0.78	-0.05	0.9019 (< 0.000)
Ash	10.96	1.22	8.90	12.85	3.95	-0.15	-0.74	0.907 (< 0.000)
Fat	1.89	0.26	1.48	2.32	0.84	0.07	-1.11	0.922 (< 0.000)
Protein	15.40	1.35	13	16.89	3.89	-0.61	-1.28	0.8237 (< 0.000)
pH	5.90	0.12	5.73	6.16	0.43	0.75	-0.59	0.9004 (< 0.000)
Ca	25219.63	3415.65	19840	31940	12100	0.27	-0.99	0.9527 (0.021)
K	14918.07	2655.66	11180	18690	7510	-0.32	-1.46	0.8532 (< 0.000)
Na	73.43	8	62.29	98.41	36.12	1.17	1.28	0.9053 (< 0.000)
Mg	4999.80	691.88	4214	6461	2247	0.83	-0.54	0.8684 (< 0.000)
Fe	149.31	31.33	108.40	211.10	102.7	0.53	-1.20	0.8884 (< 0.000)
Mn	40.15	28.08	2.88	97.01	94.13	1.16	-0.17	0.7166 (< 0.000)
TPC	361.65	19.26	326	404	78	0.61	-0.24	0.94088 (0.006)
TFC	1399.28	106.46	1226	1569	343	-0.17	-1.14	0.9137 (< 0.000)

result helps stabilize the phenolic content and enhance the mineral solubility of baobab leaves [111]. Standard deviation and range indicate biological variability that likely reflects soil components and ecological factors [14]. For the univariate case, the combined evidence from skewness and kurtosis indicates that many variables deviate from normality, primarily due to asymmetry and, in some cases, lighter or heavier tails. These findings are further supported by the Shapiro–Wilk test, which confirms a violation of the normality assumption. Multivariate normality was further assessed using Mardia’s measures of skewness and kurtosis. The estimated Mardia kurtosis was 197.11, with an associated p-value of 0.679, indicating no significant deviation from normality in tail behavior. In contrast, Mardia’s skewness statistic was 71.32, with a highly significant p-value ($p < 0.001$), providing strong evidence against multivariate normality due to asymmetry in the data. This finding is further supported by the Q-Q plot shown in Figure 4.7.

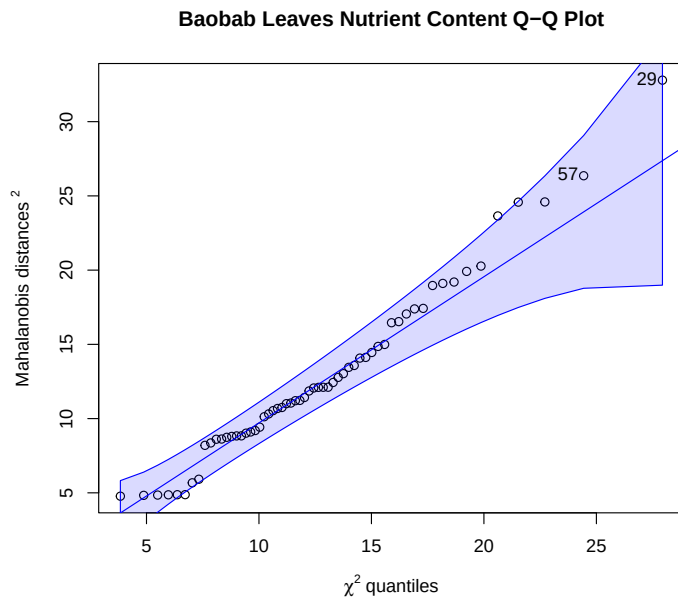


Figure 4.7: Assessing multivariate normality of the nutrient content data from Baobab leaves using a Q-Q plot.

Figure 4.8 shows the Pearson correlation coefficients among the inspected nutrients and minerals, visualized as a heatmap, where sky blue indicates stronger correlations. The figure shows multiple strong positive correlations, such as between Ash, Ca, and K, reflecting shared accumulation and mobility pathways. The co-variation among these cations suggests they are absorbed and translocated together, possibly via symplastic transport mechanisms and cation exchange processes in the leaf mesophyll [20]. In contrast, negative correlations exist between Fat and pH, and between phenolics (TPC) and minerals such as Fe and Ca, suggesting reasonable binding or appropriation. These patterns align with the well-documented metal–polyphenol complexation in plant matrices, where polyphenols chelate divalent cations, thus limiting their free ionic forms [65, 67]. The high ash–pH correlation ($r \approx 0.9$) highlights the alkaline buffering capacity given by basic cations [109], while negative correlations between Fat and minerals indicate low lipid–mineral affinity in the hydrophilic leaf tissue [93]. Though the Pearson heatmap, valuable for initial understanding, includes indirect connections, it cannot disentangle confounding influences among multiple nutrients.

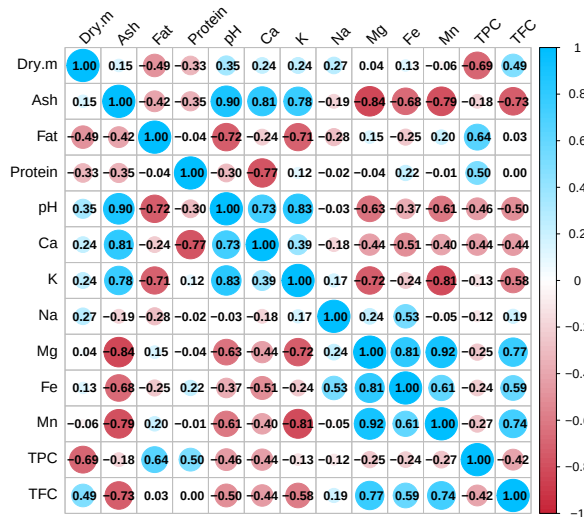


Figure 4.8: Heatmap of pairwise Pearson correlations among the analyzed elements.

With a sample size of 60 observations across 13 nutrient components, the normality test and descriptive statistics indicate that the normality assumption is violated. Consequently, the following two subsections address alternative approaches that relax the normality assumptions in identifying graphical models for this data. Therefore, EM-GLASSO_H and SGCGMs are applied.

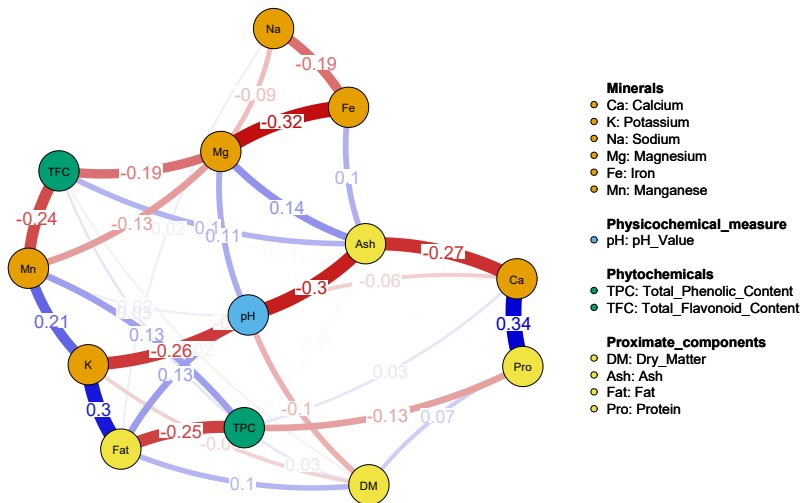


Figure 4.9: The network analysis of Baobab leaves using SGCGMs

SGCGMs: The SGCGMs approach was implemented using the **huge** R package, which provides k-fold cross-validation for selecting the penalty parameter and implicitly defines the transformation functions $f_i(\cdot)$. The optimal value of the penalty parameter (ρ) was selected using the **pulsar** R package with the Stability Approach to Regularization Selection (stars) criterion, yielding $\rho = 0.297$. As a result, Figure 4.9 shows the ability of SGCGMs to identify both linear and nonlinear associations while accommodating variables measured on different scales. Beyond the numbers, the results show that the SGCGM network for nutrient content in baobab leaves has entries close to zero, indicating weak conditional associations among many variables after controlling

for others. Still, a few clear clusters emerge that drive the network, as illustrated below.

1. **Mineral cluster:** It includes Mg, K, Mn, Fe and partly Na. Within this group, several strong links emerge: Mg shows a strong negative association with Fe (-0.317) and a moderate negative association with Mn (-0.130). Also, K connects to Mn (0.215) and Fat (0.305), while Na links negatively to Mg (-0.086) and Fe (-0.190). These relations are not random; they suggest shared pathways or competitive effects. This cluster is dense and cohesive, dominating the structure, consistent with ionomic studies demonstrating coordinated and antagonistic interactions among mineral nutrients in plant systems [112].
2. **Composition cluster:** It includes DM, Fat, and Pro. These variables show moderate positive links. DM connects to Fat (0.098) and Pro (0.066). Fat also correlates with pH (0.129) and strongly with K (0.305). Pro connects strongly to Ca (0.335). This cluster is smaller but internally consistent.
3. **pH-centered interaction layer:** pH links negatively with Ash (-0.297) and K (-0.260), and positively with Fat (0.129) and Mg (0.109). These mixed signs matter because they show that pH does not belong to a single cluster. Instead, it acts as a mediator, bridging the composition and mineral components, but with weak-to-moderate strength.
4. **Polyphenol cluster:** It includes TPC and TFC. TPC is negatively associated with Fat (-0.252) and Pro (-0.125), and positively associated with Mn (0.126). TFC connects to Mn (-0.236), Mg (-0.188), and Ash (0.100). These links are selective as they show that phenolic compounds interact more with minerals than with core composition. Still, this cluster remains loosely connected.

The role of Ash and Ca is transitional. Ash shows strong associations with pH (-0.297) and Ca (-0.274), as well as with Mg and Fe. Ca links strongly to Pro (0.335) but weakly to the others. These variables act as bridges. They connect composition to minerals, but neither dominates.

Across clusters, the strongest effects are concentrated on a few edges. The Fat–K link (0.305), Pro–Ca link (0.335), Mg–Fe link (−0.317), and Mn–TFC link (−0.236) stand out. These edges shape the network by defining how information or influence can flow.

The centrality measures shown in Figure 4.10 confirm a sparse yet well-organized structure, consistent with the network. The measures do not contradict the network; instead, they sharpen its interpretation by showing which nodes matter most once weak edges have been removed. These centrality measure results are highlighted in the following paragraphs.

Strength highlights the main actors, where pH (0.986), Mg (0.975), Ash (0.964), and K (0.869) have the highest values. These nodes carry most of the connections that remain after regularization. This result matches the network, in which these variables exhibit the strongest and most stable partial correlations. In contrast, DM (0.378) and Na (0.299) have low strength and appear peripheral, with few meaningful links.

Betweenness provides a clearer separation of roles, with K (16), Ash (15), Mg (14), Mn (14), and pH (13) dominating because they lie on many shortest paths. They connect clusters that would otherwise remain weakly linked, which is consistent with the graph structure. For example, K links the mineral cluster to Fat, while Ash connects pH, Ca, and Mg. Mg bridges to Fe and Mn, and pH links composition and mineral variables. In contrast, DM and Na have zero betweenness, as they do not act as intermediaries. Their connections are either too weak or too isolated to support paths.

Closeness shows that Ash (0.0089), pH (0.0088), and K (0.0081) are closest to all others, as they sit near the center of the network. Mg and Mn follow closely, benefiting from short paths created by the strongest edges in the graph. DM (0.0049) and Na (0.0047) again rank lowest as they remain distant from the rest of the system.

EM-GLASSO_H: Section 3.2.8 outlines our selection criteria to select between the MSGH and MPE. Based on these criteria and Mardia’s skewness and kurtosis statistics reported at the beginning of this section, we select the MSGH distribution as the base model. We estimated the

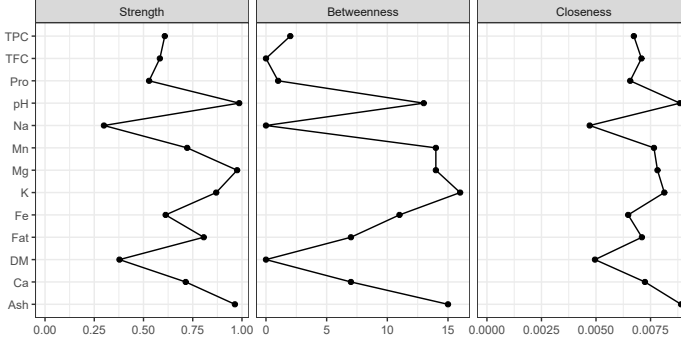


Figure 4.10: Conceptual Diagrams of Centrality Measures in SGCGMs: (a) Strength; (b) Closeness; and (c) Betweenness.

initial distributional parameters via the MLE. The estimation yielded $\lambda = -8.306286$, $\psi = 3.282104 \times 10^{-13}$, and $\chi = 14.6125$. We adopted the same optimal penalty used for the SGCGM method.

As a result, Figures 4.11 and 4.12 display the resulting graphical network and its corresponding centrality measures, respectively. The EM-GLASSO_H algorithm recovers a distinctly sparse biochemical network in which K has the highest centrality measures, operating as the primary hub connecting the overall network. The model isolates specific, strong dependencies; for example, TPC and TFC exhibit a high positive partial correlation, and Fe and Mg exhibit a strong negative correlation. Conversely, variables such as pH, DM, and fat lie on the network periphery, yielding betweenness scores of exactly zero and remaining largely isolated. The algorithm effectively shrinks weak correlations, separating core biochemical pathways from conditionally independent features.

4.3.3 Strengths and Limitations

In this section, we compared the structural networks estimated by SGCGM and EM-GLASSO_H. Because the underlying data distribution remains unknown, the strict parametric assumptions imposed by EM-GLASSO_H may distort the network structure by inducing spurious partial correlations, such as inflated conditional dependencies between DM and TFC. In contrast, SGCGM relaxes these rigid assumptions by

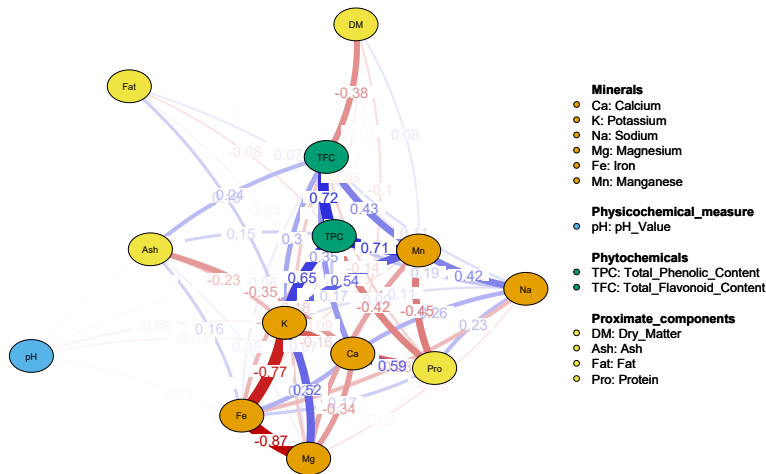


Figure 4.11: The network analysis of Baobab leaves using EM-GLASSO_H

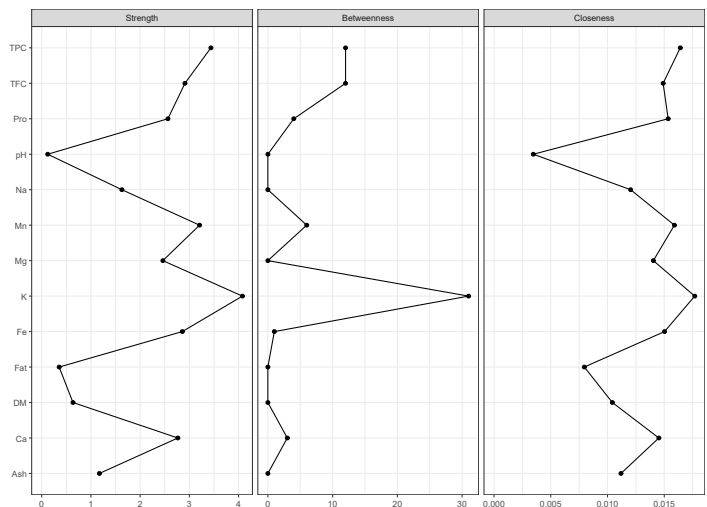


Figure 4.12: Conceptual Diagrams of Centrality Measures in EM-GLASSO_H: (a) Strength; (b) Closeness; and (c) Betweenness.

employing a semiparametric Gaussian copula to model the dependence structure without specifying the marginal distributions. This approach naturally handles skewness and unknown data-generating processes. It prevents non-normal variables from biasing the network graph. Consequently, SGCGM yields a cleaner, more robust network. It captures the true biochemical topology rather than mere distributional artifacts. In addition, limitations include restricted sampling scope, lack of causal inference, and reliance on in vitro data. Future work that combines experimental validation with multi-omics approaches is needed to confirm the mechanisms underlying these findings and extend them.

4.4 Thesis 3

Objective: Evaluate the interpretability, sparsity, and stability of the EM-GLASSO framework relative to the GLASSO benchmark using cognitive-behavioral datasets.

Key Findings: EM-GLASSO_P performs comparably to GLASSO when the Gaussian assumption is correctly specified.

Related publication: Dawod, ABA. and Terdik Gy., Mapping Cognitive-Behavioral Networks: A Graphical Lasso Approach to Dependence Structure Visualization, 2025 IEEE 16th International Conference on Cognitive Infocommunications (CogInfoCom), Wien, Austria, 2025, pp. 000109-000114, doi:10.1109/CogInfoCom66819.2025.11200813.

4.5 Thesis 4

Objective: Develop and evaluate one-sided upward and downward control charts for detecting transient process changes under the GLASSO and EM-GLASSO_P frameworks.

Key Findings: The empirical results show that the EM-GLASSO_P charts consistently outperform the standard GLASSO charts when the normality assumption is violated.

Related publications:

- Adegoke NA, Dawod ABA, Adeoti OA, Sanusi RA, Abbasi SA. (2022). Monitoring multivariate coefficient of variation for high-dimensional processes. *Qual Reliab Eng Int.* 38: 2606–2621. doi: 10.1002/qre.3094
- Dawod, ABA., Terdik, Gy. (2026). Railway Track Quality Inspection via GLasso with Zero-Crossing Segmentation. In: Zöldy, M., Szászi, I., Balasubramanian, D., Török, Á. (eds) *Proceedings of the 4th Cognitive Mobility Conference. COGMOB 2025. Lecture Notes in Networks and Systems*, vol 1768. Springer, Cham. doi: 10.1007/978-3-032-13898-928
- Dawod, ABA., Terdik, Gy. (2024). Use of Zero-crossings Segmentation for Track Quality Assessment. *Acta Technica Jaurinensis*, 17(1), 8–21. doi: 10.14513/actatechjaur.00726

4.6 Thesis 5

Objective: Infer and interpret the conditional dependence structure among nutrient components in baobab leaf samples using robust GLASSO and SGCGMs under non-Gaussian distributional assumptions.

Key Findings: When parametric assumptions fail, SGCGMs outperform EM-GLASSO_H. They strictly separate signal from noise. The resulting networks are highly interpretable and capture true conditional dependencies.

Related publication: Dawod ABA., Kovács B., Mustafa E., Ahmed A. 2026. Semiparametric Gaussian Copula Graphical Modeling for Identifying Nutrient Interaction Networks in Baobab Leaves: Implications for Healthy Nutrient Source Analysis, *PLOS One*.

5

CONCLUSION AND FUTURE DIRECTIONS

5.1 Overview of Research and Objectives

While sparse estimation techniques like GLASSO are widely used, their performance on non-Gaussian processes and their theoretical connections to conditional independence remain underexplored. Therefore, the dissertation highlights a critical need for unified frameworks that balance computational efficiency and statistical rigor across diverse data distributions without sacrificing interpretability.

These limitations are especially pronounced in practical applications such as industrial process monitoring, complex system analysis, and behavioral network studies. In these contexts, methodologies must transcend mere theoretical soundness to demonstrate robustness to data deficiencies and scalability in high-dimensional environments.

Hence, to address these challenges, this dissertation explored the effective modeling and estimation of graphical structures across both Gaussian and non-Gaussian high-dimensional datasets. The overarching objective is to develop robust, interpretable, and scalable methodologies, substantiated by rigorous theoretical analysis and validated through real-world applications.

To achieve this aim, this dissertation fulfilled the following objectives:

1. Established the theoretical link between the GLASSO framework and conditional independence.
2. Developed the EM-GLASSO algorithm to guarantee robust estimation for non-Gaussian, elliptically distributed data.

3. Applied GLASSO and EM-GLASSO to improve statistical process control charts and model behavioral networks.
4. Mapped nutrient interactions in baobab leaves using SGCGM and EM-GLASSO to advance nutritional profiling in food systems.

Together, these objectives provide a structured framework for advancing graphical modeling methodologies across diverse distributional assumptions and application domains.

5.2 Summary of Findings

This dissertation investigated the development and application of graphical modeling methodologies for Gaussian and non-Gaussian processes in high-dimensional settings. The findings provide theoretical clarification, methodological extensions, and practical validation across multiple domains.

First, the study of GLASSO clarified its relationship with conditional independence structures. The results demonstrate that, while GLASSO effectively estimates sparse precision matrices under Gaussian assumptions, careful interpretation is required when extending its conditional independence implications beyond the Gaussian setting. The analysis provides a clearer theoretical understanding of when and how sparsity patterns correspond to meaningful dependency structures.

Second, the dissertation proposed the EM-GLASSO algorithm for the elliptically distributed data. The findings show that the proposed extensions preserve sparsity and interpretability while improving robustness under heavy-tailed and non-Gaussian data. Simulation studies confirm that the extended methods outperform competing approaches when the normality assumption is violated, particularly in estimation accuracy and structural recovery.

Third, when the normality assumptions are fulfilled, GLASSO and EM-GLASSO will generate the same underlying network structure. The application of graphical modeling to SPC demonstrates that incorporating precision matrix estimation enhances monitoring performance in high-dimensional processes. The proposed approaches improve detection capability and reduce false alarm rates compared to conventional

multivariate control charts, especially in non-normal correlated systems. Similarly, in behavioral science applications, graphical models successfully capture meaningful network structures and reveal interpretable relationships among variables.

Fourth, the use of graphical models to identify network behavior in behavioral science applications reveals meaningful dependency structures among observed variables. The results show that when the data do not violate the normality assumptions, both EM-GLASSO and GLASSO can uncover interpretable relational patterns, providing insight into behavioral dynamics and latent interaction mechanisms that may not be evident with traditional modeling techniques.

Fifth, the analysis of nutrient interaction networks in baobab leaves demonstrates the versatility of graphical modeling in food systems research. The SGCGM outperformed EM-GLASSO in identifying key conditional correlations among nutrients when assumptions about the underlying distribution are violated. SGCGM offers a structured perspective on nutrient interactions and their implications for nutrient profiling and nutritional discovery. This application highlights the potential of graphical models to support evidence-based nutritional analysis and food system innovation.

Overall, the research confirms that robust and interpretable graphical models based on EM-GLASSO that extend beyond Gaussian assumptions can be effectively applied across diverse high-dimensional domains. The combined theoretical, methodological, and empirical findings strengthen both the conceptual foundation and practical implementation of graphical models in complex real-world settings.

5.3 Contributions to Knowledge

This dissertation makes significant contributions to the theory, methodology, and application of graphical modeling in high-dimensional settings. The contributions extend beyond incremental development by addressing fundamental limitations of Gaussian-based sparse estimation and demonstrating the versatility of graphical models across diverse domains.

1. **Theoretical Contributions:** This research clarifies the relationship between GLASSO estimation and conditional independence structures. While GLASSO is widely used for sparse precision matrix estimation, its interpretation in terms of conditional independence is often implicitly restricted to Gaussian settings. This dissertation provides a structured examination of the theoretical foundations underlying this connection and delineates the conditions under which sparsity patterns correspond to meaningful dependency structures.

Furthermore, by extending graphical modeling methodologies to elliptical distributions, the study advances the theoretical understanding of precision matrix estimation beyond the normal case. This extension contributes to the literature by demonstrating how conditional independence interpretations can be preserved or carefully adapted in broader distributional frameworks.

2. **Methodological Contributions:** Methodologically, the dissertation develops robust graphical modeling approaches that accommodate both Gaussian and non-Gaussian processes while maintaining sparsity, interpretability, and computational feasibility. The proposed extensions enhance the robustness of sparse precision matrix estimation for heavy-tailed, non-Gaussian data, addressing a critical limitation of conventional GLASSO-based methods. In addition, integrating graphical modeling techniques into SPC frameworks represents a methodological advancement. By embedding sparse precision matrix estimation into control chart design, the research provides an improved mechanism for monitoring high-dimensional, highly correlated processes. This integration contributes a systematic approach to enhancing detection performance while controlling false alarm rates.

3. **Applied Contributions:** The dissertation demonstrates the practical relevance of graphical models across multiple applied contexts. In behavioral science applications, the research shows how sparse graphical models can identify meaningful network behavior and reveal interpretable relational patterns among vari-

ables. Thus, the sparse graphical modelling contributes to empirical network analysis by offering a statistically principled and scalable alternative to traditional modelling approaches.

In food systems research, identifying nutrient interaction networks in baobab leaves represents a novel application of graphical modeling. By uncovering conditional dependencies among nutrients, the study provides a structured framework for nutrient profiling and nutritional discovery. This application highlights the potential of graphical models to support evidence-based decision-making in nutrition science and food innovation.

Overall, the contributions of this work strengthen the conceptual foundation of graphical modeling, expand its methodological robustness, and demonstrate its broad applicability across scientific and industrial domains.

5.4 Limitations

Despite the theoretical and applied contributions of this dissertation, several limitations should be acknowledged.

1. Although the proposed methodologies extend graphical modeling beyond the Gaussian assumption to selected elliptical distributions, the distributional scope remains restricted. Real-world data may exhibit skewness, multimodality, or complex nonlinear dependencies. While the extended frameworks improve robustness, they do not fully eliminate reliance on structural assumptions.
2. The high-dimensional settings considered in this study assume sparsity of the underlying precision matrix. Although sparsity is a reasonable assumption in many practical contexts, it may not adequately represent systems characterized by dense or weak dependency structures. The performance of sparse estimators may deteriorate when the true network deviates substantially from sparsity assumptions.

3. Computational considerations impose practical constraints. Although the proposed approaches are computationally feasible for moderately high-dimensional problems, scalability may become challenging in ultra-high-dimensional settings or in scenarios requiring real-time updating. Further optimization and parallel implementation strategies were beyond the scope of this research.

These limitations do not diminish the contributions of the dissertation but rather define the boundaries within which the results should be interpreted. They also provide a foundation for future methodological refinement and empirical expansion.

5.5 Future Research Directions

While this dissertation advances graphical modeling methodologies for Gaussian and non-Gaussian high-dimensional processes, several promising directions remain open for further investigation.

1. Broader Distributional Extensions

Future research may extend sparse graphical modeling frameworks to more general distributional families beyond elliptical distributions. In particular, investigating semiparametric or nonparametric approaches could further relax distributional assumptions while preserving the interpretability of conditional independence structures. Developing theoretical guarantees under such flexible settings would substantially broaden the applicability of graphical models in complex real-world data.

2. Dynamic and Time-Varying Networks

Many real-world systems exhibit evolving dependency structures over time. Extending sparse graphical models to dynamic or time-varying frameworks represents a valuable direction. Developing computationally efficient algorithms for estimating evolving precision matrices could enhance applications in finance, industrial monitoring, behavioral sciences, and biological systems.

3. Integration with Machine Learning Approaches

The growing intersection between statistical modeling and machine learning presents important opportunities. Future studies may explore hybrid frameworks that combine sparse graphical modeling with deep learning architectures, such as recurrent or attention-based models, to balance interpretability and predictive performance. Establishing principled connections between structured statistical models and flexible neural networks remains a promising research frontier.

Overall, future research should continue to strengthen the theoretical foundations, computational scalability, and interdisciplinary applicability of graphical models. Advancing these directions will further establish sparse graphical methodologies as powerful and versatile tools for analyzing complex high-dimensional systems.

BIBLIOGRAPHY

- [1] S. A. ABBASI, T. ABBAS, AND N. A. ADEGOKE, *Efficient cv control charts based on ranked set sampling*, IEEE Access, 7 (2019), pp. 78050–78062.
- [2] A. ABBRUZZO, I. VUJAČIĆ, A. M. MINEO, AND E. C. WIT, *Selecting the tuning parameter in penalized gaussian graphical models*, Statistics and Computing, 29 (2019), pp. 559–569.
- [3] G. M. ABDELLA, M. R. MALEKI, S. KIM, K. N. AL-KHALIFA, AND A. M. S. HAMOUDA, *Phase-i monitoring of high-dimensional covariance matrix using an adaptive thresholding lasso rule*, Computers & Industrial Engineering, 144 (2020), p. 106465.
- [4] V. ABIOYE, J. ADEJUYITAN, AND C. IDOWU, *Effects of different drying methods on the nutritional and quality attributes of baobab leaves (adansonia digitata)*, Agric. Biol. JN Am, 5 (2014), pp. 103–108.
- [5] E. G. ACHIGAN-DAKO, D. E. SOGBOHOSSOU, C. A. HOUEGBE, M. A. SALAOU, F. S. SOHINDJI, J. BLALOGOE, B. Y. CHATAIKA, H. F. ZOHOUNBOGBO, C. A. ADJE, N. V. FASSINOU HOTEgni, ET AL., *Ten years of gynandropsis gynandra research for improvement of nutrient-rich leaf consumption: lessons learnt and way forwards*, Annu Plant Rev Online, 4 (2021), pp. 767–812.
- [6] N. A. ADEGOKE, S. A. ABBASI, A. N. H. SMITH, M. J. ANDERSON, AND M. D. M. PAWLEY, *A multivariate homogeneously weighted moving average control chart*, IEEE Access, 7 (2019), pp. 9586–9597.
- [7] N. A. ADEGOKE, A. DAWOD, O. A. ADEOTI, R. A. SANUSI, AND S. A. ABBASI, *Monitoring multivariate coefficient of variation for high-dimensional processes*, Quality and Reliability Engineering International, 38 (2022), pp. 2606–2621.

-
- [8] M. A. AGUIRRE-QUEZADA AND M. P. ARANDA-RAMÍREZ, *Ir-ruption of network analysis to explain dietary, psychological and nutritional patterns and metabolic health status in metabolically healthy and unhealthy overweight and obese university students: Ecuadorian case*, *Nutrients*, 16 (2024), p. 2924.
- [9] A. E. M. AHMED, M. MOZZON, A. B. DAWOD, E. O. A. MUSTAFA, S. A. MUKARRAM, T. ELOBEID, E. A. H. MOHAMMED, AND B. KOVÁCS, *Evaluation of the nutritional impact of baobab leaves (adansonia digitata l.) as a dietary intervention to combat nutrient deficiencies and poverty-related health problems*, *Nutrients*, 16 (2024), p. 4340.
- [10] M. ALTENBUCHINGER, A. WEIHS, J. QUACKENBUSH, H. J. GRABE, AND H. U. ZACHARIAS, *Gaussian and mixed graphical models as (multi-)omics data analysis tools*, *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863 (2020), p. 194418. Transcriptional Profiles and Regulatory Gene Networks.
- [11] A. J. L. AND, *On conditional and partial correlation*, *The American Statistician*, 30 (1976), pp. 146–149.
- [12] S. A. ANDERSSON, D. MADIGAN, AND M. D. PERLMAN, *On the markov equivalence of chain graphs, undirected graphs, and acyclic digraphs*, *Scandinavian Journal of Statistics*, 24 (1997), pp. 81–102.
- [13] C. ARMOUR, E. I. FRIED, M. K. DESERNO, J. TSAI, AND R. H. PIETRZAK, *A network analysis of DSM-5 posttraumatic stress disorder symptoms and correlates in US military veterans*, *Journal of Anxiety Disorders*, 45 (2017), pp. 49–59.
- [14] I. S. ASOGWA, A. N. IBRAHIM, AND J. I. AGBAKA, *African baobab: Its role in enhancing nutrition, health, and the environment*, *Trees, Forests and People*, 3 (2021), p. 100043.
- [15] N. AUGUIN, D. MORALES-JIMENEZ, M. MCKAY, AND R. COUILLET, *Robust shrinkage m-estimators of large covariance*

- matrices*, in 2016 IEEE Statistical Signal Processing Workshop (SSP), 2016, pp. 1–4.
- [16] K. BABA, R. SHIBATA, AND M. SIBUYA, *Partial correlation and conditional correlation as measures of conditional independence*, Australian & New Zealand Journal of Statistics, 46 (2004), pp. 657–664.
- [17] O. BANERJEE, L. EL GHAOU, AND A. D’ASPREMONT, *Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data*, The Journal of Machine Learning Research, 9 (2008), pp. 485–516.
- [18] O. BARNDORFF-NIELSEN, *Exponentially decreasing distributions for the logarithm of particle size*, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 353 (1977), pp. 401–419.
- [19] A. T. BECK, *Cognitive therapy and the emotional disorders*, Penguin, 1979.
- [20] V. S. BHATI, P. K. SINGH, Y. J. PARMAR, AND P. G. ROKINS, *Foliar uptake and translocation of nanoparticles in the plants: a review*, Journal of Plant Nutrition, (2025), pp. 1–38.
- [21] M. S. BIRKELAND, T. GREENE, AND T. R. SPILLER, *The network approach to posttraumatic stress disorder: a systematic review*, European Journal of Psychotraumatology, 11 (2020), p. 1700614. PMID: 32002135.
- [22] T. F. BLANKEN, J. BATHELT, M. K. DESERNO, L. VOGEL, D. BORSBOOM, AND L. DOUW, *Connecting brain and behavior in clinical neuroscience: A network approach*, Neuroscience & Biobehavioral Reviews, 130 (2021), pp. 81–90.
- [23] J. BLASCOVICH AND W. B. MENDES, *Social psychophysiology and embodiment*, in Handbook of Social Psychology, S. Fiske, D. Gilbert, and G. Lindzey, eds., Wiley, 5 ed., 2010.

-
- [24] D. R. BRILLINGER, *Remarks concerning graphical models for time series and point processes*, Brazilian Review of Econometrics, 16 (1996), pp. 1–23.
- [25] R. P. BROWNE AND P. D. MCNICHOLAS, *A mixture of generalized hyperbolic distributions*, Canadian Journal of Statistics, 43 (2015), pp. 176–198.
- [26] J. BRUIN, I. V. STULDREHER, P. PERONE, K. HOGENELST, M. NABER, W. KAMPHUIS, AND A.-M. BROUWER, *Detection of arousal and valence from facial expressions and physiological responses evoked by different types of stressors*, Frontiers in Neuroergonomics, Volume 5 - 2024 (2024).
- [27] T. T. CAI, W. LIU, AND H. H. ZHOU, *Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation*, The Annals of Statistics, 44 (2016), pp. 455 – 488.
- [28] J. S. CARTER, D. ROSSELL, AND J. Q. SMITH, *Partial correlation graphical lasso*, Scandinavian Journal of Statistics, 51 (2024), pp. 32–63.
- [29] P. CASTAGLIOLA, A. AMDOUNI, H. TALEB, AND G. CELANO, *One-sided shewhart-type charts for monitoring the coefficient of variation in short production runs*, Quality Technology & Quantitative Management, 12 (2015), pp. 53–67.
- [30] X. CHEN, Z. FENG, AND H. PENG, *Estimation and order selection for multivariate exponential power mixture models*, Journal of Multivariate Analysis, 195 (2023), p. 105140.
- [31] Y. CHEN, A. WIESEL, AND A. O. HERO, *Robust shrinkage estimation of high-dimensional covariance matrices*, IEEE Transactions on Signal Processing, 59 (2011), pp. 4097–4107.
- [32] M. CHERTKOV, R. ABRAMS, A. M. E. SIKAROUDI, M. KRECHETOV, C. N. SLAGLE, A. EFRAT, R. FULEK, AND E. OREN, *Graphical models of pandemic*, Medrxiv, (2021), pp. 2021–02.

-
- [33] D. A. CLARK AND A. T. BECK, *Cognitive therapy of anxiety disorders: Science and practice*, Guilford Press, 2011.
- [34] I. CÓRDOBA, C. BIELZA, AND P. LARRAÑAGA, *A review of gaussian markov models for conditional independence*, Journal of Statistical Planning and Inference, 206 (2020), pp. 127–144.
- [35] P. CORTEZ, A. CERDEIRA, F. ALMEIDA, T. MATOS, AND J. REIS, *Wine Quality*. UCI Machine Learning Repository, 2009. DOI: <https://doi.org/10.24432/C56S3T>.
- [36] R. B. CROSIER, *Multivariate generalizations of cumulative sum quality-control schemes*, Technometrics, 30 (1988), pp. 291–303.
- [37] R. DAHLHAUS, *Graphical interaction models for multivariate time series*, Metrika, 51 (2000), pp. 157–172.
- [38] U. J. DANG, R. P. BROWNE, AND P. D. MCNICHOLAS, *Mixtures of multivariate power exponential distributions*, Biometrics, 71 (2015), pp. 1081–1089.
- [39] A. P. DAWID, *Conditional independence for statistical operations*, The Annals of Statistics, 8 (1980), pp. 598–617.
- [40] A. B. A. DAWOD AND GY. TERDIK, *Mapping cognitive-behavioral networks: A graphical lasso approach to dependence structure visualization*, in 2025 IEEE 16th International Conference on Cognitive Infocommunications (CogInfoCom), 2025, pp. 000109–000114.
- [41] X. DENG AND J. YAO, *On the property of multivariate generalized hyperbolic distribution and the stein-type inequality*, Communications in Statistics - Theory and Methods, 47 (2018), pp. 5346–5356.
- [42] D. DEY AND V. ZIPUNNIKOV, *Semiparametric gaussian copula regression modeling for mixed data types (sgcrm)*, arXiv preprint arXiv:2205.06868, (2022).

-
- [43] C. DIMA, E. ASSADPOUR, A. NECHIFOR, S. DIMA, Y. LI, AND S. M. JAFARI, *Oral bioavailability of bioactive compounds; modulating factors, in vitro analysis methods, and enhancing strategies*, Critical reviews in food science and nutrition, 64 (2024), pp. 8501–8539.
- [44] J. C. DONHOUEDÉ, K. V. SALAKO, A. E. ASSOGBADJO, A. I. RIBEIRO-BARROS, AND N. RIBEIRO, *The relative role of soil, climate, and genotype in the variation of nutritional value of annona senegalensis fruits and leaves*, Heliyon, 9 (2023).
- [45] B. DROPULJIĆ, S. POPOVIĆ, D. PETRINOVIĆ, AND K. ĆOSIĆ, *Estimation of emotional states enhanced by a priori knowledge*, in 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), IEEE, 2013, pp. 481–486.
- [46] S. EPSKAMP, J. KRUIS, AND M. MARSMAN, *Estimating psychopathological networks: Be careful what you wish for*, PloS one, 12 (2017), p. e0179891.
- [47] S. EPSKAMP, L. J. WALDORP, R. MÖTTUS, AND D. BORSBOOM, *The gaussian graphical model in cross-sectional and time-series data*, Multivariate behavioral research, 53 (2018), pp. 453–480.
- [48] J. FAN, Y. FENG, AND Y. WU, *Network exploration via the adaptive LASSO and SCAD penalties*, Ann. Appl. Stat., 3 (2009), pp. 521–541.
- [49] J. FAN AND R. LI, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Am. Stat. Assoc., 96 (2001), pp. 1348–1360.
- [50] K. W. FANG, S. KOTZ, AND K. W. NG, *Symmetric Multivariate and Related Distributions*, Chapman and Hall/CRC, 2017.
- [51] A. FARASAT, A. NIKOLAEV, S. N. SRIHARI, AND R. H. BLAIR, *Probabilistic graphical models in modern social network analysis*, Social Network Analysis and Mining, 5 (2015), pp. 1–18.

-
- [52] M. FINEGOLD AND M. DRTON, *Robust graphical modeling of gene networks using classical and alternative t -distributions*, The Annals of Applied Statistics, 5 (2011), pp. 1057–1080.
- [53] R. FOYGEL AND M. DRTON, *Extended bayesian information criteria for gaussian graphical models*, in Advances in Neural Information Processing Systems, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, eds., vol. 23, Curran Associates, Inc., 2010.
- [54] B. C. FRANCAZAK, R. P. BROWNE, AND P. D. MCNICHOLAS, *Mixtures of shifted asymmetric Laplace distributions*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 36 (2013), pp. 1149–1157.
- [55] J. FRIEDMAN, T. HASTIE, AND R. TIBSHIRANI, *Sparse inverse covariance estimation with the graphical lasso*, Biostatistics, 9 (2008), pp. 432–441.
- [56] M. FRYDENBERG, *The chain graph markov property*, Scandinavian journal of statistics, (1990), pp. 333–353.
- [57] E. GÓMEZ, M. GOMEZ-VIILEGAS, AND J. M. MARÍN, *A multivariate generalization of the power exponential family of distributions*, Communications in Statistics-Theory and Methods, 27 (1998), pp. 589–600.
- [58] E. GÓMEZ-SÁNCHEZ-MANZANO, M. GÓMEZ-VILLEGAS, AND J. MARÍN, *Multivariate exponential power distributions as mixtures of normal distributions with bayesian applications*, Communications in Statistics—Theory and Methods, 37 (2008), pp. 972–985.
- [59] X. HE, K. WANG, L. ZHANG, D. ZHANG, F. YANG, Q. ZHANG, S. PAN, J. LI, L. BAI, J. SUN, AND Z. LIU, *Hgatgs: Hyper-graph attention network for crop genomic selection*, Agriculture, 15 (2025).
- [60] D. HECKERMAN, *A tutorial on learning with bayesian networks*, Learning in graphical models, (1998), pp. 301–354.

-
- [61] S. G. HOFMANN, A. ASNAANI, I. J. VONK, A. T. SAWYER, AND A. FANG, *The efficacy of cognitive behavioral therapy: A review of meta-analyses*, Cognitive therapy and research, 36 (2012), pp. 427–440.
- [62] R. HOLLIDAY, N. HOLDER, C. YAMOKOSKI, J. BURDEN, L. M. SIPPEL, I. HARPAZ-ROTEM, AND N. B. SMITH, *A preliminary evaluation of the effectiveness of massed delivery of ptsd psychotherapy in va residential treatment*, General Hospital Psychiatry, 97 (2025), pp. 126–129.
- [63] H. HOTELLING, *Multivariate quality control illustrated by the air testing of sample bombsights*, in Techniques of Statistical Analysis, C. Eisenhart, M. W. Hastay, and W. A. Wallis, eds., McGraw-Hill, New York, NY, USA, 1947, pp. 111–184.
- [64] Z. HU, T. TONG, AND M. G. GENTON, *A pairwise hotelling method for testing high-dimensional mean vectors*, Statistica Sinica, 34 (2024), pp. pp. 229–256.
- [65] L. HUANG, Q. LI, Q. LI, Y. CHI, Z. XU, AND B. SHI, *Metal-phenolic networks for transformative agri-food systems: A biobased toolkit beyond molecular polyphenols*, Journal of Agricultural and Food Chemistry, (2025).
- [66] J. JANKOVÁ AND S. VAN DE GEER, *Confidence intervals for high-dimensional inverse covariance estimation*, Electronic Journal of Statistics, 9 (2015), pp. 1205 – 1229.
- [67] Z. KEJÍK, R. KAPLÁNEK, M. MASAŘÍK, P. BABULA, A. MATKOWSKI, P. FILIPENSKÝ, K. VESELÁ, J. GBUŘEK, D. SÝKORA, P. MARTÁSEK, ET AL., *Iron complexes of flavonoids-antioxidant capacity and beyond*, International journal of molecular sciences, 22 (2021), p. 646.
- [68] D. KELKER, *Distribution theory of spherical distributions and a location-scale parameter generalization*, Sankhyā: The Indian Journal of Statistics, Series A, (1970), pp. 419–430.

-
- [69] J. O. KEPHART AND S. R. WHITE, *Directed-graph epidemiological models of computer viruses*, in *Computation: the micro and the macro view*, World Scientific, 1992, pp. 71–102.
- [70] K. W. KHAW, M. B. KHOO, P. CASTAGLIOLA, AND M. RAHIM, *New adaptive control charts for monitoring the multivariate coefficient of variation*, *Computers & Industrial Engineering*, 126 (2018), pp. 595–610.
- [71] J. M. LAILO, J. SHIN, G. MENICHETTI, AND S.-A. LEE, *Network approach to evaluate the effect of diet on stroke or myocardial infarction using gaussian graphical model*, *Nutrients*, 17 (2025).
- [72] G. N. LAROCCA, E. BALDI, AND M. TOSELLI, *Understanding the role of calcium in kiwifruit: Ion transport, signaling, and fruit quality*, *Horticulturae*, 11 (2025), p. 335.
- [73] S. L. LAURITZEN, *Graphical models*, vol. 17, Clarendon Press, 1996.
- [74] F. LI AND Z. ZHAO, *Guaranteed robust large precision matrix estimation under t -distribution*, in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 3414–3419.
- [75] L. LI, K.-H. LIU, AND J. SHEEN, *Dynamic nutrient signaling networks in plants*, *Annual Review of Cell and Developmental Biology*, 37 (2021), pp. 341–367.
- [76] X. LI, R. WANG, AND M. WHANG, *Designing light for emotion: A neurophysiological approach to modeling affective responses to the interplay of color and illuminance*, *Biomimetics*, 10 (2025).
- [77] J. LINDSEY, *Multivariate elliptically contoured distributions for repeated measurements*, *Biometrics*, 55 (1999), pp. 1277–1280.
- [78] H. LIU, F. HAN, M. YUAN, J. LAFFERTY, AND L. WASSERMAN, *High-dimensional semiparametric gaussian copula graphical models*, *Annals of Statistics*, 40 (2012), pp. 2293–2326.

-
- [79] Y. LIU, J. TIAN, H. SONG, T. ZHU, C. LEI, J. DU, AND S. LI, *Osmoregulation and physiological response of largemouth bass (*micropterus salmoides*) juvenile to different salinity stresses*, International Journal of Molecular Sciences, 26 (2025), p. 3847.
- [80] C. A. LOWRY, W. H. WOODALL, C. W. CHAMP, AND S. E. RIGDON, *A multivariate exponentially weighted moving average control chart*, Technometrics, 34 (1992), pp. 46–53.
- [81] Z. LU AND L. FENG, *Robust sparse precision matrix estimation and its application*, 2025.
- [82] Z. LYU, W. M. TAI, M. KOLAR, AND B. ARAGAM, *Inconsistency of cross-validation for structure learning in gaussian graphical models*, in International Conference on Artificial Intelligence and Statistics, PMLR, 2024, pp. 3691–3699.
- [83] B. H. M., *Propriétés asymptotiques d’un estimateur bi-spectral pour un processus p -adique strictement homogène*, C. R. Acad. Sci. Paris, 326 (1998), pp. 987–990.
- [84] E. M. MABOUDOU-TCHAO AND N. DIAWARA, *A lasso chart for monitoring the covariance matrix*, Quality Technology & Quantitative Management, 10 (2013), pp. 95–114.
- [85] R. MAZUMDER AND T. HASTIE, *The graphical lasso: New insights and alternatives*, Electronic journal of statistics, 6 (2012), p. 2125.
- [86] R. J. McNALLY, D. J. ROBINAUGH, G. W. WU, L. WANG, M. K. DESERNO, AND D. BORSBOOM, *Mental disorders as causal systems: A network approach to posttraumatic stress disorder*, Clinical psychological science, 3 (2015), pp. 836–849.
- [87] A. J. McNEIL, R. FREY, AND P. EMBRECHTS, *Quantitative risk management*, Princeton Series in Finance, Princeton University Press, Princeton, NJ, 2005. Concepts, techniques and tools.

-
- [88] N. MEINSHAUSEN AND P. BÜHLMANN, *Stability selection*, Journal of the Royal Statistical Society Series B: Statistical Methodology, 72 (2010), pp. 417–473.
- [89] R. J. MUIRHEAD, *Aspects of multivariate statistical theory*, vol. 197, John Wiley & Sons, 2009.
- [90] Q. T. NGUYEN, K. P. TRAN, P. CASTAGLIOLA, G. CELANO, AND S. LARDJANE, *One-sided synthetic control charts for monitoring the multivariate coefficient of variation*, Journal of Statistical Computation and Simulation, 89 (2019), pp. 1841–1862.
- [91] M. NIKULIN AND V. VOINOV, *Unbiased estimators and their applications*, in International Encyclopedia of Statistical Science, Springer, 2025, pp. 2821–2824.
- [92] T. ONIZUKA AND S. HASHIMOTO, *Robust bayesian graphical modeling using γ -divergence*, Journal of Multivariate Analysis, (2025), p. 105461.
- [93] R. PANDEY, *Mineral nutrition of plants*, in Plant Biology and Biotechnology: Volume I: Plant Diversity, Organization, Function and Improvement, Springer, 2015, pp. 499–538.
- [94] M. S. PAOLELLA, *Intermediate probability: A computational approach*, John Wiley & Sons, 2007.
- [95] C. PATSALIS, G. IYER, M. BRANDENBURG, A. KARNOVSKY, AND G. MICHAILIDIS, *Dnea: an r package for fast and versatile data-driven network analysis of metabolomics data*, BMC bioinformatics, 25 (2024), p. 383.
- [96] K. PRAUSE, *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures*, PhD thesis, PhD thesis, University of Freiburg, 1999.
- [97] P. RAVIKUMAR, M. J. WAINWRIGHT, G. RASKUTTI, AND B. YU, *High-dimensional covariance estimation by minimizing L_1 -penalized log-determinant divergence*, Electronic Journal of Statistics, 5 (2011), pp. 935 – 980.

-
- [98] R. RICCOBELLO, M. BOGDAN, G. BONACCOLTO, P. J. KREMER, S. PATERLINI, AND P. SOBCZYK, *Sparse graphical modelling via the sorted l_1 -norm*, 2023.
- [99] R. RICCOBELLO, G. BONACCOLTO, P. J. KREMER, P. SOBCZYK, M. BOGDAN, AND S. PATERLINI, *Sparse graphical modelling for global minimum variance portfolio*, Computational Management Science, 22 (2025), p. 8.
- [100] O. ROSEN AND W. K. THOMPSON, *Bayesian semiparametric copula estimation with application to psychiatric genetics*, Biometrical Journal, 57 (2015), pp. 468–484.
- [101] D. ROSSELL AND P. ZWIERNIK, *Dependence in elliptical partial correlation graphs*, Electronic Journal of Statistics, 15 (2021), pp. 4236 – 4263.
- [102] A. J. SEDGEWICK, K. BUSCHUR, I. SHI, J. D. RAMSEY, V. K. RAGHU, D. V. MANATAKIS, Y. ZHANG, J. BON, D. CHANDRA, C. KAROLESKI, ET AL., *Mixed graphical models for integrative causal analysis with application to chronic lung disease diagnosis and prognosis*, Bioinformatics, 35 (2019), pp. 1204–1212.
- [103] W. M. SINGLETARY, *An integrative model of autism spectrum disorder: Asd as a neurobiological disorder of experienced environmental deprivation, early life stress and allostatic overload*, Neuropsychanalysis, 17 (2015), pp. 81–119.
- [104] P. TANG, H. JIANG, H. KIM, AND X. DENG, *Robust estimation of sparse precision matrix using adaptive weighted graphical lasso approach*, Journal of Nonparametric Statistics, 33 (2021), pp. 249–272.
- [105] GY. TERDIK, *Multivariate Statistical Methods; Going Beyond the Linear*, Frontiers in Probability and the Statistical Sciences, Springer International Publishing, Cham, Switzerland, 2022.
- [106] GY. TERDIK, *On Fisher information for multivariate elliptically contoured distributions*, Statistics & Probability Letters, 234 (2026), p. 110690.

-
- [107] G. TORRI, GY. TERDIK, E. TAUFER, R. GIACOMETTI, AND S. PATERLINI, *Generalized precision matrices to capture financial dependence*, Available at SSRN 4638352, (2023).
- [108] C. TRAN AND G. YU, *A completely tuning-free and robust approach to sparse precision matrix estimation*, in Proceedings of the 39th International Conference on Machine Learning, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, eds., vol. 162 of Proceedings of Machine Learning Research, PMLR, 17–23 Jul 2022, pp. 21733–21750.
- [109] E. UWIRINGIYIMANA, H.-W. LAI, N. NI, R.-Y. SHI, X.-Y. PAN, J.-N. GAO, M. R. BISWASH, J.-Y. LI, X.-M. CUI, AND R.-K. XU, *Comparative efficacy of alkaline slag, biomass ash, and biochar application for the amelioration of different acidic soils*, Plant and Soil, 504 (2024), pp. 47–61.
- [110] I. VUJAČIĆ, A. ABBRUZZO, AND E. WIT, *A computationally fast alternative to cross-validation in penalized gaussian graphical models*, Journal of Statistical Computation and Simulation, 85 (2015), pp. 3628–3640.
- [111] H. WANG, W. DI, X. GAO, Y. GUO, T. TANG, X. BAI, AND H. CAO, *Materials, syntheses and biomedical applications of nano-quercetin formulations: A comprehensive literature review*, International Journal of Nanomedicine, (2025), pp. 8729–8764.
- [112] T. WATANABE, E. MAEJIMA, T. YOSHIMURA, M. URAYAMA, A. YAMAUCHI, M. OWADANO, R. OKADA, M. OSAKI, Y. KANAYAMA, AND T. SHINANO, *The ionomic study of vegetable crops*, PLOS ONE, 11 (2016), pp. 1–19.
- [113] M. WEIBEL, D. LUETHI, AND H.-E. BREYMAN, *ghyp: Generalized Hyperbolic Distribution and Its Special Cases*, 2024. R package version 1.6.5.
- [114] J. WHITTAKER, *Graphical models in applied multivariate statistics*, Wiley Publishing, 2009.

-
- [115] L. XUE AND H. ZOU, *Regularized rank-based estimation of high-dimensional nonparanormal graphical models*, The Annals of Statistics, 40 (2012), pp. 2541–2571.
- [116] E. YANG AND A. C. LOZANO, *Robust gaussian graphical modeling with the trimmed graphical lasso*, in Advances in Neural Information Processing Systems, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, eds., vol. 28, Curran Associates, Inc., 2015.
- [117] W. C. YEONG, M. B. C. KHOO, W. L. TEOH, AND P. CASTAGLIOLA, *A control chart for the multivariate coefficient of variation*, Quality and Reliability Engineering International, 32 (2016), pp. 1213–1225.
- [118] M. YUAN AND Y. LIN, *Model selection and estimation in the gaussian graphical model*, Biometrika, 94 (2007), pp. 19–35.
- [119] T. ZHAO, K. ROEDER, AND H. LIU, *Positive semidefinite rank-based correlation matrix estimation with application to semiparametric graph estimation*, Journal of Computational and Graphical Statistics, 23 (2014), pp. 895–922.
- [120] C. ZOU AND P. QIU, *Multivariate statistical process control using lasso*, Journal of the American Statistical Association, 104 (2009), pp. 1586–1596.
- [121] G. ŠIMIĆ, M. TKALČIĆ, V. VUKIĆ, D. MULC, E. ŠPANIĆ, M. ŠAGUD, F. E. OLUCHA-BORDONAU, M. VUKŠIĆ, AND P. R. HOF, *Understanding emotions: Origins and roles of the amygdala*, Biomolecules, 11 (2021).



Registry number:

DEENK/425/2026.PL

Subject:

PhD Publication List

Candidate: Abdaljbbar Babiker Abdaljbbar Dawod

Doctoral School: Doctoral School of Informatics

List of publications related to the dissertation

Foreign language international book chapters (1)

1. **Dawod, A. B. A.**, Terdik, G.: Railway Track Quality Inspection via GLasso with Zerocrossing Segmentation.
In: Springer, New York, 244-250, 2026, (Lecture Notes in Networks and Systems, ISSN 2367-3389 ; 1768.) ISBN: 9783032138972

Foreign language scientific articles in Hungarian journals (1)

2. **Dawod, A. B. A.**, Terdik, G.: Use of Zero-crossings Segmentation for Track Quality Assessment.
Acta Tech. Jaurinensis. 17 (1), 8-21, 2024. EISSN: 2064-5228.
DOI: <http://dx.doi.org/10.14513/actatechjaur.00726>

Foreign language scientific articles in international journals (3)

3. Terdik, G., **Dawod, A. B. A.**: Extending GLASSO to Non-Gaussian Settings: Sparse Concentration Estimation via EM Algorithm.
Comput. Stat. "Accepted by Publisher", 1-33, 2026. ISSN: 0943-4062.
IF: 1.4 (2024)
4. **Dawod, A. B. A.**, Kovács, B., Eltayeb, O. A. M., Ahmed, A. E. M.: Semiparametric Gaussian Copula Graphical Modeling for Identifying Nutrient Interaction Networks in Baobab Leaves: Implications for Healthy Nutrient Source Analysis.
PLOS One. [Under review], 1-22, 2026. ISSN: 1932-6203.
IF: 2.6 (2024)
5. Adegoke, N. A., **Dawod, A. B. A.**, Adeoti, O. A., Sanusi, R. A., Abbasi, S. A.: Monitoring multivariate coefficient of variation for high-dimensional processes.
Qual. Reliab. Eng. Int. 38 (5), 2606-2621, 2022. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.3094>
IF: 2.3





Foreign language conference proceedings (1)

6. **Dawod, A. B. A.**, Terdik, G.: Mapping Cognitive-Behavioral Networks: A Graphical Lasso Approach to Dependence Structure Visualization.
In: 2025 IEEE 16th International Conference on Cognitive Infocommunications (CogInfoCom) /CogInfo (szerk.) ISBN: 9798350356922

List of other publications

Foreign language scientific articles in international journals (9)

7. Ahmed, A. E. M., Mozzon, M., **Dawod, A. B. A.**, Awad Mustafa Eltayeb, O., Shaikh, A. M., Kovács, B.: Potential of Baobab (*Adansonia digitata* L.) Dried Fruit Pulp in Contributing to Food and Nutrient Security in Sudan and Senegal.
J. Agric. Food Res. 19, 1-9, 2025. ISSN: 2666-1543.
DOI: <http://dx.doi.org/10.1016/j.jafr.2025.101681>
IF: 6.2 (2024)
8. **Dawod, A. B. A.**, Nsaif, M., Alshuwaili, D. G. H., Abdi, I. H., Neamah, H. A.: Value-at-Risk Estimation in the Tadawul Exchange: A Comparison of Recurrent Neural Networks and GARCH Models.
Systems and Soft Computing. 7, 1-14, 2025. EISSN: 2772-9419.
DOI: <https://doi.org/10.1016/j.sasc.2025.200411>
IF: 3.6 (2024)
9. Ahmed, A. E. M., Mozzon, M., **Dawod, A. B. A.**, Eltayeb, O. A. M., Shaikh, A. M., ElObeid, T., Mohammed, E. A. H., Kovács, B.: Evaluation of the Nutritional Impact of Baobab Leaves (*Adansonia digitata* L.) as a Dietary Intervention to Combat Nutrient Deficiencies and Poverty-Related Health Problems.
Nutrients. 16, 1-21, 2024. EISSN: 2072-6643.
DOI: <https://doi.org/10.3390/nu16244340>
IF: 5
10. Al, M. M., **Dawod, A. B. A.**: Model Selection and Post Selection to Improve the Estimation of the ARCH Model.
J. Risk Financ. Manag. 15 (4), 1-17, 2022. ISSN: 1911-8066.
DOI: <http://dx.doi.org/10.3390/jrfm15040174>
11. **Dawod, A. B. A.**, Adegoke, N. A., Abbasi, S. A.: Efficient linear profile schemes for monitoring bivariate correlated processes with applications in the pharmaceutical industry.
Chemometrics Intell. Lab. Syst. 206, 1-16, 2020. ISSN: 0169-7439.
DOI: <http://dx.doi.org/10.1016/j.chemolab.2020.104137>
IF: 3.491





12. Adegoke, N. A., Abbasi, S. A., **Dawod, A. B. A.**, Pawley, M. D.: Enhancing the performance of the EWMA control chart for monitoring the process mean using auxiliary information.
Qual. Reliab. Eng. Int. 35 (4), 920-933, 2018. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.2436>
IF: 1.409
13. **Dawod, A. B. A.**, Al, M. M., Abbasi, S. A.: On efficient estimation strategies in monitoring of linear profiles.
Int. J. Adv. Manuf. Technol. 96 (9-12), 3977-3991, 2018. ISSN: 0268-3768.
DOI: <http://dx.doi.org/10.1007/s00170-018-1835-y>
IF: 2.496
14. **Dawod, A. B. A.**, Abbasi, S. A., Al, M. M.: On the performance of coefficient of variation control charts in Phase I.
Qual. Reliab. Eng. Int. 34 (6), 1029-1040, 2018. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.2305>
IF: 1.409
15. **Dawod, A. B. A.**, Riaz, M., Abbasi, S. A.: On Model Selection for Autocorrelated Processes in Statistical Process Control.
Qual. Reliab. Eng. Int. 33 (4), 867-882, 2016. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.2063>
IF: 1.366

Foreign language abstracts (1)

16. **Dawod, A. B. A.**: Time series processes monitoring using statistical process control and machine learning approaches with applications on social network surveillance.
In: ICAI 2023 - 12th International Conference on Applied Informatics, Eszterházy Károly Egyetem, Eger, 1-2, 2023.

Total IF of journals (all publications): 31,271

Total IF of journals (publications related to the dissertation): 6,3

The Candidate's publication data submitted to the Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

08 June, 2026

