Development of a Probabilistic Multi-Class Model Selection Algorithm for High-Dimensional and Complex Data

Madeline Anne Ausdemore
South Dakota State University

DEVELOPMENT OF A PROBABILISTIC MULTI-CLASS MODEL SELECTION ALGORITHM FOR HIGH-DIMENSIONAL AND COMPLEX DATA

MADELINE AUSDEMORE

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Computational Science and Statistics

South Dakota State University

2021

Development of a Probabilistic Multi-Class Model Selection Algorithm for High-Dimensional and Complex Data

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Computational Science and Statistics degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

DocuSigned by:
Kurt Cogswell
DA1993DDD148408...
3/31/2021 | 05:37 PDT

Kurt Cogswell
Advisor | Date

DocuSigned by:
Kurt Cogswell
DA1993DDD148408...
3/31/2021 | 05:37 PDT

Kurt Cogswell
Department Head | Date

DocuSigned by:
Nicole B Lounsbery
1CD8E984DF2647F...
3/31/2021 | 08:22 CDT

Nicole Lounsbery, PhD
Director, Graduate School | Date

ACKNOWLEDGEMENTS

I would like to thank Dr. Kurt Cogswell, Dr. Semhar Michael, Dr. Cedric Neumann, Dr. Christopher Saunders, and Dr. Robert Thaler for their involvement as I worked through the different phases of this research, and for their guidance in working through the various model developments. In particular, I would like to thank Dr. Cogswell for providing a stimulating and supportive learning environment, and for being such a strong source of encouragement throughout my time at South Dakota State University.

I would like to thank my mom, Tiffany Ausdemore for keeping me sane and secure with her constant cheerleading and her unconditional love and support. I would like to thank my dad, David Ausdemore, for instilling a sense of determination in me, for ensuring my success as a "mathlympian", and for always encouraging me to "show my work" and "be a problem solver". I would also like to thank my brother and my sister-in-law, Max and Jayma Ausdemore, for always being there to share a laugh and a slice of pizza.

I would also like to thank the many mentors I've had the privilege to work with and learn from over the years: I would like to thank Dr. Austin Hicklin of Noblis for expanding and developing my interest in forensic statistics; I would like to thank Dr. Diana Wright of the Federal Bureau of Investigation for exposing me to new perspectives in forensic science; I would like to thank Mr. Sean Nolan of the Central Intelligence Agency for introducing me to intriguing and scintillating research questions; and I would like to thank my eighth grade mathematics teacher, Mrs. Susan LaViola of Alton C. Crews Middle School, and my Calculus II professor, Dr. Kristina Garrett of St. Olaf College, without whom I would not have realized my enjoyment or capacity for mathematics.

Finally, I would like to thank the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, who supported this research under award numbers 2014-IJ-CX-K088 and 2018-R2-CX-0010[1].

[1]The opinions, finding, and conclusions or recommendations expressed in this dissertation are those of the author and do not necessarily reflect those of the Department of Justice

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DEVELOPMENT OF A PROBABILISTIC MULTI-CLASS MODEL SELECTION ALGORITHM FOR HIGH-DIMENSIONAL AND COMPLEX DATA

MADELINE AUSDEMORE

2021

The development of quantifiable measures of uncertainty in forensic conclusions has resulted in the debut of several ad-hoc methods for approximating the weight of evidence (WoE). In particular, forensic researchers have attempted to use similarity measures, or scores, to approximate the weight of evidence characterized by high-dimensional and complex data.

Score-based methods have been proposed to approximate the WoE for numerous evidence types (e.g., fingerprints, handwriting, inks, voice analysis). In general, score-based methods consider the score as a projection onto the real line. For example, the score-based likelihood ratio evaluates and compares the likelihoods of a score calculated between two objects in two density functions, based on sampling distributions of the score under two mutually exclusive propositions. Other score-based methods have been proposed [6, 7, 31, 82], which do not rely on such a ratio.

This dissertation focuses on a class of kernel-based algorithms that fall in the latter group of score-based methods, and introduces a model that serves to complete the class of kernel-based algorithms initiated under NIJ Awards 2009-DN-BX-K234 and 2015-R2-CX-0028, which addressed the "outlier detection" and "common source" problems, by proposing a fully probabilistic model for addressing the "specific source" problem. This "specific source" problem is addressed in three progressive models: first, the problem is addressed for a pair of fixed sources; next, the two-class model is extended to consider multiple fixed sources; finally, a kernel-based model selection algorithm is developed to consider a single fixed source juxtaposed with multiple random sources.

This class of algorithms relates pairs of high-dimensional, complex objects

through a kernel function to obtain a vector of within-source and between-source scores, and capitalizes on the variability that exists within and between these sets of scores. The model makes no assumptions about the type or dimension of data to which it can be applied, and can be tailored to any type of data by modifying the kernel function at the core of the model. In addition, this algorithm provides a naturally probabilistic, multi-class, and compact alternative to current kernel-based pattern recognition methods such as support vector machines, relevance vector machines, and approximate Bayesian computation methods.

Part I

Introduction

OVERVIEW OF PART I: INTRODUCTION

The following chapters serve to review the background information that is relevant to reading and understanding the various aspects of this dissertation. Definitions, propositions, and theorems that may be useful as points of reference throughout the dissertation are highlighted in grey.

Chapter 1 presents the reader with the fundamentals of kernel theory. In particular, it introduces the development of kernel theory, starting with the first mentions of the kernel in the early 1900s, and moves to discuss the properties of the kernel functions used in today's pattern recognition and machine learning algorithms. Together, the information considered in this chapter allows for understanding the development and resulting implications of the class of algorithms discussed, developed and tested throughout this proposal.

Chapter 2 provides a brief overview of the different frameworks, methods, and models used to quantify the weight of forensic evidence to bring to light the novelty and necessity of the proposed algorithm. In addition, this chapter discusses the set of models that laid the groundwork for the class of algorithms developed in this proposal. These algorithms are of particular importance in that each builds off of the previous to allow for developing the final model presented in this proposal. Together, the information considered in this chapter allows for pondering, discussing, and addressing the problem at hand.

Chapter 1

KERNEL METHODS

To facilitate later conversation surrounding the various models overviewed in Chapter 2, we will review the relevant concepts underlying the theory of kernel functions. In particular, this chapter will discuss the origin and first use of kernel functions, will present the properties of kernels and how these properties can be leveraged in different scenarios, and will define several types of kernel functions.

1.1 DEVELOPMENT OF KERNEL THEORY

The discussion of kernel functions first began in 1904 with the publication of David Hilbert's paper [32] on integral equations (English translation available from Stewart [80]). In his development, he defines a continuous symmetric function $\kappa(\boldsymbol{x}, \boldsymbol{x}')$, which will be referred to as a *kernel function*.

Definition 1 (Hilbert's Kernel Function) *Consider the following measure,*

$$\begin{aligned} \kappa: \quad & \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}, \\ & (\boldsymbol{x}, \boldsymbol{x}') \mapsto \kappa(\boldsymbol{x}, \boldsymbol{x}'), \end{aligned}$$

where, given two observations, $\boldsymbol{x}$ and $\boldsymbol{x}'$, κ returns a real number that describes the similarity of the two objects. We call the function, κ, a kernel function.

While Hilbert assumed that the function κ was continuous and symmetric, in 1909, James Mercer [45, 89] refined the set of kernel functions to include those functions which are continuous, symmetric, and "of positive type" (e.g., positive semi-definite). In particular, he showed that functions of this type are able to be represented as inner products in another space.

Theorem 1 (Mercer's Theorem) *A continuous, symmetric function $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ in $L_2(C)$ has an expansion*

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \sum_{k=1}^{\infty} \lambda_k e_k(\boldsymbol{x}) e_k(\boldsymbol{x}')$$

if and only if

$$\int_C \int_C \kappa(\boldsymbol{x}, \boldsymbol{x}') \, g(\boldsymbol{x}) g(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \geq 0$$

for all $g \in L_2(C)$, where $\{e_k(\boldsymbol{x})\}_k$ is an orthonormal basis of C with corresponding eigenvalues $\{\lambda_k\}_k \geq 0$, C is a compact subset of $\mathbb{R}^n$, and g is a function that satisfies $\int g^2(u) du < \infty$).

In plain terms, this theorem tells us that any continuous, symmetric, positive semi-definite function $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ is guaranteed to have a representation as an inner product in some (not necessarily known) higher dimensional feature space. These higher dimensional feature spaces, in which the concept of the dot product exists, are called *Hilbert spaces* (see Capiński and Kopp [15] for a more in-depth discussion). We can use this notion to more formally define the concept of a kernel function.

Definition 2 (Mercer's Kernel Function) *Define the mapping of an observation from its original space $\mathcal{X}$ to some Hilbert space $\mathcal{H}$ by a function ϕ, such that*

$$\begin{aligned} \phi: \quad & \mathcal{X} \mapsto \mathcal{H} \\ & \boldsymbol{x} \mapsto \phi(\boldsymbol{x}). \end{aligned}$$

Then the kernel function describing the similarity between two objects $\boldsymbol{x}$ and $\boldsymbol{x}'$ is given by

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the inner product, or dot product, between two vectors.

In 1964, Aizerman, Braverman, and Rozonoer [3] extended the results of Mercer's work to the context of machine learning algorithms. In particular, they replaced the "potential function" (i.e., kernel function) in their algorithm with the inner product taken in the "linearization space" (i.e., Hilbert space) to prove that their method of potential functions converged to the linear perceptron algorithm (a supervised binary classification algorithm in a linear space), developed by Rosenblatt in 1958 [69]. By realizing the algorithmic implications of Mercer's theorem, they paved the way for what is certainly a paramount notion in kernel theory, the aptly named *kernel trick* (see [12, 33, 65, 91] for alternative definitions).

Proposition 1 (The Kernel Trick) *Consider a linear algorithm that is expressed in terms of inner products. By replacing each inner product with some kernel function, the algorithm can be executed entirely in the feature space associated with the considered kernel function without ever visiting that feature space.*

This "kernel trick" eschews the need to identify the feature space, project the observations into this feature space, and compute their inner product in this feature space. Given that the sought after feature space may very well be of infinite dimensionality (as is the case with the feature space associated with the Gaussian kernel), forgoing these calculations is certainly advantageous. Finally, the kernel trick allows for easily developing new models by replacing one kernel function with another. This is particularly convenient for applying a given classification or regression model to different types and dimensions of data.

Although the concept of replacing inner products with their kernel representations was realized in 1964, the first formal use of this mechanism to design new algorithms did not occur until 1992. In their work, Boser, Guyon, and Vapnik [13] use kernel functions to extend the results of the Generalized Portrait Method [88] to nonlinear spaces to develop the *Support Vector Machine* (SVM). This algorithm maximizes the margin between training data and a decision boundary to obtain the optimal separating hyperplane in the feature space. In their development, Boser, Guyon, and Vapnik consider various kernel functions. In particular, they consider Radial Basis Kernel Functions and Polynomial Kernel functions (the forms of these kernels, and others, can be found in Section 1.2). Although the initial development of SVMs can be traced back to the theoretical developments associated with the introduction of statistical learning theory [87, 88], the work by Boser, Guyon, and Vapnik [13] introduced SVMs in their current form. In 1995, Cortes and Vapnik [17] extended the results of Boser, Guyon, and Vapnik [13] by introducing a soft-margin classifier, which allows for considering misclassified points. Other kernel-based algorithms include kernel principal component analysis [75, 76], kernel Fisher discriminant analysis [8, 48, 70], and probabilisitic SVMs (i.e., Relevance Vector Machines) [85].

The introduction of the SVM laid the foundation for the future of kernel research: in the the past twenty years, researchers have focused on generalizing and extending the theoretical components of SVMs (e.g., [24, 57, 77–79, 81, 89]); developed tricks for implementation (e.g., [17, 25, 36, 63, 89]); and applied SVMs to various fields of research (e.g., [37, 40, 41, 73]).

1.1.1 SUPPORT VECTOR MACHINES: A CLOSER LOOK

SVMs laid the foundation for developing pattern recognition techniques, and are still considered to be cutting-edge in the fields of pattern recognition and machine learning. Because they remain a fundamental tool for pattern recognition, we will use this method as a basis for comparing the effectiveness of the models introduced in this proposal, and so we use this section to describe SVMs in more detail (for a more in-depth development, see [89]). We will also use this section to discuss some of the limitations of SVMs, which are addressed by the models considered in this proposal.

Consider a hyperplane that exists in some inner product space, $\mathcal{H}$, that is defined by $f(\boldsymbol{z}) = 0$, where

$$f(\boldsymbol{z}) = \langle \boldsymbol{w}, \phi(\boldsymbol{z}) \rangle + b, \tag{1.1}$$

$\boldsymbol{w}$ is a vector that is orthogonal to the hyperplane, $\phi(\boldsymbol{z})$ is the input vector $\boldsymbol{z}$ projected into the Hilbert space $\mathcal{H}$, and $b \in \mathbb{R}$ is a bias. Several methods have been proposed that allow for distinguishing between two classes of observations by defining some hyperplane that separates their projections into $\mathcal{H}$. For example,Vapnik and Lerner [88], and Vapnik and Chervonenkis [86, 87] proposed the Generalized Portrait Method, a learning algorithm for linearly separable problems. Boser, Guyon, and Vapnik [13] extended this algorithm to apply to nonlinear problems. Cortes and Vapnik [17] extended the concept behind these hard margin classifiers to account for points that may be misclassified in their soft margin classifier. Though subtly different in their construction, these algorithms are all based on two fundamental notions relating to the concept of the *margin*:

> **Definition 3 (Margin)** *The margin is defined as the distance from the separating hyperplane to the point that lies closest to this hyperplane.*

(1) Given a set of hyperplanes that are able to separate some data, there exists an *optimal hyperplane* that separates the data such that the margin is maximized. Mathematically, this optimal hyperplane is the solution of

$$\max_{\boldsymbol{w}\in\mathcal{H}, b\in\mathbb{R}} \min\left\{||\phi\left(\boldsymbol{z}\right)-\phi\left(\boldsymbol{x}_i\right)|| \;\middle|\; \phi\left(\boldsymbol{x}\right)\in\mathcal{H},\ \langle\boldsymbol{w},\phi\left(\boldsymbol{z}\right)\rangle+b=0,\ i\in\{1,\ldots,\ell\}\right\}.$$

That is, given a fixed feature-space transformation $\phi(\boldsymbol{x}) \in \mathcal{H}$, we want to choose b and $\boldsymbol{w}$ such that the minimum distance of any observation $\phi(\boldsymbol{x}_i)$ in the training set from the hyperplane defined by the points $\phi(\boldsymbol{z})$ is maximized.

(2) The *VC dimension* of the set of separating hyperplanes decreases as the margin increases[1]. The VC dimension is defined as the cardinality of the largest set of points that can be arranged such that a decision function, f, shatters the points[2]. The VC dimension of the set of separating hyperplanes is less than or equal to $n + 1$, where n is the dimension of the space [90]. The VC dimension of a set of functions is responsible for the generalisability of the model. More specifically,

[1]In particular, the set of hyperplanes is restricted to lying within the area between its closest points on either side. Consider two points belonging to separate classes. As the distance between these points increases, the number of potential hyperplanes that may be drawn between them also increases.

[2]Consider a set of m points with labels ± 1. There exists, at most, 2^m ways to separate the data. If the set of separating hyperplanes is able to realize each of these 2^m separations, then the set of functions *shatters* the m points. For example, a set of $m = 3$ points in $\mathbb{R}^2$ is able to be shattered by the set of separating hyperplanes, but a set of $m > 3$ points is not able to be shattered by this set of separating hyperplanes. Thus, the VC dimension of the set of separating hyperplanes in $\mathbb{R}^2$ is 3.

when the VC dimension is small relative to the number of parameters, the machine becomes more generalisable. Thus, by maximizing the margin, we are effectively defining the most generalisable learning machine for the considered set of training data.

The hyperplane which maximizes the margin can be defined by solving a straightforward quadratic optimization problem.

1.1.1.1 HARD MARGIN SVMs

We begin by considering the development of the hard margin classifier. This classifier is ideal for situations in which the considered data are completely linearly separable in some feature space. In particular, it assumes that all points can be correctly classified. The development of this classifier will allow us to discuss in more detail the shortfalls of SVMs and how these shortcomings can be addressed by the models proposed in this dissertation.

In the hard-margin classification problem, we are interested in hyperplanes of the form (1.1) for which $y_i f(\boldsymbol{x}_i) > 0,\ \forall i$. That is, for a given object, the true class, $y_i \in \pm 1$, and the predicted class, $f(\boldsymbol{x}_i)$, are of the same sign, and so the point has been correctly classified. We can use (1.1) to define the distance of a point $\boldsymbol{x}_i$ to the hyperplane by

$$\frac{y_i f(\boldsymbol{x}_i)}{||\boldsymbol{w}||} = \frac{y_i \left(\langle \boldsymbol{w}, \phi\left(\boldsymbol{x_i}\right)\rangle + b\right)}{||\boldsymbol{w}||}. \tag{1.2}$$

Using (1.2), the maximum margin solution is then found by solving

$$\underset{\boldsymbol{w},b}{\operatorname{argmax}} \left\{ \frac{1}{||\boldsymbol{w}||} \min_i \left[y_i \left(\langle \boldsymbol{w}, \phi\left(\boldsymbol{x_i}\right)\rangle + b\right)\right] \right\}. \tag{1.3}$$

Without loss of generality, we can rescale $\boldsymbol{w}$ and b so that the point closest to the hyperplane satisfies

$$y_i \left(\langle \boldsymbol{w}, \phi\left(\boldsymbol{x_i}\right)\rangle + b\right) = 1,$$

and all other points satisfy

$$y_i \left[\langle \boldsymbol{w}, \phi\left(\boldsymbol{x}_i\right)\rangle - b\right] \geq 1,\ i \in \{1, \ldots, \ell\}. \tag{1.4}$$

We call this representation the *canonical hyperplane*.

Definition 4 (Canonical Hyperplane) *The canonical hyperplane is defined such that the minimum distance between any observation, $\phi(\boldsymbol{x}_1), \ldots, \phi(\boldsymbol{x}_\ell)$ and the hyperplane defined according to (1.1) by $(\boldsymbol{w}, b) \in \mathcal{H} \times \mathbb{R}$ is given by the value $1/||\boldsymbol{w}||$. Mathematically, this consists of ensuring that the hyperplane given by $(\boldsymbol{w}, b)$ satisfies*

$$\min_{i=\{1,\ldots,\ell\}} |\langle \boldsymbol{w}, \phi(\boldsymbol{x}_i)\rangle + b| = 1.$$

By construction, there will always be at least one point for which (1.4) is true, and, upon maximizing the margin, there will be at least two points for which (1.4) is true (i.e., there will exist at least one observation lying on either side of the hyperplane whose projection onto the hyperplane has length 1). By defining this constraint, the optimization problem in (1.3) requires only that we maximize $||\boldsymbol{w}||^{-1}$, which is equivalent to solving

$$\operatorname*{argmin}_{\boldsymbol{w}} \frac{1}{2}||\boldsymbol{w}||^2.$$

To solve this quadratic optimization problem, we can introduce the Lagrange multipliers, $\alpha_i \geq 0$ for each observation $\phi(\boldsymbol{x}_i)$, $i \in \{1, \ldots, \ell\}$. Then, rather than attempt to minimize $\frac{1}{2}||\boldsymbol{w}||^2$, we can instead define and minimize the Lagrangian, $\mathcal{L}\left(\boldsymbol{w}, b, \boldsymbol{\alpha}\right)$,

$$\mathcal{L}\left(\boldsymbol{w}, b, \boldsymbol{\alpha}\right) = \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{\ell} \alpha_i \left\{y_i \left(\langle \boldsymbol{w}, \phi\left(\boldsymbol{x}_i\right)\rangle + b\right) - 1\right\}. \tag{1.5}$$

Minimizing (1.5) is equivalent to minimizing $\frac{1}{2}||\boldsymbol{w}||^2$, since the terms α_i are defined such that the additional term, $\sum_{i=1}^{\ell} \alpha_i \left\{y_i \left(\langle \boldsymbol{w}, \phi\left(\boldsymbol{x}_i\right)\rangle + b\right) - 1\right\}$, is equal to zero. In particular, if $\phi\left(\boldsymbol{x}_i\right)$ lies on the margin, then the term inside the brackets is equal to zero, and the value of α_i is irrelevant. If $\phi(\boldsymbol{x}_i)$ lies outside of the margin, then α_i is equal to zero.

So, instead of starting from scratch with $\frac{1}{2}||\boldsymbol{w}||^2$, in which we know nothing about $\boldsymbol{w}$, we can consider $\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\alpha})$, which permits us to use differentiation techniques so that we can consider the minimum value of $||\boldsymbol{w}||$ with respect to our observations, $\phi(\boldsymbol{x}_i)$, and their known labels, y_i. We thus proceed by applying a basic calculus technique, in which we calculate the derivative of (1.5) with respect to the unknown parameters, $\boldsymbol{w}$ and b, and set each of these derivatives equal to zero, such that

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{\ell} \alpha_i y_i \phi(\boldsymbol{x}_i) = 0 \iff \boldsymbol{w} = \sum_{i=1}^{\ell} \alpha_i y_i \phi(\boldsymbol{x}_i), \tag{1.6}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{\ell} \alpha_i y_i = 0. \tag{1.7}$$

We can now define the *dual problem*, in which we re-write the original Lagrangian in (1.5) using the constraints (1.6) and (1.7) that allow for identifying the optimal $\boldsymbol{w}$, such that

$$\begin{aligned}
\widetilde{\mathcal{L}}(\alpha) &= \frac{1}{2}\sum_{i=1}^{\ell}\sum_{i'=1}^{\ell}\alpha_i\alpha_{i'}y_iy_{i'}\phi(\boldsymbol{x}_i)'\phi(\boldsymbol{x}_{i'}) - \sum_{i=1}^{\ell}\alpha_i\left\{y_i\left(\left\langle\sum_{i'=1}^{\ell}\alpha_{i'}y_{i'}\phi(\boldsymbol{x}_{i'}), \phi(\boldsymbol{x}_i)\right\rangle + b\right) - 1\right\} \\
&= \frac{1}{2}\sum_{i=1}^{\ell}\sum_{i'=1}^{\ell}\alpha_i\alpha_{i'}y_iy_{i'}\langle\phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_{i'})\rangle - \sum_{i=1}^{\ell}\alpha_i\left\{y_i\left(\sum_{i'=1}^{\ell}\alpha_{i'}y_{i'}\langle\phi(\boldsymbol{x}_{i'}), \phi(\boldsymbol{x}_i)\rangle + b\right) - 1\right\} \\
&= \frac{1}{2}\sum_{i=1}^{\ell}\sum_{i'=1}^{\ell}\alpha_i\alpha_{i'}y_iy_{i'}\langle\phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_{i'})\rangle - \sum_{i=1}^{\ell}\alpha_iy_i\left(\sum_{i'=1}^{\ell}\alpha_{i'}y_{i'}\langle\phi(\boldsymbol{x}_{i'}), \phi(\boldsymbol{x}_i)\rangle + b\right) + \sum_{i=1}^{\ell}\alpha_i \\
&= \frac{1}{2}\sum_{i=1}^{\ell}\sum_{i'=1}^{\ell}\alpha_i\alpha_{i'}y_iy_{i'}\kappa(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) - \sum_{i=1}^{\ell}\sum_{i'=1}^{\ell}\alpha_i\alpha_{i'}y_iy_{i'}\kappa(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) + b\sum_{i=1}^{\ell}\alpha_iy_i + \sum_{i=1}^{\ell}\alpha_i \\
&= -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{i'=1}^{\ell}\alpha_i\alpha_{i'}y_iy_{i'}\kappa(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) + \sum_{i=1}^{\ell}\alpha_i. \qquad (1.8)
\end{aligned}$$

Examining (1.8), we see that we are able to obtain a function that is free of the unknown parameters $\boldsymbol{w}$ and b, and that does not require defining the function ϕ, and so allows for directly working with the original observations $\boldsymbol{x}_i$.

Furthermore, by defining the following constraints, we can use quadratic programming to solve (1.8), and define the optimal hyperplane using (1.12):

(1) $\alpha_i \geq 0$: this constraint follows from our original assumption used to define (1.5);

(2) $y_i f(\boldsymbol{x}_i) - 1 \geq 0$: If an observation lies on the margin, then $y_i f(\boldsymbol{x}_i) = 1$, and so the assumption is fulfilled. Likewise, if the observation lies outside of the margin, then $y_i f(\boldsymbol{x}_i) > 1$, and the assumption is fulfilled;

(3) $\alpha_i \left(y_i f(\boldsymbol{x}_i) - 1\right) = 0$: if the observation lies on the margin, the assumption is fulfilled by (2); if the observation lies outside of the margin, then $\alpha_i = 0$.

Thus, for every observation, either $\alpha_i = 0$, or $y_i f(\boldsymbol{x}_i) = 1$. We define the points for which $\alpha_i \neq 0$ to be *support vectors*, and denote this set using $\mathcal{S}$. We use the set of points, $\mathcal{S}$, for two purposes. First, we use the results of (1.6) to completely define the projection of any observation $\boldsymbol{x}^*_{i'}$ onto $\boldsymbol{w}$ using only $\boldsymbol{\alpha}$, and the set of support vectors $\mathcal{S}$, such that

$$\langle \phi(\boldsymbol{x}_{i'}), \boldsymbol{w} \rangle = \sum_{i \in \mathcal{S}} \alpha_i y_i \kappa \left(\boldsymbol{x}^*_{i'}, \boldsymbol{x}_i\right). \tag{1.9}$$

Thus, we have eliminated the need to identify $\boldsymbol{w}$ in the feature space. Second, we can define our bias term, b. By noting that any support vector $\boldsymbol{x}_{i'} \in \mathcal{S}$ satisfies $y_{i'} f(\boldsymbol{x}_{i'}) = 1$, it follows that

$$y_{i'} \left(\sum_{i \in \mathcal{S}} \alpha_i y_i \kappa \left(\boldsymbol{x}_{i'}, \boldsymbol{x}_i\right) + b \right) = 1,$$

and so, by solving for b, we obtain

$$\begin{aligned} b &= \frac{1}{y_{i'}} - \sum_{i \in \mathcal{S}} \alpha_i y_i \kappa \left(\boldsymbol{x}_{i'}, \boldsymbol{x}_i\right) \\ &= y_{i'} - \sum_{i \in \mathcal{S}} \alpha_i y_i \kappa \left(\boldsymbol{x}_{i'}, \boldsymbol{x}_i\right), \end{aligned} \tag{1.10}$$

since $y_{i'} = \pm 1, \forall \boldsymbol{x}'_i \in \mathcal{S}$. However, to get a more robust estimate of b, we can consider the average of (1.10) across all support vectors, such that

$$b = \frac{1}{\#\mathcal{S}} \sum_{i' \in \mathcal{S}} \left(y_{i'} - \sum_{i \in \mathcal{S}} \alpha_i y_i \kappa \left(\boldsymbol{x}_{i'}, \boldsymbol{x}_i\right) \right), \tag{1.11}$$

where $\#\mathcal{S}$ is the cardinality of $\mathcal{S}$, and thus is the number of support vectors in $\mathcal{S}$. Thus, by defining the Lagrangian as in (1.8), we can effectively define the optimal hyperplane, $(\boldsymbol{w}, b)$, that separates a set of observations $\boldsymbol{x}_i$ without explicitly defining ϕ.

Finally, we can use (1.6) to define the predicted class of an observation $\boldsymbol{x}^*_{i'}$. In particular, we consider that

$$\begin{aligned} f(\boldsymbol{x}^*_{i'}) &= \langle \phi(\boldsymbol{x}^*_{i'}), \boldsymbol{w} \rangle + b \\ &= \left\langle \phi(\boldsymbol{x}^*_{i'}), \sum_{i \in \mathcal{S}} \alpha_i y_i \phi(\boldsymbol{x}_i) \right\rangle + b \\ &= \sum_{i \in \mathcal{S}} \alpha_i y_i \left\langle \phi(\boldsymbol{x}^*_{i'}), \phi(\boldsymbol{x}_i) \right\rangle + b \\ &= \sum_{i \in \mathcal{S}} \alpha_i y_i \kappa \left(\boldsymbol{x}^*_{i'}, \boldsymbol{x}_i \right) + b. \end{aligned} \tag{1.12}$$

Again, we see that we are working only with the original observations, $\boldsymbol{x}_i$, $\boldsymbol{x}^*_i$, and do not need to consider the projection of the data into some feature space.

1.1.1.2 SOFT MARGIN SVMs

In the case where we consider the *soft margin classifier*, which allows for the potential of misclassifying points, we follow the same procedure described above for the hard margin classifier. However, rather than minimize $\frac{1}{2}||\boldsymbol{w}||^2$, we seek to maximize the margin while penalizing points that lie on the wrong side of the optimal hyperplane. To achieve this, we introduce a *slack variable*, ξ_i, for each observation: we define $\xi_i = 0$ for correctly classified observations that lie on or outside the margin; $0 < \xi_i \leq 1$ for observations that lie within the margin on the correct side of the decision boundary; and $\xi_i > 1$ for observations that are misclassified. This slack variable can be thought of as a correction term, and we can interpret its value as a function of how far an observation would need to be moved so that it would be correctly classified. Furthermore, $\sum_{i=1}^{\ell} \xi_i$ can be considered to be an upper bound on the number of misclassified points.

Incorporating this condition, we thus seek to minimize

$$\frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{\ell}\xi_i, , \tag{1.13}$$

subject to the constraint

$$y_i f\left(\phi(\boldsymbol{x}_i)\right) \geq 1 - \xi_i, \tag{1.14}$$

where the term $C > 0$ controls the trade-off between the penalty induced by the slack variable and the margin. The optimization of (1.13) follows the same process as that of (1.3) for the hard margin optimization, outlined above.

1.1.1.3 LIMITATIONS OF SVMs

The development of hard- and soft-margin SVMs (see, e.g., [13, 17, 86–88]) redefined the way in which the classification of objects is approached. Not only is the development relatively straightforward, but it also ensures that the best separation is achieved for the considered data given the considered kernel, and thus ensures a higher probability of correct classification for future observations. Finally, this class of algorithms is compelling in that it is not restricted by the type or size of the data. The use of kernel functions allows any type of data to be considered, regardless of the dimensionality, quality, or quantity of the data.

These algorithms, however, are susceptible to several shortcomings. First, training the hard- or soft-margin SVMs may require a large amount of data that is not readily available to the scientist, as is the case in intelligence and forensic scenarios. It is oftentimes the case that the intelligence officer or forensic examiner may have only a limited number of control observations (in fact, it is not uncommon for an investigation to have as few as three reference objects that may be used for inferring the source of an unknown object). In such scenarios, the generalizability of SVMs trained on too few samples is questionable.

Second, these algorithms do not enjoy a natural extension to the multi-class scenario: while multi-class versions of these algorithms do exist, they partition the space in such a way that the entirety of the space is not accounted for, and the resulting inference may be affected as a result. For example, consider two extensions of the SVM to a three-class scenario, as portrayed in Figure 1.1. These two classifiers attempt to identify three regions into which an observation may be classified. However, Figure 1.1 demonstrates that this is not possible without creating an enigmatic fourth region. Clearly, we run into troubles whenever an observation is made inside this region.

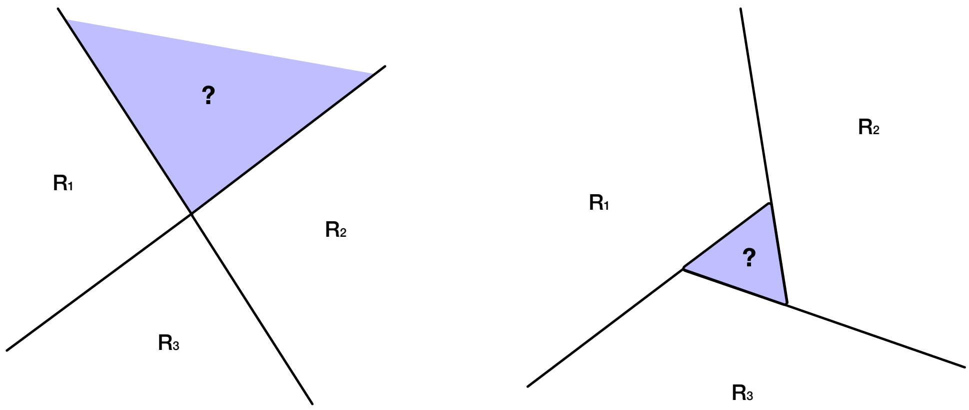


Figure 1.1: Ambiguous regions induced by extending SVMs to multi-class scenarios. Left: Two hyperplanes are used to discriminate objects into three classes. Right: Three hyperplanes are used to discriminate objects into three classes.

Finally, these algorithms do not allow for probabilistic output. While the relevance vector machine does allow for probabilistic output, it suffers from a complex design and lengthy training period, and, like its deterministic counterpart, lacks a straightforward extension to the multi-class scenario. The models proposed in Parts II and III of this dissertation address these issues.

1.2 PROPERTIES AND VARIATIONS OF KERNELS

In this section we discuss the properties and variations of kernel functions used in the types of algorithms described above. In Section 1.1, we noted that any continuous, symmetric, positive semi-definite function, κ, is a viable kernel that may be used in any kernel-based algorithm. Given this valid kernel, we can construct new kernels using this kernel as a base.

We begin by defining some kernels that arise from the basic definition of a kernel function (Definition 2): given a kernel function $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})^T\phi(\boldsymbol{x}') = \langle\phi(\boldsymbol{x}), \phi(\boldsymbol{x}')\rangle$, we can define a valid kernel by identifying a new function ϕ' in the context of the original function ϕ. For example, we can define

$$\begin{aligned}
\kappa'\left(\boldsymbol{x}, \boldsymbol{x}'\right) &:= f(\boldsymbol{x})\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right) f(\boldsymbol{x}') \\
&= f(\boldsymbol{x})\langle\phi(\boldsymbol{x}), \phi(\boldsymbol{x}')\rangle f(\boldsymbol{x}') \\
&= \langle f(\boldsymbol{x})\phi(\boldsymbol{x}), f(\boldsymbol{x}')\phi(\boldsymbol{x}')\rangle \\
&= \langle\phi'(\boldsymbol{x}), \phi'(\boldsymbol{x}')\rangle,
\end{aligned}$$

where f is any function. Similarly, we can define

$$\begin{aligned}
\kappa'\left(\boldsymbol{x}, \boldsymbol{x}'\right) &:= \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}' \\
&= \boldsymbol{x}^T\boldsymbol{A}^{1/2^T}\boldsymbol{A}^{1/2}\boldsymbol{x}' \\
&= \left(\boldsymbol{A}^{1/2}\boldsymbol{x}\right)^T\left(\boldsymbol{A}^{1/2}\boldsymbol{x}'\right) \\
&= \phi(\boldsymbol{x})^T\phi(\boldsymbol{x}) \\
&= \langle\phi(\boldsymbol{x}), \phi(\boldsymbol{x}')\rangle,
\end{aligned}$$

where $\boldsymbol{A}$ is a symmetric, positive-definite matrix, and the function $\phi = \boldsymbol{A}^{1/2}\boldsymbol{x}$. These particular examples arise from the basic properties of an inner product.

We can define an even larger class of kernels by considering the closure properties of kernel functions. Specifically, we can show that any function that is a product or sum of a set of valid kernels is itself a valid kernel [12, 33]. Proofs of these properties can be found in [9].

Proposition 2 (Closure Properties of Positive Semi-Definite Kernels) *Consider that κ_i, $i \in \{1, \ldots, N\}$ are positive definite kernels mapping from $\mathcal{X}_i \times \mathcal{X}_i$ to $\mathbb{R}$, where $\mathcal{X}_i$ is a nonempty set. Then*

(a) If $\kappa(\boldsymbol{x}, \boldsymbol{x}') := \lim_{n\to\infty} \kappa_n(\boldsymbol{x}, \boldsymbol{x}')$ exists for all $\boldsymbol{x}, \boldsymbol{x}'$, then κ is a positive definite kernel;

(b) If $\alpha_i \geq 0$, then $\sum_{i=1}^{N} \alpha_i$ is a positive definite kernel;

(c) The point-wise product, $\prod_{i=1}^{N} \kappa_i$, is a positive definite kernel;

(d) The tensor product $\kappa_i \otimes \kappa_{i'}$ is a positive definite kernel on $(\mathcal{X}_i \times \mathcal{X}_{i'}) \times (\mathcal{X}_i \times \mathcal{X}_{i'})$;

(e) The direct sum $\kappa_i \oplus \kappa_{i'}$ is a positive definite kernel on $(\mathcal{X}_i \times \mathcal{X}_{i'}) \times (\mathcal{X}_i \times \mathcal{X}_{i'})$.

These are the only functions that preserve positive definiteness. Note that, by considering closure properties (d) and (e) simultaneously, kernels can be defined as functions of other kernels.

Using the closure properties defined above, we can construct variations of some of the more popular kernels that are frequently used in various machine learning algorithms. Consider that κ_1 and κ_2 are valid kernels in $\mathcal{X} \times \mathcal{X}$. Then the following are also valid kernels on (as defined above in Proposition 2):

(i) $c\kappa_1(\boldsymbol{x}, \boldsymbol{x}')$: satisfied by closure property *(b)*, where $c \in \mathbb{R}$;

(ii) $\kappa_1(\boldsymbol{x}, \boldsymbol{x}') + \kappa_2(\boldsymbol{x}, \boldsymbol{x}')$: satisfied by closure property *(b)*;

(iii) $\kappa_1(\boldsymbol{x}, \boldsymbol{x}')\kappa_2(\boldsymbol{x}, \boldsymbol{x}')$: satisfied by closure property *(c)*;

(iv) $q(\kappa_1(\boldsymbol{x}, \boldsymbol{x}'))$: satisfied by closure properties *(b)* and *(c)*, where q is a polynomial with nonnegative coefficients;

(v) $\exp(\kappa_1(\boldsymbol{x}, \boldsymbol{x}'))$: satisfied by closure properties *(a)*, *(b)*, and *(c)*;

(vi) $\kappa_a(\boldsymbol{x}_a, \boldsymbol{x}'_a)\,\kappa_b(\boldsymbol{x}_b, \boldsymbol{x}'_b)$: satisfied by closure property *(d)*, where $\boldsymbol{x} = (\boldsymbol{x}_a, \boldsymbol{x}_b)$, and the functions κ_a and κ_b are valid kernel functions over $\mathcal{X}_a \times \mathcal{X}_a$ and $\mathcal{X}_b \times \mathcal{X}_b$;

(vii) $\kappa_a(\boldsymbol{x}_a, \boldsymbol{x}'_a) + \kappa_b(\boldsymbol{x}_b, \boldsymbol{x}'_b)$: satisfied by closure property *(e)*, where $\boldsymbol{x} = (\boldsymbol{x}_a, \boldsymbol{x}_b)$, and the functions κ_a and κ_b are valid kernel functions over $\mathcal{X}_a \times \mathcal{X}_a$ and $\mathcal{X}_b \times \mathcal{X}_b$;

(viii) $f\left(d\left(\boldsymbol{x},\boldsymbol{x}'\right)\right)$: satisfied by closure properties *(a)* and *(b)*, where f is a function on $\mathbb{R}_0^+$, and d is a metric on $\mathcal{X}$. ;

The kernel given by (viii) is called a *stationary kernel* [74]. These types of kernels allow us to explicitly consider the distance between any two observations, or between an observation and a central point of interest. The class of stationary kernels can be further restricted to define a class of *radial basis function kernels*, which consider the magnitude of the distance between two observations. Table 1.1 provides several examples of radial basis function kernels.

Definition 5 (Stationary Kernel) *A stationary kernel is a function that relies on the distance between the observations $\boldsymbol{x}$, $\boldsymbol{x}'$. These kernels are invariant to translations in the input space.*

Definition 6 (Radial Basis Function Kernel) *A radial basis function kernel is a function that relies on the magnitude of the distance between the observations $\boldsymbol{x}$, $\boldsymbol{x}'$. These kernels are a subset of the set of stationary kernels, and so are also translation invariant.*

Kernel Name	Kernel Form
Gaussian Kernel	$\exp\{-\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|^2/\sigma^2\}$, where $\sigma^2 \in \mathbb{R}^+$
Exponential Kernel	$\exp\{-\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|/2\sigma^2\}$, where $\sigma \in \mathbb{R}^+$
Laplacian Kernel	$\exp\{-\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|/\sigma\}$, where $\sigma \in \mathbb{R}^+$
Power Kernel	$-\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|^d$, where $d \in \mathbb{N}$
Log Kernel	$-\log\left(\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|^d\right)$, where $d \in \mathbb{N}$
Gaussian Spectral Kernel	$\cos\left(2\pi\mu\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|\right)\exp\left(-2\pi^2\sigma^2\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|^2\right)$, where $\mu \in \mathbb{R}$, and $\sigma^2 \in \mathbb{R}+$
Cauchy Kernel	$\left(1+\sigma^{-2}\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|\right)^{-1}$, where $\sigma^2 \in \mathbb{R}^+$
Student T Kernel	$\left(1+\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|^d\right)^{-1}$, where $d \in \mathbb{N}$
Mahalanobis Kernel	$\sqrt{\left(\boldsymbol{x}-\boldsymbol{x}'\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{x}'\right)}$, where $\boldsymbol{\Sigma}$ is a positive definite matrix

Table 1.1: Examples of Radial Basis Function Kernels

Chapter 2

QUANTIFYING THE WEIGHT OF FORENSIC EVIDENCE

Since the institutionalization of modern scientific techniques in criminal investigations in the late 19th century, there has been interest in using statistical techniques to support the inferences made on sets of forensic evidence (see [83] for a comprehensive review). From the introduction of the first identification system by Alphonse Bertillon in 1886 [10], it was only a matter of years before probabilities were being reported in court [19], and, in 1904, a move towards Bayesian reasoning was proposed [83].

For the better part of a century, legal and scientific scholars have widely advocated for the use of Bayesian reasoning in handling the uncertainty associated with determining the source of a piece of forensic evidence [2, 23]. At the core of Bayesian reasoning lies the *Bayes factor*, which allows for updating the prior beliefs surrounding two competing propositions about the source of the evidence. Given two competing propositions, $\mathcal{H}_p$ and $\mathcal{H}_d$[1], regarding the source of a piece of evidence, x_u, a Bayes factor can be assigned, in which the likelihoods under the two propositions, given by $f(\cdot)$, are compared. In the case where the parameters are known and certain, it is not necessary to integrate over these parameter spaces, and a likelihood ratio may be considered instead. Ommen and Saunders [58] show that the Bayes factor converges to the true likelihood ratio as the amount of available information increases.

[1] $\mathcal{H}_p$ and $\mathcal{H}_d$ are typically used by the legal and forensic science communities to represent the *prosecution* and *defense* propositions. Traditionally, the prosecution will take some stance as to how a piece of evidence is believed to have been generated, while the defense will claim that the evidence could have been generated by other means. These thought patterns are reflected by $\mathcal{H}_p$ and $\mathcal{H}_d$, respectively.

Definition 7 (Bayes Factor) *The Bayes factor takes the form*

$$\mathcal{BF} = \frac{\int f\left(x_u|\mathbf{\Omega}, \mathcal{H}_p\right) d\Pi\left(\mathbf{\Omega}|\mathcal{H}_p\right)}{\int f\left(x_u|\mathbf{\Omega}, \mathcal{H}_d\right) d\Pi\left(\mathbf{\Omega}|\mathcal{H}_d\right)},$$

where $f(\cdot)$ *is the likelihood of observing the evidence,* $\mathbf{\Omega}$ *is a vector of parameters,* Π *is a probability measure over the parameter spaces of* $\mathbf{\Omega}$*, and the uncertainty of the parameters under* $\mathcal{H}_p$ *and* $\mathcal{H}_d$ *is accounted for by integrating over the parameter space.*

Definition 8 (Likelihood Ratio) *The likelihood ratio takes the form*

$$\mathcal{LR} = \frac{f\left(x_u|\mathbf{\Omega}_0, \mathcal{H}_p\right)}{f\left(x_u|\mathbf{\Omega}_0, \mathcal{H}_d\right)},$$

where $\mathbf{\Omega}_0$ *denotes the set of known parameters.*

It is important to note that the Bayes factor is not an intrinsic property of the evidence in itself, and that there is no true or universal Bayes factor for a given piece of evidence. Surely, different weights will be assigned to the same evidence if different propositions are considered. Moreover, even if a fixed pair of alternative propositions is considered, different scientists may assign different weights based on their personal handling of the available evidentiary material. For example, the evidence may be characterized using different types of features or measured using different analytical techniques (e.g., glass fragments may be characterized by their refractive index, by their elemental composition, or by their chemical structure), the data may be summarized or organized in different ways (e.g., while Neumann, Evett, and Skerrett [50] describe a method to characterize the spatial relationships between fingerprint minutiae using triangles, and then use these triangles to assign probability distributions to the minutiae constellations, it is certainly possible to characterize the spatial relationships between minutiae in many other ways), or different assumptions may be used to model the distributions of the measured characteristics (e.g., given a set of observations, one scientist may elect to rely on normality assumptions, while another scientist may elect to consider a non-parametric model).

It is apparent that there exists an element of subjectivity is assigning the weight of evidence. Indeed, a probability can represent the degree of belief of an individual regarding some event (see, e.g., [30, 34, 35, 42, 72]). Nevertheless, the subjectivity of the resulting weights of evidence is not meant to suggest or justify that probability can be arbitrarily assigned, or that it reflects sloppy thinking [42, 84]. Personal probabilities must be coherent (e.g., self-consistent, free of inherent contradictions, do not impose a guaranteed loss [66, 68]) and follow the ordinary axioms of probability.

Updating prior beliefs using the Bayes factor or likelihood ratio allows for obtaining posterior beliefs about the source of a piece of evidence. These posterior beliefs are probabilistic in nature, and do not equate to categorical decisions. However, by using a loss function, these posterior probabilities can be used to reach a conclusion. This process is described in the forensic context by Biedermann, Bozza, and Taroni [11]. Proponents of Bayesian reasoning argue that this path is the only logical and coherent process that exists for making inferences and updating personal beliefs in forensic science. They do, however, go on to argue that, in casework, forensic scientists do not possess the information that would allow them to assign prior beliefs to the considered propositions, and so they should limit themselves to reporting only the Bayes factor, and leave the triers of fact (e.g., jurors, judges) to complete the remainder of the inference process. The challenge for forensic scientists, therefore, lies in assigning Bayes factors to various types of evidence (e.g., fibres, paints, dust, footwear impressions, fingerprints, tool marks).

2.1 TWO FRAMEWORKS FOR ASSIGNING THE WEIGHT OF EVIDENCE

Ommen, Saunders, and Neumann [59] describe two classes of problems in which a forensic scientist may be tasked to assign a Bayes factor: an examiner may operate within the context of the "common source" framework; or he may operate within the context of the "specific source" framework. Each of these frameworks is particular to its set of circumstances, and results in a specific Bayes factor. The next two sections serve to explain these frameworks. See Neumann and Ausdemore [49] for a comparison of these frameworks via a simple simulation scenario.

2.1.1 THE COMMON SOURCE FRAMEWORK

In the common source framework, the examiner evaluates whether it is likely that two pieces of forensic evidence have originated from the same source or from different sources, without formally specifying which sources are considered.

Definition 9 (Common Source Framework) *In the common source framework, the forensic examiner aims to make inference on the source of two sets of trace samples, $\boldsymbol{X}_{u_1}$ and $\boldsymbol{X}_{u_2}$, to determine if they originate from a common, but unknown, source. This context considers the following pair of propositions:*

$\mathcal{H}_{p_{CS}}$: $\boldsymbol{X}_{u_1}$ *and* $\boldsymbol{X}_{u_2}$ *originate from a common, but unknown, source;*

$\mathcal{H}_{d_{CS}}$: $\boldsymbol{X}_{u_1}$ *and* $\boldsymbol{X}_{u_2}$ *originate from two different, but unknown, sources.*

In this scenario, the true source of each piece of evidence is considered to be a random source from a population of potential sources. Under $\mathcal{H}_{p_{CS}}$, the source of the two pieces of evidence is the same random source, while under $\mathcal{H}_{d_{CS}}$, the two pieces of evidence correspond to their own distinct random sources.

As an example of this scenario, Neumann and Ausdemore [49] consider a simple univariate setting to explore some of the properties of this likelihood ratio. In particular, they define a pair of generative models using hierarchical random effects models, such that

$$\begin{aligned} x_{u_1} &= \mu + d_1 + u_1, \text{ where } d_1 \sim \mathcal{N}\left(0, \sigma_d^2\right) \text{ and } u_1 \sim \mathcal{N}\left(0, \sigma_{u_1}^2\right); \\ x_{u_2} &= \mu + d_2 + u_2, \text{ where } d_2 \sim \mathcal{N}\left(0, \sigma_d^2\right) \text{ and } u_2 \sim \mathcal{N}\left(0, \sigma_{u_2}^2\right), \end{aligned} \tag{2.1}$$

where μ is the mean of the population of potential sources, d_1 and d_2 are random effects due to the sources, and u_1 and u_2 are random effects due to objects within sources.

Consider that we have observed two sets of red automotive paint chips at two different crime scenes, and that we are interested in determining if these two sets of red paint chips originated from the same car (which remains unidentified), or if they originated from two different cars in the population of potential cars. In this scenario, μ represents the mean of the distribution of the characteristics of the paint of all cars in the population; d_1 and d_2 represent the deviations between the overall mean of the population, μ, and the characteristics of the unknown first and second sources of red paint; and u_1 and u_2 are random effects that affect the final presentation of the characteristics of the paint chips that result from the two different transfers of paint. The random effects u_1 and u_2 may be distinct since the transfer of the two evidentiary objects x_{u_1} and x_{u_2} may be affected by different sets of factors. For example, consider that two different sections on the car were damaged (for example, the hood of the car, and the rear bumper). It is likely that the paint in these two locations exhibit different characteristics (e.g., there may exist subtle changes in the paint due to varied sun exposure at these two locations on the car).

Under $\mathcal{H}_{p_{CS}}$, the two pieces of evidence originate from the same unknown source, and so $d_1 = d_2$. However, given that the paint chips may have been transferred under varying conditions, it is not necessarily true that $\sigma^2_{u_1} = \sigma^2_{u_2}$. Under $\mathcal{H}_{d_{CS}}$, the two pieces of evidence originate from two different, but unknown, sources, and so $d_1 \neq d_2$. Furthermore, given that x_{u_1} and x_{u_2} did not originate from the same source, they are independent. Given the above generative models, the joint distributions of x_{u_1} and x_{u_2} under $\mathcal{H}_{p_{CS}}$ and $\mathcal{H}_{d_{CS}}$ are distributed according to a Multivariate Normal distribution, and are respectively given by

$$\begin{aligned} \begin{pmatrix} x_{u_1} \\ x_{u_2} \end{pmatrix} \Bigg| \mathcal{H}_{p_{CS}} \sim \quad & \mathcal{MVN}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_d^2 + \sigma_{u_1}^2 & \sigma_d^2 \\ \sigma_d^2 & \sigma_d^2 + \sigma_{u_2}^2 \end{pmatrix} \right); \\ \begin{pmatrix} x_{u_1} \\ x_{u_2} \end{pmatrix} \Bigg| \mathcal{H}_{d_{CS}} \sim \quad & \mathcal{MVN}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_d^2 + \sigma_{u_1}^2 & 0 \\ 0 & \sigma_d^2 + \sigma_{u_2}^2 \end{pmatrix} \right). \end{aligned} \tag{2.2}$$

Taking the view that forensic evidence must be evaluated within a Bayesian paradigm, the weight of evidence should be quantified using a Bayes factor (or, when parameters are known, using a likelihood ratio). The common source likelihood ratio takes the form

$$\mathcal{LR}_{CS} = \frac{f\left(x_{u_1}, x_{u_2} | \mathbf{\Omega}_0, \mathcal{H}_{p_{CS}}\right)}{f\left(x_{u_1}, x_{u_2} | \mathbf{\Omega}_0, \mathcal{H}_{d_{CS}}\right)} = \frac{f\left(x_{u_1}, x_{u_2} | \mathbf{\Omega}_0, \mathcal{H}_{p_{CS}}\right)}{f\left(x_{u_1} | \mathbf{\Omega}_0, \mathcal{H}_{d_{CS}}\right) f\left(x_{u_2} | \mathcal{H}_{d_{CS}}\right)}. \tag{2.3}$$

2.1.2 THE SPECIFIC SOURCE FRAMEWORK

In the specific source framework, the examiner evaluates whether it is likely that a piece of forensic evidence has originated from a particular known source.

Definition 10 (Specific Source Framework) *In the specific source framework, the forensic examiner aims to make inference on the source of a set of trace samples $\boldsymbol{X}_u$ to determine if it originates from the specified known source that produced a set of control material $\boldsymbol{X}_s$. This context considers the following pair of propositions:*

$\mathcal{H}_{p_{SS}}$: $\boldsymbol{X}_u$ *originates from the known source that produced* $\boldsymbol{X}_s$;

$\mathcal{H}_{d_{SS}}$: $\boldsymbol{X}_u$ *originates from an unknown source, different from the source that produced* $\boldsymbol{X}_s$.

In this scenario, the source that produced the control material is identified and so can be considered fixed. Under $\mathcal{H}_{p_{SS}}$, the source that produced the evidence is considered to be the source that produced the control material, while under $\mathcal{H}_{d_{SS}}$, the source that produced x_u is unknown, and is considered to be a random source from a population of potential sources, different from the undisputed source that produced the control material.

As in the common source scenario, Neumann and Ausdemore [49] consider a simple univariate setting to explore some of the properties of this likelihood ratio. However, in the specific source scenario, the generative models differ depending on whether $\mathcal{H}_{p_{SS}}$ or $\mathcal{H}_{d_{SS}}$ is considered. Under $\mathcal{H}_{p_{SS}}$, when the evidence originates from the same source as the considered control material, the generative models take the form of two simple random effects models, such that

$$\begin{aligned} x_u = &\quad \mu_d + u, \text{ where } u \sim \mathcal{N}\left(0, \sigma_u^2\right); \\ x_s = &\quad \mu_d + s, \text{ where } s \sim \mathcal{N}\left(0, \sigma_s^2\right), \end{aligned} \tag{2.4}$$

where μ_d represents the mean for the considered specific source, and u and s are random effects corresponding to the trace and control samples.

Under $\mathcal{H}_{d_{SS}}$, the generative model for the control material is the same as under $\mathcal{H}_{p_{SS}}$ (indeed, there is no dispute that x_s originates from the considered known source!); however, the generative model for the trace material is now given by a hierarchical random effects model to reflect that the true source of x_u is unknown in a population of potential sources. Thus, under $\mathcal{H}_{d_{SS}}$, the generative models are given by

$$\begin{aligned} x_u &= \mu + d + u \text{ where } d \sim \mathcal{N}\left(0, \sigma_d^2\right), \text{ and } u \sim \mathcal{N}\left(0, \sigma_u^2\right); \\ x_s &= \mu_d + s, \text{ where } s \sim \mathcal{N}\left(0, \sigma_s^2\right), \end{aligned} \tag{2.5}$$

where μ is the mean of the population of potential sources, d is a random effect due to the unknown source, and s and u are random effects corresponding to the trace and control samples.

Consider that we have now observed a single set of red automotive paint chips. In this scenario, we are interested in determining if this set of paint chips originated from a particular car, from which we have observed a control sample. In this scenario, μ_d represents the mean of the distribution of the characteristics of the considered car; μ represents the mean of the distribution of the characteristics of all cars in the population; d represents the deviation between the overall mean of the population, μ, and the characteristics of the unknown car; and u and s are random effects that affect the final presentation of the characteristics of the paint chips that result from the two different transfers of paint.

Under $\mathcal{H}_{p_{SS}}$, the trace and control materials are independent, given μ_d. Under $\mathcal{H}_{d_{SS}}$, the trace and control materials are independent since they are not from the same source. Given the above generative models for the specific source scenario, the joint distributions of x_u and x_s under $\mathcal{H}_{p_{SS}}$ and $\mathcal{H}_{d_{SS}}$ are distributed according to a Multivariate

Normal distribution, and are respectively given by

$$\begin{pmatrix} x_u \\ x_s \end{pmatrix} \Big| \mathcal{H}_{p_{SS}} \sim \mathcal{MVN}\left(\begin{pmatrix} \mu_d \\ \mu_d \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_s^2 \end{pmatrix}\right); \\ \begin{pmatrix} x_u \\ x_s \end{pmatrix} \Big| \mathcal{H}_{d_{SS}} \sim \mathcal{MVN}\left(\begin{pmatrix} \mu \\ \mu_d \end{pmatrix}, \begin{pmatrix} \sigma_d^2 + \sigma_u^2 & 0 \\ 0 & \sigma_s^2 \end{pmatrix}\right). \tag{2.6}$$

As in the common source framework, the weight of evidence should be quantified using a Bayes factor (or, when the parameters are known, using a likelihood ratio). The specific source likelihood ratio takes the form

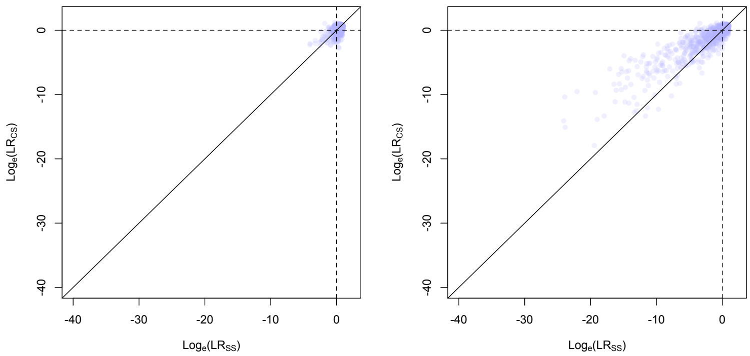
$$\mathcal{LR}_{SS} = \frac{f\left(x_u, x_s|\boldsymbol{\Omega}_0, \mathcal{H}_{p_{SS}}\right)}{f\left(x_u, x_s|\boldsymbol{\Omega}_0, \mathcal{H}_{d_{SS}}\right)} = \frac{f\left(x_u|\boldsymbol{\Omega}_0, \mathcal{H}_{p_{SS}}\right)}{f\left(x_u|\boldsymbol{\Omega}_0, \mathcal{H}_{d_{SS}}\right)}. \tag{2.7}$$

2.1.3 COMPARING THE TWO FRAMEWORKS

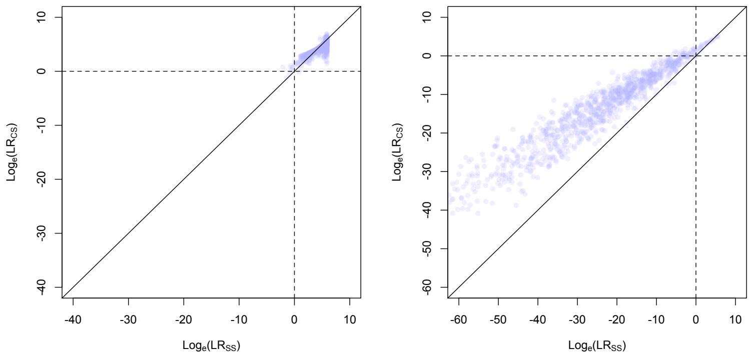
It is important to note that the Bayes factors that result from the common source and specific source frameworks in (2.3) and (2.7) are indeed different, even when the same information is considered, and thus lead to different interpretations of the results of forensic examinations.

Neumann and Ausdemore [49] show that the likelihood ratios in (2.3) and (2.7) converge only when the variance of the source of either x_{u_1} or x_{u_2} in (2.3) and the variance of the suspect in (2.7) are negligible; otherwise, substituting one framework for the other is inappropriate. This assumption is satisfied in very few scenarios. Figure 1 in [49] is reproduced as Figure 2.1 below. These experiments demonstrate that the likelihood ratios of the common and specific source scenarios are undeniably different!

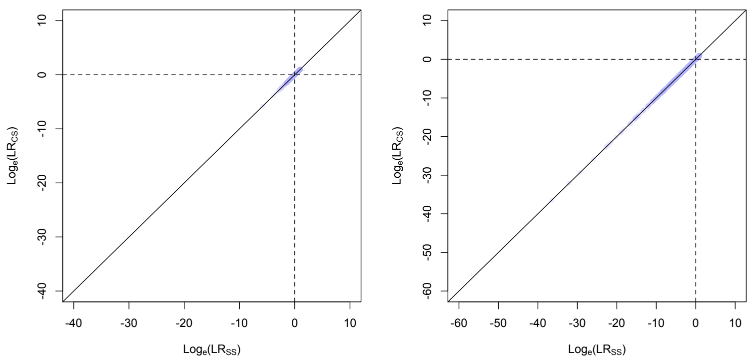
Regardless of the context in which the examiner is operating, it is critical that the resulting Bayes factor be relevant to the scenario at hand, and be appropriately defined for the evidence being considered. In the vast majority of cases, the interests of the criminal justice system fall under the span of the specific source scenario, and the aim of this thesis is to develop a probabilistic model for this context. This is not to say that the common source scenario is not relevant. Indeed, determining whether two pieces of evidence were made by the same unknown source may be pertinent to intelligence investigations which aim to connect multiple crime scenes. Armstrong [5] develops a useful kernel-based model for making inference on the source of two sets of trace objects under the common source scenario.



(a)



(b)



(c)

Figure 2.1: Convergence of the common source likelihood ratio to the specific source likelihood ratio. Columns: the left column reports the results when x_u and x_s have been sampled under $\mathcal{H}_{p_{SS}}$; the right column reports the results when x_u and x_s have been sampled under $\mathcal{H}_{d_{SS}}$. Rows: (a) the source of the control object is common in the population and has some variance; (b) the source of the control object is rare in the population and has some variance; (c) the source of the control objects is common and has virtually no variance.

2.2 A SCORE-BASED APPROACH

While forensic scientists have been assigning Bayes factors to simple forms of forensic evidence (e.g., single DNA profiles) for many years, only anecdotal attempts have been made to assign Bayes factors to complex forms of forensic evidence (e.g., handwriting evidence, fingerprint evidence). Assigning Bayes factors to complex forms of evidence requires defining reasonable and appropriate likelihood functions to represent the joint distributions of heterogeneous and high-dimensional feature vectors. Unfortunately, it is often impossible to assign probability measures for these sorts of data, and so assigning Bayes factors or performing any other likelihood-based inference is impossible. Consequently, the forensic scientists reporting on these types of evidence are left without means to support their assessments of the probative value of the evidence.

In an attempt to bypass the need to work with intractable likelihood functions, researchers have concentrated on the use of scores to reduce the complexity and dimensionality of the problem. A score can have two interpretations: it can be interpreted as a summary statistic resulting from the comparison of two objects; or it can be interpreted as the scalar projection that results from the inner product of two vectors. In the first case, we talk about (dis)similarity functions, $\delta(\cdot, \cdot)$ with real-valued output, while in the second case, we talk about kernel functions, $\kappa(\cdot, \cdot)$, as described above in Section 1.1. The primary difference between δ and κ is that δ can be any function with real-valued output, while κ must be a positive semi-definite, continuous, symmetric function (see Theorem 1 and Definition 2 in Section 1.1). Both functions map complex random vectors from their natural space to the real line, $\mathbb{R}$, and both offer great flexibility to researchers. These scores allow researchers to measure the "distance" between two objects, such that the value representing the distance is minimized (or maximized) when two objects originate from the same source, and is maximized (or minimized) when they originate from different sources, and to express the level of (dis)similarity between pairs of objects as a univariate continuous random variable, whose probability distribution is significantly more convenient to model than the distribution of the original vectors of observations.

2.2.1 SCORE-BASED LIKELIHOOD RATIOS

Initial attempts to circumvent the complexity of the original data gave rise to a family of ad-hoc methods, dubbed “score-based likelihood ratios”. The first members of this family were presented in the late 1990’s and early 2000’s in the fields of speaker recognition and fingerprint evidence (see, e.g., [20, 21, 27–29, 46, 47, 51, 52]). Different constructions of score-based likelihood ratios have been proposed over the years, and their use in case work has been advocated for, especially in Europe [22]. These models first compare two objects to obtain a score using some (dis)similarity metric. The likelihood of this score is then evaluated using a pair of density functions based on the sampling distributions of the score under two mutually exclusive propositions. The ratio of these densities is then reported as the “score-based likelihood ratio”.

Figure 2.2 shows two sampling distributions for a pair of propositions formulated under the specific source scenario. Given a (dis)similarity score, $\delta\left(x_u, x_s\right) = 3$, the density of this score is evaluated under the sampling distribution defined for $\mathcal{H}_p$ (solid black line) and under the sampling distribution defined for $\mathcal{H}_d$ (dashed black line). In this scenario, the resulting score-based likelihood ratio would support $\mathcal{H}_p$, since the density of $\delta\left(x_u, x_s\right) = 3$ is greater for the sampling distribution defined under $\mathcal{H}_p$ than under the sampling distribution defined under $\mathcal{H}_d$.

Some commonly described score-based likelihood ratios are the common source score-based likelihood ratio, the asymmetric specific source score-based likelihood ratio, the trace-anchored specific source score-based likelihood ratio, and the suspect-anchored specific source score-based likelihood ratio. Unfortunately, each of these proposed methods suffers from limitations, and so is inappropriate for reporting the weight of evidence: the common source score-based likelihood ratio does not account for the rarity of the trace objects; the asymmetric specific source score-based likelihood ratio does not consider the same evidence in the numerator and denominator; the trace-anchored specific source score-based likelihood ratio reduces to the actual likelihood (and so is not so score-based afterall); and the suspect-anchored score-based likelihood ratio is not coherent[2]. Neumann and Ausdemore [49] present these score-based likelihoods and their limitations in more detail. Furthermore, they show that convergence of the score-based likelihood ratios to the "true" likelihood ratio only occurs under very specific conditions (e.g., when the variance of the considered source in the specific source scenario is negligible.)

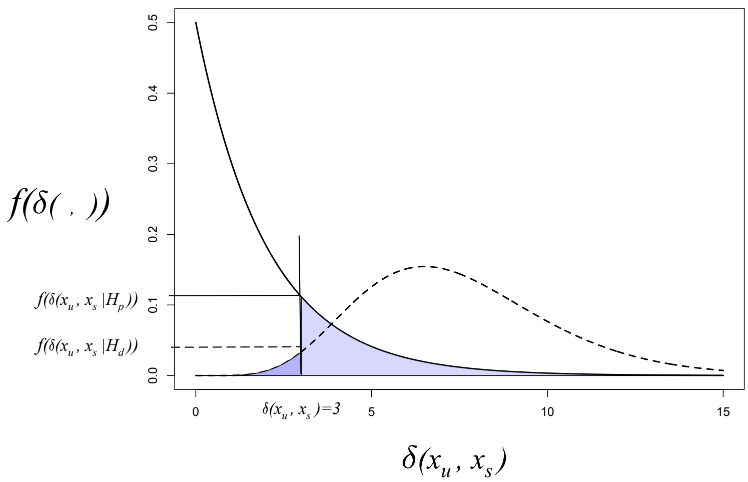


Figure 2.2: Score-based likelihood ratio obtained by calculating the ratio of the densities of a summary statistic for an observed pair of trace and control objects, $\delta(x_u, x_s)$, in the sampling distributions defined under $\mathcal{H}_p$ and $\mathcal{H}_d$.

2.2.2 OTHER SCORE-BASED METHODS

Using kernels to express the similarity of pairs of objects and to reduce the dimension of the data is not unique to score-based likelihood ratios. Other nonparametric methods that use these scores for model selection have been proposed [33], and include approximate Bayesian computation (ABC) methods [18, 67], SVMs [89, 90] (introduced in Chapter 1), and relevance vector machines (RVMs) [64, 85]. Unfortunately, applying these models to the types of data encountered in forensic science, and, more

[2]The inverse of the Bayes factor corresponds to the amount of support for the alternative proposition. This is not the case for specific source score-based likelihood ratios.

generally, in pattern recognition, is not straightforward. ABC methods require assuming generative models and using sufficient summary statistics, and ideally rely on an infinite number of resampling simulations. In practice, these assumptions are not realistic: generative models for high-dimensional, heterogenous data may not exist, or may be too complex to implement; it is not possible to define what "sufficiency" means for these types of data; and the number of simulations is necessarily limited. In addition, most ABC algorithms require defining arbitrary thresholds that allow for assessing whether resampled datasets are sufficiently similar to the observed dataset. However, as the dimension of the data increases, the curse of dimensionality may imply that no observations are "similar enough" to the original data. RVM and SVM algorithms require complex and computationally expensive optimization procedures (as seen in Chapter 1), in part due to the repeated inversion of the Gram matrix [12, 33], and do not have a natural extension to the multi-class scenario. Finally, SVMs do not allow for likelihood-based inference [33, 89, 90].

2.3 A NEW CLASS OF SCORE-BASED MODELS

In the last decade, a new class of algorithms has been introduced. While this class of algorithms is score-based in nature due to its reliance on kernel functions, it maintains the ability to make appropriate likelihood based inference. We wish to capitalize on these properties to calculate Bayes factors for complex data, particularly when few data are observed.

In 2014, Gantz and Saunders [26] developed a method that allows for assigning a parametric model to the joint distribution of pairwise scores between multiple objects that are known to originate from a single source. In practice, this algorithm can be used to make inference on whether a new (set of) object(s) originates from the same source as a set of control objects, as is the case in one-class classification or anomaly detection algorithms [1, 7].

Although this proposed model cannot be used directly for model selection, the authors showed that this new class of algorithms is particularly suited for high-dimensional hypothesis testing (Armstrong et al. [6] demonstrated the usefulness of this model). In addition, this approach allows for a limited number of assumptions, each of which are usually satisfied in high-dimensional situations, and relies on a minimal number of parameters, which may be estimated with closed-form solutions. Finally, the approach allows for simultaneously classifying a set of objects. While the aforementioned algorithms use pairwise comparisons between multiple objects from each considered source to train or learn the parameters of the algorithms (as in SVMs or RVMs), they consider objects to be classified as being independent from one another. The class of algorithms introduced by Gantz and Saunders [26] and implemented by Armstrong et al. [6] makes it possible to perform model selection on a joint set of observations[3].

2.3.1 A COMMON SOURCE MODEL-SELECTION ALGORITHM

In 2018, Armstrong [5] expanded this class of algorithms by introducing a method that enables model selection in the common source scenario for stationary kernels. In particular, this model is used to determine whether two sets of trace objects are simple random samples from a common, but unknown, source in a population of potential sources (see Definition 9 for a formal statement of the propositions considered in the common source framework). The model enables a fully Bayesian treatment of non-nested model selection, and, in particular, can be used to quantify the weight of high-dimensional and heterogeneous forms of forensic evidence (e.g., chemical spectra, pattern evidence). Specifically, given two sets of trace objects, $\boldsymbol{x}_{u_1}$ and $\boldsymbol{x}_{u_2}$, this method allows for selecting between the models

$\mathcal{H}_p$: . $\boldsymbol{x}_{u_1}$ and $\boldsymbol{x}_{u_2}$ originate from a common but unknown source in the population of potential sources;

$\mathcal{H}_d$: $\boldsymbol{x}_{u_1}$ and $\boldsymbol{x}_{u_2}$ originate from two different but unknown sources in the population of potential sources.

[3] Computational developments initially explored in [53].

Proceeding in the manner of Gantz and Saunders [26], Armstrong [5] assigns a parametric model to the joint distribution of pairwise scores, $\boldsymbol{s}$, between multiple objects that are known to originate from a single source. More specifically, an observation $s_{ij,i'j'}$ of $\boldsymbol{s}$ is given by $s_{ij,i'j'} = \kappa\left(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}\right)$, where $\kappa\left(\cdot,\cdot\right)$ is a kernel function as defined in Definition 2 in Section 1.1, and

$$s_{ij,i'j'} = \begin{cases} \theta_b + b_i + b_i' + d_{ii'} + t_{i:ij} + t_{i:i'j'} + t_{i':ij} + t_{i':i'j} + w_{ij} + w_{i'j'} + e_{iji'k'}, & i \neq i' \\ \theta_w + w_{ij} + w_{i'j'} + e_{iji'j'}, & i = i', \end{cases}$$

such that $\boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \Sigma\right)$, where $\Sigma := \boldsymbol{BB}'\sigma_b^2 + \boldsymbol{DD}'\sigma_d^2 + \boldsymbol{TT}'\sigma_t^2 + \boldsymbol{WW}'\sigma_b^2 + \boldsymbol{I}\sigma_e^2$. The design matrices given by $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{T}$ and $\boldsymbol{W}$ capture distinct random effects that exist within the score model. The rows of each of the design matrices consist of all $\binom{nn_0}{2}$ possible pairwise comparisons of ij and $i'j'$, where n is the number of random sources considered in the population, and n_0 is the number of objects observed per source. The columns of $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{T}$ and $\boldsymbol{W}$, however, vary according to the effect being described: the design matrix $\boldsymbol{B}$ describes the between-source effects $(b_i, b_{i'})$; the design matrix $\boldsymbol{D}$ describes the source interaction that exists between-source comparisons, $(d_{ii'})$; the design matrix $\boldsymbol{T}$ describes the interaction that exists between the sources and their samples $(t_{i:ij}, t_{i:i'j'}, t_{i':ij}, t_{i':i'j'})$; the design matrix $\boldsymbol{W}$ describes the within-source effects $(w_{ij}, w_{i'j'})$. Armstrong [5] provides a more in-depth look at these matrices.

To select between the two models presented above, Armstrong [5] proposes to evaluate a Kernel Bayes Factor, $\mathcal{K}$.

Definition 11 (Kernel Bayes Factor) *The Kernel Bayes Factor allows for selecting between two mutually exclusive models based on a set of scores obtained by a kernel function, and is given by*

$$\mathcal{K}_{CS} = \frac{\int_{\boldsymbol{\Omega}} f\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_p\right) d\Pi\left(\boldsymbol{\Omega}|\mathcal{H}_p\right)}{\int_{\boldsymbol{\Omega}} f\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_d\right) d\Pi\left(\boldsymbol{\Omega}|\mathcal{H}_d\right)} \tag{2.8}$$

where $\boldsymbol{\Omega}$ is the set of parameters characterizing the likelihood functions of $\boldsymbol{s}$ under $\mathcal{H}_p$ and $\mathcal{H}_d$, and Π is a probability measure over the parameter spaces of $\boldsymbol{\Omega}$.

Evaluating (2.8) then requires studying $\boldsymbol{\Omega} := \{\theta_b, \theta_w, \sigma_b^2, \sigma_d^2, \sigma_t^2, \sigma_w^2, \sigma_e^2\}$ under the two models, $\mathcal{H}_p$, and $\mathcal{H}_d$. However, following Ommen, Saunders, and Neumann [59], Armstrong [5] shows that, in the common source scenario and for a perfectly balanced sample, this Bayes factor can be evaluated using (2.9), which requires estimating the parameters only under $\mathcal{H}_d$,

$$\mathcal{K}_{CS} = \int_{\boldsymbol{\Omega}} \frac{f_p\left(\boldsymbol{s}_m|\boldsymbol{s}_n, \boldsymbol{\Omega}\right)}{f_d\left(\boldsymbol{s}_m|\boldsymbol{s}_n, \boldsymbol{\Omega}\right)} d\Pi\left(\boldsymbol{\Omega}|\boldsymbol{s}, \mathcal{H}_d\right), \tag{2.9}$$

where $\boldsymbol{s}_m$ is vector of scores that considers comparisons that include at least one trace object and $\boldsymbol{s}_n$ is the vector of scores that consider comparisons between control objects.

Armstrong [5] shows that the likelihood can be decomposed into independent sums of squares. He shows that closed-form solutions exist for the parameters of the model, and shows that the distributions of the parameters can be studied by defining a Markov Chain Monte Carlo (MCMC) sampler. Using this technique, posterior samples of the parameters are obtained by sampling from $\pi\left(\boldsymbol{\Omega}|\boldsymbol{s}, \mathcal{H}_d\right)$. This sampling technique is described in detail in [5, Section 5.6]. More recently, Ausdemore et al. [7] made the process more computationally efficient.

By developing this common source model, Armstrong [5] bolstered the work of Gantz and Saunders [26] to allow for selecting between models via a kernel Bayes factor, and laid the necessary foundation for developing the specific source model. In particular, this development will be particularly useful in developing the final algorithm in the class of algorithms constructed throughout this dissertation.

2.3.2 A TWO-STAGE APPROACH

In 2020, Ausdemore et al. [7] implemented the work of Gantz and Saunders [26] using a two-stage framework. This model is the first step towards defining a specific-source model selection algorithm and lays the groundwork for developing the class of algorithms discussed in this dissertation in that it allows for obtaining a proxy of the Bayes factor in the specific source scenario. Given a set of M trace objects, $\boldsymbol{x}_u$, assumed to originate from a single, unknown source, a set of N control objects $\boldsymbol{x}_s$, originating from a specific source of interest, $\mathcal{S}$, and a sample of sources from a relevant population of potential sources, $\mathcal{A}$, they are interested in making inference on the source of the set of trace objects. Formally, they consider the following mutually exclusive propositions:

$\mathcal{H}_p$: $\boldsymbol{x}_u$ is a simple random sample from $\mathcal{S}$;

$\mathcal{H}_d$: $\boldsymbol{x}_u$ is a simple random sample from another source in the population represented by $\mathcal{A}$.

To make inference on the source of $\boldsymbol{x}_u$, they revisit the two-stage approach, formally introduced by Parker [60, 61], and Parker and Holford [62]. This two-stage inference framework consists of a similarity stage and a discrimination stage. In the similarity stage, the goal is to compare the characteristics of the trace and control objects, and determine if they are indistinguishable. In the discrimination stage, the goal is to determine the rarity of the characteristics observed in the first stage in the population of potential sources (i.e., determine the match probability).

To address the similarity stage, Ausdemore et al. [7] develop a generic $\alpha-$level test that relies on a kernel function to measure the level of similarity between pairs of high-dimensional, heterogeneous, and complex data. Given two vectors of measurements, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, representing the observations made on two objects i, j, sampled from the same random source, a kernel function κ, is used to measure their level of similarity, and report it as a score, s_{ij}. In the spirit of Gantz and Saunders [26] and Armstrong et al. [6], Ausdemore et al. [7] choose to represent this score as a linear random effects model, given by

$$s_{ij} = \kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \theta + a_i + a_j + \varepsilon_{ij},$$

where θ is the expected value of the score between any two objects from the same considered source, $a_i, a_j \overset{iid}{\sim} N(0, \sigma_a^2)$ are random effects representing the contributions of the i^{th} and j^{th} objects, and $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$ is a lack of fit term. Given M trace objects and N control objects from a source of interest, they define the vector of scores $\boldsymbol{s}_{m+n} = (\boldsymbol{s}_m, \boldsymbol{s}_n)'$, where $\boldsymbol{s}_m$ and $\boldsymbol{s}_n$ are defined as in Section 2.3.1. Given that all N control objects originate from a single source, the multivariate representation and distribution of $\boldsymbol{s}_n$ are given by

$$\boldsymbol{s}_n = \theta \mathbf{1}_n + \boldsymbol{P}\boldsymbol{a} + \boldsymbol{\varepsilon}, \quad \boldsymbol{s}_n \sim \mathcal{MVN}\left(\theta \mathbf{1}_n, \boldsymbol{\Sigma}_{n\times n}\right),$$

where $\mathbf{1}_n$ is a one vector of length n, $\boldsymbol{P}$ is an $n \times N$ design matrix (where each row represents an i, j combination consisting of ones in the i^{th} and j^{th} columns and zeros elsewhere), $\boldsymbol{a}$ is the vector of random effects for the considered objects, $\boldsymbol{\varepsilon}$ is the vector of ε_{ij} corresponding to each pair of objects, and $\boldsymbol{\Sigma}_{n\times n} = \boldsymbol{P}\boldsymbol{P}'\sigma_a^2. + \boldsymbol{I}_n\sigma_e^2.$ Using this definition and distributional assumption, they define they hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ as:

$\mathcal{H}_0$: The M trace objects have the same characteristics as the N control objects, and so

$$\begin{aligned}\begin{pmatrix}\boldsymbol{s}_m\\ \boldsymbol{s}_n\end{pmatrix}|\boldsymbol{\Psi} &\sim \mathcal{MVN}\left(\theta \mathbf{1}_{m+n}, \boldsymbol{Q}\boldsymbol{Q}'\sigma_a^2 + \boldsymbol{I}_{m+n}\sigma_e^2\right).\\ &= \mathcal{MVN}\left(\theta \mathbf{1}_{m+n}, \begin{bmatrix}\boldsymbol{\Sigma}_{m\times m} & \boldsymbol{\Sigma}_{m\times n}\\ \boldsymbol{\Sigma}_{n\times m} & \boldsymbol{\Sigma}_{n\times n}\end{bmatrix}\right)\end{aligned}$$

where $\boldsymbol{\Psi} = \{\theta, \sigma_a^2, \sigma_e^2\}$, and $\boldsymbol{Q}$ is a design matrix of the same construction as $\boldsymbol{P}$, but with dimensions corresponding to the length of the vector $(\boldsymbol{s}_m, \boldsymbol{s}_n)'$.

$\mathcal{H}_1$: The M trace objects have different characteristics than the N control objects, and so $\boldsymbol{s}_m$ has some other distribution.

To design their test statistic, Ausdemore et al. [7] begin by considering the conditional likelihood of the vector of scores involving at least one trace object, given the vector of scores involving only control objects, $\mathscr{L}\left(\boldsymbol{s}_m|\boldsymbol{s}_n, \boldsymbol{\Psi}\right)$. The test statistic is then defined by

$$T\left(\boldsymbol{s}_m, \boldsymbol{s}_n, \boldsymbol{\Psi}\right) = Pr\left(\mathscr{L}\left(\boldsymbol{s}_m|\boldsymbol{s}_n, \boldsymbol{\Psi}\right) \geq \mathscr{L}\left(\boldsymbol{s}_m^*|\boldsymbol{s}_n, \boldsymbol{\Psi}\right)\right).$$

where $\boldsymbol{s}_m^*$ is a random vector of scores calculated between pairs of objects involving at least one trace object when the trace and control objects *truly* originate from the source of interest. Using this test statistic, they decide to reject H_0 at a specific

$\alpha-$level if $\int T\left(\boldsymbol{s}_m, \boldsymbol{s}_n, \boldsymbol{\Psi}\right) d\pi\left(\boldsymbol{\Psi}|\boldsymbol{s}_n\right) \leq c(\alpha)$, in which the uncertainty associated with the model parameters has been integrated out, and $c(\alpha)$ is a constant chosen to satisfy $Pr\left(T\left(\boldsymbol{s}_m, \boldsymbol{s}_n, \boldsymbol{\Psi}\right) \leq c(\alpha)\right) \leq \alpha$. The processes for estimating $\int T\left(\boldsymbol{s}_m, \boldsymbol{s}_n, \boldsymbol{\Psi}\right) d\pi\left(\boldsymbol{\Psi}|\boldsymbol{s}_n\right)$ and for obtaining $c\left(\alpha\right)$ are found in more detail in [7].

By applying their model to a set of Fourier-Transform Infrared spectroscopy (FTIR) spectra of paint chips from cans of common household paint, Ausdemore et al. [7] demonstrate that their proposed approach works well with the number of samples commonly encountered in forensic science.

2.3.3 A SPECIFIC SOURCE MODEL SELECTION ALGORITHM

A specific source model selection algorithm is missing from the class of models outlined above. This dissertation introduces a model that serves to complete this class of kernel-based algorithms (initiated under NIJ Awards 2009-DN-BX-K234, which addressed the outlier detection problem [26], and 2015-R2-CX-0028, which addressed the common source problem [5]), by proposing a fully probabilistic model for addressing the specific source problem in forensic science.

To define the model necessary for assigning a Bayes factor, we proceed by designing a series of three progressive models: first, we develop the model for a fixed pair of sources; we then extend the two-class model to consider a set (with cardinality greater than 2) of fixed sources; finally, we design a kernel-based model selection algorithm that considers a single suspected source against a population of multiple random sources and that allows for assigning the Bayes factor in (2.8). The remainder of this dissertation will use the concepts introduced in Chapters 1 and 2 to develop the set of models described in Section 2.3.

Part II

A Two-Class Model Selection Algorithm for High-Dimensional and Complex Data

OVERVIEW OF PART II: A TWO-CLASS MODEL SELECTION ALGORITHM FOR HIGH-DIMENSIONAL AND COMPLEX DATA

In this part, we develop the theory and implementation for a two-class kernel-based model-selection algorithm. In Chapter 3, we define the problem and develop the algorithm that allows for determining which of two classes is more likely to have produced a set of trace objects. In addition, we propose a method for studying the parameters of the proposed model, and a sampling algorithm that can be used to study the distributions of the considered parameters.

In Chapter 4, we implement the proposed model on the MNIST hand-written digits data that is commonly used to evaluate the performance of pattern recognition algorithms.

In Chapter 5 we discuss the benefits and limitations of the proposed two-class model.

Chapter 3

DEFINING THE TWO-CLASS MODEL-SELECTION PROBLEM

We begin by considering a two-class scenario, in which we are interested in defining which of two specific, fixed sources is more likely to have produced a set of objects of unknown but common origin. That is, given two sets of objects, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, where each set of objects is known to have originated from one of two distinct sources, and a set of objects, $\boldsymbol{X}_u$, of unknown origin but known to have originated from one of the two sources that produced the objects observed in $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, we are interested in determining which of the two sources is most likely to have generated the set of objects observed in $\boldsymbol{X}_u$. Formally, we are interested in determining if

$\mathcal{H}_1$: $\boldsymbol{X}_u$ is a simple random sample from the source that produced $\boldsymbol{X}_1$;

$\mathcal{H}_2$: $\boldsymbol{X}_u$ is a simple random sample from the source that produced $\boldsymbol{X}_2$.

Oftentimes, differentiating between these propositions can be reduced to a simple classification or model-selection problem that can be addressed using machine learning or likelihood-based techniques. However, when we are faced with high-dimensional, complex, heterogenous data and limited sample sizes, the process is not so straightforward: small sample sizes rule out many machine learning techniques, and high-dimensional, complex, or heterogenous data make it impossible to assign the necessary probability measures for assigning Bayes factors or performing any other likelihood-based inference.

We propose a model that leverages the properties of kernel functions (see Chapter 1) to obtain a vector of scores, $\boldsymbol{s}$, that characterizes pairwise comparisons of all objects observed in $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_u$. This vector consists of within-source scores, which arise when compared objects originate from a common source, and between-source scores, which arise when compared objects originate from different sources. The model capitalizes on the variability that exists within and between these sets of scores to address the above inference question. Because the method relies on a kernel function, the method can be tailored to any type of data by merely modifying this function, and the remaining inference process remains the same. Furthermore, the model makes only one assumption, which can be satisfied through the design of the kernel function.

3.1 PROBLEM STATEMENT

Consider two sets of exchangeable observations, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, made on two distinguishable sets of objects, and the set of exchangeable observations, $\boldsymbol{X}_u$, made on objects of common but unknown origin. The sets $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are considered to be sets of control objects, while the set $\boldsymbol{X}_u$ is considered to be a set of test objects. We define the sets $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_u$ as being simple random samples,

$$\boldsymbol{X}_1 := \left\{\boldsymbol{x}_{1,1},\ \boldsymbol{x}_{1,2},\ \ldots,\ \boldsymbol{x}_{1,n_0}\right\},$$
$$\boldsymbol{X}_2 := \left\{\boldsymbol{x}_{2,1},\ \boldsymbol{x}_{2,2},\ \ldots,\ \boldsymbol{x}_{2,n_0}\right\},$$
$$\boldsymbol{X}_u := \left\{\boldsymbol{x}_{u,1},\ \boldsymbol{x}_{u,2},\ \ldots,\ \boldsymbol{x}_{u,n_u}\right\},$$

where the sets of control objects consist in n_0 objects from their respective sources, and the set of test objects consists in n_u objects known to originate from one of the two sources represented by the observations in $\boldsymbol{X}_1$ or $\boldsymbol{X}_2$. We are interested in quantifying the extent of support provided to $\mathcal{H}_1$ and $\mathcal{H}_2$ above.

Rather than consider the observations themselves, we instead consider the vector of all pairwise *scores*, $\boldsymbol{s} \in \mathbb{R}^N$, $N = \binom{2n_0+n_u}{2}$, obtained by comparing the m-dimensional observations in the sets $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_u$ via some kernel,

$$\kappa : \mathbb{R}^m \mapsto \mathbb{R},$$
$$s_{ij,i'j'} := \kappa\left(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}\right) = \left\langle \phi\left(\boldsymbol{x}_{ij}\right), \phi\left(\boldsymbol{x}_{i'j'}\right)\right\rangle \qquad i, i' \in \{1, 2, u\},\ j, j' \in \{1,\ \ldots,\ \max\{n_0, n_u\}\},$$

where ϕ is a mapping into some separable, high-dimensional Hilbert space [12, 65, 71, 74]. That is, $s_{ij,i'j'}$ is the score obtained by comparing object $\boldsymbol{x}_{ij}$ to object $\boldsymbol{x}_{i'j'}$ using some kernel function, κ, as defined in Definition 2 in Section 1.1.

We define our kernel such that our vector of scores is distributed according to a Multivariate Normal distribution, with

$$\boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right), \tag{3.1}$$

where $\boldsymbol{\theta}$ is the vector of mean terms, and $\boldsymbol{\Sigma}$ is the covariance matrix associated with the vector of scores (see Section 9.5 and Chapter 10 for a discussion on the validity of this assumption, and the implications when this assumption does not hold). These parameters will collectively be referred to as $\boldsymbol{\Omega} := \{\boldsymbol{\theta}, \boldsymbol{\Sigma}\}$, and will be more explicitly defined as we move through the chapter. We can define the Bayes factor in terms of the Multivariate Normal Likelihood and the associated parameter $\boldsymbol{\Omega}$, such that

$$\begin{aligned} \Lambda &= \frac{\int_{\boldsymbol{\Omega}} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_1\right) d\Pi\left(\boldsymbol{\Omega}|\mathcal{H}_1\right)}{\int_{\boldsymbol{\Omega}} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_2\right) d\Pi\left(\boldsymbol{\Omega}|\mathcal{H}_2\right)} \\ &:= \frac{\int_{\boldsymbol{\Omega}_1} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}_1\right) d\Pi\left(\boldsymbol{\Omega}_1\right)}{\int_{\boldsymbol{\Omega}_2} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}_2\right) d\Pi\left(\boldsymbol{\Omega}_2\right)}. \end{aligned} \tag{3.2}$$

It is worth noting that there exist differences between $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$. While the individual elements of each of the parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ are restricted to the same set of potential values, the structures of the mean vectors and covariance matrices depend on whether $\mathcal{H}_1$ or $\mathcal{H}_2$ is being considered: if $\mathcal{H}_1$ is being considered, then the parameter $\boldsymbol{\Omega}_1$ considers that the set of unknown object $\boldsymbol{X}_u$ originates from the source that produced $\boldsymbol{X}_1$, and so scores of the form $s_{uj,1j'}$, $s_{uj,uj'}$ will all be considered as "within-source 1" scores, and scores of the form $s_{uj,2j'}$ will be considered as "between-source" scores; likewise, if $\mathcal{H}_2$ is being considered, then the parameter $\boldsymbol{\Omega}_2$ considers that the set of unknown object $\boldsymbol{X}_u$ originates from the source that produced $\boldsymbol{X}_2$, and so scores of the form $s_{uj,2j'}$, $s_{uj,uj'}$ will all be considered as "within-source 2" scores, and scores of the form $s_{uj,1j'}$ will be considered as "between-source" scores. Figure 3.1

shows the differences in the covariance structures when $\mathcal{H}_1$ is considered versus when $\mathcal{H}_2$ is considered for a non-stationary kernel (3.5) when $n_0 = 5$ and $n_u = 3$. We see that the elements take on different values in different positions. In addition, the similar color palettes between the two images indicate that the individual elements in the two matrices take on values from the same set of parameter values.

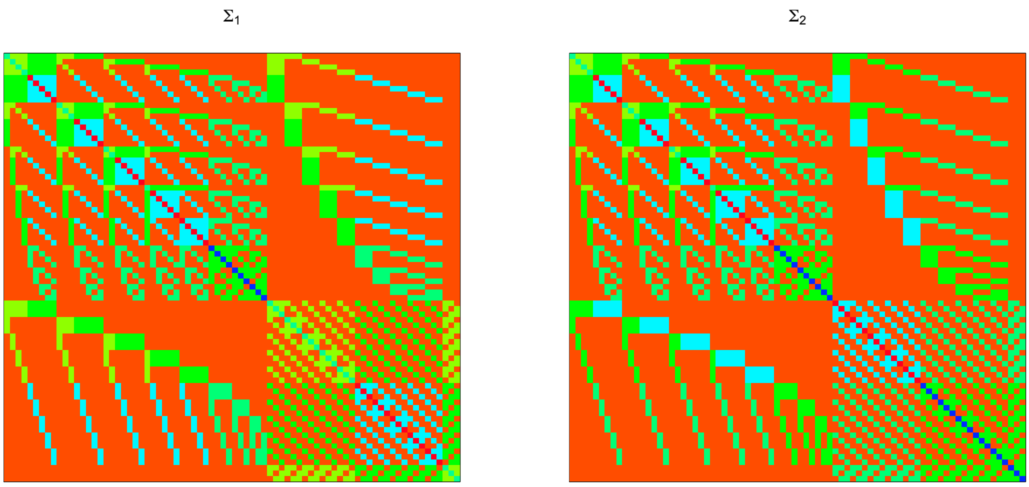


Figure 3.1: Covariance matrix structure under $\mathcal{H}_1$ (left) and $\mathcal{H}_2$ (right) when scores are calculated using a non-stationary kernel (3.5) when $n_0 = 5$ and $n_u = 3$. Comparisons that consider scores between control objects appear in the upper left corner. Comparisons that consider scores between trace objects are given in the lower right corner. Comparisons that involve trace and control objects are given in the upper right and lower left corners. Like colors indicate like values between the two covariance matrices, and so we see that the objects that consider scores between control objects have identical covariances under Σ_1 and Σ_2, while the scores that consider trace objects have different covariances under Σ_1 and Σ_2.

3.1.1 COVARIANCE STRUCTURE FOR THE OBJECT MODEL

We can investigate the covariance structure for a vector of scores by considering a univariate object-based model. Consider an object x_{ij} defined in terms of the linear model given by

$$x_{ij} = \mu_i + e_{ij}, \tag{3.3}$$

where μ_i is the overall mean of class $i \in \{1, 2\}$, and $e_{ij} \sim \mathcal{N}(0, \sigma^2)$ when $i = 1$, and $e_{ij} \sim \mathcal{N}(0, \tau^2)$ when $i = 2$. Scores can be studied by choosing the following stationary and non-stationary kernels, given by $s'_{ij,i'j'}$ and $s^*_{ij,i'j'}$, respectively:

$$s'_{ij,i'j'} := (x_{ij} - x_{i'j'})^2 \tag{3.4}$$

$$s^*_{ij,i'j'} := x_{ij}x_{i'j'}. \tag{3.5}$$

We can directly examine the mean and covariance terms associated with these two kernels by calculating the various terms that arise from the different possible score combinations. Table 3.1 provides the different parameter values under the two kernel functions given in (3.4) and (3.5) above.

Description of Considered Scores	Score 1	Score 2	Stationary Kernel (3.4)	Non-Stationary Kernel (3.5)
Expected Value Terms				
Within Source 1	1112	–	$2\sigma^2$	μ_1^2
Within Source 2	2122	–	$2\tau^2$	μ_2^2
Between Source	1221	–	$(\mu_1 - \mu_2)^2 + \sigma^2 + \tau^2$	$\mu_1\mu_2$
Covariance Terms				
Both Within Source 1, Two Common Objects	1112	1112	$8\sigma^4$	$2\mu_1^2\sigma^2 + \sigma^4$
Both Within Source 2, Two Common Objects	2122	2122	$8\tau^4$	$2\mu_2^2\tau^2 + \tau^4$
Both Within Source 1, One Common Object	1112	1113	$2\sigma^4$	$\mu_1^2\sigma^2$
Both Within Source 2, One Common Object	2122	2123	$2\tau^4$	$\mu_2^2\tau^2$
Both Within Source 1, No Common Objects	1112	1314	0	0
Both Within Source 2, No Common Objects	2122	2324	0	0
Both Between Source, Two Common Objects	1121	1121	$4(\mu_1 - \mu_2)^2(\sigma^2 + \tau^2) + 2(\sigma^2 + \tau^2)^2$	$\mu_1^2\tau^2 + \mu_2^2\sigma^2 + \sigma^2\tau^2$
Both Between Source, One Common Object from Source 1	1121	1122	$4(\mu_1 - \mu_2)^2\sigma^2 + 2\sigma^4$	$\mu_2^2\sigma^2$
Both Between Source, One Common Object from Source 2	1121	1221	$4(\mu_1 - \mu_2)^2\tau^2 + 2\tau^4$	$\mu_1^2\tau^2$
Both Between Source, No Common Objects	1121	1222	0	0
Within Source 1, Between Source, One Common Object from Source 1	1112	1121	$2\sigma^4$	$\mu_1\mu_2\sigma^2$
Within Source 2, Between Source, One Common Object from Source 2	2122	1121	$2\tau^4$	$\mu_1\mu_2\tau^2$
Within Source 1, Between Source, No Common Objects	1112	1321	0	0
Within Source 1, Within Source 2, No Common Objects	1112	2122	0	0

Table 3.1: Expected value and covariance terms obtained for the object-based model for each type of score comparison when a stationary kernel (e.g., (3.4)) and non-stationary kernel (e.g., (3.5)) are considered. Column one provides descriptions of each type of comparison that may be observed; columns two and three provide examples of indices for scores that could be compared in each situation; columns four and five present the parameter values obtained under the stationary and non-stationary kernels given by (3.4) and (3.5).

Table 3.1 indicates that the covariance structure varies depending on whether a stationary or non-stationary kernel is used to obtain the vector of scores, $\boldsymbol{s}$. For example, we see that there are ten unique terms that arise when a stationary kernel is considered, versus twelve unique terms that arise when a non-stationary kernel in considered. Furthermore, we have that some of the stationary terms are relatively straightforward functions of each other (e.g., $2\sigma^4$ is a fraction of $8\sigma^4$). In addition, the covariance terms that arise when a non-stationary kernel is considered depend much more on the means of the different sources. Finally, we note that the zeros occur in the same positions for the stationary and non-stationary kernels.

3.1.2 DEFINING A SCORE-BASED MODEL

Suppose, now, that we expand upon (3.1) and define our vector of scores such that

$$\begin{aligned} \boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right) \implies & s_{ij,i'j'} \sim \mathcal{N}\left(\theta_{ii'}, \sigma^2_{ii'}\right) \\ \implies & \frac{s_{ij,i'j'} - \theta_{ii'}}{\sigma_{ii'}} \sim \mathcal{N}\left(0, 1\right), \end{aligned} \tag{3.6}$$

where the parameters $\theta_{ii'}$ and $\sigma_{ii'}$ are the means and variances associated with the different comparisons that may be considered by a given score. That is, each score in the vector of scores is either a within-source 1 comparison, a within-source 2 comparison, or a between-source 1 and 2 comparison. For example, consider a score $s_{ij,i'j'}$ in which $i = i' = 1$, and so $\theta_{ii'}$ gives us the expected value of scores that compare any two objects in $\boldsymbol{X}_1$. Likewise, when $i = i' = 2$, $\theta_{ii'}$ gives us the expected value of scores that compare any two objects in $\boldsymbol{X}_2$. Finally, when $i \neq i'$, $\theta_{ii'}$ gives us the expected value of the scores that compare an object in $\boldsymbol{X}_1$ to an object in $\boldsymbol{X}_2$. We can thus consider $\theta_{ii'} \in \{\theta_{11}, \theta_{22}, \theta_{12}\}$, corresponding respectively to the three scenarios previously described. The parameter $\sigma_{ii'}$ can similarly be defined for standard deviations, such that $\sigma_{ii'} \in \{\sigma_{11}, \sigma_{22}, \sigma_{12}\}$.

From here, following the work of Gantz and Saunders, and Armstrong [5, 6, 26], we choose to define the standardized scores from (3.6) according to a random effects model

$$\frac{s_{ij,i'j'} - \theta_{ii'}}{\sigma_{ii'}} = a_{ij} + a_{i'j'} + \varepsilon_{ij,i'j'}, \tag{3.7}$$

where a_{ij} and $a_{i'j'}$ are random source effects, such that $a_{ij}, a_{i'j'} \overset{iid}{\sim} \mathcal{N}(0, \sigma_a^2)$, and $\varepsilon_{ij,i'j'}$ is a lack-of-fit term, such that $\varepsilon_{ij,i'j'} \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$. Furthermore, from (3.6), we have that $2\sigma_a^2 + \sigma_e^2 = 1$. Finally, we rewrite the model in terms of $s_{ij,i'j'}$, such that

$$s_{ij,i'j'} = \theta_{ii'} + \sigma_{ii'} \left(a_{ij} + a_{i'j'} + \varepsilon_{ij,i'j'} \right),$$

and so, given the distributional assumptions associated with (3.6) and (3.7), we define the distribution of our vector of scores to be

$$\boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta},\ \boldsymbol{\Delta} \left(\boldsymbol{P}\boldsymbol{P}'\sigma_a^2 + \boldsymbol{I}\sigma_e^2 \right) \boldsymbol{\Delta}' \right), \tag{3.8}$$

where $\boldsymbol{\theta}$ is a vector of length N of the mean terms given by $\theta_{ii'}$, and $\boldsymbol{\Delta}$ is an $N \times N$ diagonal matrix of the standard deviations given by $\sigma_{ii'}$. The design matrix $\boldsymbol{P}$ describes the effects of the objects being compared on the score for each score considered in the vector $\boldsymbol{s}$. For each of the $2n_0 + n_u$ rows of $\boldsymbol{P}$, a one is placed in the columns associated with the labels of the objects being compared in that row, and zeros are placed in the remaining columns. For example, given a score $s_{12,21}$, the columns of $\boldsymbol{P}$ corresponding to the second object from source 1, and the first object from source 2 would be assigned a value of one, while the remaining positions in the row would be assigned a value of zero. Finally, $\boldsymbol{I}$ is the $N \times N$ identity matrix.

The likelihood function for $\boldsymbol{s}$ in the numerator and denominator of (3.2) can be represented using the distribution given in (3.8). As explained in the introduction of Section 3.1, we have that the structure of the mean vector and covariance matrix depend on whether $\mathcal{H}_p$ or $\mathcal{H}_d$ is being considered (see, e.g., Figure 3.1).

It is worth noting that the covariance matrix defined in (3.8) is not equivalent to that of the object model described in Section 3.1.1. First, the covariance matrix in

Section 3.1.1 includes 7 unique covariance terms for the stationary kernel and 9 unique covariance terms for the non-stationary kernel, while the covariance matrix given in (3.8) includes 8 unique covariance terms (see Table 3.2). This is due to the fact that the covariance matrix in Section 3.1.1 considers a single term, σ_{12} to describe the relationship that occurs when a score involves an object from source 1 and an object from source 2. For the covariance matrix in (3.8) to coincide with that defined in Section 3.1.1, we would need to define two terms, σ_{12} and σ_{21}, that describe the effect when the object in common between two scores comes from source 1 versus from source 2. For example, consider a pair of scores, $s_{11,12}$ and $s_{11,21}$. To appropriately capture the covariance that exists between these two scores would require defining a term σ_{12}, since the common object between the scores comes from source 1. Likewise, a pair of scores, $s_{11,21}$ and $s_{21,22}$, would require defining a term σ_{21}, since the common object between the scores comes from source 2. Thus, the covariance terms of the score model in rows 7-9, 11, and 12 of Table 3.2 do not necessarily have a direct counterpart in the object model.

However, despite these discrepancies, we choose to move forward with the model given by (3.8). While the covariance matrices of the object and score models may not be exactly the same, their structures under $\mathcal{H}_1$ and $\mathcal{H}_2$ remain sufficiently similar. Furthermore, as we will see below, an elegant solution exists for studying the parameters of the model given by (3.8).

Description of Considered Scores	Score 1	Score 2	Object Model (3.5)	Score Model (3.8)
Covariance Terms				
Both Within Source 1, Two Common Objects	1112	1112	$2\mu_1^2\sigma^2 + \sigma^4$	$\sigma_{11}^2\,(2\sigma_a^2 + \sigma_e^2)$
Both Within Source 2, Two Common Objects	2122	2122	$2\mu_2^2\tau^2 + \tau^4$	$\sigma_{22}^2\,(2\sigma_a^2 + \sigma_e^2)$
Both Within Source 1, One Common Object	1112	1113	$\mu_1^2\sigma^2$	$\sigma_{11}^2\sigma_a^2$
Both Within Source 2, One Common Object	2122	2123	$\mu_2^2\tau^2$	$\sigma_{22}^2\sigma_a^2$
Both Within Source 1, No Common Objects	1112	1314	0	0
Both Within Source 2, No Common Objects	2122	2324	0	0
Both Between Source, Two Common Objects	1121	1121	$\mu_1^2\tau^2 + \mu_2^2\sigma^2 + \sigma^2\tau^2$	$\sigma_{12}^2\,(2\sigma_a^2 + \sigma_e^2)$
Both Between Source, One Common Object from Source 1	1121	1122	$\mu_2^2\sigma^2$	$\sigma_{12}^2\sigma_a^2$
Both Between Source, One Common Object from Source 2	1121	1221	$\mu_1^2\tau^2$	$\sigma_{12}^2\sigma_a^2$
Both Between Source, No Common Objects	1121	1222	0	0
Within Source 1, Between Source, One Common Object from Source 1	1112	1121	$\mu_1\mu_2\sigma^2$	$\sigma_{12}\sigma_{11}\sigma_a^2$
Within Source 2, Between Source, One Common Object from Source 2	2122	1121	$\mu_1\mu_2\tau^2$	$\sigma_{12}\sigma_{22}\sigma_a^2$
Within Source 1, Between Source, No Common Objects	1112	1321	0	0
Within Source 1, Within Source 2, No Common Objects	1112	2122	0	0

Table 3.2: Comparison of Covariance terms in Object Model defined according to (3.5), and Score Model defined according to (3.8). We see that there are two obvious scenarios (rows 8 and 9) in which the covariance term given by the score model is not equivalent to that given by the object model.

3.2 MODEL DEVELOPMENT

Evaluating the likelihood ratio in (3.2) requires estimating the parameters θ_{11}, θ_{22}, θ_{12}, σ_{11}, σ_{22}, σ_{12}, σ_a^2, and σ_e^2 using the information contained in the vector of scores, $\boldsymbol{s}$. To study these parameters, we subset the vector of scores to define $\boldsymbol{s}_c$, which includes only the comparisons between control objects in $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and so is a vector of length $N_c = \binom{2n_0}{2}$. We can then use $\boldsymbol{s}_c$ to define the total sums of squares

$$\begin{aligned} SS_{Tot} &= (\boldsymbol{s}_c - \boldsymbol{\theta}_c)'\,[\boldsymbol{\Delta}_c\boldsymbol{\Delta}_c']^{-1}\,(\boldsymbol{s}_c - \boldsymbol{\theta}_c), \\ &= \left(\boldsymbol{\Delta}_c^{-1}(\boldsymbol{s}_c - \boldsymbol{\theta}_c)\right)'\left(\boldsymbol{\Delta}_c^{-1}(\boldsymbol{s}_c - \boldsymbol{\theta}_c)\right) \end{aligned} \tag{3.9}$$

where $\boldsymbol{\theta}_c$ is the N_c vector of score means, $\theta_{ii'}$, and $\boldsymbol{\Delta}_c$ is the $N_c \times N_c$ diagonal matrix of the score standard deviations, $\sigma_{ii'}$, associated with the scores $\boldsymbol{s}_c$.

Now, Cochran's theorem [16] provides us with a means to decompose this sum of squares into several independent sums of squares.

Theorem 2 (Cochran's Theorem) *Let $\boldsymbol{x}$ be a $p \times 1$ random vector distributed according to a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}$. That is, $\boldsymbol{x} \sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{I})$. In addition, let*

$$\sum_{i=1}^{p} x_i^2 = q_1 + \cdots + q_K,$$

where the q_k are quadratic forms in $\boldsymbol{x}$, where $q_k = \boldsymbol{x}'\boldsymbol{A}_k\boldsymbol{x} \in \mathbb{R}$, where $\sum_{k=1}^{K} \boldsymbol{A}_k = \boldsymbol{I}$, and r_k is the rank of $\boldsymbol{A}_k$.

Then a necessary and sufficient condition that $q_1, \ldots, q_K$ are independently distributed χ^2 distributions with degrees of freedom $r_1, \ldots, r_K$ is that

$$p = r_1 + \cdots + r_K.$$

Cochran [16] proved this result for central χ^2 distributions. Notable developments of this theorem include: the extension to the non-central case, i.e., $\boldsymbol{x} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{I})$, by Madow [43]; the extension to positive definite covariance matrices, i.e., $\boldsymbol{x} \sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$, by Ogawa [55, 56]; the extension to consider non-zero mean vectors alongside positive definite covariance matrices, i.e., $\boldsymbol{x} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, by Ogasawara and Takahashi [54]. Anderson and Styan [4] survey these results, as well as other implications and consequences of these theorems.

The extension proven by Ogawa [54] allows us to apply Cochran's Theorem to $\widetilde{\boldsymbol{s}} = \left(\boldsymbol{\Delta}_c^{-1}\left(\boldsymbol{s}_c - \boldsymbol{\theta}_c\right)\right)$, and rewrite (3.9) as

$$SS_{Tot} = \widetilde{\boldsymbol{s}}'\boldsymbol{I}\widetilde{\boldsymbol{s}} = \widetilde{\boldsymbol{s}}'\left[\sum_{l=1}^{N_c} \boldsymbol{v}_l\boldsymbol{v}_l'\right]\widetilde{\boldsymbol{s}} \tag{3.10}$$

where $\{\boldsymbol{v}_l\}_l$, $l = 1, \ldots, N_c$ is any orthonormal basis for $\mathbb{R}^{N_c}$. Furthermore, consider the following three idempotent design matrices $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{22}$, and $\boldsymbol{B}_{12}$:

$\boldsymbol{B}_{11}$: The $N_c \times N_c$ diagonal matrix with one's in the first $\binom{n_0}{2}$ positions of the diagonal and zero's in the remaining $N_c - \binom{n_0}{2}$ positions;

$\boldsymbol{B}_{22}$: The $N_c \times N_c$ diagonal matrix with zero's in the first $\binom{n_0}{2}$ positions of the diagonal, one's in the second $\binom{n_0}{2}$ positions, and zero's in the remaining n_0^2 positions;

$\boldsymbol{B}_{12}$: The $N_c \times N_c$ diagonal matrix with zero's in the first $n\binom{n_0}{2}$ positions and one's in the remaining n_0^2 positions.

Since $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{22}$, and $\boldsymbol{B}_{12}$ sum to the identity matrix, we have that

$$\begin{aligned}
SS_{Tot} &= \widetilde{\boldsymbol{s}}'\left(\boldsymbol{B}_{11} + \boldsymbol{B}_{22} + \boldsymbol{B}_{12}\right)\boldsymbol{I}\left(\boldsymbol{B}_{11} + \boldsymbol{B}_{22} + \boldsymbol{B}_{12}\right)\widetilde{\boldsymbol{s}} \\
&= \widetilde{\boldsymbol{s}}'\boldsymbol{B}_{11}\boldsymbol{I}\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}} + \widetilde{\boldsymbol{s}}'\boldsymbol{B}_{22}\boldsymbol{I}\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}} + \widetilde{\boldsymbol{s}}'\boldsymbol{B}_{12}\boldsymbol{I}\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}} \\
&= \widetilde{\boldsymbol{s}}'\boldsymbol{B}_{11}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'\right]\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}} + \widetilde{\boldsymbol{s}}'\boldsymbol{B}_{22}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}'\right]\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}} + \widetilde{\boldsymbol{s}}'\boldsymbol{B}_{12}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}'\right]\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}} \\
&= SS_{11} + SS_{22} + SS_{12} \qquad (3.11)
\end{aligned}$$

where $\{\boldsymbol{v}_{11_l}\}_l$, $\{\boldsymbol{v}_{22_l}\}_l$, and $\{\boldsymbol{v}_{12_l}\}_l$, $l = 1, \ldots, N_c$, are different orthonormal bases spanning $\mathbb{R}^{N_c}$, and will be discussed in more detail in later sections. The matrices $\boldsymbol{B}_{ii'}$, $ii' \in \{11, 22, 12\}$, effectively activate different parts of the vector $\widetilde{\boldsymbol{s}}_c$ according to the different source comparisons. In particular, we have that

(1) $\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to the within-source one comparisons, so that $\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}}_c$ gives us the sum of squares for within-source one comparisons. Recall that $\boldsymbol{B}_{11}\boldsymbol{I}\boldsymbol{B}_{11} = \boldsymbol{B}_{11}$, and note that $\boldsymbol{B}_{11}$ has rank $r_{11} = \binom{n_0}{2}$.

(2) $\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to the within-source two comparisons, so that $\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}}_c$ gives us the sum of squares for within-source two comparisons. Recall that $\boldsymbol{B}_{22}\boldsymbol{I}\boldsymbol{B}_{22} = \boldsymbol{B}_{22}$, and note that $\boldsymbol{B}_{22}$ has rank $r_{22} = \binom{n_0}{2}$.

(3) $\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to the between-source one and two comparisons, so that $\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}}_c$ gives us the sum of squares for

between-source comparisons. Recall that $\boldsymbol{B}_{12}\boldsymbol{I}\boldsymbol{B}_{12} = \boldsymbol{B}_{12}$, and note that $\boldsymbol{B}_{12}$ has rank $r_{12} = n_0^2$.

Thus, we have defined the total sums of squares in terms of the various source comparisons that exist within our vector of scores. Bearing in mind that the goal is to find a way to estimate the parameters of the distribution given in (3.8), we note that this decomposition of the total sums of squares allows us to independently study the mean and standard deviation terms, θ_{ii} and σ_{ii}, associated with their respective source comparisons. Note that we can choose the orthonormal bases in (3.11) to be any orthonormal bases, and, in particular, we can choose these orthonormal bases to be the normalized eigenvectors of the following three matrices:

$$\boldsymbol{V}_{11} := \boldsymbol{B}_{11}\left(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{I}_c\sigma_e^2\right)\boldsymbol{B}_{11}, \tag{3.12}$$

$$\boldsymbol{V}_{22} := \boldsymbol{B}_{22}\left(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{I}_c\sigma_e^2\right)\boldsymbol{B}_{22}, \tag{3.13}$$

$$\boldsymbol{V}_{12} := \boldsymbol{B}_{12}\left(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{I}_c\sigma_e^2\right)\boldsymbol{B}_{12}. \tag{3.14}$$

First, choosing $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{22}$, and $\boldsymbol{V}_{12}$ to be any function of $\boldsymbol{\Sigma}_c := \boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{I}_c\sigma_e^2$ is advantageous in that it introduces the parameters σ_a^2 and σ_e^2, and so provides a means for studying these parameters. Second, defining $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{22}$, and $\boldsymbol{V}_{12}$ in terms of $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{22}$, and $\boldsymbol{B}_{12}$ allows us to take the relevant parts of $\boldsymbol{\Sigma}_c$ with respect to each source comparison by activating only the rows and columns of $\boldsymbol{\Sigma}_c$ corresponding to the considered source comparison.

$\boldsymbol{V}_{11}$		
Eigenvalue (ν_{11_l})	Multiplicity ($m_{\nu_{11_l}}$)	Eigenvectors ($\boldsymbol{v}_{11_l}$)
$2\left(n_0-1\right)\sigma_a^2+\sigma_e^2$	1	$\boldsymbol{v}_{11_l}$ such that $\boldsymbol{V}_{11}\boldsymbol{v}_{11_l}=\nu_{11_1}\boldsymbol{v}_{11_l}$
$\left(n_0-2\right)\sigma_a^2+\sigma_e^2$	n_0-1	$\boldsymbol{v}_{11_l}$ such that $\boldsymbol{V}_{11}\boldsymbol{v}_{11_l}=\nu_{11_2}\boldsymbol{v}_{11_l}$
σ_e^2	$\binom{n_0}{2}-n_0$	$\boldsymbol{v}_{11_l}$ such that $\boldsymbol{V}_{11}\boldsymbol{v}_{11_l}=\nu_{11_3}\boldsymbol{v}_{11_l}$
0	$N_c-\binom{n_0}{2}$	$\boldsymbol{v}_{11_l}$ such that $\boldsymbol{V}_{11}\boldsymbol{v}_{11_l}=\nu_{11_4}\boldsymbol{v}_{11_l}$
$\boldsymbol{V}_{22}$		
Eigenvalue (ν_{22_l})	Multiplicity ($m_{\nu_{22_l}}$)	Eigenvectors ($\boldsymbol{v}_{22_i}$)
$2\left(n_0-1\right)\sigma_a^2+\sigma_e^2$	1	$\boldsymbol{v}_{22_l}$ such that $\boldsymbol{V}_{22}\boldsymbol{v}_{22_l}=\nu_{22_1}\boldsymbol{v}_{22_l}$
$\left(n_0-2\right)\sigma_a^2+\sigma_e^2$	n_0-1	$\boldsymbol{v}_{22_l}$ such that $\boldsymbol{V}_{22}\boldsymbol{v}_{22_l}=\nu_{22_2}\boldsymbol{v}_{22_l}$
σ_e^2	$\binom{n_0}{2}-n_0$	$\boldsymbol{v}_{22_i}$ such that $\boldsymbol{V}_{22}\boldsymbol{v}_{22_l}=\nu_{22_3}\boldsymbol{v}_{22_l}$
0	$N_c-\binom{n_0}{2}$	$\boldsymbol{v}_{22_i}$ such that $\boldsymbol{V}_{22}\boldsymbol{v}_{22_l}=\nu_{22_4}\boldsymbol{v}_{22_l}$
$\boldsymbol{V}_{12}$		
Eigenvalue (ν_{12_l})	Multiplicity ($m_{\nu_{12_l}}$)	Eigenvectors ($\boldsymbol{v}_{12_l}$)
$2n_0\sigma_a^2+\sigma_e^2$	1	$\boldsymbol{v}_{12_l}$ such that $\boldsymbol{V}_{12}\boldsymbol{v}_{12_l}=\nu_{12_1}\boldsymbol{v}_{12_l}$
$n_0\sigma_a^2+\sigma_e^2$	$2n_0-2$	$\boldsymbol{v}_{12_l}$ such that $\boldsymbol{V}_{12}\boldsymbol{v}_{12_l}=\nu_{12_2}\boldsymbol{v}_{12_l}$
σ_e^2	$\left(n_0-1\right)^2$	$\boldsymbol{v}_{12_l}$ such that $\boldsymbol{V}_{12}\boldsymbol{v}_{12_l}=\nu_{12_3}\boldsymbol{v}_{12_l}$
0	$N_c-n_0^2$	$\boldsymbol{v}_{12_l}$ such that $\boldsymbol{V}_{12}\boldsymbol{v}_{12_l}=\nu_{12_4}\boldsymbol{v}_{12_l}$

Table 3.3: Eigenstructure of design matrices $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{22}$, and $\boldsymbol{V}_{12}$ in (3.12), (3.13), and (3.14)

We can study the eigenstructure of $\boldsymbol{B}_{ii'}\left(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2+\boldsymbol{I}_c\sigma_e^2\right)\boldsymbol{B}_{ii'}$, $ii'\in\{11,22,12\}$, for each source comparison (see Table 3.3). This study reveals the presence of multiple subspaces for each of the considered eigenspaces, and allows us to decompose each of the sums of squares in (3.11) as another sum of squares. In particular, we have that

$$
\begin{aligned}
\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{11}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'\right]\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{11}\left[\boldsymbol{v}_{11_1}\boldsymbol{v}_{11_1}'+\sum_{l=2}^{n_0}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'+\sum_{l=n_0+1}^{\binom{n_0}{2}}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'+\sum_{l=\binom{n_0}{2}+1}^{N_c}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'\right]\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}}_c \\
&= \widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{11}\left[\boldsymbol{v}_{11_1}\boldsymbol{v}_{11_1}'+\sum_{l=2}^{n_0}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'+\sum_{l=n_0+1}^{N_{11}}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'\right]\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}}_c,
\end{aligned}
$$

where $N_{11}=\binom{n_0}{2}$ is the number of eigenvectors associated with non-zero eigenvalues, and is also the number of "interesting" eigenvectors, $\boldsymbol{v}_{11_l}$. In particular, we have that the first N_{11} eigenvectors have N_{11} non-zero elements in their first N_{11} rows, while the remaining N_c-N_{11} eigenvectors are vectors of zeros, each with one element equal to one. The elements that are equal to one correspond to the dimensions whose associated eigenvalues are zero. These vectors form the standard basis for the null space of the corresponding matrix, and correspond to the rows of $\boldsymbol{B}_{11}$ that are equal to the zero

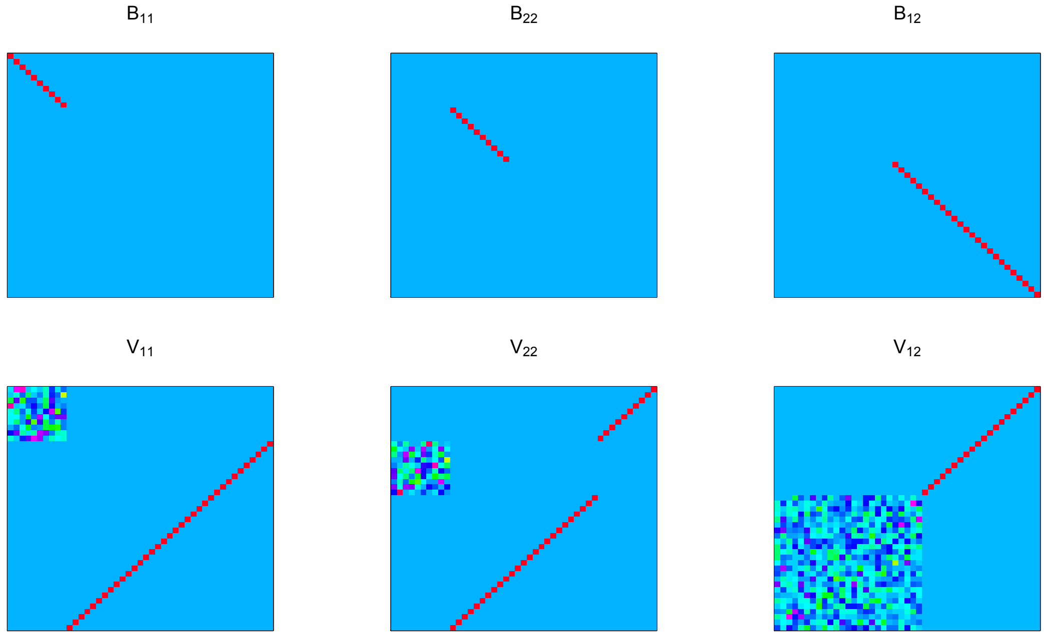


Figure 3.2: Heat maps for $B_{ii'}$ matrices (top) and eigenvectors of associated $\boldsymbol{V}_{ii'}$ matrices (bottom). The $\binom{n_0}{2} \times \binom{n_0}{2}$ and $n_0^2 \times n_0^2$ patchworks in the first $\binom{n_0}{2}$ and n_0^2 columns correspond to eigenvectors with non-zero elements. The remaining columns correspond to vectors with one non-zero element (in red), and are associated with zero-valued eigenvalues.

vector. The first column of Figure 3.2 displays the heat maps of $\boldsymbol{B}_{11}$ (top) and of the matrix of eigenvectors of $\boldsymbol{V}_{11}$ (bottom). Note that the $N_{11} \times N_{11}$ patchwork matrix corresponds to the non-zero rows of $\boldsymbol{B}_{11}$, and the orthonormal basis for the null space of $\boldsymbol{V}_{11}$ corresponds to the zero-valued rows of $\boldsymbol{B}_{11}$. Because the placements of the nonzero elements in these eigenvectors correspond to the zero elements of the diagonal in $\boldsymbol{B}_{11}$, the product of $\boldsymbol{B}_{11}$ with this set of eigenvectors results in a zero-valued sum of squares.

Similarly, for $\boldsymbol{B}_{22}$, we have

$$\begin{aligned}
\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{22}\left[\sum_{l=1}^{N_n}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}'\right]\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{22}\left[\boldsymbol{v}_{22_1}\boldsymbol{v}_{22_1}' + \sum_{l=2}^{n_0}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}' + \sum_{l=n_0+1}^{\binom{n_0}{2}}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}' + \sum_{l=\binom{n_0}{2}+1}^{N_c}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}'\right]\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}}_c \\
&= \widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{22}\left[\boldsymbol{v}_{22_1}\boldsymbol{v}_{22_1}' + \sum_{l=2}^{n_0}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}' + \sum_{l=n_0+1}^{N_{22}}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}'\right]\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}}_c,
\end{aligned}$$

where $N_{22} = \binom{n_0}{2}$ is the number of eigenvectors associated with non-zero eigenvalues, and is also the number of "interesting" eigenvectors, $\boldsymbol{v}_{22_l}$. In particular, we have that the first N_{22} eigenvectors have N_{22} non-zero elements in their $N_{11}+1$ through $N_{11}+N_{22}$ rows, while the remaining $N_c - N_{22}$ eigenvectors are vectors of zeros, each with one element equal to one. The elements that are equal to one correspond to the dimensions whose associated eigenvalues are zero. These vectors form the standard basis for the null space of the corresponding matrix, and correspond to the rows of $\boldsymbol{B}_{22}$ that are equal to the zero vector. The second column of Figure 3.2 displays the heat maps of $\boldsymbol{B}_{22}$ (top) and of the matrix of eigenvectors of $\boldsymbol{V}_{22}$ (bottom). Note that the $N_{22} \times N_{22}$ patchwork matrix corresponds to the non-zero rows of $\boldsymbol{B}_{22}$, and the orthonormal basis for the null space of $\boldsymbol{V}_{22}$ corresponds to the zero-valued rows of $\boldsymbol{B}_{22}$. Because the placements of the nonzero elements in these eigenvectors correspond to the zero elements of the diagonal in $\boldsymbol{B}_{22}$, the product of $\boldsymbol{B}_{22}$ with this set of eigenvectors results in a zero-valued sum of squares.

Finally, for $\boldsymbol{B}_{12}$, we have

$$\begin{aligned}
\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{12}\left[\sum_{l=1}^{N_n}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}'\right]\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{12}\left[\boldsymbol{v}_{12_1}\boldsymbol{v}_{12_1}' + \sum_{l=2}^{\binom{n_0}{2}-1}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}' + \sum_{l=\binom{n_0}{2}}^{n_0^2}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}' + \sum_{l=n_0^2+1}^{N_c}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}'\right]\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}}_c \\
&= \widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{12}\left[\boldsymbol{v}_{12_1}\boldsymbol{v}_{12_1}' + \sum_{l=2}^{\binom{n_0}{2}-1}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}' + \sum_{l=\binom{n_0}{2}}^{N_{12}}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}'\right]\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}}_c.
\end{aligned}$$

where $N_{12} = n_0^2$ is the number of eigenvectors associated with non-zero eigenvalues, and is also the number of "interesting" eigenvectors, $\boldsymbol{v}_{12_l}$. In particular, we have that the first N_{12} eigenvectors have N_{12} non-zero elements in their last N_{12} rows, while the remaining $N_c - N_{12}$ eigenvectors are vectors of zeros, each with one element equal to onefavorable. The elements that are equal to one correspond to the dimensions whose associated eigenvalues are zero. These vectors form the standard basis for the null space of the corresponding matrix, and correspond to the rows of $\boldsymbol{B}_{12}$ that are equal to the zero vector. The third column of Figure 3.2 displays the heat maps of $\boldsymbol{B}_{12}$ (top) and of the matrix of eigenvectors of $\boldsymbol{V}_{12}$ (bottom). Note that the $N_{12} \times N_{12}$ patchwork matrix corresponds to the non-zero rows of $\boldsymbol{B}_{12}$, and the orthonormal basis for the null space of $\boldsymbol{V}_{12}$ corresponds to the zero-valued rows of $\boldsymbol{B}_{12}$. Because the placements of the nonzero elements in these eigenvectors correspond to the zero elements of the diagonal in $\boldsymbol{B}_{12}$, the product of $\boldsymbol{B}_{12}$ with this set of eigenvectors results in a zero-valued sum of squares. favorable This decomposition is favorable in that studying the relevant parts of the eigen-decomposition of $\boldsymbol{\Sigma}_c$ is equivalent to studying the eigen-decomposition of the relevant parts of $\boldsymbol{\Sigma}_c$. That is,

$$\begin{aligned}
\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{11}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{11_l}\boldsymbol{v}_{11_l}'\right]\boldsymbol{B}_{11}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_{11}'\left[\sum_{l=1}^{N_{11}}\boldsymbol{v}_{11_l}^*\boldsymbol{v}_{11_l}^{*\prime}\right]\widetilde{\boldsymbol{s}}_{11} \\
\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{22}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{22_l}\boldsymbol{v}_{22_l}'\right]\boldsymbol{B}_{22}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_{22}'\left[\sum_{l=1}^{N_{22}}\boldsymbol{v}_{22_l}^*\boldsymbol{v}_{22_l}^{*\prime}\right]\widetilde{\boldsymbol{s}}_{22} \\
\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{12}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{12_l}\boldsymbol{v}_{12_l}'\right]\boldsymbol{B}_{12}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_{22}'\left[\sum_{l=1}^{N_{12}}\boldsymbol{v}_{12_l}^*\boldsymbol{v}_{12_l}^{*\prime}\right]\widetilde{\boldsymbol{s}}_{12}
\end{aligned}$$

where $\boldsymbol{v}_{11_l}^*$ are the eigenvectors of the $N_{11} \times N_{11}$ matrix formed by considering the non-zero rows of the columns associated with the non-zero eigenvalues of $\boldsymbol{V}_{11}$ (given by the $1 : N_{11}$ rows and $1 : N_{11}$ columns of $\boldsymbol{V}_{11}$), $\boldsymbol{v}_{22_l}^*$ are the eigenvectors of the $N_{22} \times N_{22}$ matrix formed by considering the non-zero portions of the columns associated with the non-zero eigenvalues of $\boldsymbol{V}_{22}$ (given by the $(N_{11}+1) : (N_{11}+N_{22})$ rows and $1 : N_{22}$ columns of $\boldsymbol{V}_{22}$), and $\boldsymbol{v}_{12_l}^*$ are the eigenvectors of the $N_{12} \times N_{12}$ matrix formed by considering the non-zero portions of the columns associated with the non-zero eigenvalues of $\boldsymbol{V}_{12}$ (given by the $(N_{22}+1) : N_n$ rows and $1 : N_{12}$ columns of $\boldsymbol{V}_{12}$). This is equivalent to considering only the indices of $\boldsymbol{\Sigma}_c$ that correspond to each source comparison. That is, $\boldsymbol{v}_{11_l}^*$ are the eigenvectors of

$$\boldsymbol{\Sigma}_{11} := \boldsymbol{P}_{11}\boldsymbol{P}_{11}'\sigma_a^2 + \boldsymbol{I}_{11}\sigma_e^2, \tag{3.15}$$

the matrix formed by considering the rows of $\boldsymbol{\Sigma}_c$ associated with within-source one comparisons; $\boldsymbol{v}_{22_l}^*$ are the eigenvectors of

$$\boldsymbol{\Sigma}_{22} := \boldsymbol{P}_{22}\boldsymbol{P}_{22}'\sigma_a^2 + \boldsymbol{I}_{22}\sigma_e^2, \tag{3.16}$$

the matrix formed by considering the rows of $\boldsymbol{\Sigma}_c$ associated with within-source two comparisons; and $\boldsymbol{v}^*_{12_l}$ are the eigenvectors of

$$\boldsymbol{\Sigma}_{12} := \boldsymbol{P}_{12}\boldsymbol{P}'_{12}\sigma_a^2 + \boldsymbol{I}_{12}\sigma_e^2, \tag{3.17}$$

the matrix formed by considering the rows of $\boldsymbol{\Sigma}_c$ associated with between source comparisons. In addition, we have that $\widetilde{\boldsymbol{s}}_{11} = \left(\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11} - \theta_{11}\mathbf{1}_{N_{11}}\right)\right)$, where $\boldsymbol{\Delta}_{11}$ is the $N_{11} \times N_{11}$ portion of $\boldsymbol{\Delta}_n$ that considers σ_{11}^2, $\widetilde{\boldsymbol{s}}_{22} = \left(\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22} - \theta_{22}\mathbf{1}_{N_{22}}\right)\right)$, where $\boldsymbol{\Delta}_{22}$ is the $N_{22} \times N_{22}$ portion of $\boldsymbol{\Delta}_n$ that considers σ_{22}^2, and $\widetilde{\boldsymbol{s}}_{12} = (\boldsymbol{\Delta}_{12}^{-1}(\boldsymbol{s}_{12} - \theta_{12}\mathbf{1}_{N_{12}}))$, where $\boldsymbol{\Delta}_{12}$ is the $N_{12} \times N_{12}$ portion of $\boldsymbol{\Delta}_n$ that considers σ_{12}^2.

$\boldsymbol{\Sigma}_{11}$		
Eigenvalue (ν_{11_l})	Multiplicity ($m_{\nu_{11_l}}$)	Eigenvectors ($\boldsymbol{v}^*_{11_l}$)
$2(n_0-1)\sigma_a^2 + \sigma_e^2$	1	$\boldsymbol{v}^*_{11_1} := \frac{\mathbf{1}_{N_{11}}}{\sqrt{N_{11}}}$
$(n_0-2)\sigma_a^2 + \sigma_e^2$	$n_0 - 1$	$\boldsymbol{v}^*_{11_l}$ such that $\boldsymbol{\Sigma}_{11}\boldsymbol{v}^*_{11_l} = \nu_{11_2}\boldsymbol{v}^*_{11_l}$
σ_e^2	$N_{11} - n_0$	$\boldsymbol{v}^*_{11_l}$ such that $\boldsymbol{\Sigma}_{11}\boldsymbol{v}^*_{11_l} = \nu_{11_3}\boldsymbol{v}^*_{11_l}$
$\boldsymbol{\Sigma}_{22}$		
Eigenvalue (ν_{22_l})	Multiplicity ($m_{\nu_{22_l}}$)	Eigenvectors ($\boldsymbol{v}^*_{22_l}$)
$2(n_0-1)\sigma_a^2 + \sigma_e^2$	1	$\boldsymbol{v}^*_{22_1} := \frac{\mathbf{1}_{N_{22}}}{\sqrt{N_{22}}}$
$(n_0-2)\sigma_a^2 + \sigma_e^2$	$n_0 - 1$	$\boldsymbol{v}^*_{22_l}$ such that $\boldsymbol{\Sigma}_{22}\boldsymbol{v}^*_{22_l} = \nu_{22_2}\boldsymbol{v}^*_{22_l}$
σ_e^2	$\binom{n_0}{2} - n_0$	$\boldsymbol{v}^*_{22_l}$ such that $\boldsymbol{\Sigma}_{22}\boldsymbol{v}^*_{22_l} = \nu_{22_3}\boldsymbol{v}^*_{22_l}$
$\boldsymbol{\Sigma}_{12}$		
Eigenvalue (ν_{12_l})	Multiplicity ($m_{\nu_{12_l}}$)	Eigenvectors ($\boldsymbol{v}^*_{12_l}$)
$2n_0\sigma_a^2 + \sigma_e^2$	1	$\boldsymbol{v}^*_{12_1} := \frac{\mathbf{1}_{N_{12}}}{\sqrt{N_{12}}}$
$n_0\sigma_a^2 + \sigma_e^2$	$2n_0 - 2$	$\boldsymbol{v}^*_{12_l}$ such that $\boldsymbol{\Sigma}_{12}\boldsymbol{v}^*_{12_l} = \nu_{12_2}\boldsymbol{v}^*_{12_l}$
σ_e^2	$(n_0-1)^2$	$\boldsymbol{v}^*_{12_l}$ such that $\boldsymbol{\Sigma}_{12}\boldsymbol{v}^*_{12_l} = \nu_{12_3}\boldsymbol{v}^*_{12_l}$

Table 3.4: Eigenstructure of design matrices $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$, and $\boldsymbol{\Sigma}_{12}$ in (3.15), (3.16), and (3.17)

These results follow from using the $\boldsymbol{B}_{ii'}$ matrices to activate certain areas of the vector $\widetilde{\boldsymbol{s}}_c$ and the matrices $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{22}$, and $\boldsymbol{V}_{12}$, i.e., introducing $\boldsymbol{B}_{ii'}$ allows us to activate the parts of $\widetilde{\boldsymbol{s}}_c$ and $\boldsymbol{\Sigma}_c$ that correspond to the different source comparisons. Rather than consider a sparse N_c vector alongside a sparse $N_c \times N_c$ matrix, we can directly consider the interesting parts of the vector and matrix by considering the associated $N_{ii'}$-dimensional vector and $N_{ii'} \times N_{ii'}$ matrix, $ii' \in \{11, 22, 12\}$. Thus, we can explicitly define the source comparison sums of squares such that

$$\begin{aligned} SS_{11} &= \left(\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)\right)'\left(\boldsymbol{v}_{11_1}^{*}\boldsymbol{v}_{11_1}^{*}{}'\right)\left(\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)\right) + \\ &\quad \left(\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)\right)'\left(\sum_{l=2}^{n_0}\boldsymbol{v}_{11_l}^{*}\boldsymbol{v}_{11_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)\right) + \\ &\quad \left(\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)\right)'\left(\sum_{l=n_0+1}^{N_{11}}\boldsymbol{v}_{11_l}^{*}\boldsymbol{v}_{11_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)\right), \end{aligned} \tag{3.18}$$

$$\begin{aligned} SS_{22} &= \left(\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)\right)'\left(\boldsymbol{v}_{22_1}^{*}\boldsymbol{v}_{22_1}^{*}{}'\right)\left(\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)\right) + \\ &\quad \left(\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)\right)'\left(\sum_{l=2}^{n_0}\boldsymbol{v}_{22_l}^{*}\boldsymbol{v}_{22_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)\right) + \\ &\quad \left(\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)\right)'\left(\sum_{l=n_0+1}^{N_{22}}\boldsymbol{v}_{22_l}^{*}\boldsymbol{v}_{22_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)\right), \end{aligned} \tag{3.19}$$

and

$$\begin{aligned} SS_{12} &= \left(\boldsymbol{\Delta}_{12}^{-1}\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)\right)'\left(\boldsymbol{v}_{12_1}^{*}\boldsymbol{v}_{12_1}^{*}{}'\right)\left(\boldsymbol{\Delta}_{12}^{-1}\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)\right) + \\ &\quad \left(\boldsymbol{\Delta}_{12}^{-1}\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)\right)'\left(\sum_{l=2}^{\binom{n_0}{2}-1}\boldsymbol{v}_{12_l}^{*}\boldsymbol{v}_{12_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{12}^{-1}\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)\right) + \\ &\quad \left(\boldsymbol{\Delta}_{12}^{-1}\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)\right)'\left(\sum_{l=\binom{n_0}{2}}^{N_{12}}\boldsymbol{v}_{12_l}^{*}\boldsymbol{v}_{12_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{12}^{-1}\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)\right), \end{aligned} \tag{3.20}$$

where the degrees of freedom for each line of (3.18), (3.19), and (3.20) are equal to the multiplicities of the associated eigenvalues in Table 3.4, and the total sum of squares remains as in (3.9). It is trivial to show that $\sum_{l=1}^{3} m_{\nu_{11_l}} + \sum_{l=1}^{3} m_{\nu_{22_l}} + \sum_{l=1}^{3} m_{\nu_{12_l}} = N_c$.

In the following sections, we analyze the three terms that make up each of the sums of squares defined in (3.18), (3.19), and (3.20), so that we can write each term without the use of eigenvectors, and more efficiently estimate the model parameters.

3.2.1 ALTERNATIVE REPRESENTATION OF SS_{11}

We begin by studying the terms in SS_{11}, given by (3.18). All developments can be found in Appendix A. We can re-write the first term as

$$\left(\boldsymbol{s}_{11} - \theta_{11}\mathbf{1}_{N_{11}}\right)' \boldsymbol{\Delta}_{11_1}^{-1'} \boldsymbol{v}_{11_1}^* \boldsymbol{v}_{11_1}^{*}{}' \boldsymbol{\Delta}_{11}^{-1} \left(\boldsymbol{s}_{11} - \theta_{11}\mathbf{1}_{N_{11}}\right) = \frac{N_{11}}{\sigma_{11}^2} \left(\overline{s}_{11} - \theta_{11}\right)^2, \tag{3.21}$$

where $\overline{s}_{11}$ is the average score observed for within-source comparisons from source 1.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{11}^{-1'} \left[\sum_{l=2}^{n_0} \boldsymbol{v}_{11_l}^* \boldsymbol{v}_{11_l}^{*}{}'\right] \boldsymbol{\Delta}_{11}^{-1}$. Following the development by [6], we can write this second sum of squares as

$$\begin{aligned} \boldsymbol{s}_{11}' \boldsymbol{\Delta}_{11}^{-1'} \left[\sum_{l=2}^{n_0} \boldsymbol{v}_{11_l}^* \boldsymbol{v}_{11_l}^{*}{}'\right] \boldsymbol{\Delta}_{11}^{-1} \boldsymbol{s}_{11} &= \frac{(n_0-1)^2}{\sigma_{11}^2 (n_0-2)} \sum_{j=1}^{n_0} \left(\overline{s}_{11}^{(1_j)} - \overline{s}_{11}\right)^2, \\ &:= SS_{W_{11}} \end{aligned} \tag{3.22}$$

where $\overline{s}_{11}^{(1_j)}$, $j \in \{1, \ldots, n_0\}$ is the mean value of scores that compare object j in source 1 to any other object in source 1, and $\overline{s}_{11}$ is as in (3.21). The final result, given by (3.22), gives the within-source sum of squares for the within-source-one model, $SS_{W_{11}}$. By considering this term in conjunction with the total sums of squares for the considered model, $SS_{Tot_{11}} = \frac{1}{\sigma_{11}^2} \sum_{j=1}^{n_0-1} \sum_{j'=j+1}^{n_0} \left(s_{1j,1j'} - \overline{s}_{11}\right)^2$ (this is the sum of the last two terms in (3.18)), we can obtain an eigenvector-free estimate of the last bracketed term in SS_{11} by considering $SS_{Tot_{11}} - SS_{W_{11}}$ (see Table 3.5).

Following Cochran's theorem [16] and the development in Section 3.2, we have that each sum of squares in SS_{11} is an independent central χ^2-distribution with degrees of freedom equal to the multiplicity of the eigenvalue, ν_{11_l}, associated with the considered sum (see Table 3.4). We can study these terms to obtain the expected mean sums of squares. Generally speaking, if we have any sum of squares term, SS, associated with the eigenvalue λ whose multiplicity is m_λ, then we have

$$\frac{SS}{Var[SS]} \sim \chi^2_{df_{SS}} \Longrightarrow \frac{SS}{\lambda} \sim \chi^2_{df=m_\lambda},$$

such that,

$$\mathbb{E}\left[\frac{SS}{\lambda}\right] = m_\lambda \iff \mathbb{E}\left[SS\right] = \lambda m_\lambda$$
$$\iff \mathbb{E}\left[MS\right] = \lambda.$$

Thus, the expected value of a sum of squares divided by its degrees of freedom corresponds to the corresponding eigenvalue. Applying these results to the three terms that comprise SS_{11} in (3.18), we obtain the results presented in Table 3.5.

Source of Variance	df	SS	MS	E(MS)
Within Source	$n_0 - 1$	$SS_{W_{11}}$	$MS_{W_{11}} = \frac{SS_{W_{11}}}{n_0-1}$	$(n_0-2)\sigma_a^2 + \sigma_e^2$
Error	$N_{11} - n_0$	$SS_{E_{11}} = SS_{Tot_{11}} - SS_{W_{11}}$	$MS_{E_{11}} = \frac{SS_{E_{11}}}{N_{11}-n_0}$	σ_e^2

Table 3.5: ANOVA table corresponding to SS_{11}

3.2.2 ALTERNATIVE REPRESENTATION OF SS_{22}

Next, we move to study the terms in SS_{22}, given by (3.19). All developments can be found in Appendix B. We proceed in the same manner as in Section 3.2.1. We re-write the first term as

$$(\boldsymbol{s}_{22} - \theta_{22}\mathbf{1}_{N_{22}})'\,\boldsymbol{\Delta}_{22}^{-1'}\boldsymbol{v}_{22_1}^{*}\boldsymbol{v}_{22_1}^{*}{}'\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22} - \theta_{22}\mathbf{1}_{N_{22}}\right) = \frac{N_{22}}{\sigma_{22}^2}\left(\bar{s}_{22} - \theta_{22}\right)^2, \tag{3.23}$$

where $\bar{s}_{22}$ is the average score observed for within-source comparisons from source 2.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{22}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{22_l}^{*}\boldsymbol{v}_{22_l}^{*}{}'\right]\boldsymbol{\Delta}_{22}^{-1}$. Again, we can write this second sum of squares as the within-source sum of squares, $SS_{W_{22}}$, for the the within-source-two model

$$\begin{aligned}\boldsymbol{s}_{22}'\boldsymbol{\Delta}_{22}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{22_l}^{*}\boldsymbol{v}_{22_l}^{*}{}'\right]\boldsymbol{\Delta}_{22}^{-1}\boldsymbol{s}_{22} &= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\sum_{j=1}^{n_0}\left(\bar{s}_{22}^{(2_j)} - \bar{s}_{22}\right)^2 \\ &:= SS_{W_{22}}\end{aligned} \tag{3.24}$$

where $\bar{s}_{22}^{(2_j)}$, $j \in \{1, \ldots, n_0\}$ is the mean value of scores that compare object j in source 2 to any other object in source 2, and $\bar{s}_{22}$ is as in (3.23). Again, by considering this term in conjunction with the total sum of squares for the considered model, given by $SS_{Tot_{22}} = \frac{1}{\sigma_{22}^2}\sum_{j=1}^{n_0-1}\sum_{j'=j+1}^{n_0}(s_{2j,2j'} - \bar{s}_{22})^2$ (this is the sum of the last two terms in (3.19)), we can obtain an eigenvector-free estimate of the last bracketed term in SS_{22}. By using the results presented in Section 3.2.1, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	$n_0 - 1$	$SS_{W_{22}}$	$MS_{W_{22}} = \frac{SS_{W_{22}}}{n_0-1}$	$(n_0-2)\sigma_a^2 + \sigma_e^2$
Error	$N_{22} - n_0$	$SS_{E_{22}} = SS_{Tot_{22}} - SS_{W_{22}}$	$MS_{E_{22}} = \frac{SS_{E_{22}}}{N_{22}-n_0}$	σ_e^2

Table 3.6: ANOVA table corresponding to SS_{22}

Note that the results presented in table 3.6 are identical to those presented in table 3.5. This phenomenon arises from the balanced nature of the design.

3.2.3 ALTERNATIVE REPRESENTATION OF SS_{12}

Finally, we consider the terms in SS_{12}, given by (3.20). All developments can be found Appendix C. We proceed in the same manner as in Sections 3.2.1 and 3.2.2. As before, we begin by re-writing the first term as

$$(\boldsymbol{s}_{12} - \theta_{12}\mathbf{1}_{N_{12}})' \boldsymbol{\Delta}_{12}^{-1'} \boldsymbol{v}_{12_1}^* \boldsymbol{v}_{12_1}^{*\prime} \boldsymbol{\Delta}_{12}^{-1} (\boldsymbol{s}_{12} - \theta_{12}\mathbf{1}_{N_{12}}) = \frac{N_{12}}{\sigma_{12}^2} (\bar{s}_{12} - \theta_{12})^2, \qquad (3.25)$$

where $\bar{s}_{12}$ is the average score observed for between-source comparisons between sources 1 and 2.

Next, we consider the structure of $\boldsymbol{\Delta}_{12}^{-1'} \left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}_{12_l}^* \boldsymbol{v}_{12_l}^{*\prime}\right] \boldsymbol{\Delta}_{12}^{-1}$. Again, we can write this second sum of squares as the within-source sum of squares, $SS_{W_{12}}$ for the between-source model

$$\begin{aligned} \boldsymbol{s}_{12}' \boldsymbol{\Delta}_{12}^{-1'} \left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}_{12_l}^* \boldsymbol{v}_{12_l}^{*\prime}\right] \boldsymbol{\Delta}_{12}^{-1} \boldsymbol{s}_{12} &= \frac{n_0^2}{\sigma_{12}^2 n_0} \sum_{j=1}^{n_0} \left[\left(\bar{s}_{12}^{(1_j)} - \bar{s}_{12}\right)^2 + \left(\bar{s}_{12}^{(2_{j'})} - \bar{s}_{12}\right)^2\right] \\ &= SS_{W_{12}} \end{aligned} \qquad (3.26)$$

where $\bar{s}_{12}^{(1_j)}$, $j \in \{1, \ldots, n_0\}$, is the mean value of scores that compare object j in source 1 to any other object in source 2, $\bar{s}_{12}^{(2_{j'})}$, $j' \in \{1, \ldots, n_0\}$, is the mean value of scores that compare object j' in source 2 to any other object in source 1, and $\bar{s}_{12}$ is as in (3.25). Finally, by considering this term in conjunction with the total sum of squares for the considered model, given by $SS_{Tot_{12}} = \frac{1}{\sigma_{12}} \sum_{j=1}^{n_0} \sum_{j'=1}^{n_0} \left(s_{1j,2j'} - \overline{s}_{12}\right)^2$ (this is the sum of the last two terms in (3.20)), we can obtain an eigenvector-free estimate of the final bracketed term in SS_{12}. By using the results of Cochran's theorem presented in Section 3.2.1, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	$2n_0 - 2$	$SS_{W_{12}}$	$MS_{W_{12}} = \frac{SS_{W_{12}}}{2n_0-2}$	$n_0\sigma_a^2 + \sigma_e^2$
Error	$(n_0-1)^2$	$SS_{E_{12}} = SS_{Tot_{12}} - SS_{W_{12}}$	$MS_{E_{12}} = \frac{SS_{E_{12}}}{(n_0-1)^2}$	σ_e^2

Table 3.7: ANOVA table corresponding to SS_{12}

3.3 PARAMETER ESTIMATION

At this point, we would like to use the results presented in Sections 3.2.1, 3.2.2, and 3.2.3 to estimate the parameters of our model; however, given the dependencies that exist between the various developments, we must resort to sampling methods to obtain posterior samples of the model parameters. In particular, we use a Gibbs sampler with a Metropolis-Hastings step to study the distributions of our various parameters [14, 44]. Before defining the Gibbs sampler, we must first assign posterior distributions to the model parameters (development of posterior distributions for σ_e^2, $\theta_{ii'}$, and $\sigma_{ii'}^2$ can be found in Appendix F).

We begin by assigning posterior distributions for the variance terms, σ_a^2 and σ_e^2. Because we have the constraint that $2\sigma_a^2 + \sigma_e^2 = 1$ (see (3.6) in Section 3.1.2), we can define a posterior distribution for one variance term, obtain posterior samples from this distribution, and directly obtain the associated value of the other. In this case, we choose to obtain posterior samples of σ_e^2, so as to exploit all information available in Tables 3.5, 3.6, and 3.7. The value of $\sigma_a^2 := (1 - \sigma_e^2)/2$ then follows directly.

To define the posterior distribution of σ_e^2, we begin by collecting all sums of squares terms defined in Tables 3.5, 3.6, and 3.7 to capitalize on all information related to the value of σ_e^2. We have that

$$\begin{aligned}
\frac{SS_{W_{11}}}{(n_0-2)\sigma_a^2+\sigma_e^2} &= \frac{SS_{W_{11}}}{(n_0-2)\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2} \sim \chi^2_{df=n_0-1} & \frac{SS_{E_{11}}}{\sigma_e^2} &\sim \chi^2_{df=N_{11}-n_0}\\
\frac{SS_{W_{22}}}{(n_0-2)\sigma_a^2+\sigma_e^2} &= \frac{SS_{W_{22}}}{(n_0-2)\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2} \sim \chi^2_{df=n_0-1} & \frac{SS_{E_{22}}}{\sigma_e^2} &\sim \chi^2_{df=N_{22}-n_0}\\
\frac{SS_{W_{12}}}{n_0\sigma_a^2+\sigma_e^2} &= \frac{SS_{W_{12}}}{n_0\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2} \sim \chi^2_{df=2n_0-2} & \frac{SS_{E_{12}}}{\sigma_e^2} &\sim \chi^2_{df=(n_0-1)^2},
\end{aligned}$$

and so we define

$$\begin{aligned}
MS_e &= \frac{SS_{W_{11}}}{C_2}+\frac{SS_{W_{22}}}{C_2}+\frac{SS_{W_{12}}}{C_1}+\frac{SS_{E_{11}}}{C_3}+\frac{SS_{E_{22}}}{C_3}+\frac{SS_{E_{12}}}{C_3}\\
&\sim \chi^2_{df=(n_0-1)+(n_0-1)+(2n_0-2)+(N_{11}-n_0)+(N_{22}-n_0)+(n_0-1)^2},
\end{aligned} \tag{3.27}$$

where

$$C_1 = n_0\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2 \qquad C_2 = (n_0-2)\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2 \qquad C_3 = \sigma_e^2.$$

We can simplify (3.27) by considering a common denominator, such that

$$\begin{aligned}
MS_e &= \frac{C_1C_3\left(SS_{W_{11}}+SS_{W_{22}}\right)+C_2C_3\left(SS_{W_{12}}\right)+C_1C_2\left(SS_{E_{11}}+SS_{E_{22}}+SS_{E_{12}}\right)}{C_1C_2C_3}\\
&\sim \chi^2_{df=\binom{2n_0}{2}-3}.
\end{aligned}$$

We now define the posterior distribution for the variance term, σ_e^2 by considering a χ^2 likelihood for the MS_e term, and assuming a Beta prior (since we have the constraint that $\sigma_e^2 \leq 1$), such that

$$\pi\left(\sigma_e^2|MS_e,\alpha_e,\beta_e\right) \propto \chi^2\left(MS_e|\boldsymbol{\sigma},\sigma_e^2,\alpha_e,\beta_e\right)\mathcal{B}\left(\sigma_e^2|\alpha_e,\beta_e\right). \tag{3.28}$$

where the dependence on $\boldsymbol{\sigma} := \{\sigma_{ii'}\}_{ii'}$ in (3.28) is inherent in the construction of MS_e as a sum of the various sum of squares terms defined in Section 3.2.

Next, we assign the posterior distributions for the mean parameters, θ_{11}, θ_{22}, and

θ_{12}, by considering Multivariate Normal likelihoods and assuming Normal priors, with mean $\phi_{ii'}$ and variance $\omega_{ii'}$, such that

$$\pi\left(\theta_{11}|\boldsymbol{s}_{11},\sigma_{11},\sigma_a^2,\sigma_e^2,\phi_{11},\omega_{11}\right) \propto \mathcal{MVN}\left(\boldsymbol{s}_{11}|\theta_{11},\sigma_{11},\sigma_a^2,\sigma_e^2,\phi_{11},\omega_{11}\right)\mathcal{N}\left(\theta_{11}|\phi_{11},\omega_{11}\right); \quad (3.29)$$

$$\pi\left(\theta_{22}|\boldsymbol{s}_{22},\sigma_{22},\sigma_a^2,\sigma_e^2,\phi_{22},\omega_{22}\right) \propto \mathcal{MVN}\left(\boldsymbol{s}_{22}|\theta_{22},\sigma_{22},\sigma_a^2,\sigma_e^2,\phi_{22},\omega_{22}\right)\mathcal{N}\left(\theta_{22}|\phi_{22},\omega_{22}\right); \quad (3.30)$$

$$\pi\left(\theta_{12}|\boldsymbol{s}_{12},\sigma_{12},\sigma_a^2,\sigma_e^2,\phi_{12},\omega_{12}\right) \propto \mathcal{MVN}\left(\boldsymbol{s}_{12}|\theta_{12},\sigma_{12},\sigma_a^2,\sigma_e^2,\phi_{12},\omega_{12}\right)\mathcal{N}\left(\theta_{12}|\phi_{12},\omega_{12}\right), \quad (3.31)$$

where the resulting posterior distributions of equations (3.29), (3.30), and (3.31) are Normally distributed. The parameters of the posterior distribution of θ_{11} are given by

$$\mu_{11_p} = \frac{\mathbf{1}'_{N_{11}}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{s}_{11} + \phi_{11}}{\mathbf{1}'_{N_{11}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{1}_{N_{11}}\omega_{11} + 1} \qquad \sigma^2_{11_p} = \frac{\omega_{11}}{\mathbf{1}'_{N_{11}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{1}_{N_{11}}\omega_{11} + 1},$$

the parameters of the posterior distribution of θ_{22} are given by

$$\mu_{22_p} = \frac{\mathbf{1}'_{N_{22}}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{s}_{22} + \phi_{22}}{\mathbf{1}'_{N_{22}}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{1}_{N_{22}}\omega_{22} + 1} \qquad \sigma^2_{22_p} = \frac{\omega_{22}}{\mathbf{1}'_{N_{22}}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{1}_{N_{22}}\omega_{22} + 1},$$

and the parameters of the posterior distribution of θ_{12} are given by

$$\mu_{12_p} = \frac{\mathbf{1}'_{N_{12}}\boldsymbol{\Sigma}_{12}^{-1}\boldsymbol{s}_{12} + \phi_{12}}{\mathbf{1}'_{N_{12}}\boldsymbol{\Sigma}_{12}^{-1}\mathbf{1}_{N_{12}}\omega_{12} + 1} \qquad \sigma^2_{12_p} = \frac{\omega_{12}}{\mathbf{1}'_{N_{12}}\boldsymbol{\Sigma}_{12}^{-1}\mathbf{1}_{N_{12}}\omega_{12} + 1}.$$

Finally, we define the posterior distributions for the variance terms, σ^2_{11}, σ^2_{22}, and σ^2_{12}, by considering Multivariate Normal likelihoods and assuming Inverse-Gamma priors with hyperparameters $\alpha_{ii'}$ and $\beta_{ii'}$, such that

$$\pi(\sigma^2_{11}|\boldsymbol{s}_{11},\theta_{11},\sigma_a^2,\sigma_e^2,\alpha_{11},\beta_{11}) \propto \mathcal{MVN}(\boldsymbol{s}_{11}|\sigma_{11},\theta_{11},\sigma_a^2,\sigma_e^2,\alpha_{11},\beta_{11})\mathcal{IG}(\sigma^2_{11}|\alpha_{11},\beta_{11}); \quad (3.32)$$

$$\pi(\sigma^2_{22}|\boldsymbol{s}_{22},\theta_{22},\sigma_a^2,\sigma_e^2,\alpha_{22},\beta_{22}) \propto \mathcal{MVN}(\boldsymbol{s}_{22}|\sigma_{22},\theta_{22},\sigma_a^2,\sigma_e^2,\alpha_{22},\beta_{22})\mathcal{IG}(\sigma^2_{22}|\alpha_{22},\beta_{22}); \quad (3.33)$$

$$\pi(\sigma^2_{12}|\boldsymbol{s}_{12},\theta_{12},\sigma_a^2,\sigma_e^2,\alpha_{12},\beta_{12}) \propto \mathcal{MVN}(\boldsymbol{s}_{12}|\sigma_{12},\theta_{12},\sigma_a^2,\sigma_e^2,\alpha_{12},\beta_{12})\mathcal{IG}(\sigma^2_{12}|\alpha_{12},\beta_{12}), \quad (3.34)$$

where the resulting posterior distributions of (3.32), (3.33), and (3.34) are distributed according to Inverse Gamma distributions. The parameters of the posterior distribution of σ_{11}^2 are given by

$$\alpha_{11_p} = \frac{N_{11}}{2} + \alpha_{11} \qquad \beta_{11_p} = \frac{1}{2}\left(\boldsymbol{s}_{11} - \theta_{11}\mathbf{1}_{N_{11}}\right)' \boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{s}_{11} - \theta_{11}\mathbf{1}_{N_{11}}\right) + \beta_{11},$$

the parameters of the posterior distribution of σ_{22}^2 are given by

$$\alpha_{22_p} = \frac{N_{22}}{2} + \alpha_{22} \qquad \beta_{22_p} = \frac{1}{2}\left(\boldsymbol{s}_{22} - \theta_{22}\mathbf{1}_{N_{22}}\right)' \boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{s}_{22} - \theta_{22}\mathbf{1}_{N_{22}}\right) + \beta_{22},$$

and the parameters of the posterior distribution of σ_{12}^2 are given by

$$\alpha_{12_p} = \frac{N_{12}}{2} + \alpha_{12} \qquad \beta_{12_p} = \frac{1}{2}\left(\boldsymbol{s}_{12} - \theta_{12}\mathbf{1}_{N_{12}}\right)' \boldsymbol{\Sigma}_{12}^{-1}\left(\boldsymbol{s}_{12} - \theta_{12}\mathbf{1}_{N_{12}}\right) + \beta_{12}.$$

The equations given by (3.32), (3.33), and (3.34) provide us with samples from the posterior distribution of the variances, σ_{11}^2, σ_{22}^2, and σ_{12}^2. We, however, are interested in the standard deviation terms, σ_{11}, σ_{22}, and σ_{12}, and so we simply take the square root of the sampled variance terms to obtain samples of the standard deviation terms. The resulting inference is not affected.

Note that each of the distributions given in (3.28) through (3.33) depends on the value of at least one other parameter; therefore, we must rely on sampling techniques to study the distributions of the model parameters. We construct the following Gibbs sampler, described in Algorithm 1.

Algorithm 1: Gibbs sampler for generating posterior samples from the distributions of the model parameters

Data: Initial values for all parameters at $t = 0$; values for hyperparameters
Result: Posterior samples for all parameters
for $t \in 1 : T$ iterations **do**

1. Draw $\sigma_e^{2(t)}|\boldsymbol{s}, MS_e^{(t-1)}, \boldsymbol{\sigma}^{(t-1)}, \alpha_e, \beta_e$ from the distribution defined in (3.28);
2. Define $\sigma_a^{2(t)} = (1 - \sigma_e^{2(t)})/2$;
3. Draw $\theta_{11}^{(t)}|\boldsymbol{s}_{11}, \sigma_{11}^{(t-1)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \phi_{11}, \omega_{11}$ from the distribution defined in (3.29);
4. Draw $\theta_{22}^{(t)}|\boldsymbol{s}_{22}, \sigma_{22}^{(t-1)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \phi_{22}, \omega_{22}$ from the distribution defined in (3.30);
5. Draw $\theta_{12}^{(t)}|\boldsymbol{s}_{12}, \sigma_{12}^{(t-1)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \phi_{12}, \omega_{12}$ from the distribution defined in (3.31);
6. Draw $\sigma_{11}^{2(t)}|\boldsymbol{s}_{11}, \theta_{11}^{(t)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \alpha_{11}, \beta_{11}$ from the distribution defined in (3.32);
7. Draw $\sigma_{22}^{2(t)}|\boldsymbol{s}_{22}, \theta_{22}^{(t)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \alpha_{22}, \beta_{22}$ from the distribution defined in (3.33);
8. Draw $\sigma_{12}^{2(t)}|\boldsymbol{s}_{12}, \theta_{12}^{(t)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \alpha_{12}, \beta_{12}$ from the distribution defined in (3.34);

end

Because the distribution defined for σ_e^2 in (3.28) is not readily known, we cannot directly sample from this distribution. As a result, the first step in Algorithm 1 is not so straightforward - indeed, obtaining a sample from the posterior distribution of σ_e^2 requires introducing a Metropolis-Hastings algorithm [14, 44]. This procedure is summarized in Algorithm 2.

Algorithm 2: Metropolis-Hastings algorithm for obtaining a sample from the posterior distribution of σ_e^2

Data: Value of $\sigma_e^{2(t-1)}$; values for hyperparameters α_e, β_e
Result: Posterior sample of $\sigma_e^{2(t)}$

1. Sample a candidate value, $\sigma_e^{2*} \sim \mathcal{B}(2, \sigma_e^{-2(t-1)})$;
2. Calculate the value of MS_e^* using the candidate value σ_e^{2*} using (6.22);
3. Calculate the value of $MS_e^{(t-1)}$ using the current value $\sigma_e^{2(t-1)}$ (6.22);
4. Evaluate the density, f^*, at σ_e^{2*} using the hyperparameters α_e and β_e, the value of MS_e^*, and (3.28);
5. Evaluate the density, $f^{(t-1)}$, at $\sigma_e^{2(t-1)}$ using the hyperparameters α_e and β_e, the value of $MS_e^{(t-1)}$, and (3.28);
6. Calculate the probability of acceptance, $p_{acc} = \frac{f^*}{f^{(t-1)}} \frac{\mathcal{B}\left(\sigma^{2(t-1)}|2,\sigma^{-2*}\right)}{\mathcal{B}\left(\sigma^{2*}|2,\sigma^{-2(t-1)}\right)}$
7. Generate a random probability, $p^* \sim \mathcal{U}(0, 1)$;
8. If $p_{acc} \geq p^*$, then define $\sigma_e^{2(t)} := \sigma_e^{2*}$; otherwise define $\sigma_e^{2(t)} := \sigma_e^{2(t-1)}$;

Now that we have identified a method for obtaining samples of the parameters used to define $\boldsymbol{\theta}, \boldsymbol{\Delta}, \sigma_a^2$, and σ_e^2, we can assign the kernel Bayes factor,

$$\mathcal{K} = \frac{\int \ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_1\right)\pi\left(\boldsymbol{\Omega}_1\right)}{\int \ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_2\right)\pi\left(\boldsymbol{\Omega}_2\right)}, \tag{3.35}$$

$$\approx \frac{\frac{1}{T}\sum_{t=1}^{T}\ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_1^{(t)}\right)}{\frac{1}{T}\sum_{t'=1}^{T}\ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_2^{(t')}\right)} \tag{3.36}$$

where $\boldsymbol{s}_t$ is the vector of scores that consider at least one trace object, the subscripts on $\boldsymbol{\Omega}$ correspond to the model being considered, and $\boldsymbol{\Omega}^{(t)}$ are posterior samples of the parameters obtained using Algorithm 1. We consider the conditional log-likelihood of the scores that consider objects of unknown origin from the set $\boldsymbol{X}_u$, $\boldsymbol{s}_t$, rather than the joint log likelihood of $\boldsymbol{s}$, so as to not recycle the information contained in the scores $\boldsymbol{s}_c$, which are used to sample the parameter values.

By constructing the vector of scores such that (3.8) is satisfied, we can define $\kappa\left(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}\right)$ using virtually any type of kernel, so long as we adhere to the constraints outlined in Section 3.1.

3.4 RECOVERING THE MODEL PARAMETERS

We now move to study the ability of the Gibbs sampler defined in Algorithm 1 to recover the true model parameters in an ideal situation. We consider the performance of the model when objects are generated using the object model and the associated scores are calculated using the non-stationary kernel (product of pairs of objects) given in (3.5), and when scores are sampled directly from the proposed model given in (3.8).

We begin by considering the scenario in which scores are obtained by calculating the product of pairs of objects generated from the object model described in Section 3.1.1. Doing so requires defining two mean terms, μ_1 and μ_2, and two variance terms σ^2 and τ^2, associated with the error terms, e_{1j} and e_{2j} in (3.3). As an example, we consider $\mu_1 = 3$, $\mu_2 = 5$, $\sigma^2 = 0.15$, and $\tau^2 = 0.23$. From Table 3.2 we have that $\mu_{11} = 9$, $\mu_{22} = 25$, $\mu_{12} = 15$, $\sigma_{11}^2 = 2.72$ and $\sigma_{22}^2 = 11.55$.

We sample objects according to (3.3) to obtain 100 vectors of scores using (3.5).

Then, using the model proposed above, we attempt to recover the model parameters (μ_{11}, μ_{22}, μ_{12}, σ_{11}, and σ_{22}). Table 3.8 (column 1) shows the model's ability to recover the model parameters in this scenario.

We also consider the scenario in which the scores are sampled directly according to (3.8). In this case, we directly build the mean vector and the covariance matrix by using the values of the parameters μ_{11}, μ_{22}, μ_{12}, σ_{11} and σ_{22} reported in the previous paragraph. Note that we do not include a "true" value for σ_{12}, given the discrepancies described in Section 3.1.2. Instead, we use the parameters recovered for the object model to reasonably define $\sigma_{12}^2 = 3$. After defining these parameters, we directly sample 1000 score vectors from the model given in (3.8). The results of these experiments are presented below in Table 3.8 (column 2).

Parameter	n_0	Object Model	Proposed Model
$\mu_{11} = 9$	5	9.00	8.94
	10	8.95	8.97
	15	8.96	9.00
$\mu_{22} = 25$	5	25.09	25.01
	10	24.94	25.27
	15	25.05	25.08
$\mu_{12} = 15$	5	15.02	14.95
	10	14.94	15.04
	15	14.98	15.02
$\sigma_{11}^2 = 2.72$	5	2.44	2.72
	10	3.04	2.68
	15	3.26	2.78
$\sigma_{22}^2 = 11.55$	5	8.93	11.52
	10	8.51	11.00
	15	8.63	11.86

Table 3.8: Parameters recovered for the object model (see Section 3.1.1) and the proposed model (3.8) for $n_0 = 5, 10,$ and 15 control objects.

Table 3.8 demonstrates that the parameters are better recovered when the scores are directly generated from the model given by (3.8), than when they are generated from the object model. This is to be expected. This is particularly true when we consider the different variance terms, σ_{11}^2 and σ_{22}^2 (overall, there does not seem to be any issues with the model's ability to recover the mean parameters, μ_{11}, μ_{22}, and μ_{12}, under either scenario). There are two items worth noting when we recover the parameters for scores generated using the object model. First, the initial variances defined by σ and τ do influence the ability of the model to accurately recover the parameters. That is, when there exists large variation within a group of objects, the model does experience more difficulties in recovering the variance terms σ_{11}^2 and σ_{22}^2.

Chapter 4

IMPLEMENTING THE TWO-CLASS MODEL-SELECTION ALGORITHM

4.1 CONSIDERING THE MNIST HANDWRITTEN DIGIT DATA

In this section, we apply the proposed model to the pervasive MNIST Handwritten Digit Data [39] to obtain some preliminary results. The MNIST Handwritten Digit Data consists in approximately 70,000 observations of handwritten digits. Each observation is a 28×28 pixel image of an integer, 0 through 9.

We analyze the performance of our model and compare its performance to that of the classic Support Vector Machine (SVM) (we use the $ksvm()$ function from the $kernlab$ package [38]), typically used for binary classification [12, 65, 74, 89, 90]. The methods are inherently similar in that they both rely on calculating the inner products between vectors in some feature space, and reduce the dimension of the information that is stored and used to make a decision. However, while an SVM uses optimization to characterize an *optimal hyperplane* in some high-dimensional feature space using a set of *support vectors*, and classifies new observations according to which side of the hyperplane they lie on (see Section 1.1.1), the proposed model directly studies the parameters of the Multivariate Normal distribution that characterizes the vector of scores, and evaluates a Bayes Factor that compares the likelihood of observing the vector of trace scores under each of the considered models, $\mathcal{H}_1$ and $\mathcal{H}_2$. Finally, the SVM considers some subset of the original vectors (i.e., the *support vectors*) to classify a new point, while the proposed model considers a set of only 8 parameters, regardless of the dimension, type, or quantity of data considered.

According to the literature, the digit pairs 3 and 5, 2 and 6, and 7 and 9 are, typically, the most difficult for the SVM to distinguish. Thus, we have elected to compare the abilities of the proposed model and an SVM to differentiate between the digits 3 and 5. For each model, we consider a radial basis function kernel, given that this kernel has been shown to work well for SVMs on this data set.

4.1.1 ASSESSING MODEL PERFORMANCE

To assess the performance of the model, we consider a series of simulations in which we consider $n_0 = 5$, 10, and 15 control objects per source, and $n_u = 3$ trace objects. For a series of simulations, we consider the performance of the models when a fixed set of control objects is considered alongside 200 sets of trace objects (100 from each source). That is, for a single iteration, we sample n_0 control objects from sources 1 and 2 (digits 3 and 5), and 100 sets of n_u trace objects from each source, for a total of 200 sets of trace objects. The two sets of n_0 control objects are used to determine the source of the 200 sets of trace objects. This process is repeated 100 times, and the average performance is assessed. The results are presented in Table 4.1.

n_0	SVM (Overall Performance)	SVM (Triplet Performance)	SVM (Voting Performance)	Proposed Model Performance
5	73.18%	48.22%	74.90%	68.22%
10	78.34%	55.85%	85.68%	81.87%
15	83.50%	62.48%	89.32%	87.35%

Table 4.1: Performance of SVM versus two-class Model when $n_0 = 5, 10$, and 15 control objects and $n_u = 3$ trace objects for 100 iterations of the experiment. The overall performance of the SVM gives the average percentage of correct classifications. The triplet performance of the SVM gives the average percentage of sets of n_u that were correctly classified in their entirety. The voting performance of the SVM gives the percentage of sets of n_u that would be correctly classified if a voting system were used to determine the class of the set of trace objects. The performance of the Proposed model gives the average percentage of sets of n_u that were correctly classified.

The results in Table 4.1 present the results of the experiment described above. We see that, as the number of control objects increases, the rates of correct classification increase across all columns. Notably, Table 4.1 indicates that when we consider $n_0 = 5$ control objects, the overall performance of the SVM slightly out-performs the performance of the proposed model. However, when we compare the triplet performance of the SVM to that of the proposed model, we see that the proposed model drastically out-performs the SVM. When we move to consider $n_0 = 10$ and 15 control objects, the proposed model outperforms the SVM both in terms of overall performance, and in terms of the triplet performance. We do note, however, that when a voting system is used to classify the set of trace objects (For example, suppose that the SVM classifies the three trace objects as 3, 3, 5. Under the voting system, the entire set of trace objects would be assigned class 3. Likewise, should the SVM classify the three trace objects as 5, 3, 5, the entire set of trace objects would be assigned class 5), we see a more comparable performance. In particular, we see that the difference between the voting performance of the SVM and the performance of the proposed model decreases as we move from $n_0 = 5$ control objects to $n_0 = 15$ control objects.

Finally, it is worth mentioning that the computational cost associated with each model is different. Overall, the SVM is more computationally efficient. At this time, we cannot directly compare the two models: first, the SVM package used in this experiment is coded in C, while the proposed model is coded in R; second, the SVM is not Bayesian, while the proposed model is, and so some inherent computational costs exist. That being said, if a Bayesian alternative of the SVM were considered alongside the proposed model, we can speculate that the SVM would remain more computationally efficient, since the proposed model involves inverting a covariance matrix, which is a step that is not required by the SVM.

Chapter 5

DISCUSSION ON THE TWO-CLASS MODEL SELECTION ALGORITHM

In this part, we considered the development for a two-class classification algorithm that allows for making inference on the source of a set of test objects known to originate from one of two potential sources. This method is novel in that it allows for classifying the complete set of objects at once, rather than classifying each object in turn. This method relies on a kernel function, which allows for considering virtually any set of high-dimensional, complex, heterogeneous data as a single vector of real-valued scores between observations by merely modifying the kernel to accommodate the considered data. In addition, our method is particularly well-suited for scenarios in which a limited number of observations are available for consideration.

An evaluation of the performance of the proposed model indicates that the model performs well when as few as 5 control objects are considered for each source. As we increase the number of control objects per source, the model's performance continues to improve, though at an increased computational cost. Operationally speaking, there is not much benefit from considering more than 10 control objects per source in that the prediction ability of the model does not significantly improve in considering more control objects, and that the computational cost of the algorithm remains manageable at $n_0 = 10$ control objects. Finally, comparing the performance of the proposed model to that of the traditional SVM indicates that the proposed method has superior classification ability. This performance indicates that the model works well in the two class scenario, and that extension to the multi-class scenario is reasonable.

Part III

A Multi-Class Model-Selection Algorithm for High-Dimensional and Complex Data

OVERVIEW OF PART III: A MULTI-CLASS MODEL SELECTION ALGORITHM FOR HIGH-DIMENSIONAL AND COMPLEX DATA

In this part, we develop the theory and implementation for an n-class kernel-based model-selection algorithm. In Chapter 6, we define the problem and develop the algorithm that allows for determining which of n classes is more likely to have produced a set of trace objects. In addition, we propose a method for studying the parameters of the proposed model, and a sampling algorithm that can be used to study the distributions of the considered parameters.

In Chapter 7, we implement the proposed model on the MNIST hand-written digits data that is commonly used to evaluate the performance of pattern recognition algorithms.

In Chapter 8 we discuss the benefits and limitations of the proposed n-class model.

Chapter 6

DEFINING THE MULTI-CLASS MODEL-SELECTION PROBLEM

We move to extend the results of the two-class model-selection algorithm to propose an n-class model selection algorithm that allows for simultaneously determining the class of a set of objects. Given n sets of n_0 objects, $\{\boldsymbol{X}_i\}_{i=1}^n$, where each set of objects is known to have originated from n distinct sources, and a set of n_u objects, $\boldsymbol{X}_u$, known to have originated from one of the n sources that produced the objects observed in $\{\boldsymbol{X}_i\}_{i=1}^n$, we are interested in determining which of the n sources is most likely to have generated the set of objects observed in $\boldsymbol{X}_u$. Formally, we are interested in determining if

$$\begin{aligned}
\mathcal{H}_1 &: \boldsymbol{X}_u \text{ is a simple random sample from the source that produced } \boldsymbol{X}_1;\\
\mathcal{H}_2 &: \boldsymbol{X}_u \text{ is a simple random sample from the source that produced } \boldsymbol{X}_2;\\
&\vdots\\
\mathcal{H}_n &: \boldsymbol{X}_u \text{ is a simple random sample from the source that produced } \boldsymbol{X}_n.
\end{aligned}$$

As discussed in Chapter 3, differentiating between these propositions cannot be reduced down to a simple classification or model-selection problem that can be addressed using machine learning or likelihood-based techniques. As before, small sample sizes rule out many machine learning techniques, and high-dimensional, complex, or heterogenous data make it impossible to assign the necessary probability measures for assigning Bayes factors or performing likelihood-based inference.

We propose a model that leverages the properties of kernel functions (see Chapter 1) to obtain a vector of scores, $\boldsymbol{s}$, that characterizes pairwise comparisons of all objects observed in $\{\boldsymbol{X}_i\}_{i=1}^n$ and $\boldsymbol{X}_u$. This vector consists of within-source scores, which arise when compared objects originate from a common source, and between-source scores, which arise when compared objects originate from different sources. The model capitalizes on the variability that exists within and between these sets of scores to address the above inference question. Because the method relies on a kernel function, the method can be tailored to any type of data by merely modifying this function, and the overall inference process remains the same. Furthermore, the model relies only one assumption, which can be satisfied through the design of the kernel function.

6.1 PROBLEM STATEMENT

Consider n sets of exchangeable observations, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$, made on n distinguishable sets of objects, and the set of exchangeable observations, $\boldsymbol{X}_u$, made on objects of common but unknown origin. The sets $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are considered to be sets of control objects, while the set $\boldsymbol{X}_u$ is considered to be a set of test objects. We define the sets $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ and $\boldsymbol{X}_u$ as being simple random samples,

$$
\begin{aligned}
\boldsymbol{X}_1 &:= \{\boldsymbol{x}_{1,1}, \boldsymbol{x}_{1,2}, \ldots, \boldsymbol{x}_{1,n_0}\}, \\
\boldsymbol{X}_2 &:= \{\boldsymbol{x}_{2,1}, \boldsymbol{x}_{2,2}, \ldots, \boldsymbol{x}_{2,n_0}\}, \\
&\vdots \\
\boldsymbol{X}_n &:= \{\boldsymbol{x}_{n,1}, \boldsymbol{x}_{n,2}, \ldots, \boldsymbol{x}_{n,n_0}\}, \\
\boldsymbol{X}_u &:= \{\boldsymbol{x}_{u,1}, \boldsymbol{x}_{u,2}, \ldots, \boldsymbol{x}_{1,n_u}\},
\end{aligned}
$$

where the sets of control objects consist of n_0 objects from their respective sources, and the set of test objects consists in n_u objects known to originate from one of the n sources represented by the observations in $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, or $\boldsymbol{X}_n$. We are interested in quantifying the extent of support provided to $\mathcal{H}_1, \ldots, \mathcal{H}_n$ above.

Rather than consider the observations themselves, we instead consider the vector of all pairwise *scores*, $\boldsymbol{s} \in \mathbb{R}^N$, $N = \binom{nn_0+n_u}{2}$, obtained by comparing the m-

dimensional observations in the sets $\{\boldsymbol{X}_i\}_{i=1}^{n}$ and $\boldsymbol{X}_u$ via some kernel,

$$\kappa : \mathbb{R}^m \mapsto \mathbb{R},$$

$$\kappa\left(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}\right) = \left\langle \phi\left(\boldsymbol{x}_{ij}\right), \phi\left(\boldsymbol{x}_{i'j'}\right)\right\rangle \quad i, i' \in \{1, \ldots, n, u\},\ j, j' \in \{1,\ \ldots,\ \max\{n_0, n_u\}\},$$

where ϕ is a mapping into some separable, high-dimensional Hilbert space [12, 65, 71, 74]. As before, $s_{ij,i'j'}$ is the score obtained by comparing object $\boldsymbol{x}_{ij}$ to object $\boldsymbol{x}_{i'j'}$ using some kernel function κ (see Definition 2 in Section 1.1).

We define our kernel function such that our vector of scores is distributed according to a Multivariate Normal distribution, with

$$\boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right) \tag{6.1}$$

where $\boldsymbol{\theta}$ is the vector of the mean terms, and $\boldsymbol{\Sigma}$ is the covariance matrix associated with the vector of scores (see Section 9.5 and Chapter 10 for a discussion on the validity of this assumption, and the implications when this assumption does not hold). These parameters will collectively be referred to as $\boldsymbol{\Omega} := \{\boldsymbol{\theta}, \boldsymbol{\Sigma}\}$, and we will more explicitly define $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ as we move through the chapter. We can assign the posterior probability that the set of trace objects $\boldsymbol{X}_u$ was generated by the source characterized by the objects in the set $\boldsymbol{X}_i$ in terms of the Multivariate Normal Likelihood and the associated parameter $\boldsymbol{\Omega}$, such that

$$\begin{aligned}
\pi\left(\mathcal{H}_i|\boldsymbol{s}, \boldsymbol{\Omega}\right) &= \frac{\pi\left(\mathcal{H}_i\right)\int_{\boldsymbol{\Omega}} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_i\right) d\Pi\left(\boldsymbol{\Omega}\right)}{\sum_{i'=1}^{n} \pi\left(\mathcal{H}_{i'}\right)\int_{\boldsymbol{\Omega}} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_{i'}\right) d\Pi\left(\boldsymbol{\Omega}\right)} \\
&= \frac{\pi\left(\mathcal{H}_i\right)\int_{\boldsymbol{\Omega}_i} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}_i\right) d\Pi\left(\boldsymbol{\Omega}_i\right)}{\sum_{i'=1}^{n} \pi\left(\mathcal{H}_{i'}\right)\int_{\boldsymbol{\Omega}_{i'}} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}_{i'}\right) d\Pi\left(\boldsymbol{\Omega}_{i'}\right)}
\end{aligned} \tag{6.2}$$

It is worth noting that there exist differences between $\boldsymbol{\Omega}_i$, $i = \{1, \ldots, n\}$. While the individual elements of each of the parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\Sigma}_i$ are restricted to the same set of potential values, the structures of the mean vectors and covariance matrices depend on which $\mathcal{H}_i$ is being considered. That is, when $\mathcal{H}_i$ is being considered, then the parameter $\boldsymbol{\Omega}_i$ considers that the set of unknown objects $\boldsymbol{X}_u$ originates from the source of the objects contained in $\boldsymbol{X}_i$. Then, scores of the form $s_{uj,uj'}$, and $s_{uj,ij'}$ will all be considered as "within-source" scores, and scores $s_{uj,i'j'}$ will be considered as "between-source" scores.

6.1.1 COVARIANCE STRUCTURE FOR THE OBJECT MODEL

As before, we can investigate the covariance structure for a vector of scores by considering a univariate object-based model. Consider an object x_{ij} defined in terms of the linear model given by

$$x_{ij} = \mu_i + e_{ij}, \tag{6.3}$$

where μ_i is the overall mean of class $i \in \{1, 2, \ldots, n\}$, and $e_{ij} \sim \mathcal{N}(0, \sigma_i^2)$. Scores can be studied by choosing the stationary and non-stationary kernels (first presented in Section 3.1.1 as (3.4) and (3.5)), given by $s'_{ij,i'j'}$ and $s^*_{ij,i'j'}$, respectively:

$$s'_{ij,i'j'} \quad := \quad (x_{ij} - x_{i'j'})^2 \tag{6.4}$$

$$s^*_{ij,i'j'} \quad := \quad x_{ij}x_{i'j'}. \tag{6.5}$$

As before, we can directly examine the mean and covariance terms associated with these two kernels by calculating the various terms that arise from the different possible score combinations. Table 6.1 provides the different parameter values under the two kernel functions, (6.4) and (6.5) given above.

Table 6.1 indicates that the covariance structure varies depending on whether a stationary or non-stationary kernel is used to obtain the vector of scores, $\boldsymbol{s}$. For example, we see that there are five unique covariance scenarios that can occur when a stationary kernel is considered, versus six unique covariance scenarios that can occur when a non-stationary kernel is considered. Furthermore, we have that some of the terms associated with the stationary kernel are relatively straightforward functions of each other (e.g., $2\sigma_i^4$ is a fraction of $8\sigma_i^4$). In addition, the covariance terms that arise when a non-stationary kernel is considered depend much more on the means of the different sources.

Finally, we note that the zeros occur in the same positions for the stationary and non-stationary kernels.

Description of Considered Scores	Score 1	Score 2	Stationary Kernel (6.4)	Non-Stationary Kernel (6.5)
Expected Value Terms				
Within Source	$i1i2$	–	$2\sigma_i^2$	μ_i^2
Between Source	$i1i'1$	–	$(\mu_i - \mu_{i'})^2 + \sigma_i^2 + \tau_i^2$	$\mu_i \mu_{i'}$
Covariance Terms				
Both Within Source, Two Common Objects	$i1i2$	$i1i2$	$8\sigma_i^4$	$2\mu_i^2\sigma_i^2 + \sigma_i^4$
Both Within Source, One Common Object	$i1i2$	$i1i3$	$2\sigma_i^4$	$\mu_i^2\sigma_i^2$
Both Within Source, No Common Objects	$i1i2$	$i3i4$	0	0
Both Between Source, Two Common Objects	$i1i'1$	$i1i'1$	$4\left(\mu_i - \mu_{i'}\right)^2\left(\sigma_i^2 + \sigma_{i'}^2\right) + 2\left(\sigma_i^2 + \sigma_{i'}^2\right)^2$	$\mu_i^2\sigma_{i'}^2 + \mu_{i'}^2\sigma_i^2 + \sigma_i^2\sigma_{i'}^2$
Both Between Source, Two Common Sources, One Common Object from Source i	$i1i'1$	$i1i'2$	$4\left(\mu_i - \mu_{i'}\right)^2\sigma_i^2 + 2\sigma_i^4$	$\mu_{i'}^2\sigma_i^2$
Both Between Source, One Common Source, One Common Object from Source i	$i1i'1$	$i1i''1$	$4(\mu_i - \mu_{i'})(\mu_i - \mu_{i''})\sigma_i^2 + 2\sigma_i^4$	$\mu_{i'}\mu_{i''}\sigma_i^2$
Both Between Source, No Common Objects	$i1i'1$	$i2i'2$	0	0
Both Between Source, No Common Sources	$i1i'1$	$i''1i'''1$	0	0
Within Source i, Between Source, One Common Object from Source i	$i1i2$	$i1i'1$	$2\sigma_i^4$	$\mu_i\mu_{i'}\sigma_i^2$
Within Source i, Between Source, No Common Objects	$i1i2$	$i3i'1$	0	0
Within Source i, Between Source, No Common Sources	$i1i2$	$i'1i''1$	0	0
Within Source i, Within Source i', No Common Objects	$i1i2$	$i'1i'2$	0	0

Table 6.1: Expected value and covariance terms obtained for the object-based model for each type of score comparison when a stationary kernel (e.g., (6.4)) and non-stationary kernel (e.g., (6.5)) are considered. Column one provides descriptions of each type of comparison that may be observed; columns two and three provide examples of indices for scores that could be compared in each situation; columns four and five present the parameter values obtained under the stationary and non-stationary kernels given by (6.4) and (6.5).

6.1.2 DEFINING A SCORE-BASED MODEL

Suppose, now, that we expand upon (6.1) and define our vector of scores such that

$$\begin{aligned} \boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right) \quad &\Longrightarrow \quad s_{ij,i'j'} \sim \mathcal{N}\left(\theta_{ii'}, \sigma_{ii'}^2\right) \\ &\Longrightarrow \quad \frac{s_{ij,i'j'} - \theta_{ii'}}{\sigma_{ii'}} \sim \mathcal{N}\left(0, 1\right) \end{aligned} \tag{6.6}$$

where the parameters $\theta_{ii'}$ and $\sigma_{ii'}$ are the means and variances associated with the different comparisons that may be considered by a given score. That is, each score in the vector of scores is either a within-source i comparison, or a between source i and i'

comparison. For example, consider a score $s_{ij,i'j'}$ in which $i = i'$, and so $\theta_{ii'}$ gives us the expected value of scores that compare any two objects in $\boldsymbol{X}_i$. Likewise, when $i \neq i'$, $\theta_{ii'}$ gives us the expected value of the scores that compare an object in $\boldsymbol{X}_i$ to an object in $\boldsymbol{X}_{i'}$. The parameter $\sigma_{ii'}$ can be similarly defined for the standard deviation terms.

From here, following the work of Gantz and Saunders, and Armstrong [5, 6, 26], and the development in Section 3.1.2, we choose to define the standardized scores from (6.6) according to a random effects model

$$\frac{s_{ij,i'j'} - \theta_{ii'}}{\sigma_{ii'}} = a_{ij} + a_{i'j'} + \varepsilon_{ij,i'j'}, \tag{6.7}$$

where, as before, a_{ij} and $a_{i'j'}$ are random effects such that a_{ij}, $a_{i'j'} \sim \mathcal{N}\left(0, \sigma_a^2\right)$, and $\varepsilon_{ij,i'j'}$ is a lack-of-fit term, such that $\varepsilon_{ij,i'j'} \sim \mathcal{N}\left(0, \sigma_e^2\right)$. Furthermore, from (6.6), we have that $2\sigma_a^2 + \sigma_e^2 = 1$. Finally, we rewrite the model in terms of $s_{ij,i'j'}$, such that

$$s_{ij,i'j'} = \theta_{ii'} + \sigma_{ii'}\left(a_{ij} + a_{i'j'} + \varepsilon_{ij,i'j'}\right),$$

and so, given the distributional assumptions associated with (6.6) and (6.7), we define the distribution of our vector of scores to be

$$\boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Delta}\left(\boldsymbol{P}\boldsymbol{P}'\sigma_a^2 + \boldsymbol{I}\sigma_e^2\right)\boldsymbol{\Delta}'\right), \tag{6.8}$$

where $\boldsymbol{\theta}$ is a vector of length N of the mean terms given by $\theta_{ii'}$, and $\boldsymbol{\Delta}$ is an $N \times N$ diagonal matrix of the standard deviation terms given by $\sigma_{ii'}$. The design matrix $\boldsymbol{P}$ describes the effects of the objects being compared for each score considered in the vector $\boldsymbol{s}$. As before, for each of the $nn_0 + n_u$ rows of $\boldsymbol{P}$, a one is placed in the columns associated with the labels of the objects being compared in that row, and zeros are placed in the remaining columns.

The likelihood function in the numerator and denominator of (6.2) can be represented using the distribution given in (6.8). As explained in the introduction of Section 6.1, we have that the structure of the mean vector and covariance matrix depend on which $\mathcal{H}_i$ is being considered. See Figure 3.1 for an example when $n = 2$ classes.

Description of Considered Scores	Score 1	Score 2	Object Model (6.5)	Score Model (6.8)
Covariance Terms				
Both Within Source, Two Common Objects	$i1i2$	$i1i2$	$2\mu_i^2\sigma_i^2 + \sigma_i^4$	$\sigma_{ii}^2\left(2\sigma_a^2 + \sigma_e^2\right)$
Both Within Source, One Common Object	$i1i2$	$i1i3$	$\mu_i^2\sigma_i^2$	$\sigma_{ii}^2\sigma_a^2$
Both Within Source, No Common Objects	$i1i2$	$i3i4$	0	0
Both Between Source, Two Common Objects	$i1i'1$	$i1i'1$	$\mu_i^2\sigma_{i'}^2 + \mu_{i'}^2\sigma_i^2 + \sigma_i^2\sigma_{i'}^2$	$\sigma_{ii'}^2\left(2\sigma_a^2 + \sigma_e^2\right)$
Both Between Source, Two Common Sources, One Common Object from Source i	$i1i'1$	$i1i'2$	$\mu_{i'}^2\sigma_i^2$	$\sigma_{ii'}^2\sigma_a^2$
Both Between Source, One Common Source, One Common Object from Source i	$i1i'1$	$i1i''2$	$\mu_{i'}\mu_{i''}\sigma_i^2$	$\sigma_{ii'}\sigma_{ii''}\sigma_a^2$
Both Between Source, No Common Objects	$i1i'1$	$i2i'2$	0	0
Both Between Source, No Common Sources	$i1i'1$	$i''1i'''1$	0	0
Within Source i, Between Source, One Common Object from Source i	$i1i2$	$i1i'1$	$\mu_i\mu_{i'}\sigma_i^2$	$\sigma_{ii'}\sigma_{ii}\sigma_a^2$
Within Source i, Between Source, No Common Objects	$i1i2$	$i3i'1$	0	0
Within Source i, Between Source, No Common Sources	$i1i2$	$i'1i''1$	0	0
Within Source i, Within Source i', No Common Objects	$i1i2$	$i'1i'2$	0	0

Table 6.2: Comparison of Covariance terms in Object Model defined according to (6.5), and Score Model defined according to (6.8).

As in the two class model, it is worth noting that the covariance matrix defined in (6.8) is not equivalent to that of the object model described in Section 6.1.1. This is due to the fact that the covariance matrix in Section 6.1.1 considers a single term $\sigma_{ii'}$ to describe the relationship that occurs when a score involves an object from source i and an object from source i'. For the covariance matrix in (6.8) to coincide with that defined in Section 6.1.1, we would need to define two terms, $\sigma_{ii'}$ and $\sigma_{i'i}$, that describe the effect when the object in common between two scores comes from source i versus from source i'. For example, consider a pair of scores $s_{i1,i2}$ and $s_{i1,i'1}$. To appropriately capture the covariance that exists between these two scores would require defining a term $\sigma_{ii'}$, since the common object between the scores comes from source i. Likewise, a pair of scores, $s_{i1,i'1}$ and $s_{i'1,i'2}$, would require defining a term $\sigma_{i'i}$, since the common object between the scores comes from source i'. Note that such a pair of standard deviation terms would need to be defined for each of the $\binom{n}{2}$ possible combinations of sources. As a result, the covariance terms of the score model in rows 4-6 and 9 of Table 6.2 do not necessarily have a direct counterpart in the object model.

However, despite these discrepancies, we choose to move forward with the model given by (6.8). While the covariance matrices of the object and score models may not be exactly the same, their structures under each $\mathcal{H}_i$ remain sufficiently similar. Furthermore, as we will see below, an elegant solution exists for studying the parameters of the model given by (6.8).

6.2 MODEL DEVELOPMENT

Assigning the posterior probability in (6.2) requires estimating the parameters $\{\theta_{ii'}\}_{ii'}$, $\{\sigma_{ii'}\}_{ii'}$, σ_a^2, and σ_e^2 using the information contained in the vector of scores $\boldsymbol{s}$. To study these parameters, we follow the development described in Part II, and subset the vector of scores to define $\boldsymbol{s}_c$, which includes only the comparisons between the control objects contained in the sets $\{\boldsymbol{X}_i\}_{i=1}^n$, and so is a vector of length $N_c = \binom{nn_0}{2}$. We can then use $\boldsymbol{s}_c$ to define the total sum of squares

$$\begin{aligned} SS_{Tot} &= (\boldsymbol{s}_c - \boldsymbol{\theta}_c)' \left[\boldsymbol{\Delta}_c \boldsymbol{\Delta}_c'\right]^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c) \\ &= \left(\boldsymbol{\Delta}_c^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c)\right)' \left(\boldsymbol{\Delta}_c^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c)\right) \end{aligned} \tag{6.9}$$

where $\boldsymbol{\theta}_c$ is the N_c vector of score means, $\theta_{ii'}$, and $\boldsymbol{\Delta}_c$ is the $N_c \times N_c$ diagonal matrix of the score standard deviations, $\sigma_{ii'}$, associated with the scores $\boldsymbol{s}_c$.

Following the development in Section 3.2, we apply Cochran's theorem (see Theorem 2 in Section 3.2) to $\widetilde{\boldsymbol{s}} = (\boldsymbol{\Delta}_c^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c))$, and rewrite (6.9) as

$$SS_{Tot} = \widetilde{\boldsymbol{s}}' \boldsymbol{I} \widetilde{\boldsymbol{s}} = \widetilde{\boldsymbol{s}}' \left[\sum_{l=1}^{N_c} \boldsymbol{v}_l \boldsymbol{v}_l'\right] \widetilde{\boldsymbol{s}} \tag{6.10}$$

where $\{\boldsymbol{v}_l\}_l$, $l = 1, \ldots, N_c$ is any orthonormal basis for $\mathbb{R}^{N_c}$. Furthermore, we consider a set of $n + \binom{n}{2}$ diagonal design matrices, $\boldsymbol{B}_{ii'}, i, i' \in \{1, \ldots, n\}$. Each of these matrices is an idempotent $N_c \times N_c$ matrix whose diagonal matrix can be partitioned into $n + \binom{n}{2}$

segments.

The first n segments are each of length $\binom{n_0}{2}$, and are each associated with one considered source. The remaining $\binom{n}{2}$ segments are each of length n_0^2, and are each associated with one of the possible combinations of sources. Note that $n\binom{n_0}{2}+\binom{n}{2}n_0^2 = N_c$. When we are considering a within-source comparison ($i = i'$), the matrix $\boldsymbol{B}_{ii}$ has ones in the segment of length $\binom{n_0}{2}$ that correspond to the i^{th} source, and zeros elsewhere. When we are considering a between-source comparison ($i \neq i'$), the matrix $\boldsymbol{B}_{ii'}$ has ones in the segment of length n_0^2 corresponding to the comparison between source i and i', and zeros elsewhere.

As an example, let $n = 3$, $n_0 = 4$. Then we have $n + \binom{n}{2} = 6$ matrices, consisting of $n = 3$ within source matrices, $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{22}$, $\boldsymbol{B}_{33}$, and $\binom{n}{2} = 3$ between source matrices, $\boldsymbol{B}_{12}$, $\boldsymbol{B}_{13}$, $\boldsymbol{B}_{23}$. These six matrices are displayed in Figures 6.1 and 6.2. Since $\boldsymbol{B}_{11}, \boldsymbol{B}_{12}, \ldots, \boldsymbol{B}_{22}, \ldots, \boldsymbol{B}_{nn}$ sum to the identity matrix, we have that

$$\begin{aligned} SS_{Tot} &= \widetilde{\boldsymbol{s}}'\left(\sum_{i,i'}\boldsymbol{B}_{ii'}\right)\boldsymbol{I}\left(\sum_{i,i'}\boldsymbol{B}_{ii'}\right)\widetilde{\boldsymbol{s}} \\ &= \widetilde{\boldsymbol{s}}'\left(\sum_{i,i'}\boldsymbol{B}_{ii'}\boldsymbol{I}\boldsymbol{B}_{ii'}\right)\widetilde{\boldsymbol{s}} \\ &= \widetilde{\boldsymbol{s}}'\left(\sum_{i,i'}\boldsymbol{B}_{ii'}\left[\sum_{l=1}^{N_c} v_{ii'}v_{ii'}'\right]\boldsymbol{B}_{ii'}\right)\widetilde{\boldsymbol{s}} \\ &= \sum_{i\in\{1,\ldots,n\}} SS_{ii} + \sum_{i<i'\in\{1,\ldots,n\}} SS_{ii'} \end{aligned} \tag{6.11}$$

where $\{\boldsymbol{v}_{ii'_l}\}$, $l = 1, \ldots, N_c$ are different orthonormal bases spanning $\mathbb{R}^{N_c}$, and will be discussed in more detail in later sections. The matrices $\boldsymbol{B}_{ii'}$ effectively activate different parts of the vector $\widetilde{\boldsymbol{s}}_c$ according to the different source comparisons. In particular, we have:

(1) $\boldsymbol{B}_{ii}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to some within-source comparison, so that $\widetilde{\boldsymbol{s}}_c\boldsymbol{B}_{ii}\widetilde{\boldsymbol{s}}_c$ gives us the corresponding within-source sum of squares. Recall that $\boldsymbol{B}_{ii}\boldsymbol{I}\boldsymbol{B}_{ii} = \boldsymbol{B}_{ii}$, and note that $\boldsymbol{B}_{ii}$ has rank $r_{ii} = \binom{n_0}{2}$.

(2) $\boldsymbol{B}_{ii'}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to some between-

source comparison, so that $\widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'} \widetilde{\boldsymbol{s}}_c$ gives us the corresponding between-source sum of squares. Recall that $\boldsymbol{B}_{ii'} \boldsymbol{I} \boldsymbol{B}_{ii'} = \boldsymbol{B}_{ii'}$, and note that $\boldsymbol{B}_{ii'}$ has rank $r_{ii'} = n_0^2$.

Thus, we have defined the total sums of squares in terms of the various source comparisons that exist within our vector of scores. Bearing in mind that the goal is to find a way to estimate the parameters of the distribution given in (6.8), we note that this decomposition of the total sums of squares allows us to independently study the mean and variance terms, $\theta_{ii'}$ and $\sigma_{ii'}$, associated with their respective source comparisons. Note that we can choose the orthonormal bases in (6.11) to be any orthonormal bases, and, in particular, we can choose these orthonormal bases to be the normalized eigenvectors for the following matrices. We choose to define

$$\boldsymbol{V}_{ii} \quad := \quad \boldsymbol{B}_{ii} \left(\boldsymbol{P}_c \boldsymbol{P}_c' \sigma_a^2 + \boldsymbol{I}_c \sigma_e^2 \right) \boldsymbol{B}_{ii} \tag{6.12}$$

for the matrices $\boldsymbol{V}_{ii}$, corresponding to within-source comparisons, and

$$\boldsymbol{V}_{ii'} \quad := \quad \boldsymbol{B}_{ii'} \left(\boldsymbol{P}_c \boldsymbol{P}_c' \sigma_a^2 + \boldsymbol{I}_c \sigma_e^2 \right) \boldsymbol{B}_{ii'} \tag{6.13}$$

for the matrices $\boldsymbol{V}_{ii'}$ corresponding to between-source comparisons.

As in Section 3.2, choosing $\boldsymbol{V}_{ii}$ and $\boldsymbol{V}_{ii'}$ to be a function of $\boldsymbol{\Sigma}_c := \boldsymbol{P}_c \boldsymbol{P}_c' \sigma_a^2 + \boldsymbol{I}_c \sigma_e^2$ is advantageous in that it introduces the parameters σ_a^2 and σ_e^2, and so provides a means for studying these parameters. Second, defining $\boldsymbol{V}_{ii}$ and $\boldsymbol{V}_{ii'}$ in terms of $\boldsymbol{B}_{ii}$ and $\boldsymbol{B}_{ii'}$ allows us to take the relevant parts of $\boldsymbol{\Sigma}_c$ with respect to each source comparison by activating only the rows and columns of $\boldsymbol{\Sigma}_c$ corresponding to the considered source comparison.

$\boldsymbol{V}_{ii}$		
Eigenvalue (ν_{ii_l})	Multiplicity ($m_{\nu_{ii_l}}$)	Eigenvectors ($\boldsymbol{v}_{ii_l}$)
$2\left(n_0-1\right)\sigma_a^2+\sigma_e^2$	1	$\boldsymbol{v}_{ii_l}$ such that $\boldsymbol{V}_{ii}\boldsymbol{v}_{ii_l}=\nu_{ii_1}\boldsymbol{v}_{ii_l}$
$\left(n_0-2\right)\sigma_a^2+\sigma_e^2$	n_0-1	$\boldsymbol{v}_{ii_l}$ such that $\boldsymbol{V}_{ii}\boldsymbol{v}_{ii_l}=\nu_{ii_2}\boldsymbol{v}_{ii_l}$
σ_e^2	$\binom{n_0}{2}-n_0$	$\boldsymbol{v}_{ii_l}$ such that $\boldsymbol{V}_{ii}\boldsymbol{v}_{ii_l}=\nu_{ii_3}\boldsymbol{v}_{ii_l}$
0	$N_c-\binom{n_0}{2}$	$\boldsymbol{v}_{ii_l}$ such that $\boldsymbol{V}_{ii}\boldsymbol{v}_{ii_l}=\nu_{ii_4}\boldsymbol{v}_{ii_l}$
$\boldsymbol{V}_{ii'}$		
Eigenvalue ($\nu_{ii'_l}$)	Multiplicity ($m_{\nu_{ii'_l}}$)	Eigenvectors ($\boldsymbol{v}_{ii'_l}$)
$2n_0\sigma_a^2+\sigma_e^2$	1	$\boldsymbol{v}_{ii'_l}$ such that $\boldsymbol{V}_{ii'}\boldsymbol{v}_{ii'_l}=\nu_{ii'_1}\boldsymbol{v}_{ii'_l}$
$n_0\sigma_a^2+\sigma_e^2$	$2n_0-2$	$\boldsymbol{v}_{ii'_l}$ such that $\boldsymbol{V}_{ii'}\boldsymbol{v}_{ii'_l}=\nu_{ii'_2}\boldsymbol{v}_{ii'_l}$
σ_e^2	$\left(n_0-1\right)^2$	$\boldsymbol{v}_{ii'_l}$ such that $\boldsymbol{V}_{ii'}\boldsymbol{v}_{ii'_l}=\nu_{ii'_3}\boldsymbol{v}_{ii'_l}$
0	$N_c-n_0^2$	$\boldsymbol{v}_{ii'_l}$ such that $\boldsymbol{V}_{ii'}\boldsymbol{v}_{ii'_l}=\nu_{ii'_4}\boldsymbol{v}_{ii'_l}$

Table 6.3: Eigenstructure of design matrices for within-source comparisons, $\boldsymbol{V}_{ii}$, and between-source comparisons, $\boldsymbol{V}_{ii'}$ in (6.12) and (6.13)

We can study the eigenstructure of the matrices $\boldsymbol{B}_{ii'}\left(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2+\boldsymbol{I}_c\sigma_e^2\right)\boldsymbol{B}_{ii'}$ for each source comparison (see Table 6.3). This study reveals the presence of multiple subspaces for each of the considered eigenspaces. This allows us to decompose each of the sums of squares in (6.11) as another sum of squares. For within-source comparisons, we have that

$$\begin{aligned}
\widetilde{\boldsymbol{s}}_c\boldsymbol{B}_{ii}\left[\sum_{l=1}^{N_c}\boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}'\right]\boldsymbol{B}_{ii}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_c\boldsymbol{B}_{ii}\left[\boldsymbol{v}_{ii_1}\boldsymbol{v}_{ii_1}'+\sum_{l=2}^{n_0}\boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}'+\sum_{l=n_0+1}^{\binom{n_0}{2}}\boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}'+\sum_{l=\binom{n_0}{2}}^{N_c}\boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}'\right]\boldsymbol{B}_{ii}\widetilde{\boldsymbol{s}}_c \\
&= \widetilde{\boldsymbol{s}}_c\boldsymbol{B}_{ii}\left[\boldsymbol{v}_{ii_1}\boldsymbol{v}_{ii_1}'+\sum_{l=2}^{n_0}\boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}'+\sum_{l=n_0+1}^{N_{ii}}\boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}'\right]\boldsymbol{B}_{ii}\widetilde{\boldsymbol{s}}_c,
\end{aligned}$$

where $N_{ii}=\binom{n_0}{2}\forall i$ is the number of eigenvectors associated with non-zero eigenvalues, and is also the number of "interesting" eigenvectors $\boldsymbol{v}_{ii_l}$. In particular, we have that the elements that are equal to one correspond to the dimensions whose associated eigenvalues are zero. These vectors form the standard basis for the null space of the corresponding matrix, and correspond to the rows of $\boldsymbol{B}_{ii}$ that are equal to the zero vector. As an example, we consider the matrices that result when $n=3$ and $n_0=4$. The first row of Figure 6.1 displays the heat maps of $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{22}$, and $\boldsymbol{B}_{33}$. The second row displays the heat maps of the matrices of eigenvectors of $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{22}$, and $\boldsymbol{V}_{33}$. Note that the $N_{ii}\times N_{ii}$ patchwork matrices within each of the $\boldsymbol{V}_{ii}$ matrices correspond to the non-zero rows of the corresponding $\boldsymbol{B}_{ii}$ matrices. Because the placements of the nonzero elements in these eigenvectors correspond to the zero elements of the diagonals in the associated $\boldsymbol{B}_{ii}$ matrices, the product of the $\boldsymbol{B}_{ii}$ matrix with these sets of eigenvectors results in a zero-valued sum of squares.

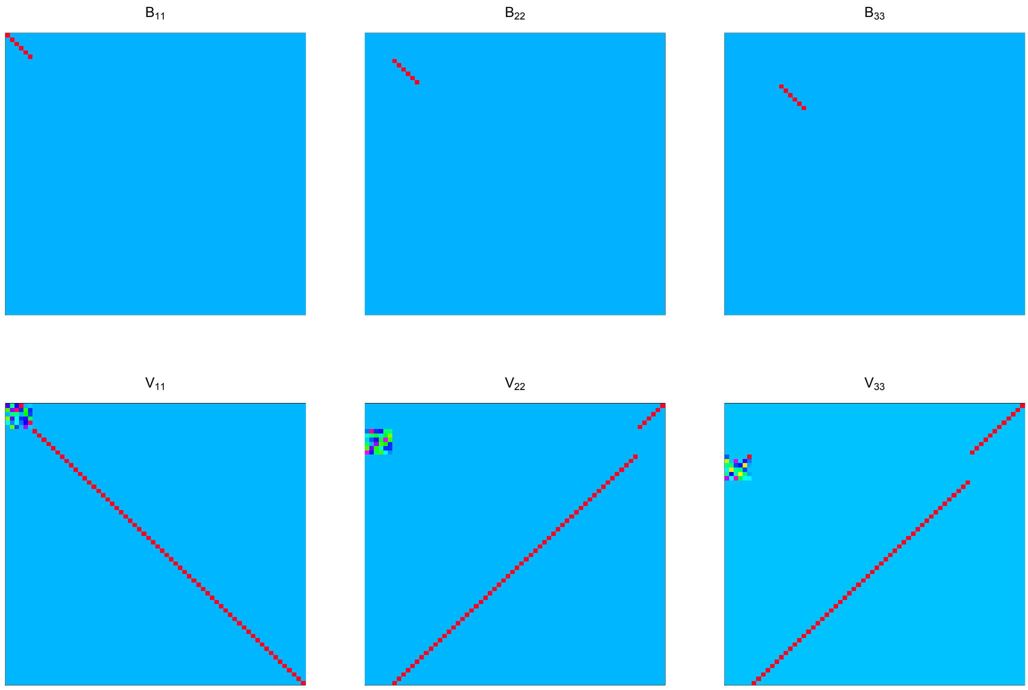


Figure 6.1: Heat maps for within-source B_{ii} matrices (top) and eigenvectors of associated $\boldsymbol{V}_{ii}$ matrices (bottom) when $n = 3$ and $n_0 = 4$. The $N_{ii} \times N_{ii}$ patchworks correspond to eigenvectors with non-zero elements. The remaining columns correspond to vectors with one non-zero element (in red), and are associated with zero-valued eigenvalues.

Similarly, for between-source comparisons, we have that

$$\begin{aligned}
\widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'} \left[\sum_{l=1}^{N_c} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} \right] \boldsymbol{B}_{ii'} \widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'} \left[\boldsymbol{v}_{ii'_1} \boldsymbol{v}'_{ii'_1} + \sum_{l=2}^{\binom{n_0}{2}-1} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} + \sum_{l=\binom{n_0}{2}}^{n_0^2} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} + \sum_{l=n_0^2+1}^{N_c} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} \right] \boldsymbol{B}_{ii'} \widetilde{\boldsymbol{s}}_c \\
&= \widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'} \left[\boldsymbol{v}_{ii'_1} \boldsymbol{v}'_{ii'_1} + \sum_{l=2}^{\binom{n_0}{2}-1} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} + \sum_{l=\binom{n_0}{2}}^{N_{ii'}} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} \right] \boldsymbol{B}_{ii'} \widetilde{\boldsymbol{s}}_c
\end{aligned}$$

where $N_{ii'} = n_0^2 \forall ii'$ is the number of eigenvectors associated with non-zero eigenvalues, and is also the number of "interesting" eigenvectors $\boldsymbol{v}_{ii'_l}$. In particular, we have that the elements that are equal to one correspond to the dimensions whose associated eigenvalues are zero. These vectors form the standard basis for the null space of the corresponding matrix, and correspond to the rows of $\boldsymbol{B}_{ii'}$ that are equal to the zero vector.

The first row of Figure 6.2 displays the heat maps of $\boldsymbol{B}_{12}$, $\boldsymbol{B}_{13}$, and $\boldsymbol{B}_{23}$. The second row displays the heat maps of the matrices of eigenvectors of $\boldsymbol{V}_{12}$, $\boldsymbol{V}_{13}$, and $\boldsymbol{V}_{23}$. Note that the $N_{ii'} \times N_{ii'}$ patchwork matrices within each of the $\boldsymbol{V}_{ii'}$ matrices correspond to the non-zero rows of the corresponding $\boldsymbol{B}_{ii'}$ matrices. Because the placements of the nonzero elements in these eigenvectors correspond to the zero elements of the diagonals in the associated $\boldsymbol{B}_{ii'}$ matrices, the product of the $\boldsymbol{B}_{ii'}$ matrix with these sets of eigenvectors results in a zero-valued sum of squares.

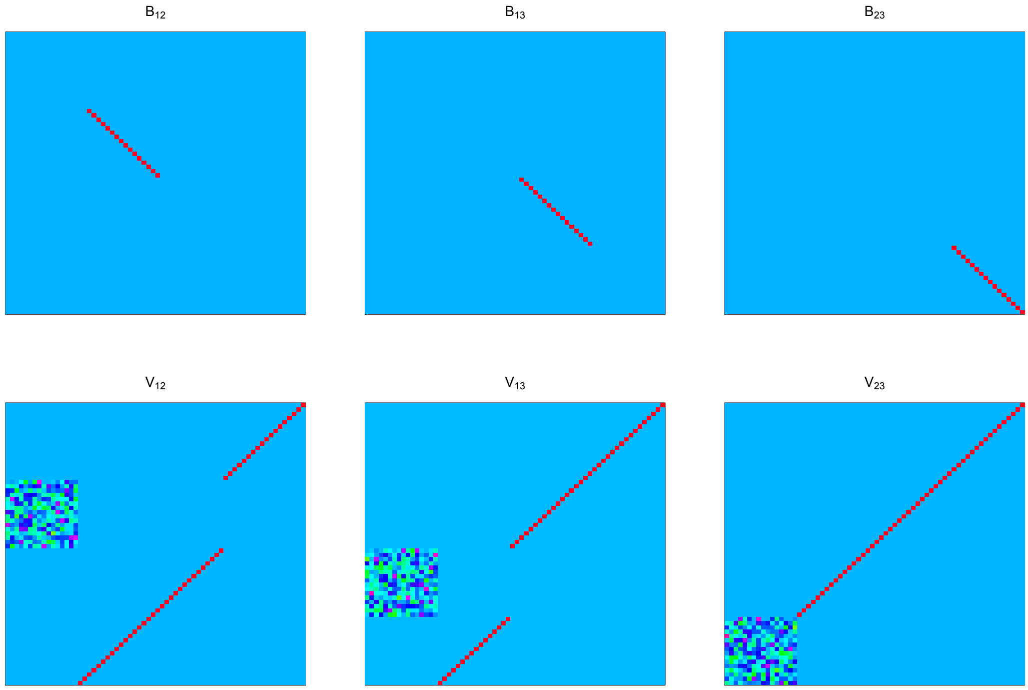


Figure 6.2: Heat maps for the between-source $B_{ii'}$ matrices (top) and eigenvectors of the associated $\boldsymbol{V}_{ii'}$ matrices (bottom) when $n = 3$ and $n_0 = 4$. The $N_{ii'} \times N_{ii'}$ patchworks in the correspond to eigenvectors with non-zero elements. The remaining columns correspond to vectors with one non-zero element (in red), and are associated with zero-valued eigenvalues.

This decomposition is favorable in that studying the relevant parts of the eigen-decomposision of $\boldsymbol{\Sigma}_c$ is equivalent to studying the eigen-decomposision of the relevant parts of $\boldsymbol{\Sigma}_c$. That is,

$$\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{ii}\left[\sum_{l=1}^{N_n}\boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}'\right]\boldsymbol{B}_{ii}\widetilde{\boldsymbol{s}}_c \quad = \quad \widetilde{\boldsymbol{s}}_{ii}'\left[\sum_{l=1}^{N_{ii}}\boldsymbol{v}_{ii_l}^{*}\boldsymbol{v}_{ii_l}^{*'}\right]\widetilde{\boldsymbol{s}}_{ii}$$

$$\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{ii'}\left[\sum_{l=1}^{N_n}\boldsymbol{v}_{ii_l'}\boldsymbol{v}_{ii_l'}'\right]\boldsymbol{B}_{ii'}\widetilde{\boldsymbol{s}}_c \quad = \quad \widetilde{\boldsymbol{s}}_{ii'}'\left[\sum_{l=1}^{N_{ii'}}\boldsymbol{v}_{ii_l'}^{*}\boldsymbol{v}_{ii_l'}^{*'}\right]\widetilde{\boldsymbol{s}}_{ii'}$$

where $\boldsymbol{v}_{ii_l}^{*}$ are the eigenvectors of the $N_{ii} \times N_{ii}$ matrix formed by considering the non-zero rows of the columns associated with the non-zero eigenvalues of $\boldsymbol{V}_{ii}$, and $\boldsymbol{v}_{ii'}^{*}$ are the eigenvectors of the $N_{ii'} \times N_{ii'}$ matrix formed by considering the non-zero portions of the columns associated with the non-zero eigenvalues of $\boldsymbol{V}_{ii'}$. This is equivalent to considering only the indices of $\boldsymbol{\Sigma}_c$ that correspond to each source comparison. That is, $\boldsymbol{v}_{ii_l}^{*}$ are the eigenvectors of

$$\boldsymbol{\Sigma}_{ii} := \boldsymbol{P}_{ii}\boldsymbol{P}_{ii}'\sigma_a^2 + \boldsymbol{I}_{ii}\sigma_e^2, \tag{6.14}$$

the matrix formed by considering the rows of $\boldsymbol{\Sigma}_c$ associated with some within-source comparison, and $\boldsymbol{v}_{ii_l'}^{*}$ are the eigenvectors of

$$\boldsymbol{\Sigma}_{ii'} := \boldsymbol{P}_{ii'}\boldsymbol{P}_{ii'}\sigma_a^2 + \boldsymbol{I}_{ii'}\sigma_e^2, \tag{6.15}$$

the matrix formed by considering the rows of $\boldsymbol{\Sigma}_c$ associated with some between-source comparison. In addition, we have that $\widetilde{\boldsymbol{s}}_{ii} = (\boldsymbol{\Delta}_{ii}(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}))$, where $\boldsymbol{\Delta}_{ii}$ is the $N_{ii} \times N_{ii}$ portion of $\boldsymbol{\Delta}_c$ that considers σ_{ii}, and $\widetilde{\boldsymbol{s}}_{ii'} = \left(\boldsymbol{\Delta}_{ii'}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right)$, where $\boldsymbol{\Delta}_{ii'}$ is the $N_{ii'} \times N_{ii'}$ portion of $\boldsymbol{\Delta}_n$ that considers $\sigma_{ii'}$.

$\boldsymbol{\Sigma}_{ii}$		
Eigenvalue (ν_{ii_l})	Multiplicity ($m_{\nu_{ii_l}}$)	Eigenvectors ($\boldsymbol{v}^*_{ii_l}$)
$2\,(n_0-1)\,\sigma_a^2+\sigma_e^2$	1	$\boldsymbol{v}^*_{ii_1} := \frac{\mathbf{1}_{N_{ii}}}{\sqrt{N_{ii}}}$
$(n_0-2)\,\sigma_a^2+\sigma_e^2$	n_0-1	$\boldsymbol{v}^*_{ii_l}$ such that $\boldsymbol{\Sigma}_{ii}\boldsymbol{v}^*_{ii_l}=\nu_{ii_2}\boldsymbol{v}^*_{ii_l}$
σ_e^2	$N_{ii}-n_0$	$\boldsymbol{v}^*_{ii_l}$ such that $\boldsymbol{\Sigma}_{ii}\boldsymbol{v}^*_{ii_l}=\nu_{ii_3}\boldsymbol{v}^*_{ii_l}$
$\boldsymbol{\Sigma}_{ii'}$		
Eigenvalue ($\nu_{ii'_l}$)	Multiplicity ($m_{\nu_{ii'_l}}$)	Eigenvectors ($\boldsymbol{v}^*_{ii'_l}$)
$2n_0\sigma_a^2+\sigma_e^2$	1	$\boldsymbol{v}^*_{ii'_1} := \frac{\mathbf{1}_{N_{ii'}}}{\sqrt{N_{ii'}}}$
$n_0\sigma_a^2+\sigma_e^2$	$2n_0-2$	$\boldsymbol{v}^*_{ii'_l}$ such that $\boldsymbol{\Sigma}_{ii'}\boldsymbol{v}^*_{ii'_l}=\nu_{ii'_2}\boldsymbol{v}^*_{ii'_l}$
σ_e^2	$(n_0-1)^2$	$\boldsymbol{v}^*_{ii'_l}$ such that $\boldsymbol{\Sigma}_{ii'}\boldsymbol{v}^*_{ii'_l}=\nu_{ii'_3}\boldsymbol{v}^*_{ii'_l}$

Table 6.4: Eigenstructure of design matrices $\boldsymbol{\Sigma}_{ii}$, and $\boldsymbol{\Sigma}_{ii'}$ in (6.14) and (6.15)

These results follow from using the $\boldsymbol{B}_{ii}$ and $\boldsymbol{B}_{ii'}$ matrices to activate certain areas of the vector $\widetilde{\boldsymbol{s}}_c$ and the matrices $\boldsymbol{V}_{ii}$, $\boldsymbol{V}_{ii'}$, i.e., introducing the matrices $\boldsymbol{B}_{ii}$ and $\boldsymbol{B}_{ii'}$ allows us to activate the parts of $\widetilde{\boldsymbol{s}}_c$ and $\boldsymbol{\Sigma}_c$ that correspond to the different source comparisons. Rather than considering a sparse N_c vector alongside a sparse $N_c \times N_c$ matrix, we can directly consider the interesting parts of the vector and matrix by considering the associated $N_{ii}-$ or $N_{ii'}-$dimensional vector and $N_{ii} \times N_{ii}$ or $N_{ii'} \times N_{ii'}$ dimensional matrix. Thus, we can explicitly define the sums of squares such that

$$\begin{aligned} SS_{ii} \;=\; & \left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right)'\left(\boldsymbol{v}^*_{ii_1}\boldsymbol{v}^{*\,\prime}_{ii_1}\right)\left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right) + \\ & \left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right)'\left(\sum_{l=2}^{n_0}\boldsymbol{v}^*_{ii_l}\boldsymbol{v}^{*\,\prime}_{ii_l}\right)\left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right) + \\ & \left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{22}\mathbf{1}_{N_{ii}}\right)\right)'\left(\sum_{l=n_0+1}^{N_{ii}}\boldsymbol{v}^*_{ii_l}\boldsymbol{v}^{*\,\prime}_{ii_l}\right)\left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right), \end{aligned} \tag{6.16}$$

for within-source comparisons and

$$\begin{aligned} SS_{ii'} \;=\; & \left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right)'\left(\boldsymbol{v}^*_{ii'_1}\boldsymbol{v}^{*\,\prime}_{ii'_1}\right)\left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right) + \\ & \left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right)'\left(\sum_{l=2}^{\binom{n_0}{2}-1}\boldsymbol{v}^*_{ii'_l}\boldsymbol{v}^{*\,\prime}_{ii'_l}\right)\left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right) + \\ & \left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{12}\mathbf{1}_{N_{ii'}}\right)\right)'\left(\sum_{l=\binom{n_0}{2}}^{N_{ii'_l}}\boldsymbol{v}^*_{ii'_l}\boldsymbol{v}^{*\,\prime}_{ii'_l}\right)\left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right), \end{aligned} \tag{6.17}$$

for between-source comparisons, where the degrees of freedom for each line of (6.16) and (6.17) are equal to the multiplicities of the associated eigenvalues in Table 6.4, and the total sum of squares remains as in (6.9). It is trivial to show that

$$\sum_{i=1}^{n}\left[\sum_{l=1}^{3} m_{\nu_{ii_l}}\right]+\sum_{i=1}^{n-1}\sum_{i'=i+1}^{n}\left[\sum_{l=1}^{3} m_{\nu_{ii'_l}}\right]=N_c$$

In the following sections, we analyse the three terms that make up the sums of squares defined in (6.16) and (6.17) so that we can write each term without the use of eigenvectors.

6.2.1 ALTERNATIVE REPRESENTATION OF WITHIN-SOURCE SUMS OF SQUARES

We begin by studying the individual terms in the within-source sums of squares terms of the form given by (6.16). All developments can be found in Appendix D. We re-write the first term as

$$\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)'\boldsymbol{\Delta}_{ii}^{-1'}\boldsymbol{v}_{ii_1}^{*}\boldsymbol{v}_{ii_1}^{*'}\boldsymbol{\Delta}_{ii}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)=\frac{N_{ii}}{\sigma_{ii}^{2}}\left(\overline{s}_{ii}-\theta_{ii}\right)^{2}, \tag{6.18}$$

where $\overline{s}_{ii}$ is the average score observed for the within-source comparisons from source i. Recall that $N_{ii}=\binom{n_0}{2}$, $\forall i\in\{1,\dots,n\}$.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{ii}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{ii_l}^{*}\boldsymbol{v}_{ii_l}^{*'}\right]\boldsymbol{\Delta}_{ii}^{-1}$. Following the development by [6], we can write this second sum of square as

$$\begin{aligned}\boldsymbol{s}_{ii}'\boldsymbol{\Delta}_{ii}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{ii_l}^{*}\boldsymbol{v}_{ii_l}^{*}{}'\right]\boldsymbol{\Delta}_{ii}^{-1}\boldsymbol{s}_{ii} &= \frac{\left(n_0-1\right)^2}{\sigma_{ii}^{2}\left(n_0-2\right)}\sum_{j=1}^{n_0}\left(\overline{s}_{ii}^{(i_j)}-\overline{s}_{ii}\right)^2 \\ &:= SS_{W_{ii}}\end{aligned} \tag{6.19}$$

where $\overline{s}_{i_j}^{(ii)}$, $i\in\{1,\dots,n\}$, $j\in\{1,\dots,n_0\}$ is the mean value of scores that compare object j in source i to any other object in source i, and $\overline{s}_{ii}$ is as in (6.18). The final result, given by (6.19), gives the within-source sum of squares for within-source i model. By considering this term in conjunction with the total sum of squares for the considered model, $SS_{Tot_{ii}} = \frac{1}{\sigma^2_{ii}} \sum_{j=1}^{n_0-1} \sum_{j'=j+1}^{n_0} \left(s_{ij,ij'} - \overline{s}_{ii}\right)^2$ (this is the sum of the last two terms in (6.16)), we can obtain an eigenvector-free estimate of the last term in SS_{ii} by considering $SS_{Tot_{ii}} - SS_{W_{ii}}$ (see Table 6.5). By using the results of Cochran's theorem presented in Section 3.2.1, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	$n_0 - 1$	$SS_{W_{ii}}$	$MS_{W_{ii}} = \frac{SS_{W_{ii}}}{n_0-1}$	$(n_0-2)\sigma^2_a + \sigma^2_e$
Error	$N_{ii} - n_0$	$SS_{E_{ii}} = SS_{Tot_{ii}} - SS_{W_{ii}}$	$MS_{E_{ii}} = \frac{SS_{E_{ii}}}{N_{ii}-n_0}$	σ^2_e

Table 6.5: ANOVA table corresponding to within-source sums of squares, SS_{ii}

6.2.2 ALTERNATIVE REPRESENTATION OF BETWEEN-SOURCE SUMS OF SQUARES

Finally, we move to consider the terms in the between-source sums of squares terms of the form given by (6.17). All developments can be found in Appendix E. As in the common-source sums of squares development, we rewrite the first term as

$$\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)' \boldsymbol{\Delta}_{ii'}^{-1'} \boldsymbol{v}^*_{ii'_1} \boldsymbol{v}^{*'}_{ii'_1} \boldsymbol{\Delta}_{ii'}^{-1} \left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right) = \frac{N_{ii'}}{\sigma^2_{ii'}} \left(\overline{s}_{ii'} - \theta_{ii'}\right)^2, \tag{6.20}$$

where $\overline{s}_{ii'}$ is the average score observed for between-source comparisons between sources i and i'.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{ii'}^{-1}\left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}^*_{ii'_l} \boldsymbol{v}^{*'}_{ii'_l}\right]\boldsymbol{\Delta}_{ii'}^{-1}$. As before, we can write this second sum of squares as

$$\begin{aligned} \boldsymbol{s}'_{ii'}\boldsymbol{\Delta}_{ii'}^{-1'}\left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}^*_{ii'_l} \boldsymbol{v}^{*'}_{ii'_l}\right]\boldsymbol{\Delta}_{ii'}^{-1}\boldsymbol{s}_{ii'} &= \frac{n_0^2}{\sigma^2_{ii'} n_0}\left(\sum_{j'=1}^{n_0}\left(\overline{s}_{ii'}^{(i_j)} - \overline{s}_{ii'}\right)^2 + \sum_{j=1}^{n_0}\left(\overline{s}_{ii'}^{(i'_{j'})} - \overline{s}_{ii'}\right)^2\right) \\ &:= SS_{W_{ii'}}, \end{aligned} \tag{6.21}$$

where $\overline{s}_{ii'}^{(i_j)}$, $i \in \{1, \ldots n\}$, $j \in \{1, \ldots, n_0\}$ is the mean value of scores that compare object j in source i to any object in source i', $\overline{s}_{ii'}^{(i'_{j'})}$ is the mean value of scores that compare object j' in source i' to any object in source i, and $\overline{s}_{ii'}$ is as in (6.20). The final result, given by (6.21), gives the within-source sum of squares for between-source i, i' comparisons. By considering this term in conjunction with the total sum of squares for the considered model, $SS_{Tot_{ii'}} = \frac{1}{\sigma^2_{ii'}} \sum_{j=1}^{n_0-1} \sum_{j'=1}^{n_0} \left(s_{ij,i'j'} - \overline{s}_{ii'}\right)^2$, we can define an eigenvector-free estimate of the last term in $SS_{ii'}$ by considering $SS_{Tot_{ii'}} - SS_{W_{ii'}}$. By using the results of Cochran's theorem presented in Section 3.2.1, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	$2n_0 - 2$	$SS_{W_{ii'}}$	$MS_{W_{ii'}} = \frac{SS_{W_{ii'}}}{2n_0-2}$	$n_0\sigma^2_a + \sigma^2_e$
Error	$(n_0-1)^2$	$SS_{E_{ii'}} = SS_{Tot_{ii'}} - SS_{W_{ii'}}$	$MS_{E_{ii'}} = \frac{SS_{E_{ii'}}}{(n_0-1)^2}$	σ^2_e

Table 6.6: ANOVA table corresponding to between-source sums of squares, $SS_{ii'}$

6.3 PARAMETER ESTIMATION

At this point, we would like to use the results presented in Sections 6.2.1 and 6.2.2 to estimate the parameters of our model; however given the dependencies that exist between the various parameters, we must resort to sampling methods to obtain posterior samples of the model parameters. In particular, we use a Gibbs sampler with a Metropolis-Hastings step to study the distributions of our various parameters [14, 44]. Before defining the Gibbs sampler, we must first assign posterior distributions to the model parameters (development of posterior distributions for $\{\theta_{ii'}\}_{ii'}$, $\{\sigma_{ii'}\}_{ii'}$, and σ^2_e can be found in Appendix F).

We begin by assigning posterior distributions for the variance terms, σ^2_a and σ^2_e. Because we have the constraint that $2\sigma^2_a + \sigma^2_e = 1$ (see (6.6) in Section 6.1.2), we can define a posterior distribution for one variance term, obtain posterior samples from this distribution, and directly obtain the associated value of the other. In this case, we choose to obtain posterior samples of σ^2_e, so as to exploit all information available in Tables 6.5 and 6.6. The value of $\sigma^2_a = (1 - \sigma^2_e)/2$ follows directly.

To define the posterior distribution of σ^2_e, we begin by collecting all sums of squares terms defined in Tables 6.5 and 6.6 to capitalize on all information related to the value of σ_e^2. We have that

$$\frac{SS_{W_{ii}}}{(n_0-2)\,\sigma_a^2+\sigma_e^2} = \frac{SS_{W_{ii}}}{(n_0-2)\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2} \sim \chi^2_{df=n_0-1} \qquad \frac{SS_{E_{ii}}}{\sigma_e^2} \sim \chi^2_{df=N_{ii'}-n_0},$$

for each of the n within-source comparison sums of squares terms and

$$\frac{SS_{W_{ii'}}}{n_0\sigma^2 a+\sigma_e^2} = \frac{SS_{W_{ii'}}}{n_0\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2} \sim \chi^2_{df=2n_0-2} \qquad \frac{SS_{E_{ii'}}}{\sigma_e^2} \sim \chi^2_{df=(n_0-1)^2} \quad ,$$

for each of the $\binom{n}{2}$ between-source comparison sums of squares terms, and so we define

$$\begin{aligned} MS_e &= \sum_{i=1}^{n}\left(\frac{SS_{W_{ii}}}{C_2}+\frac{SS_{E_{ii}}}{C_3}\right)+\sum_{i=1}^{n-1}\sum_{i'=2}^{n}\left(\frac{SS_{W_{ii'}}}{C_1}+\frac{SS_{E_{ii'}}}{C_3}\right) \\ &\sim \chi^2_{df=n((n_0-1)+(N_{ii'}-n_0))+\binom{n}{2}\left((2n_0-2)+(n_0-1)^2\right)}, \end{aligned} \tag{6.22}$$

where

$$C_1 = n_0\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2 \qquad C_2 = (n_0-2)\left(\frac{1-\sigma_e^2}{2}\right)+\sigma_e^2 \qquad C_3 = \sigma_e^2$$

We can simplify (6.22) by considering a common denominator, such that

$$\begin{aligned} MS_e &= \frac{C_1C_3\left(\sum_{i=1}^{n} SS_{W_{ii}}\right)+C_2C_3\left(\sum_{i=1}^{n-1}\sum_{i'=2}^{n} SS_{W_{ii'}}\right)+C_1C_2\left(\sum_{i=1}^{n} SS_{E_{ii}}+\sum_{i=1}^{n-1}\sum_{i'=2}^{n} SS_{E_{ii'}}\right)}{C_1C_2C_3} \\ &\sim \chi^2_{df=\binom{nn_0}{2}-\binom{n}{2}-n}. \end{aligned}$$

We now find the posterior distribution for the variance term σ_e^2 by considering a χ^2 likelihood for the MS_e term, and assuming a Beta prior (since we have the constraint that $\sigma_e^2 \le 1$), such that

$$\pi\left(\sigma_e^2|MS_e,\boldsymbol{\sigma},\alpha_e,\beta_e\right) \propto \chi^2\left(MS_e|\sigma_e^2,\boldsymbol{\sigma},\alpha_e,\beta_e\right)\mathcal{B}\left(\sigma_e^2|\alpha_e,\beta_e\right), \tag{6.23}$$

where the dependence of MS_e on $\boldsymbol{\sigma} := \{\sigma_{ii'}\}_{ii'}, i, i' \in \{1,\ldots,n\}$, in (6.23) results from the construction of MS_e as a sum of the various sum of squares terms defined in

Section 6.2.

Next, we assign the posterior distributions for each of the mean parameters $\theta_{ii'}$, $i, i' \in \{1, \ldots, n\}$, by considering a Multivariate Normal likelihood, and assuming a Normal prior with mean $\phi_{ii'}$ and variance $\omega_{ii'}$ such that

$$\pi\left(\theta_{ii'}|\boldsymbol{s}_{ii'}, \sigma_{ii'}, \sigma_a^2, \sigma_e^2, \phi_{ii'}, \omega_{ii'}\right) \propto \mathcal{MVN}\left(\boldsymbol{s}_{ii'}|\theta_{ii'}, \sigma_{ii'}, \sigma_a^2, \sigma_e^2, \phi_{ii'}, \omega_{ii'}\right)\mathcal{N}\left(\theta_{ii'}|\phi_{ii'}\omega_{ii'}\right) \quad (6.24)$$

where the resulting posterior distribution is Normally distributed. The parameters of the posterior distribution of $\theta_{ii'}$ are given by

$$\mu_{ii'_p} = \frac{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \phi_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1} \qquad \sigma^2_{ii'_p} = \frac{\omega_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1}.$$

Finally, we find the posterior distributions for each of the variance terms $\sigma_{ii'}$ by considering a Multivariate Normal likelihood, and assuming an Inverse-Gamma prior such that

$$\pi\left(\sigma_{ii'}|\boldsymbol{s}_{ii'}, \theta_{ii'}, \sigma_a^2, \sigma_e^2, \alpha_{ii'}, \beta_{ii'}\right) \propto \mathcal{MVN}\left(\boldsymbol{s}_{ii'}|\sigma_{ii'}, \theta_{ii'}, \sigma_a^2, \sigma_e^2, \alpha_{ii'}, \beta_{ii'}\right)\mathcal{IG}\left(\sigma^2_{ii'}|\alpha_{ii'}, \beta_{ii'}\right) \quad (6.25)$$

where the resulting posterior distribution is distributed according to an Inverse Gamma distribution. The parameters of the posterior distribution of $\sigma^2_{ii'}$ are given by

$$\alpha_{ii'_p} = \frac{N_{ii'}}{2} + \alpha_{ii'} \qquad \beta_{ii'_p} = \frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Sigma}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right) + \beta_{ii'}.$$

The equation given by (6.25) provides us with samples from the posterior distribution of the variance term, $\sigma^2_{ii'}$. We, however, are interested in the standard deviation term, $\sigma_{ii'}$, and so we simply take the square root of the sampled variance term to obtain samples of the standard deviation term. The resulting inference is not affected.

We note that each of the distributions described in (6.23), (6.24), and (6.25) depends on the value of at least one other parameter; therefore, we must rely on sampling techniques to study the distributions of the model parameters. We construct the following Gibbs sampler, described in Algorithm 3.

Algorithm 3: Gibbs sampler for generating posterior samples from the distributions of the model parameters

Data: Initial values for all parameters at $t = 0$; values for hyperparameters
Result: Posterior samples for all parameters
for $t \in 1 : T$ iterations **do**
  1. Draw $\sigma_e^{2(t)} | \boldsymbol{s}, MS_e^{(t-1)}, \boldsymbol{\sigma}^{(t-1)}, \alpha_e, \beta_e$ from the distribution defined in (6.23);
  2. Calcualte $\sigma_a^{2(t)} = (1 - \sigma_e^{2(t)})/2$;
  for $i, i' \in \{1, \ldots, n\}$ sources **do**
    3. Draw $\theta_{ii'}^{(t)} | \boldsymbol{s}_{ii'}, \sigma_{ii'}^{(t-1)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \phi_{ii'}, \omega_{ii'}$ from the distribution defined in (6.24);
    4. Draw $\sigma_{ii'}^{2(t)} | \boldsymbol{s}_{ii'}, \theta_{ii'}^{(t)}, \sigma_a^{2(t)}, \sigma_e^{2(t)}, \alpha_{ii'}, \beta_{ii'}$ from the distribution defined in (6.25);
  end
end

Because the posterior distribution defined for σ_e^2 in (6.23) is not readily available, we cannot directly sample from this distribution. As a result, the first step in Algorithm 3 is not so straightforward - indeed, obtaining a sample from the posterior distribution of σ_e^2 requires introducing a Metropolis-Hastings algorithm [14, 44]. This procedure is summarized in Algorithm 4.

Algorithm 4: Metropolis-Hastings algorithm for obtaining a sample from the posterior distribution of σ_e^2

Data: Value of $\sigma_e^{2(t-1)}$; values for hyperparameters α_e, β_e
Result: Posterior sample of $\sigma_e^{2(t)}$
1. Sample a candidate value, $\sigma_e^{2*} \sim \mathcal{B}(2, \sigma_e^{-2(t-1)})$;
2. Calculate the value of MS_e^* using the candidate value σ_e^{2*} using (6.22);
3. Calculate the value of $MS_e^{(t-1)}$ using the current value $\sigma_e^{2(t-1)}$ using (6.22);
4. Evaluate the posterior density of σ_e^{2*}, f^*, using the hyperparameters α_e and β_e, the value of MS_e^*, and (6.23);
5. Evaluate the posterior density of $\sigma_e^{2(t-1)}$, $f^{(t-1)}$, using the hyperparameters α_e and β_e, the value of $MS_e^{(t-1)}$, and (6.23);
6. Calculate the probability of acceptance, $p_{acc} = \frac{f^*}{f^{(t-1)}} \frac{\mathcal{B}\left(\sigma^{2(t-1)}|2, \sigma^{-2*}\right)}{\mathcal{B}\left(\sigma^{2*}|2, \sigma^{-2(t-1)}\right)}$
7. Generate a random probability, $p^* \sim \mathcal{U}(0, 1)$;
8. If $p_{acc} \geq p^*$, then define $\sigma_e^{2(t)} := \sigma_e^{2*}$; otherwise define $\sigma_e^{2(t)} := \sigma_e^{2(t-1)}$;

Now that we have identified a method for obtaining samples of the parameters used to define $\boldsymbol{\theta}$, $\boldsymbol{\Delta}$, σ_a^2, and σ_e^2, we can assign a posterior probability,

$$\begin{aligned} \pi\left(\mathcal{H}_i|\boldsymbol{s}_t,\boldsymbol{s}_c,\boldsymbol{\Omega}\right) &= \frac{\pi\left(\mathcal{H}_i\right)\int_{\boldsymbol{\Omega}_i}\ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_i,\boldsymbol{s}_c\right)d\Pi\left(\boldsymbol{\Omega}_i|\boldsymbol{s}_c\right)}{\sum_{i'=1}^{n}\pi\left(\mathcal{H}_{i'}\right)\int_{\boldsymbol{\Omega}_{i'}}\ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_{i'},\boldsymbol{s}_c\right)d\Pi\left(\boldsymbol{\Omega}_{i'}|\boldsymbol{s}_c\right)} && (6.26) \\ &\approx \frac{\pi\left(\mathcal{H}_i\right)\frac{1}{T}\sum_{t=1}^{T}\ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_i^{(t)},\boldsymbol{s}_c\right)}{\sum_{i'=1}^{n}\left(\pi\left(\mathcal{H}_{i'}\right)\frac{1}{T}\sum_{t_{i'}=1}^{T}\ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_{i'}^{(t_{i'}),\boldsymbol{s}_c}\right)\right)} && (6.27) \end{aligned}$$

where $\boldsymbol{s}_t$ is the vector of scores that consider at least one trace object, the subscripts on $\mathcal{H}$ and $\boldsymbol{\Omega}$ correspond to the model being considered, and $\boldsymbol{\Omega}^{(t)}$ are posterior samples of the parameters obtained using Algorithm 3. We consider the conditional posterior probability of the scores that consider objects of unknown origin from the set $\boldsymbol{X}_u$, $\boldsymbol{s}_t$, rather than the joint posterior probability of $\boldsymbol{s}$, so as not to recycle the information contained in the vector of score $\boldsymbol{s}_c$, which are used to sample the parameter values.

Chapter 7

IMPLEMENTING THE MULTI-CLASS MODEL SELECTION ALGORITHM

In this section, we again apply the proposed model to the MNIST Handwritten Digit Data [39]. The MNIST Handwritten Digit Data consists of approximately 70,000 observations of handwritten digits. Each observation is a 28×28 pixel image of an integer, 0 through 9.

As in the two-class scenario, we analyse the performance of our model and compare its performance to an SVM (we use the $ksvm()$ function from the $kernlab$ package [38]), typically used for binary classification [12, 65, 74, 89, 90]. We have elected to compare the abilities of the proposed model and an SVM to differentiate between the digits 3, 5, 6, 8, and 9. For each model, we consider a radial basis function kernel, given that this kernel has been shown to work well for SVMs on this data set.

To assess the performance of the model, we consider a series of simulations in which we consider $n_0 = 5$, 10, and 15 control objects per source, and $n_u = 3$ trace objects. For a series of simulations, we consider the performance of the models when a fixed set of control objects is considered alongside 125 sets of trace objects (25 from each source). That is, for a single iteration, we sample n_0 control objects from sources 1 through 5 (digits 3, 5, 6, 8 and 9), and 25 sets of n_u trace objects from each source, for a total of 125 sets of trace objects. The two sets of n_0 control objects are used to determine the source of the 125 sets of trace objects. This process is repeated 100 times, and the average performance is assessed. The results are presented in Table 7.1.

n_0	SVM (Overall Performance)	SVM (Triplet Performance)	SVM (Voting Performance)	Proposed Model Performance
5	63.65%	9.58%	23.88%	61.56%
10	73.07%	15.30%	28.00%	77.90%
15	76.67%	16.68%	29.08%	81.26%

Table 7.1: Performance of SVM versus multi-class model when $n_0 = 5, 10$, and 15 control objects and $n_u = 3$ trace objects for 125 iterations of the experiment. The overall performance of the SVM gives the average percentage of correct classifications. The triplet performance of the SVM gives the average percentage of sets of n_u that were entirely correctly classified. The voting performance of the SVM gives the percentage of sets of n_u that would be correctly classified if a voting system were used to determine the class of the set of trace objects. The performance of the Proposed model gives the average percentage of sets of n_u that were correctly classified.

The results in Table 7.1 present the results of the experiment described above. We see that, as the number of control objects increases, the rates of correct classification increase across all columns. Notably, Table 7.1 indicates that when we consider $n_0 = 5$ control objects, the overall performance of the SVM is approximately the same as the performance of the proposed model. However, when we compare the triplet performance or voting performance of the SVM to that of the proposed model, we see that the proposed model drastically out-performs the SVM. When we move to consider $n_0 = 10$ and $n_0 = 15$ control objects, the proposed model outperforms the SVM in terms of overall performance, in terms of triplet performance and in terms of voting performance. Thus, we see that the proposed model far outperforms the SVM when it comes to classifying the entire set of trace objects.

Finally, as in the two-class scenario, it is worth mentioning that the computational cost associated with each model is different. Overall, the SVM is more computationally efficient. At this time, we cannot directly compare the two models: first, the SVM package used in this experiment is coded in C, while the proposed model is coded in R; second, the SVM is not Bayesian, while the proposed model is, and so some inherent computational costs exist. That being said, if a Bayesian alternative of the SVM were considered alongside the proposed model, we can speculate that the SVM would be more computationally efficient, since the proposed model involves inverting a covariance matrix, which is a step that is not required by the SVM.

Chapter 8

EVALUATING THE MULTI-CLASS MODEL SELECTION ALGORITHM

In this part, we considered the development for a multi-class, $n > 2$ classification algorithm that allows for making inference on the source of a set of test objects known to originate from one of n potential sources. This method is novel in that it allows for classifying the complete set of objects at once, rather than classifying each object in turn. This method relies on a kernel function, which allows for considering virtually any set of high-dimensional, complex, heterogeneous data as a single vector of real-values scores between observations by merely modifying the kernel to accommodate the considered data. In addition, our method is particularly well-suited for scenarios in which a limited number of observations are available for consideration, as is oftentimes the case in forensic scenarios.

An evaluation of this performance of the proposed model indicates that the model performs just as well as the SVM when $n_0 = 5$ control objects are considered, and surpasses the performance of the SVM when $n_0 = 10$ control objects are considered. In addition, the performance of the model is not affected as the number of considered sources increases. We do note, however, that the computational time increases drastically as n increases, more-so than when n_0 or n_u increase.

The performance of this model indicates that the model works well in the multi-class scenario, and that it is reasonable to move on to consider the full model, in which we wish to determine whether an object is more likely to originate from a given source than from a random source in a population of potential sources.

Part IV

A Population-Based Model Selection Algorithm

OVERVIEW OF PART IV: A POPULATION-BASED MODEL SELECTION ALGORITHM FOR HIGH-DIMENSIONAL AND COMPLEX DATA

In this part, we develop the theory and implementation for a model-selection algorithm that considers a putative source versus a population of random sources. In Chapter 9, we define the problem and develop the algorithm that allows for determining whether a specific putative source is more likely to have produced a set of trace objects than some other random source in a population of potential sources. In addition, we propose a method for studying the parameters of the proposed model, and a sampling algorithm that can be used to study the distributions of the considered parameters. In addition, we consider the ability of the model to recover the parameters under a fixed scenario, and we investigate the scenarios in which the Normality assumption becomes reasonable.

In Chapter 10, we conduct a series of simulations to assess the performance of the model as we vary the number of random sources used to characterize the population and the number of objects considered per random source. In addition, we evaluate the proposed model using a forensic dataset consisting in FTIR spectra of paint chips.

In Chapter 11 we discuss the benefits and limitations of the proposed population-based model.

Chapter 9

DEFINING THE POPULATION-BASED MODEL-SELECTION PROBLEM

We conclude this dissertation by considering a population-based scenario in which we are interested in determining whether a specific, fixed source is more likely to have produced a set of objects of unknown but common origin. That is, given a set of n_0 objects known to have originated from a known source of interest, $\boldsymbol{X}_k$, a set, $\mathscr{P}$, consisting of several sets of n_0 objects, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$, known to have originated from some other random source in a population of potential sources, and a set of n_u objects, $\boldsymbol{X}_u$ of common but unknown origin, we are interested in determining whether the source that produced the objects in $\boldsymbol{X}_k$ is more likely to have produced the set of trace objects $\boldsymbol{X}_u$ than some other random source in the population of potential sources. Formally, we are interested in determining if

$\mathcal{H}_p$: $\boldsymbol{X}_u$ is a simple random sample from the source that produced $\boldsymbol{X}_k$;

$\mathcal{H}_d$: $\boldsymbol{X}_u$ is a simple random sample from some other random source in a population of potential sources characterized by $\mathscr{P}$.

As discussed in Chapter 3, differentiating between these propositions cannot be reduced to a simple classification or model-selection problem that can be addressed using machine learning or likelihood-based techniques. As before, small sample sizes rule out many machine learning techniques, and high-dimensional, complex, or heterogenous data make it impossible to assign the necessary probability measures for assigning Bayes factors, or performing likelihood-based inference.

We propose a model that leverages the properties of kernel functions (see Chapter 1) to obtain a vector of scores, $\boldsymbol{s}$, that consists in all pairwise comparisons of all objects observed in $\boldsymbol{X}_k$, $\mathscr{P}$, and $\boldsymbol{X}_u$. This vector consists of within-source scores, which arise when compared objects originate from a common source, and between-source scores, which arise when compared objects originate from two different sources. The model capitalizes on the variability that exists within and between these sets of scores to address the above inference question. Because the method relies on a kernel function, the method can be tailored to any type of data by merely modifying this function, and the overall inference process remains the same. Furthermore, the model relies on a single assumption, which can be satisfied through the design of the kernel function.

9.1 PROBLEM STATEMENT

Consider a set of exchangeable observations, $\boldsymbol{X}_k$, made on objects known to have been produced by a known, suspected source, a set, $\mathscr{P}$, consisting in r sets, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$, of exchangeable observations from r random sources from the population of potential sources, and the set of exchangeable observations, $\boldsymbol{X}_u$, made on objects of common but unknown origin. The sets $\boldsymbol{X}_k$ and $\mathscr{P}$ are considered to be sets of control objects, while the set $\boldsymbol{X}_u$ is considered to be a set of test objects. We define the sets $\boldsymbol{X}_k$, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$, and $\boldsymbol{X}_u$ as being simple random samples,

$$
\begin{aligned}
\boldsymbol{X}_k &:= \{\boldsymbol{x}_{k,1}, \boldsymbol{x}_{k,2}, \ldots, \boldsymbol{x}_{k,n_0}\};\\
\boldsymbol{X}_1 &:= \{\boldsymbol{x}_{1,1}, \boldsymbol{x}_{1,2}, \ldots, \boldsymbol{x}_{1,n_0}\};\\
\boldsymbol{X}_2 &:= \{\boldsymbol{x}_{2,1}, \boldsymbol{x}_{2,2}, \ldots, \boldsymbol{x}_{2,n_0}\};\\
&\vdots\\
\boldsymbol{X}_p &:= \{\boldsymbol{x}_{p,1}, \boldsymbol{x}_{p,2}, \ldots, \boldsymbol{x}_{p,n_0}\};\\
&\vdots\\
\boldsymbol{X}_r &:= \{\boldsymbol{x}_{r,1}, \boldsymbol{x}_{r,2}, \ldots, \boldsymbol{x}_{r,n_0}\};\\
\boldsymbol{X}_u &:= \{\boldsymbol{x}_{u,1}, \boldsymbol{x}_{u,2}, \ldots, \boldsymbol{x}_{u,n_u}\};
\end{aligned}
$$

where the sets of control objects consist in n_0 objects from their respective sources, and the set of test objects consists in n_u objects known to originate from either the putative source characterized by $\boldsymbol{X}_k$, or by some other source in a population of potential sources, characterized by $\mathscr{P} := \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r\}$. We are interested in quantifying the extent of the support provided to $\mathcal{H}_p$ and $\mathcal{H}_d$ above.

Rather than consider the observations themselves, we instead consider the vector of all pairwise *scores*, $\boldsymbol{s} \in \mathbb{R}^N, N = \binom{(r+1)n_0+n_u}{2}$, obtained by comparing the observations in the sets $\boldsymbol{X}_k$, $\mathscr{P}$, and $\boldsymbol{X}_u$ via some kernel,

$$
\begin{aligned}
&\kappa : \mathbb{R}^m \mapsto \mathbb{R},\\
&\kappa\left(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}\right) = \left\langle \phi\left(\boldsymbol{x}_{ij}\right), \phi\left(\boldsymbol{x}_{i'j'}\right)\right\rangle \quad i, i' \in \{1, \ldots, r, k, u\},\ j, j' \in \{1,\ \ldots,\ \max\{n_0, n_u\}\},
\end{aligned}
$$

where ϕ is a mapping into some separable, high-dimensional Hilbert space [12, 65, 71, 74]. As before, $s_{ij,i'j'}$ is the score obtained by comparing object $\boldsymbol{x}_{ij}$ to object $\boldsymbol{x}_{i'j'}$ using some kernel function, κ (see Definition 2 in Section 1.1).

We define our kernel function such that our vector of scores is distributed according to a Multivariate Normal distribution, with

$$\boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right), \tag{9.1}$$

where $\boldsymbol{\theta}$ is the vector of the mean terms, and $\boldsymbol{\Sigma}$ is the covariance matrix associated with the vector of scores (see Section 9.5 and Chapter 10 for a discussion on the validity of this assumption, and the implications when this assumption does not hold). These parameters will collectively be referred to as $\boldsymbol{\Omega} := \{\boldsymbol{\theta}, \boldsymbol{\Sigma}\}$, and we will more explicitly define $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ as we move through the chapter. We can assign a Bayes Factor in terms of the Multivariate Normal Likelihood and the associated parameter $\boldsymbol{\Omega}$, such that

$$\begin{aligned} \Lambda &= \frac{\int_{\boldsymbol{\Omega}} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_d\right) d\Pi\left(\boldsymbol{\Omega}|\mathcal{H}_p\right)}{\int_{\boldsymbol{\Omega}} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}, \mathcal{H}_d\right) d\Pi\left(\boldsymbol{\Omega}|\mathcal{H}_d\right)} \\ &:= \frac{\int_{\boldsymbol{\Omega}_p} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}_p\right) d\Pi\left(\boldsymbol{\Omega}_p\right)}{\int_{\boldsymbol{\Omega}_d} \ell\left(\boldsymbol{s}|\boldsymbol{\Omega}_d\right) d\Pi\left(\boldsymbol{\Omega}_d\right)}. \end{aligned} \tag{9.2}$$

It is worth noting that there exist differences between $\boldsymbol{\Omega}_p$ and $\boldsymbol{\Omega}_d$. While the individual elements of each of the parameters ($\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_d$, and $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_d$) are restricted to the same set of potential values, the structures of the mean vectors and covariance matrices depend on which proposition is being considered. That is, when $\mathcal{H}_p$ is being considered, then the parameter $\boldsymbol{\Omega}_p$ considers that the set of unknown objects, $\boldsymbol{X}_u$, originates from the putative source that produced the objects in $\boldsymbol{X}_k$. Likewise, when $\mathcal{H}_d$ is being considered, then the parameter $\boldsymbol{\Omega}_d$ considers that the set of unknown objects originates from some other random source in a population of potential sources characterized by $\mathscr{P}$. Under $\mathcal{H}_p$, scores that consider at least one trace object and that are of the form $s_{uj,uj'}$ or $s_{kj,uj'}$, are considered to be within-source scores, while those of the form $s_{uj,pj'}$, $p \in \{1, \ldots, r\}$, are considered to be between-source scores. In the same way, under $\mathcal{H}_d$, scores that consider at least one trace object and that are of the form $s_{uj,uj'}$ are considered to be within-source scores, while those of the form $s_{kj,uj'}$ or $s_{uj,pj'}$ are considered to be between-source scores.

Due to the nature of this problem, we consider two types of within- and between-source scores. That is, rather than consider within-source i, as in the previous two models, we consider that scores may be within the putative source (in which two objects from the putative source are compared), within random sources from the population (in which two objects from the same random source in the population are compared), between the putative source and a random source from the population (in which an object from the putative source is compared to an object from a random source used to characterize the population), or between population source (in which two objects arise from different random sources used to characterize the population). Bearing this in mind, then, under $\mathcal{H}_p$, scores of the form $s_{uj,uj'}$ or $s_{kj,uj'}$ are considered to be scores within the putative source, and scores of the form $s_{uj,pj'}$ are considered to be between the putative source and the population, while under $\mathcal{H}_d$, scores of the form $s_{uj,uj'}$ are considered to be within a random source from the population, scores of the form $s_{kj,uj'}$ are considered to be between the putative source and a random source from the population, and scores of the form $s_{uj,pj'}$ are considered to be between two random sources from the population.

9.1.1 COVARIANCE STRUCTURE OF THE OBJECT MODEL

In the previous model, we investigated the covariance structure for a vector of scores by considering a single univariate object-based model. In this scenario, however, we consider a pair of univariate object-based models, in which the linear model is contingent upon whether the object, x_{ij}, is randomly sampled from the population, or from the known and fixed putative source. In the case where an object is sampled from a randomly selected source in a population of potential sources, we consider a term x_{pj} defined in terms of the linear model given by

$$x_{pj} = \mu + \theta_p + \varepsilon_{pj}, \tag{9.3}$$

where μ is the overall mean of the population, θ_p is the difference between the overall mean of the population, μ, and that of the individual source randomly sampled from the population, $\mu + \theta_p$, $p \in \{1, \ldots, r\}$, such that $\theta \sim \mathcal{N}(0, \tau^2)$, and ε_{pj} is a lack of fit term, such that $\varepsilon_{pj} \sim \mathcal{N}(0, \sigma^2)$. In the case where an object is sampled from the fixed, putative source, we consider a term x_{kj} defined in terms of the linear model given by

$$x_{kj} = \mu_k + \varepsilon_{kj}, \tag{9.4}$$

where μ_k is the overall mean of the putative source, and ε_{kj} is a lack of fit term, such that $\varepsilon_{kj} \sim \mathcal{N}(0, \rho^2)$. We proceed by studying covariance structure of scores obtained under the non-stationary kernel (first presented in Sections 3.1.1 and 6.1.1 as (3.5) and (6.5), respectively), given by

$$s^*_{ij,i'j'} := x_{ij}x_{i'j'}. \tag{9.5}$$

As before, we can directly examine the mean and covariance terms associated with this kernel by calculating the various terms that arise from the different possible score combinations. Table 9.1 provides the different parameter values under (9.5). In particular, Table 9.1 indicates that we have 4 unique expected value terms and 23 unique covariance terms under the considered non-stationary kernel. Note that we consider only the stationary kernel for this model. Given that the stationary kernel yielded duplicate term in Parts II and III, and so failed to capture all covariance elements under a more complex kernel, we elect to move forward using only the non-stationary kernel described in (9.5).

Description of Considered Scores	Score 1	Score 2	Non-Stationary Kernel (9.5)
Expected Value Terms			
Within Two Random Sources	$p1p2$	-	$\mu^2 + \tau^2$
Within Putative Source	$k1k2$	-	μ_k^2
Between Two Random Sources	$p1p'2$	-	μ^2
Between Putative and Random Sources	$k1p1$	-	$\mu\mu_p$
Covariance Terms			
Both Within Putative Source, Two Common Objects	$k1k2$	$k1k2$	$2\mu_k^2\rho^2 + \rho^4$
Both Within Putative Source, One Common Object	$k1k2$	$k1k3$	$\mu_k^2\rho^2$
Both Within Putative Source, No Common Objects	$k1k2$	$k3k4$	0
Both Within Random Source, Two Common Objects, Two Common Sources	$p1p2$	$p1p2$	$2\tau^2\left(\tau^2 + 2\mu^2 + \sigma^2\right) + 2\mu^2\sigma^2$
Both Within Random Source, One Common Object, Two Common Sources	$p1p2$	$p1p3$	$2\tau^2\left(\tau^2 + \sigma^2\right) + \mu^2\left(4\tau^2 + \sigma^2\right)$
Both Within Random Source, No Common Objects, Two Common Sources	$p1p2$	$p3p4$	$2\tau^4 + 4\mu^2\tau^2$
Both Within Random Sources, No Common Objects, No Common Sources	$p1p2$	$p'1p'2$	0
Both Between Putative and Random Sources, Two Common Objects, Two Common Sources	$k1p1$	$k1p1$	$\left(\rho^2 + \mu_k^2\right)\left(\sigma^2 + \tau^2\right) + \mu^2\rho^2$
Both Between Putative and Random Sources, One Common Object from Source k, Same Random Source	$k1p1$	$k1p2$	$\mu^2\rho^2 + \mu_k^2\tau^2$
Both Between Putative and Random Source, One Common Object from Source k, Different Random Sources	$k1p1$	$k1p'1$	$\mu^2\rho^2$
Both Between Putative and Random Source, One Common Object from Source p	$k1p1$	$k2p1$	$\mu_k^2\left(\tau^2 + \sigma^2\right)$
Both Between Putative and Random Sources, No Common Objects, Two Common Sources	$k1p1$	$k2p2$	$\mu_k^2\tau^2$
Both Between Putative and Random Sources, No Common Objects, Putative Source in Common	$k1p1$	$k2p'2$	0
Both Between Random Sources, Two Common Objects	$p1p'1$	$p1p'1$	$\left(\tau^2 + \sigma^2 + 2\mu^2\right)\left(\tau^2 + \sigma^2\right)$
Both Between Random Sources, One Common Object, Two Common Sources	$p1p'1$	$p1p'2$	$2\mu^2\tau^2 + \mu^2\sigma^2 + \tau^4$
Both Between Random Sources, One Common Object, One Common Source	$p1p'2$	$p1p''1$	$\mu^2\left(\tau^2 + \sigma^2\right)$
Both Between Random Sources, No Common Objects, Two Common Sources	$p1p'1$	$p2p'2$	$2\mu^2\tau^2$
Both Between Random Sources, No Common Objects, One Common Source	$p1p'1$	$p2p''1$	$\mu^2\tau^2$
Both Between Random Sources, No Common Objects, No Common Sources	$p1p'1$	$p''1p'''1$	0
Within Putative Source, Within Random Source, No Common Objects	$k1k2$	$p1p2$	0
Within Putative Source, Between Putative and Random Source, One Common Object	$k1k2$	$k1p1$	$\mu\mu_k\rho^2$
Within Putative Source, Between Putative and Random Source, No Common Objects	$k1k2$	$k3p1$	0
Within Random Source, Between Putative and Random Source, One Common Object, One Common Source	$p1p2$	$k1p1$	$\mu\mu_k\left(2\tau^2 + \sigma^2\right)$
Within Random Source, Between Putative and Random Source, No Common Objects, One Common Source	$p1p2$	$k1p3$	$2\mu\mu_k\tau^2$
Within Random Source, Between Putative and Random Source, No Common Objects, No Common Sources	$p1p2$	$k1p'1$	0
Between Putative and Random Source, Between Random Sources, One Common Object, One Common Source	$k1p1$	$p1p'1$	$\mu\mu_k\left(\tau^2 + \sigma^2\right)$
Between Putative and Random Source, Between Random Sources, No Common Objects, One Common Source	$k1p1$	$p2p'1$	$\mu\mu_k\tau^2$
Between Putative and Random Source, Between Random Sources, No Common Objects, No Common Sources	$k1p1$	$p'1p''1$	0
Within Random Sources, Between Random Sources, One Common Object, One Common Source	$p1p2$	$p1p'1$	$\mu^2\left(2\tau^2 + \sigma^2\right)$
Within Random Sources, Between Random Sources, No Common Objects, One Common Source	$p1p2$	$p3p'1$	$2\mu^2\tau^2$
Within Random Sources, Between Random Sources, No Common Objects, No Common Sources	$p1p2$	$p'1p''1$	0

Table 9.1: Expected value and covariance terms obtained for the object-based model for each type of score comparison when a non-stationary kernel (e.g., (9.5)) is considered. Column one provides descriptions of each type of comparison that may be observed; columns two and three provide examples of indices for scores that could be compared in each situation; column four presents the parameter values obtained under the non-stationary kernel given by (9.5).

9.1.2 DEFINING A SCORE-BASED MODEL

Suppose, now, that we expand upon (9.1) and define our vector of scores such that

$$\begin{aligned} \boldsymbol{s} \sim \mathcal{N}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right) &\implies s_{ij,i'j'} \sim \mathcal{MVN}\left(\theta_{ii'}, \sigma^2_{ii'}\right) \\ &\implies \frac{s_{ij,i'j'} - \theta_{ii'}}{\sigma_{ii'}} \sim \mathcal{N}\left(0,1\right) \end{aligned} \tag{9.6}$$

where the parameters $\theta_{ii'}$ and $\sigma_{ii'}$, $ii' \in \{kk, k\mathscr{P}, \mathscr{P}\mathscr{P}, \mathscr{P}\mathscr{P}'\}$[1], are the means and standard deviations associated with the different comparisons that may be considered by a given score. That is, each score in the vector of scores is either one of two possible within-source scores, or one of two possible between-source scores. For example, consider a score $s_{ij,i'j'}$ in which $i = i' = k$, and so $\theta_{ii'}$ gives us the expected value of scores that compare any two objects in $\boldsymbol{X}_k$. Likewise, when $i = i' = \mathscr{P}$, $\theta_{ii'}$ gives us the expected value of scores that compare any two objects in $\boldsymbol{X}_p \in \mathscr{P}$. When $i \neq i'$, $i = k$, $i' = \mathscr{P}$, $\theta_{ii'}$ gives us the expected value of scores that compare an object in $\boldsymbol{X}_k$ to an object in $\mathscr{P}$. Finally, when $i \neq i'$, $i = \mathscr{P}, i' = \mathscr{P}'$, $\theta_{ii'}$ gives us the expected value of scores that compare two objects from different random sources in $\mathscr{P}$. Thus, we consider $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$. The parameter $\sigma_{ii'}$ can be equivalently defined for the standard deviation terms.

From here, following the work of Gantz and Saunders, and Armstrong [5, 6, 26], and the development in Sections 3.1.2 and 6.1.2, we choose to define the standardized scores from (9.6) according to a random effects model

$$\frac{s_{ij,i'j'} - \theta_{ii'}}{\sigma_{ii'}} = a_{ij} + a_{i'j'} + b_i + b_{i'} + c_i + d_{i:ij} + d_{i:i'j'} + d_{i':ij} + d_{i':i'j'} + e_{ij,i'j'}, \tag{9.7}$$

[1]The set $\mathscr{P}$ is the set of r random sources, $\{X_1, X_2, \ldots, X_r\}$ used to characterize the population. When used as a subscript, it indicates that we are considering all random sources $p \in \{1, \ldots, r\}$ simultaneously, as opposed to considering a single random source, denoted by p. For example, we could consider the vector of scores $\boldsymbol{s}_p p$, which considers only the within-source-p comparisons. Alternatively, we could consider the vector of scores $\boldsymbol{s}_{\mathscr{P}\mathscr{P}}$, which considers all within-source comparisons in the population. Likewise, we could consider the vector of scores $\boldsymbol{s}_{kp}$, which would compare observations from source k to the observations from source p. In the same way, we could consider the vector of scores $\boldsymbol{s}_{k\mathscr{P}}$, which would compare the observations from source k to any other source in the population of potential sources.

where a_{ij} and $a_{i'j'}$ are random object effects with $a_{ij}, a_{i'j'} \sim \mathcal{N}(0, \sigma_a^2)$, b_i, $b_{i'}$ are random population source interaction effects with $b_i, b_{i'} \sim \mathcal{N}(0, \sigma_b^2)$, c_i is a random putative source interaction effect with $c_i \sim \mathcal{N}(0, \sigma_c^2)$, $d_{i:ij}$, $d_{i:i'j'}$, $d_{i':ij}$, and $d_{i':i'j'}$ are random source-object interaction effects with $d_{i:ij}, d_{i:i'j'}, d_{i':ij}, d_{i':i'j'} \sim \mathcal{N}(0, \sigma_d^2)$, and $e_{ij,i'j'}$ is a lack-of-fit term with $e_{ij,i'j'} \sim \mathcal{N}(0, \sigma_e^2)$. Considering the structure of these matrices (see Figures 9.1, 9.2, 9.3, and 9.4), and following (9.6), we have that $2\sigma_a^2 + \sigma_c^2 + 4\sigma_d^2 + \sigma_e^2 = 2\sigma_a^2 + 2\sigma_b^2 + 4\sigma_d^2 + \sigma_e^2 = 1$. Finally, we rewrite the model in terms of $s_{ij,i'j'}$, such that

$$s_{ij,i'j'} = \theta_{ii'} + \sigma_{ii'} \left(a_{ij} + a_{i'j'} + b_i + b_{i'} + c_i + d_{i:ij} + d_{i:i'j'} + d_{i':ij} + d_{i':i'j'} + e_{ij,i'j'} \right),$$

and so, given the distributional assumptions associated with (9.6) and (6.7), we define the distribution of our vector of scores to be

$$\boldsymbol{s} \sim \mathcal{MVN}\left(\boldsymbol{\theta}, \boldsymbol{\Delta}\left(\boldsymbol{P}\boldsymbol{P}'\sigma_a^2 + \boldsymbol{Q}\boldsymbol{Q}'\sigma_b^2 + \boldsymbol{R}\boldsymbol{R}'\sigma_c^2 + \boldsymbol{T}\boldsymbol{T}'\sigma_d^2 + \boldsymbol{I}\sigma_e^2\right)\boldsymbol{\Delta}'\right), \qquad (9.8)$$

where $\boldsymbol{\theta}$ is a vector of length N of the mean terms given by $\theta_{ii'}$, and $\boldsymbol{\Delta}$ is an $N \times N$ diagonal matrix of the standard deviations given by $\sigma_{ii'}$. The matrix $\boldsymbol{I}$ is the $N \times N$ identity matrix. The design matrices $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{T}$ each describe an effect the different scores considered in the vector $\boldsymbol{s}$. Each of these design matrices is constructed by individually considering the potential source combinations that are of interest for the effect being considered. The following sections outline the construction of these matrices.

9.1.2.1 DESIGN MATRIX $\boldsymbol{P}$

The $((r+1)n_0 + n_u) \times \binom{(r+1)n_0}{2}$ design matrix $\boldsymbol{P}$ describes the effects of the objects being compared on the score for each score considered in the vector $\boldsymbol{s}$. The rows of $\boldsymbol{P}$ consist in all pairwise combinations of all $((r+1)n_0 + n_u)$ objects, while the columns of $\boldsymbol{P}$ consist in the objects themselves (e.g., $k1, k2, \ldots, kn_0, 11, 12, \ldots, 1n_0, \ldots, r1, \ldots, rn_0$) . To construct the design matrix $\boldsymbol{P}$ requires considering three subcomponents of the overall $\boldsymbol{P}$ matrix. We will refer to these as $\boldsymbol{P}_k$, $\boldsymbol{P}_{\mathscr{P}}$, and $\boldsymbol{P}_{k,\mathscr{P}}$. Each of these three matrices is of the same dimension as $\boldsymbol{P}$, and the columns of the matrices are of the same organizational structure. Considering different combinations of these matrices allows us to construct the full design matrix , $\boldsymbol{P}\boldsymbol{P}'$.

The design matrix $\boldsymbol{P}_k$ is constructed by considering only those rows whose associated scores consider two objects from the putative source. For the rows of $\boldsymbol{P}_k$ that consider a within-putative source score, a value of 1 is placed in the columns corresponding to the considered objects, and a value of 0 is placed in the remaining columns. For example, given a score $s_{k1,k2}$, the columns associated with the first and second objects from the putative sources (i.e., "$k1$", "$k2$") are assigned a value of 1, and all other columns are assigned a value of 0.

The design matrix $\boldsymbol{P}_{\mathscr{P}}$ is constructed by considering only those rows whose associated scores consider two objects from the population, regardless of whether or not those objects arise from the same source within the population. For the rows of $\boldsymbol{P}_{\mathscr{P}}$ that consider two objects from the population, a value of 1 is placed in the columns corresponding to the considered objects, and a value of 0 is placed in the remaining columns. For example, given a score $s_{p1,p2}$, the columns associated with the first and second objects from source p in the population (i.e., "$p1$", "$p2$") are assigned a value of 1, and all other columns are assigned a value of 0. Likewise, given a score $s_{p1,p'1}$, the columns associated with the first objects from sources p and p' in the population (i.e., "$p1$", "$p'1$") are assigned a value of 1, and all other columns are assigned a value of 0.

Finally, the design matrix $\boldsymbol{P}_{k,\mathscr{P}}$ is constructed by considering only those rows whose associated scores consider an object from the putative source alongside an object from the population. For the rows of $\boldsymbol{P}_{k,\mathscr{P}}$ that consider scores of this type, a value of 1 is placed in the columns corresponding to the considered objects, and a value of 0 is placed in the remaining columns. For example, given a score $s_{k1,p1}$, the columns associated with the first objects from the putative source and from source p in the population (i.e., "$k1$", "$p1$") are assigned a value of 1, and all other columns are assigned a value of 0.

We can then use these three matrices to construct our final design matrix, given by $\boldsymbol{PP}'$. Specifically, $\boldsymbol{PP}'$ is defined by considering all pairwise combinations of $\boldsymbol{P}_k$, $\boldsymbol{P}_{\mathscr{P}}$, and $\boldsymbol{P}_{k,\mathscr{P}}$. That is,

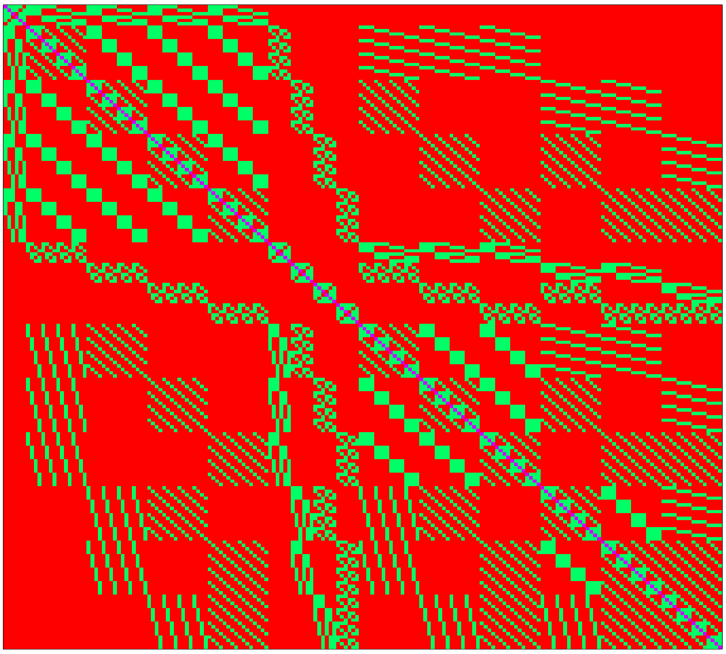
$$\boldsymbol{PP}' := \boldsymbol{P}_k\boldsymbol{P}_k' + \boldsymbol{P}_{\mathscr{P}}\boldsymbol{P}_{\mathscr{P}}' + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}_{k,\mathscr{P}}' + \left(\boldsymbol{P}_k\boldsymbol{P}_{k,\mathscr{P}}' + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}_k'\right) + \left(\boldsymbol{P}_{\mathscr{P}}\boldsymbol{P}_{k,\mathscr{P}}' + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}_{\mathscr{P}}'\right).$$

Figure 9.1 portrays the resulting $\boldsymbol{PP}'$ matrix, along with the five different combinations of the three sub-matrices matrices used to construct $\boldsymbol{PP}'$.

9.1.2.2 DESIGN MATRIX Q

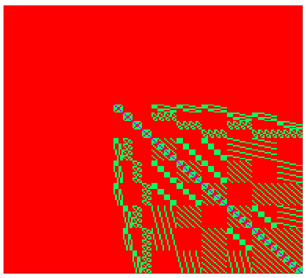
The $((r+1)n_0 + n_u) \times (r+1)$ design matrix $\boldsymbol{Q}$ describes the effect of random sources from the population of potential sources on the score for each score considered in the vector $\boldsymbol{s}$. The rows of $\boldsymbol{Q}$ consist in all $((r+1)n_0 + n_u)$ objects, while the columns of $\boldsymbol{Q}$ consist in the $r+1$ sources being considered (e.g., $k, 1, \ldots, p$). To construct the design matrix $\boldsymbol{Q}$ requires considering two sub-design matrices of the overall $\boldsymbol{Q}$ matrix. We will refer to these as $\boldsymbol{Q}_{\mathscr{P}}$ and $\boldsymbol{Q}_{k,\mathscr{P}}$. Note that we do not need to consider $\boldsymbol{Q}_k$, since we are looking at the effect of the random sources from the population of potential sources. This pair of matrices is of the same dimension as $\boldsymbol{Q}$, and the columns of the matrices are of the same organizational structure. Considering different combinations of these matrices allows us to construct the full design matrix, $\boldsymbol{QQ}'$.

The design matrix $\boldsymbol{Q}_{\mathscr{P}}$ is constructed by considering only those rows whose associated scores consider two objects from the population, regardless of whether or not those objects arise from the same source within the population. For the rows of $\boldsymbol{Q}_{\mathscr{P}}$ that consider two objects from the same random source in the population of potential sources, a value of $\sqrt{2}$ is placed in the columns corresponding to the source of the considered objects, and a value of 0 is placed in the remaining columns. For the rows of $\boldsymbol{Q}_{\mathscr{P}}$ that consider two objects from different random sources in the population of potential sources, a value of 1 is placed in the columns corresponding to the sources of the considered objects, and a value of 0 is placed in the remaining columns. For example, given a score $s_{p1,p2}$, where both objects are from random source p in the

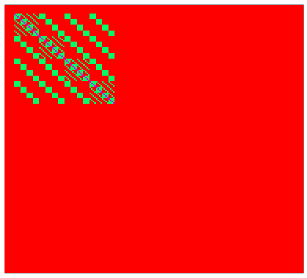


(a) $\boldsymbol{PP'}$

(b) $\boldsymbol{P}_k\boldsymbol{P}'_k$

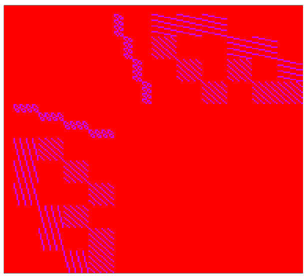


(c) $\boldsymbol{P}_{\mathscr{P}}\boldsymbol{P}'_{\mathscr{P}}$



(d) $\boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}'_{k,\mathscr{P}}$

(e) $\boldsymbol{P}_k\boldsymbol{P}'_{k,\mathscr{P}} + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}'_k$



(f) $\boldsymbol{P}_{\mathscr{P}}\boldsymbol{P}'_{k,\mathscr{P}} + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}'_{\mathscr{P}}$

Figure 9.1: Heat maps of design matrix $\boldsymbol{PP'}$ and sub-design matrices used to construct $\boldsymbol{PP'}$. (a) Final design matrix, $\boldsymbol{PP'}$. (b) Sub-design matrix, $\boldsymbol{P}_k\boldsymbol{P}'_k$, depicting the effect of considering two pairs of objects from the putative source. (c) Sub-design matrix, $\boldsymbol{P}_{\mathscr{P}}\boldsymbol{P}'_{\mathscr{P}}$, depicting the effect of considering two pairs of objects from the population. (d) Sub-design matrix, $\boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}'_{k,\mathscr{P}}$, depicting the effect of considering two pairs of objects that consider an object from the putative source alongside an object from the population. (e) Sub-design matrix, $\boldsymbol{P}_k\boldsymbol{P}'_{k,\mathscr{P}} + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}'_k$, depicting the effect of considering a pair of objects from the putative source alongside a pair of objects that considers one object from the putative source and one object from the population. (f) Sub-design matrix, $\boldsymbol{P}_{\mathscr{P}}\boldsymbol{P}'_{k,\mathscr{P}} + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}'_{\mathscr{P}}$, depicting the effect of considering a pair of objects from the population alongside a pair of objects that considers one object from the population and one object from the putative source.

population, the columns associated with source p are assigned a value of $\sqrt{2}$, and all other columns are assigned a value of 0. Similarly, given a score $s_{p1,p'1}$, where the two objects come from different random sources in the population, the columns associated with sources p and p' are assigned a value of 1, and all other columns are assigned a value of 0. Note that when the two considered objects arise from the same random source, only one position in the row is non-zero, while when the two considered objects arise from different random sources, two positions in the row are non-zero.

The design matrix $\boldsymbol{Q}_{k,\mathscr{P}}$ is constructed by considering only those rows whose associated scores consider an object from the putative source alongside an object from the population. For the rows of $\boldsymbol{Q}_{k,\mathscr{P}}$ that consider an object from the putative source alongside an object from the population, a value of 1 is placed in the columns corresponding to the sources of the two objects, and a value of 0 is placed in the remaining columns. For example, given a score $s_{k1,p1}$, the rows associated with sources k and p are assigned a value of 1, and all other columns are assigned a value of 0.

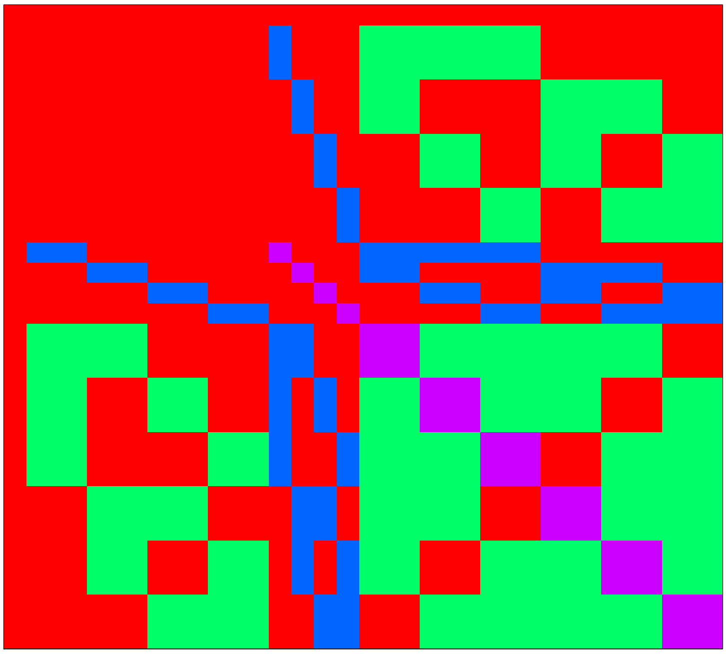
We can use these two matrices to construct our final design matrix, given by $\boldsymbol{QQ}'$. Specifically $\boldsymbol{QQ}'$ is defined by considering combinations of $\boldsymbol{Q}_{\mathscr{P}}$ and $\boldsymbol{Q}_{k,\mathscr{P}}$. That is,

$$\boldsymbol{QQ}' := \boldsymbol{Q}_{\mathscr{P}}\boldsymbol{Q}'_{\mathscr{P}} + \left(\boldsymbol{Q}_{\mathscr{P}}\boldsymbol{Q}'_{k,\mathscr{P}} + \boldsymbol{Q}_{k,\mathscr{P}}\boldsymbol{Q}'_{\mathscr{P}}\right)$$

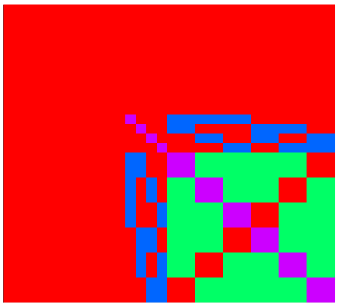
Figure 9.2 portrays the resulting $\boldsymbol{QQ}'$ matrix, along with the two different combinations of the sub-matrices matrices used to construct $\boldsymbol{QQ}'$.

9.1.2.3 DESIGN MATRIX $\boldsymbol{R}$

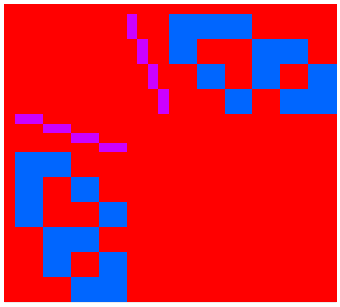
The $((r+1)n_0 + n_u) \times \left(\binom{r+1}{2} + (r+1)\right)$ design matrix $\boldsymbol{R}$ describes the effect of the putative source on the score for each score considered in the vector $\boldsymbol{s}$. The rows of $\boldsymbol{R}$ consist in all $((r+1)n_0 + n_u)$ objects, while the columns of $\boldsymbol{R}$ consist in the $\left(\binom{r+1}{2} + (r+1)\right)$ potential combinations of sources being considered (e.g., $kk, k1, k2, \ldots, k.r, 1.1, 2.2, \ldots, r.r, 1.2, 1.3, \ldots, 1.r, \ldots, r-1.r$). To construct the design matrix $\boldsymbol{R}$ requires considering two sub-design matrices of the overall $\boldsymbol{R}$ matrix.



(a) $\boldsymbol{QQ'}$



(b) $\boldsymbol{Q}_{\mathscr{P}}\boldsymbol{Q}'_{\mathscr{P}}$



(c) $\boldsymbol{Q}_{\mathscr{P}}\boldsymbol{Q}'_{k,\mathscr{P}} + \boldsymbol{Q}_{k,\mathscr{P}}\boldsymbol{Q}'_{\mathscr{P}}$

Figure 9.2: Heat maps of design matrix $\boldsymbol{QQ'}$ and sub-design matrices used to construct $\boldsymbol{QQ'}$. (a) Final design matrix, $\boldsymbol{QQ'}$. (b) Sub-design matrix, $\boldsymbol{Q}_{\mathscr{P}}\boldsymbol{Q}'_{\mathscr{P}}$, depicting the effect of considering two pairs of objects from the population. (c) Sub-design matrix, $\boldsymbol{P}_{\mathscr{P}}\boldsymbol{P}'_{k,\mathscr{P}} + \boldsymbol{P}_{k,\mathscr{P}}\boldsymbol{P}'_{\mathscr{P}}$, depicting the effect of considering a pair of objects from the population alongside a pair of objects that considers one object from the population and one object from the putative source.

We will refer to these as $\boldsymbol{R}_k$ and $\boldsymbol{R}_{k,\mathscr{P}}$. Note that we do not need to consider $\boldsymbol{R}_{\mathscr{P}}$, since we are looking at the effect of the putative source. This pair of matrices is of the same dimension as $\boldsymbol{R}$, and columns of the matrices are of the same organizational structure. Considering different combinations of these matrices allows us to construct the full design matrix $\boldsymbol{R}\boldsymbol{R}'$.

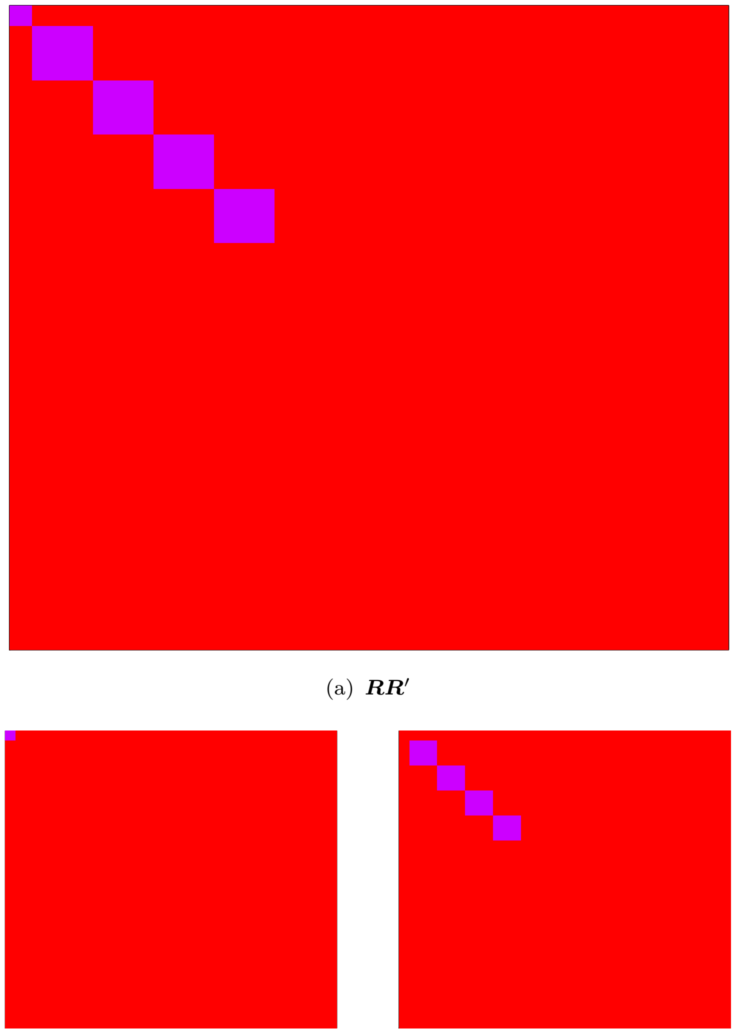
The design matrix $\boldsymbol{R}_k$ is constructed by considering only those rows whose associated scores consider two objects from the putative source. For the rows of $\boldsymbol{R}_k$ that consider two objects from the putative source, a value of 1 is placed in the columns corresponding to the source combination of the considered objects, and a value of 0 is placed in the remaining columns. For example, given a score $s_{k1,k2}$, the column associated with the within-putative source combination (i.e., "$k.k''$) is assigned a value of 1, and all other columns are assigned a value of 0. Note that all scores that are considered by this matrix will result in a value of 1 in this particular column, with a value of 0 in all other columns.

The design matrix $\boldsymbol{R}_{k,\mathscr{P}}$ is constructed by considering only those rows whose associated scores consider an object from the putative source alongside an object from the population of potential sources. For the rows of $\boldsymbol{R}_{k,\mathscr{P}}$ that consider an object from the putative source alongside an object from the population, a value of 1 is placed in the columns corresponding to the sources of the two objects, and a value of 0 is placed in the remaining columns. For example, given a score $s_{k1,p1}$, the column associated with the source combination of k and p (e.g, "$k.p''$) is assigned a value of 1, and all other columns are assigned a value of 0.

We can use these two matrices to construct our final design matrix, given by $\boldsymbol{R}\boldsymbol{R}'$. Specifically, $\boldsymbol{R}\boldsymbol{R}'$ is defined by considering combinations of $\boldsymbol{R}_k$ and $\boldsymbol{R}_{k,\mathscr{P}}$. That is

$$\boldsymbol{R}\boldsymbol{R}' := \boldsymbol{R}_k\boldsymbol{R}_k' + \boldsymbol{R}_{k,\mathscr{P}}\boldsymbol{R}_{k,\mathscr{P}}'$$

Figure 9.3 portrays the resulting $\boldsymbol{R}\boldsymbol{R}'$ matrix, along with the two different combinations of the sub-matrices matrices used to construct $\boldsymbol{R}\boldsymbol{R}'$.



(a) $\boldsymbol{RR}'$

(b) $\boldsymbol{R}_k\boldsymbol{R}_k'$

(c) $\boldsymbol{R}_{k,\mathscr{P}}\boldsymbol{R}_{k,\mathscr{P}}'$

Figure 9.3: Heat maps of design matrix $\boldsymbol{RR}'$ and sub-design matrices used to construct $\boldsymbol{RR}'$. (a) Final design matrix, $\boldsymbol{RR}'$. (b) Sub-design matrix, $\boldsymbol{R}_k\boldsymbol{R}_k'$, depicting the effect of considering two pairs of objects from the putative source. (c) Sub-design matrix, $\boldsymbol{R}_{k,\mathscr{P}}\boldsymbol{R}_{k,\mathscr{P}}'$, depicting the effect of considering two pairs of objects that consider an object from the putative source alongside an object from the population.

9.1.2.4 DESIGN MATRIX $\boldsymbol{T}$

The $((r+1)n_0 + n_u) \times ((r+2)\,((r+1)(n_0) + n_u))$ design matrix $\boldsymbol{T}$ describes the interaction effect of the objects and the sources on the score for each score considered in the vector $\boldsymbol{s}$. The rows of $\boldsymbol{T}$ consist in all $((r+1)n_0 + n_u)$ objects, while the columns of $\boldsymbol{T}$ consist in the $((r+2)\,((r+1)(n_0) + n_u))$ combinations of sources and objects being considered (e.g., $k : k1, k : k2, \ldots, k : kn_0, k : 11, \ldots, k : rn_0, 1 : k1, \ldots 1 : rn_0, \ldots, p : k1, \ldots, p : rn_0, \ldots, r : k1, \ldots, r : rn_0$). To construct the design matrix $\boldsymbol{T}$ requires considering four sub-design matrices of the overall $\boldsymbol{T}$ matrix. We will refer to these as $\boldsymbol{T}_k$, $\boldsymbol{T}_{\mathscr{P}_w}$, $\boldsymbol{T}_{\mathscr{P}_b}$, and $\boldsymbol{T}_{k,\mathscr{P}}$. These matrices are of the same dimension as $\boldsymbol{T}$, and the columns are of the same organizational structure. Considering different combinations of these matrices allows us to construct the full design matrix $\boldsymbol{T}\boldsymbol{T}'$.

The design matrix $\boldsymbol{T}_k$ is constructed by considering only those rows whose associated scores consider two objects from the putative source. For the rows of $\boldsymbol{T}_k$ that consider two objects from the putative source, a value of $\sqrt{2}$ is placed in the columns associated with the source and object combinations of the score being considered, and a value of 0 is placed in the remaining columns. For example, given a score $s_{k1,k2}$, the columns associated with the considered source and objects (i.e., $k : k1$, $k : k2$) are assigned a value of $\sqrt{2}$, and all other columns are assigned a value of 0.

The design matrix $\boldsymbol{T}_{\mathscr{P}_w}$ is constructed by considering only those rows whose associated scores consider two objects from the same random source in the population of potential sources. For the rows of $\boldsymbol{T}_{\mathscr{P}_w}$ that consider two objects from the same random source, a value of $\sqrt{2}$ is placed in the columns associated with the source and object combinations of the score being considered, and a value of 0 is placed in the remaining columns. For example, given a score s_{p1p2}, the columns associated with the considered source and objects (i.e., $p : p1$, $p : p2$) are assigned a value of $\sqrt{2}$, and all other columns are assigned a value of 0.

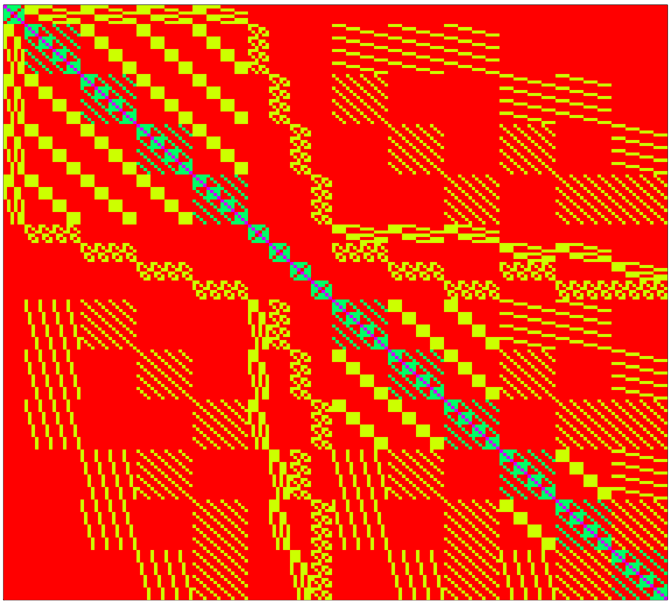
The design matrix $\boldsymbol{T}_{\mathscr{P}_b}$ is constructed by considering only those rows whose associated scores consider two objects from two different random sources in the population of potential sources. For the rows of $\boldsymbol{T}_{\mathscr{P}_b}$ that consider two objects from two different random sources, a value of 1 is placed in the columns associated with the source and object combinations of the score being considered, and a value of 0 is placed in the remaining columns. For example, given a score $s_{p1p'1}$, the columns associated with the considered sources and objects (i.e., $p : p1$, $p : p'2$, $p' : p1$, $p' : p'2$) are assigned a value of 1, and all other columns are assigned a value of 0.

Finally, the design matrix $\boldsymbol{T}_{k,\mathscr{P}}$ is constructed by considering only those rows whose associated scores consider an object from the putative source alongside an object from the population of potential sources. For the rows of $\boldsymbol{T}_{k,\mathscr{P}}$ that consider an object from the putative source alongside an object from the population, a value of 1 is placed in the columns associated with the source and object combinations of the score being considered, and a value of 0 is placed in the remaining columns. For example, given a score $s_{k1,p1}$, the columns associated with the considered sources and objects (i.e., $k : k1$, $k : p1$, $p : k1$, $p : p1$) are assigned a value of 1, and all other columns are assigned a value of 0.

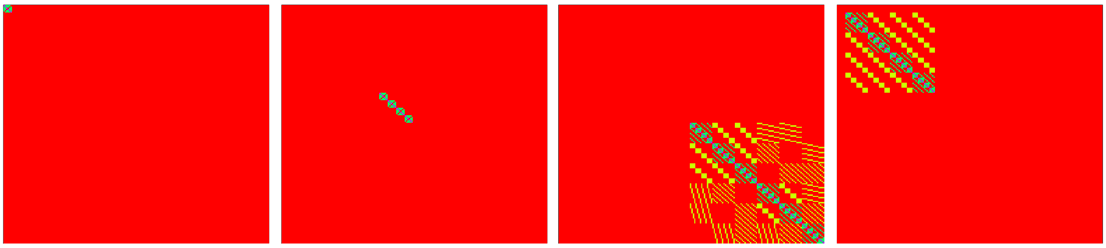
We can use these matrices to construct our final design matrix, given by $\boldsymbol{TT}'$. Specifically, $\boldsymbol{TT}'$ is defined by considering combinations of $\boldsymbol{T}_k$, $\boldsymbol{T}_{\mathscr{P}_w}$, $\boldsymbol{T}_{\mathscr{P}_b}$, and $\boldsymbol{T}_{k,\mathscr{P}}$. That is

$$\begin{aligned}\boldsymbol{TT}' \quad := \quad & \boldsymbol{T}_k\boldsymbol{T}_k' + \boldsymbol{T}_{\mathscr{P}_w}\boldsymbol{T}_{\mathscr{P}_w}' + \boldsymbol{T}_{\mathscr{P}_b}\boldsymbol{T}_{\mathscr{P}_b}' + \boldsymbol{T}_{k,\mathscr{P}}\boldsymbol{T}_{k,\mathscr{P}}' + \left(\boldsymbol{T}_k\boldsymbol{T}_{k,\mathscr{P}}' + \boldsymbol{T}_{k,\mathscr{P}}\boldsymbol{T}_k'\right) + \\ & \left(\left(\boldsymbol{T}_{\mathscr{P}_w} + \boldsymbol{T}_{\mathscr{P}_b}\right)\boldsymbol{T}_{k,\mathscr{P}}' + \boldsymbol{T}_{k,\mathscr{P}}\left(\boldsymbol{T}_{\mathscr{P}_w} + \boldsymbol{T}_{\mathscr{P}_b}\right)\right) + \left(\boldsymbol{T}_{\mathscr{P}_w}\boldsymbol{T}_{\mathscr{P}_b}' + \boldsymbol{T}_{\mathscr{P}_b}\boldsymbol{T}_{\mathscr{P}_w}'\right)\end{aligned}$$

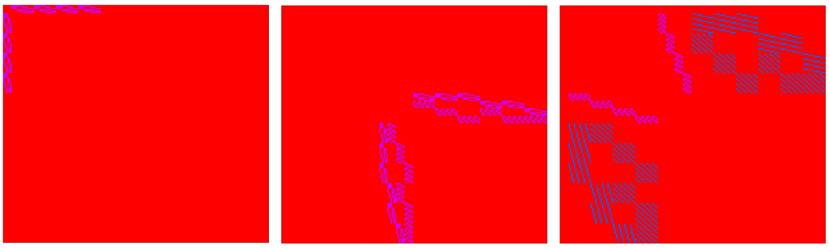
Figure 9.4 portrays the resulting $\boldsymbol{TT}'$ matrix, along with the different combinations of the sub-matrices matrices used to construct $\boldsymbol{TT}'$.



(a) $\boldsymbol{T}\boldsymbol{T}'$



(b) $\boldsymbol{T}_k\boldsymbol{T}_k'$ (c) $\boldsymbol{T}_{\mathscr{P}_w}\boldsymbol{T}_{\mathscr{P}_w}'$ (d) $\boldsymbol{T}_{\mathscr{P}_b}\boldsymbol{T}_{\mathscr{P}_b}'$ (e) $\boldsymbol{T}_{k,\mathscr{P}}\boldsymbol{T}_{k,\mathscr{P}}'$



(f) $\boldsymbol{T}_k\boldsymbol{T}_{k,\mathscr{P}}' + \boldsymbol{T}_{k,\mathscr{P}}\boldsymbol{T}_k'$ (g) $\boldsymbol{T}_{\mathscr{P}_w}\boldsymbol{T}_{\mathscr{P}_b}' + \boldsymbol{T}_{\mathscr{P}_b}\boldsymbol{T}_{\mathscr{P}_w}'$ (h) $\left(\boldsymbol{T}_{\mathscr{P}_w} + \boldsymbol{T}_{\mathscr{P}_b}\right)\boldsymbol{T}_{k,\mathscr{P}}' + \boldsymbol{T}_{k,\mathscr{P}}\left(\boldsymbol{T}_{\mathscr{P}_w} + \boldsymbol{T}_{\mathscr{P}_b}\right)'$

Figure 9.4: Heat maps of design matrix $\boldsymbol{T}\boldsymbol{T}'$ and sub-design matrices used to construct $\boldsymbol{T}\boldsymbol{T}'$. (a) Final design matrix, $\boldsymbol{T}\boldsymbol{T}'$. (b) Sub-design matrix, $\boldsymbol{T}_k\boldsymbol{T}_k'$, depicting the effect of considering two pairs of objects from the putative source. (c) Sub-design matrix, $\boldsymbol{T}_{\mathscr{P}_w}\boldsymbol{T}_{\mathscr{P}_w}'$, depicting the effect of considering two pairs of objects from the same random source in the population. (d) Sub-design matrix, $\boldsymbol{T}_{\mathscr{P}_b}\boldsymbol{T}_{\mathscr{P}_b}'$, depicting the effect of considering two pairs of objects from different random sources in the population. (e) Sub-design matrix, $\boldsymbol{T}_{k,\mathscr{P}}\boldsymbol{T}_{k,\mathscr{P}}'$, depicting the effect of considering two pairs of objects that consider an object from the putative source alongside an object from the population. (f) Sub-design matrix, $\boldsymbol{T}_{\mathscr{P}_w}\boldsymbol{T}_{\mathscr{P}_b}' + \boldsymbol{T}_{\mathscr{P}_b}\boldsymbol{T}_{\mathscr{P}_w}'$, depicting the effect of considering a pair of objects from the same source in the population alongside a pair of objects from two different sources in the population. (g) Sub-design matrix, $\left(\boldsymbol{T}_{\mathscr{P}_w} + \boldsymbol{T}_{\mathscr{P}_b}\right)\boldsymbol{T}_{k,\mathscr{P}}' + \boldsymbol{T}_{k,\mathscr{P}}\left(\boldsymbol{T}_{\mathscr{P}_w} + \boldsymbol{T}_{\mathscr{P}_b}\right)$, depicting the effect of considering a pair of objects from the population alongside a pair of objects that considers one object from the putative source and one object from the population.

The likelihood function in the numerator and denominator of (9.2) can be represented using the distribution given in (6.1). As explained in the introduction of Sections 3.1.1 and 6.1.1, we have that the structure of the mean vector and covariance matrix depend on whether $\mathcal{H}_p$ or $\mathcal{H}_d$ is being considered. See Figure 3.1 in Section 3.1.1 for an example when we consider two fixed classes.

As in the two class model, it is worth noting that the covariance matrix defined in (9.8) is not equivalent to that of the object model described in Section 9.1.1. This is due to the fact that the covariance matrix in Section 9.1.1 considers a single term $\sigma_{ii'}$ to describe the relationship that occurs when a score involves an object from source i and an object from source i'. For the covariance matrix in (9.8) to coincide with that defined in Section 9.1.1, we would need to define two terms, $\sigma_{ii'}$ and $\sigma_{i'i}$, that describe the effect when the object in common between two scores comes from source i versus from source i'. For example, consider a pair of scores $s_{k1,p1}$ and $s_{k1,p2}$. To appropriately capture the covariance that exists between these two scores would require defining a term σ_{kp}, since the common object between the two scores comes from the putative source k. Likewise, a pair of scores, $s_{k1,p2}$ and $s_{k2,p1}$, would require defining a term σ_{pk}, since the common object between the scores comes from the random source, p, from the population. As a result, the covariance terms of the score model in rows 9 and 11 of Table 9.2 do not necessarily have a direct counterpart in the model. Due to a similar phenomenon, we have that the covariance terms of the score model in rows 17 and 18 in Table 9.2 are also lacking a direct counterpart.

However, despite these discrepancies, we choose to move forward with the model given by (9.8). While the covariance matrices of the object and score models may not be exactly the same, their structures under $\mathcal{H}_p$ and $\mathcal{H}_d$ remain sufficiently similar. Furthermore, as we will see below, an elegant solution exists for studying the parameters of the model given by (9.8).

Description of Considered Scores	Score 1	Score 2	Object Model (9.5)	Score Model (9.8)
Covariance Terms				
Both Within Putative Source, Two Common Objects	$k1k2$	$k1k2$	$2\mu_k^2\rho^2 + \rho^4$	$\sigma_{kk}\left(2\sigma_a^2 + \sigma_c^2 + 4\sigma_d^2 + \sigma_e^2\right)\sigma_{kk}$
Both Within Putative Source, One Common Object	$k1k2$	$k1k3$	$\mu_k^2\rho^2$	$\sigma_{kk}\left(\sigma_a^2 + \sigma_c^2 + 2\sigma_d^2\right)\sigma_{kk}^2$
Both Within Putative Source, No Common Objects	$k1k2$	$k3k4$	0	0
Both Within Random Source, Two Common Objects, Two Common Sources	$p1p2$	$p1p2$	$2\tau^2\left(\tau^2 + 2\mu^2 + \sigma^2\right) + 2\mu^2\sigma^2$	$\sigma_{pp}\left(2\sigma_a^2 + 2\sigma_b^2 + 4\sigma_d^2 + \sigma_e^2\right)\sigma_{pp}$
Both Within Random Source, One Common Object, Two Common Sources	$p1p2$	$p1p3$	$2\tau^2\left(\tau^2 + \sigma^2\right) + \mu^2\left(4\tau^2 + \sigma^2\right)$	$\sigma_{pp}\left(\sigma_a^2 + 2\sigma_b^2 + 2\sigma_d^2\right)\sigma_{pp}$
Both Within Random Source, No Common Objects, Two Common Sources	$p1p2$	$p3p4$	$2\tau^4 + 4\mu^2\tau^2$	$\sigma_{pp}\left(2\sigma_b^2\right)\sigma_{pp}$
Both Within Random Sources, No Common Objects, No Common Sources	$p1p2$	$p'1p'2$	0	0
Both Between Putative and Random Sources, Two Common Objects, Two Common Sources	$k1p1$	$k1p1$	$\left(\rho^2 + \mu_k^2\right)\left(\sigma^2 + \tau^2\right) + \mu^2\rho^2$	$\sigma_{kp}\left(2\sigma_a^2 + \sigma_c^2 + 4\sigma_d^2 + \sigma_e^2\right)\sigma_{kp}$
Both Between Putative and Random Sources, One Common Object from Source k, Same Random Source	$k1p1$	$k1p2$	$\mu^2\rho^2 + \mu_k^2\tau^2$	$\sigma_{kp}\left(\sigma_a^2 + \sigma_c^2 + 2\sigma_d^2\right)\sigma_{kp}$
Both Between Putative and Random Source, One Common Object from Source k, Different Random Sources	$k1p1$	$k1p'1$	$\mu^2\rho^2$	$\sigma_{kp}\left(\sigma_a^2 + \sigma_d^2\right)\sigma_{kp}$
Both Between Putative and Random Source, One Common Object from Source p	$k1p1$	$k2p1$	$\mu_k^2\left(\tau^2 + \sigma^2\right)$	$\sigma_{kp}\left(\sigma_a^2 + \sigma_c^2 + 2\sigma_d^2\right)\sigma_{kp}$
Both Between Putative and Random Sources, No Common Objects, Two Common Sources	$k1p1$	$k2p2$	$\mu_k^2\tau^2$	$\sigma_{kp}\left(\sigma_c^2\right)\sigma_{kp}$
Both Between Putative and Random Sources, No Common Objects, Putative Source in Common	$k1p1$	$k2p'2$	0	0
Both Between Random Sources, Two Common Objects	$p1p'1$	$p1p'1$	$\left(\tau^2 + \sigma^2 + 2\mu^2\right)\left(\tau^2 + \sigma^2\right)$	$\sigma_{pp'}\left(2\sigma_a^2 + 2\sigma_b^2 + 4\sigma_d^2 + \sigma_e^2\right)\sigma_{pp'}$
Both Between Random Sources, One Common Object, Two Common Sources	$p1p'1$	$p1p'2$	$2\mu^2\tau^2 + \mu^2\sigma^2 + \tau^4$	$\sigma_{pp'}\left(\sigma_a^2 + 2\sigma_b^2 + 2\sigma_d^2\right)\sigma_{pp'}$
Both Between Random Sources, One Common Object, One Common Source	$p1p'2$	$p1p''1$	$\mu^2\left(\tau^2 + \sigma^2\right)$	$\sigma_{pp'}\left(\sigma_a^2 + \sigma_b^2 + \sigma_d^2\right)\sigma_{pp''}$
Both Between Random Sources, No Common Objects, Two Common Sources	$p1p'1$	$p2p'2$	$2\mu^2\tau^2$	$\sigma_{pp'}\left(2\sigma_b^2\right)\sigma_{pp'}$
Both Between Random Sources, No Common Objects, One Common Source	$p1p'1$	$p2p''1$	$\mu^2\tau^2$	$\sigma_{pp'}\left(2\sigma_b^2\right)\sigma_{pp'}$
Both Between Random Sources, No Common Objects, No Common Sources	$p1p'1$	$p''1p'''1$	0	0
Within Putative Source, Within Random Source, No Common Objects	$k1k2$	$p1p2$	0	0
Within Putative Source, Between Putative and Random Source, One Common Object	$k1k2$	$k1p1$	$\mu\mu_k\rho^2$	$\sigma_{kk}\left(\sigma_a^2 + \sqrt{2}\sigma_d^2\right)\sigma_{kp}$
Within Putative Source, Between Putative and Random Source, No Common Objects	$k1k2$	$k3p1$	0	0
Within Random Source, Between Putative and Random Source, One Common Object, One Common Source	$p1p2$	$k1p1$	$\mu\mu_k\left(2\tau^2 + \sigma^2\right)$	$\sigma_{pp}\left(\sigma_a^2 + \sqrt{2}\sigma_b^2 + \sqrt{2}\sigma_d^2\right)\sigma_{kp}$
Within Random Source, Between Putative and Random Source, No Common Objects, One Common Source	$p1p2$	$k1p3$	$2\mu\mu_k\tau^2$	$\sigma_{pp}\left(\sqrt{2}\sigma_b^2\right)\sigma_{kp}$
Within Random Source, Between Putative and Random Source, No Common Objects, No Common Sources	$p1p2$	$k1p'1$	0	0
Between Putative and Random Source, Between Random Sources, One Common Object, One Common Source	$k1p1$	$p1p'1$	$\mu\mu_k\left(\tau^2 + \sigma^2\right)$	$\sigma_{kp}\left(\sigma_a^2 + \sigma_d^2\right)\sigma_{pp'}$
Between Putative and Random Source, Between Random Sources, No Common Objects, One Common Source	$k1p1$	$p2p'1$	$\mu\mu_k\tau^2$	$\sigma_{kp}\left(\sigma_b^2\right)\sigma_{pp'}$
Between Putative and Random Source, Between Random Sources, No Common Objects, No Common Sources	$k1p1$	$p'1p''1$	0	0
Within Random Sources, Between Random Sources, One Common Object, One Common Source	$p1p2$	$p1p'1$	$\mu^2\left(2\tau^2 + \sigma^2\right)$	$\sigma_{pp}\left(2\sigma_a^2 + 2\sigma_b^2 + 4\sigma_d^2 + \sigma_e^2\right)\sigma_{pp'}$
Within Random Sources, Between Random Sources, No Common Objects, One Common Source	$p1p2$	$p3p'1$	$2\mu^2\tau^2$	$\sigma_{pp}\left(\sqrt{2}\sigma_b^2\right)\sigma_{pp'}$
Within Random Sources, Between Random Sources, No Common Objects, No Common Sources	$p1p2$	$p'1p''1$	0	0

Table 9.2: Comparison of Covariance terms in Object Model defined according to (9.5), and Score Model defined according to (9.8).

9.2 MODEL DEVELOPMENT

Assigning the Kernel Bayes Factor in (9.2) requires estimating the parameters $\{\theta_{ii'}\}_{ii'}$, $\{\sigma_{ii'}\}_{ii'}$, σ_a^2, σ_b^2, σ_c^2, σ_d^2 and σ_e^2 using the information contained in the vector of scores, $\boldsymbol{s}$. To study these parameters, we follow the development described in Parts II and III, and subset the vector of scores to define $\boldsymbol{s}_c$, which includes only the comparisons between the control objects contained in the sets $\boldsymbol{X}_k$ and $\mathscr{P}$, and so is a vector of length

$N_c = \binom{(r+1)n_0}{2}$. We can then use $\boldsymbol{s}_c$ to define the total sum of squares

$$\begin{aligned} SS_{Tot} &= (\boldsymbol{s}_c - \boldsymbol{\theta}_c)' \left[\boldsymbol{\Delta}_c \boldsymbol{\Delta}_c'\right]^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c) \\ &= \left(\boldsymbol{\Delta}_c^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c)\right)' \left(\boldsymbol{\Delta}_c^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c)\right) \end{aligned} \tag{9.9}$$

where $\boldsymbol{\theta}_c$ is the $\boldsymbol{N}_c$ vector of score means, $\theta_{ii'}$, and $\boldsymbol{\Delta}_c$ is the $N_c \times N_c$ diagonal matrix of the score standard deviations, $\sigma_{ii'}$, associated with the scores $\boldsymbol{s}_c$.

Following the development in Sections 3.2 and 6.2, we apply Cochran's theorem (see Theorem 2 in Section 3.2) to $\widetilde{\boldsymbol{s}} = (\boldsymbol{\Delta}_c^{-1} (\boldsymbol{s}_c - \boldsymbol{\theta}_c))$, and rewrite (9.9) as

$$SS_{Tot} = \widetilde{\boldsymbol{s}} \boldsymbol{I} \widetilde{\boldsymbol{s}} = \widetilde{\boldsymbol{s}} \left[\sum_{l=1}^{N_c} \boldsymbol{v}_l \boldsymbol{v}_l'\right] \widetilde{\boldsymbol{s}} \tag{9.10}$$

where $\{\boldsymbol{v}_l\}_l$, $l = 1, \ldots, N_c$ is any orthonormal basis for $\mathbb{R}^{N_c}$. Furthermore, we consider a set of $(r+1)+\binom{r+1}{2}$ diagonal design matrices, $\boldsymbol{B}_{ii'}$, $i, i' \in \{1, \ldots, r, k\}$. Each of these matrices is an idempotent $N_c \times N_c$ matrix whose diagonal matrix can be partitioned into $(r+1)+\binom{r+1}{2}$ segments. Of these segments, $(r+1)$ segments are of length $\binom{n_0}{2}$, and are each associated with a single source, $1, \ldots, r, k$. The remaining $\binom{r+1}{2}$ segments are each of length n_0^2, and are each associated with one of the possible combinations of sources. Note that $(r+1)\binom{n_0}{2} + \binom{r+1}{2} n_0^2 = N_c$. When we are considering a within-source comparison ($i = i'$), the matrix $\boldsymbol{B}_{ii}$ has ones in the segment of length $\binom{n_0}{2}$ along the diagonal that corresponds to the i^{th} source, and zeros elsewhere. When we are considering a between-source comparison ($i \neq i'$), the matrix $\boldsymbol{B}_{ii'}$ has ones in the segment of length n_0^2 along the diagonal corresponding the the comparison between source i and i', and zeros elsewhere[2].

As an example, let $r = 3$, $n_0 = 5$. Then we have $(r+1)+\binom{r+1}{2} = 10$ matrices consisting of $r+1 = 4$ within source matrices, $\boldsymbol{B}_{kk}$, $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{22}$, $\boldsymbol{B}_{33}$, and

[2]Note that the indices i, i' are not considered jointly, as in (9.6). In this case, we are considering matrices $\boldsymbol{B}_{ii'}$ for all possible source combinations, rather than just the different types of source comparisons that exist. That is, instead of consider a single matrix for all $\binom{r}{2}$ possible combinations between the random sources, we consider $\binom{r}{2}$ individual $\boldsymbol{B}_{ii'}$ matrices. Deconstructing the sums of squares in this manner allows us to obtain an elegant eigen-decomposition in which the eigenvalues are straightforward functions of the terms σ_a^2, σ_b^2, σ_c^2, σ_d^2, and σ_e^2, as we will see later in the chapter.

$\binom{r+1}{2} = 6$ between source matrices, $\boldsymbol{B}_{k1}$, $\boldsymbol{B}_{k2}$, $\boldsymbol{B}_{k3}$, $\boldsymbol{B}_{12}$, $\boldsymbol{B}_{13}$, $\boldsymbol{B}_{23}$. These 10 matrices are displayed in Figures 9.5 and 9.6. Since $\boldsymbol{B}_{k1}, \ldots, \boldsymbol{B}_{kr}, \boldsymbol{B}_{11}, \boldsymbol{B}_{12}, \ldots, \boldsymbol{B}_{rr}$ sum to the identity matrix, we have that

$$\begin{aligned}
SS_{Tot} &= \widetilde{\boldsymbol{s}}' \left(\sum_{i,i'} \boldsymbol{B}_{ii'} \right) \boldsymbol{I} \left(\sum_{i,i'} \boldsymbol{B}_{ii'} \right) \widetilde{\boldsymbol{s}} \\
&= \widetilde{\boldsymbol{s}}' \left(\sum_{i,i'} \boldsymbol{B}_{ii'} \boldsymbol{I} \sum_{i,i'} \boldsymbol{B}_{ii'} \right) \widetilde{\boldsymbol{s}} \\
&= \widetilde{\boldsymbol{s}}' \left(\sum_{i,i'} \boldsymbol{B}_{ii'} \left[\sum_{l=1}^{N_c} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} \right] \sum_{i,i'} \boldsymbol{B}_{ii'} \right) \widetilde{\boldsymbol{s}} \\
&= \sum_{i=k,1,\ldots,r} SS_{ii} + \sum_{i<i' \in \{k,1\ldots,r\}} SS_{ii'} \qquad (9.11)
\end{aligned}$$

where $\{\boldsymbol{v}_{ii'_l}\}_l$, $l = 1, \ldots, N_c$, are different orthonormal bases spanning $\mathbb{R}^{N_c}$, and will be discussed in more detail in later sections. The matrices $\boldsymbol{B}_{ii'}$ effectively activate different parts of the vector $\widetilde{\boldsymbol{s}}_c$ according to the different source comparisons. In particular, we have:

(1) $\boldsymbol{B}_{kk}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to comparisons that exist within the putative source, so that $\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{kk}\widetilde{\boldsymbol{s}}_c$ gives us the within-source sum of squares corresponding to the putative source. Recall that $\boldsymbol{B}_{kk}\boldsymbol{I}\boldsymbol{B}_{kk} = \boldsymbol{B}_{kk}$, and note that $\boldsymbol{B}_{kk}$ has rank $r_{kk} = \binom{n_0}{2}$.

(2) $\boldsymbol{B}_{pp}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to comparisons that exist within one of the random sources from the population of potential sources that is being considered, so that $\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{pp}\widetilde{\boldsymbol{s}}_c$ gives us the within-source sum of squares corresponding to that particular random source. Recall that $\boldsymbol{B}_{pp}\boldsymbol{I}\boldsymbol{B}_{pp} = \boldsymbol{B}_{pp}$. Note that we have r different $\boldsymbol{B}_{pp}$ matrices, $p \in \{1, \ldots, r\}$, where each matrix $\boldsymbol{B}_{pp}$ has rank $r_{pp} = \binom{n_0}{2}$.

(3) $\boldsymbol{B}_{kp}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to comparisons that exist between objects from the putative source and from one of the random sources from the population of potential sources that is being considered, so that $\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{kp}\widetilde{\boldsymbol{s}}_c$ gives us the between-source sum of squares corresponding to the

putative source and the considered random source. Recall that $\boldsymbol{B}_{kp}\boldsymbol{I}\boldsymbol{B}_{kp} = \boldsymbol{B}_{kp}$. Note that we have r different $\boldsymbol{B}_{kp}$ matrices, $p \in \{1, \ldots, r\}$, where each matrix $\boldsymbol{B}_{kp}$ has rank $r_{kp} = n_0^2$.

(4) $\boldsymbol{B}_{pp'}\widetilde{\boldsymbol{s}}_c$ allows us to consider only the positions of $\widetilde{\boldsymbol{s}}_c$ that correspond to comparisons that exist between objects from two different random sources from the population of potential sources, so that $\widetilde{\boldsymbol{s}}_c'\boldsymbol{B}_{pp'}\widetilde{\boldsymbol{s}}_c$ gives us the between-source sum of squares corresponding to the two considered random sources. Recall that $\boldsymbol{B}_{pp'}\boldsymbol{I}\boldsymbol{B}_{pp'} = \boldsymbol{B}_{pp'}$. Note that we have $\binom{r}{2}$ different $\boldsymbol{B}_{pp'}$ matrices, $p \neq p' \in \{1, \ldots, r\}$, where. each matrix $\boldsymbol{B}_{pp'}$ has rank $r_{pp'} = n_0^2$.

Thus, we have defined the total sums of squares in terms of all source comparisons that exist within our vector of scores. Bearing in mind that the goal is to find a way to estimate the parameters of the distribution given in (9.8), we note that this decomposition of the total sums of squares allows us to independently study the mean and standard deviation terms, $\theta_{ii'}$ and $\sigma_{ii'}$, associated with their respective source comparisons. Note that we can choose the orthonormal bases in (9.11) to be any orthonormal bases, and, in particular, we can choose these orthonormal bases to be the normalized eigenvectors for the following matrices, $\boldsymbol{V}$. We choose to define

$$\boldsymbol{V}_{ii} = \boldsymbol{B}_{ii}\left(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{Q}_c\boldsymbol{Q}_c'\sigma_b^2 + \boldsymbol{R}_c\boldsymbol{R}_c'\sigma_c^2 + \boldsymbol{T}_c\boldsymbol{T}_c'\sigma_d^2 + \boldsymbol{I}_c\sigma_e^2\right)\boldsymbol{B}_{ii} \tag{9.12}$$

for the matrices $\boldsymbol{B}_{ii}$, $i \in \{k, 1, \ldots, r\}$, corresponding to the within-source comparisons, and

$$\boldsymbol{V}_{ii'} = \boldsymbol{B}_{ii'}\left(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{Q}_c\boldsymbol{Q}_c'\sigma_b^2 + \boldsymbol{R}_c\boldsymbol{R}_c'\sigma_c^2 + \boldsymbol{T}_c\boldsymbol{T}_c'\sigma_d^2 + \boldsymbol{I}_c\sigma_e^2\right)\boldsymbol{B}_{ii'} \tag{9.13}$$

for the matrices $\boldsymbol{B}_{ii'}$, $i \neq i' \in \{k, 1, \ldots, r\}$, corresponding to the between-source comparisons, where the subscript c on the matrices in (9.12) and (9.13) are of the same structure as the design matrices described in Section 9.1.2, but have a dimension corresponding to that of the score vector $\boldsymbol{s}_c$ that considers only comparisons between two control objects.

As in Section 3.2 and 6.2, choosing $\boldsymbol{V}_{ii}$ and $\boldsymbol{V}_{ii'}$ to be a function of $\boldsymbol{\Sigma}_c := \boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{Q}_c\boldsymbol{Q}_c'\sigma_b^2 + \boldsymbol{R}_c\boldsymbol{R}_c'\sigma_c^2 + \boldsymbol{T}_c\boldsymbol{T}_c'\sigma_d^2 + \boldsymbol{I}_c\sigma_e^2$ is advantageous in that it introduces the parameters σ_a^2, σ_b^2, σ_c^2, σ_d^2, and σ_e^2, and so provides a means for studying these parameters. Second, defining $\boldsymbol{V}_{ii}$ and $\boldsymbol{V}_{ii'}$ in terms of $\boldsymbol{B}_{ii}$ and $\boldsymbol{B}_{ii'}$ allows us to take advantage of the relevant parts of $\boldsymbol{\Sigma}_c$ with respect to each source comparison by activating only the rows and columns of $\boldsymbol{\Sigma}$ corresponding to the considered source comparison.

	$\boldsymbol{V}_{kk}$	
Eigenvalue (ν_{kk_l})	Multiplicity ($m_{\nu_{kk_l}}$)	Eigenvectors ($\boldsymbol{v}_{kk_l}$)
$2\,(n_0-1)\,\sigma_a^2 + \binom{n_0}{2}\sigma_c^2 + 2\,(2n_0-2)\,\sigma_d^2 + \sigma_e^2$	1	$\boldsymbol{v}_{kk_l}$ such that $\boldsymbol{V}_{kk}\boldsymbol{v}_{kk_l} = \nu_{kk_1}\boldsymbol{v}_{kk_l}$
$(n_0-2)\,\sigma_a^2 + 2\,(n_0-2)\,\sigma_d^2 + \sigma_e^2$	n_0-1	$\boldsymbol{v}_{kk_l}$ such that $\boldsymbol{V}_{kk}\boldsymbol{v}_{kk_l} = \nu_{kk_2}\boldsymbol{v}_{kk_l}$
σ_e^2	$\binom{n_0}{2} - n_0$	$\boldsymbol{v}_{ii_l}$ such that $\boldsymbol{V}_{kk}\boldsymbol{v}_{kk_l} = \nu_{kk_3}\boldsymbol{v}_{kk_l}$
0	$N_c - \binom{n_0}{2}$	$\boldsymbol{v}_{kk_l}$ such that $\boldsymbol{V}_{kk}\boldsymbol{v}_{kk_l} = \nu_{kk_4}\boldsymbol{v}_{kk_l}$
	$\boldsymbol{V}_{kp}$	
Eigenvalue (ν_{kp_l})	Multiplicity ($m_{\nu_{kp_l}}$)	Eigenvectors ($\boldsymbol{v}_{kp_l}$)
$2n_0\sigma_a^2 + n_0^2\sigma_c^2 + 4n_0\sigma_d^2 + \sigma_e^2$	1	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{kp}\boldsymbol{v}_{kp_l} = \nu_{kp_1}\boldsymbol{v}_{kp_l}$
$n_0\sigma_a^2 + 2n_0\sigma_d^2 + \sigma_e^2$	$2n_0-2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{kp}\boldsymbol{v}_{kp_l} = \nu_{kp_2}\boldsymbol{v}_{kp_l}$
σ_e^2	$(n_0-1)^2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{kp}\boldsymbol{v}_{kp_l} = \nu_{kp_3}\boldsymbol{v}_{kp_l}$
0	$N_c - n_0^2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{kp}\boldsymbol{v}_{kp_l} = \nu_{kp_4}\boldsymbol{v}_{kp_l}$
	$\boldsymbol{V}_{pp}$	
Eigenvalue (ν_{pp_l})	Multiplicity ($m_{\nu_{pp_l}}$)	Eigenvectors ($\boldsymbol{v}_{pp_l}$)
$2(n_0-1)\sigma_a^2 + \binom{n_0}{2}\sigma_c^2 + 2(2n_0-2)\sigma_d^2 + \sigma_e^2$	1	$\boldsymbol{v}_{pp_l}$ such that $\boldsymbol{V}_{pp}\boldsymbol{v}_{pp_l} = \nu_{pp_1}\boldsymbol{v}_{pp_l}$
$(n_0-2)\sigma_a^2 + 2(n_0-2)\sigma_d^2 + \sigma_e^2$	n_0-1	$\boldsymbol{v}_{pp_l}$ such that $\boldsymbol{V}_{pp}\boldsymbol{v}_{pp_l} = \nu_{pp_2}\boldsymbol{v}_{pp_l}$
σ_e^2	$\binom{n_0}{2} - n_0$	$\boldsymbol{v}_{ii'_l}$ such that $\boldsymbol{V}_{ii'}\boldsymbol{v}_{ii'_l} = \nu_{ii'_3}\boldsymbol{v}_{ii'_l}$
0	$N_c - \binom{n_0}{2}$	$\boldsymbol{v}_{ii'_l}$ such that $\boldsymbol{V}_{ii'}\boldsymbol{v}_{ii'_l} = \nu_{ii'_4}\boldsymbol{v}_{ii'_l}$
	$\boldsymbol{V}_{pp'}$	
Eigenvalue ($\nu_{pp'_l}$)	Multiplicity ($m_{\nu_{pp'_l}}$)	Eigenvectors ($\boldsymbol{v}_{pp'_l}$)
$2n_0\sigma_a^2 + n_0^2\sigma_c^2 + 4n_0\sigma_d^2 + \sigma_e^2$	1	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{pp'}\boldsymbol{v}_{pp'_l} = \nu_{pp'_1}\boldsymbol{v}_{pp'_l}$
$n_0\sigma_a^2 + 2n_0\sigma_d^2 + \sigma_e^2$	$2n_0-2$	$\boldsymbol{v}_{pp'_l}$ such that $\boldsymbol{V}_{pp'}\boldsymbol{v}_{pp'_l} = \nu_{pp'_2}\boldsymbol{v}_{pp'_l}$
σ_e^2	$(n_0-1)^2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{pp'}\boldsymbol{v}_{pp'_l} = \nu_{pp'_3}\boldsymbol{v}_{pp'_l}$
0	$N_c - n_0^2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{pp'}\boldsymbol{v}_{pp'_l} = \nu_{pp'_4}\boldsymbol{v}_{pp'_l}$

Table 9.3: Eigenstructure of design matrices for within-source comparisons, $\boldsymbol{V}_{kk}$ and $\boldsymbol{V}_{pp}$, and between-source comparisons, $\boldsymbol{V}_{kp}$ and $\boldsymbol{V}_{pp'}$, as described in (9.12) and (9.13)

We can study the eigenstructure of the matrices $\boldsymbol{B}_{ii'}(\boldsymbol{P}_c\boldsymbol{P}_c'\sigma_a^2 + \boldsymbol{Q}_c\boldsymbol{Q}_c'\sigma_b^2 + \boldsymbol{R}_c\boldsymbol{R}_c'\sigma_c^2 + \boldsymbol{T}_c\boldsymbol{T}_c'\sigma_d^2 + \boldsymbol{I}_c\sigma_e^2)\boldsymbol{B}_{ii'}$ for each source comparison (see Table 9.3). This study reveals the presence of multiple subspaces for each of the considered eigenspaces. This allows us to further decompose each of the sums of squares in (9.11) as another sum of squares. For those within-source comparisons, we have that

$$\begin{aligned}
\widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii} \left[\sum_{l=1}^{N_c} \boldsymbol{v}_{ii_l} \boldsymbol{v}'_{ii_l}\right] \boldsymbol{B}_{ii} \widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii} \left[\boldsymbol{v}_{11_l} + \sum_{l=2}^{n_0} \boldsymbol{v}_{ii_l} \boldsymbol{v}'_{ii_l} + \sum_{l=n_0+1}^{\binom{n_0}{2}} \boldsymbol{v}_{ii_l} \boldsymbol{v}_{ii_l} + \sum_{l=\binom{n_0}{2}}^{N_c} \boldsymbol{v}_{ii_l} \boldsymbol{v}_{ii_l}\right] \boldsymbol{B}_{ii} \widetilde{\boldsymbol{s}}_c \\
&= \widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii} \left[\boldsymbol{v}_{11_l} + \sum_{l=2}^{n_0} \boldsymbol{v}_{ii_l} \boldsymbol{v}'_{ii_l} + \sum_{l=n_0+1}^{N_{ii}} \boldsymbol{v}_{ii_l} \boldsymbol{v}_{ii_l}\right] \boldsymbol{B}_{ii} \widetilde{\boldsymbol{s}}_c
\end{aligned}$$

where $N_{ii} = \binom{n_0}{2}$ is the number of eigenvectors associated with non-zero eigenvalues, and is also the number of "interesting" eigenvectors $\boldsymbol{v}_{ii}$. In particular, we have that the elements that are equal to one correspond to the dimensions whose associated eigenvalues are zero. These vectors form the standard basis for the null space of the corresponding matrix, and correspond to the rows of $\boldsymbol{B}_{ii}$ that are equal to the zero vector. As an example, we consider the matrices that result when $n = 3$ and $n_0 = 4$. The first column of Figure 9.5 and the first three columns of Figure 9.6 display the heat maps of $\boldsymbol{B}_{kk}$ and of the matrix of eigenvectors $\boldsymbol{V}_{kk}$ for comparisons within the putative source, and of $\boldsymbol{B}_{11}$, $\boldsymbol{B}_{22}$ and $\boldsymbol{B}_{33}$, and of the matrices of eigenvectors $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{22}$ and $\boldsymbol{V}_{33}$ for comparisons within the random sources, respectively. Note that the $N_{ii} \times N_{ii}$ patchwork matrices with each of the $\boldsymbol{V}_{ii}$ matrices correspond to the nonzero rows of the corresponding $\boldsymbol{B}_{ii}$ matrices. Because the placements of the nonzero elements in these eigenvectors correspond to the zero elements of the diagonals in the associated $\boldsymbol{B}_i$ matrices, the product of the $\boldsymbol{B}_{ii}$ matrix with these sets of eigenvectors results in a zero-valued sum of squares.

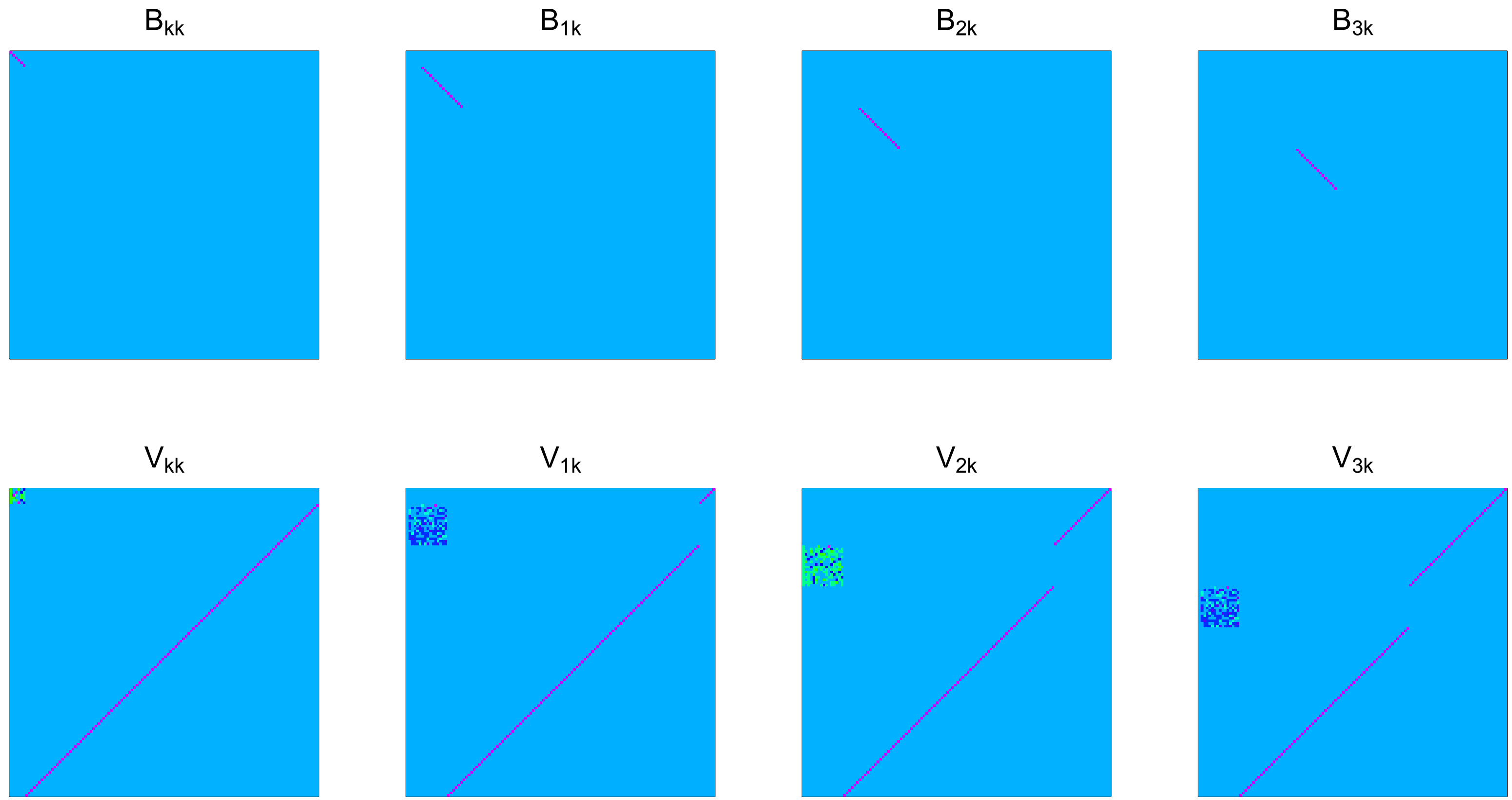


Figure 9.5: Heat maps for the matrices of the eigenvectors associated with putative source matrices B_{ki} (top) and matrices of associated eigenvectors of $\boldsymbol{V}_{ki}$ (bottom) when $i \in \{k, 1, \ldots, r\}$, $n = 3$ and $n_0 = 4$. The $N_{ki} \times N_{ki}$ patchworks correspond to eigenvectors with non-zero elements. The remaining columns correspond to vectors with one non-zero element (in red), and are associated with zero-valued eigenvalues.

Similarly, for between-source comparisons, we have

$$\begin{aligned}
\widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'} \left[\sum_{l=1}^{N_c} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l}\right] \boldsymbol{B}_{ii'} \widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'} \left[\boldsymbol{v}_{11_l} + \sum_{l=2}^{\binom{n_0}{2}-1} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} + \sum_{l=\binom{n_0}{2}}^{n_0^2} \boldsymbol{v}_{ii'_l} \boldsymbol{v}_{ii'_l} + \sum_{l=n_0^2}^{N_c} \boldsymbol{v}_{ii'_l} \boldsymbol{v}_{ii'_l}\right] \boldsymbol{B}_{ii'} \widetilde{\boldsymbol{s}}_c \\
&= \widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'} \left[\boldsymbol{v}_{11_l} + \sum_{l=2}^{\binom{n_0}{2}-1} \boldsymbol{v}_{ii'_l} \boldsymbol{v}'_{ii'_l} + \sum_{l=\binom{n_0}{2}}^{N_{ii'}} \boldsymbol{v}_{ii'_l} \boldsymbol{v}_{ii'_l}\right] \boldsymbol{B}_{ii'} \widetilde{\boldsymbol{s}}_c
\end{aligned}$$

where $N_{ii'} = n_0^2$ is the number of eigenvectors associated with non-zero eigenvalues, and is also the number of "interesting" eigenvectors $\boldsymbol{v}_{ii'_l}$. In particular, we have that the elements that are equal to one correspond to the dimensions whose associated eigenvalues are zero. These vectors form the standard basis for the null space of the corresponding matrix, and correspond to the rows of $\boldsymbol{B}_{ii'}$ that are equal to the zero vector. The second, third, and fourth columns of Figure 9.5 and the last three columns of Figure 9.6 display the heat maps of $\boldsymbol{B}_{k1}$, $\boldsymbol{B}_{k2}$ and $\boldsymbol{B}_{k3}$ and of the matrix of eigenvectors $\boldsymbol{V}_{k1}$, $\boldsymbol{V}_{k2}$ and $\boldsymbol{V}_{k3}$ for comparisons within the putative source, and of $\boldsymbol{V}_{12}$, $\boldsymbol{B}_{13}$ and $\boldsymbol{B}_{23}$, and of the matrices of eigenvectors $\boldsymbol{V}_{12}$, $\boldsymbol{V}_{13}$ and $\boldsymbol{V}_{23}$ for comparisons within the random sources, respectively. Note that the $N_{ii'} \times N_{ii'}$ patchwork matrices within each of the $\boldsymbol{V}_{ii'}$ matrices correspond to the nonzero rows of the corresponding $\boldsymbol{B}_{ii'}$ matrices. Because the placements of the nonzero elements in these eigenvectors correspond to the zero elements of the diagonals in the associated $\boldsymbol{B}_{ii'}$ matrices, the product of the $\boldsymbol{B}_{ii'}$ matrix with these sets of eigenvectors results in a zero-valued sum of squares.

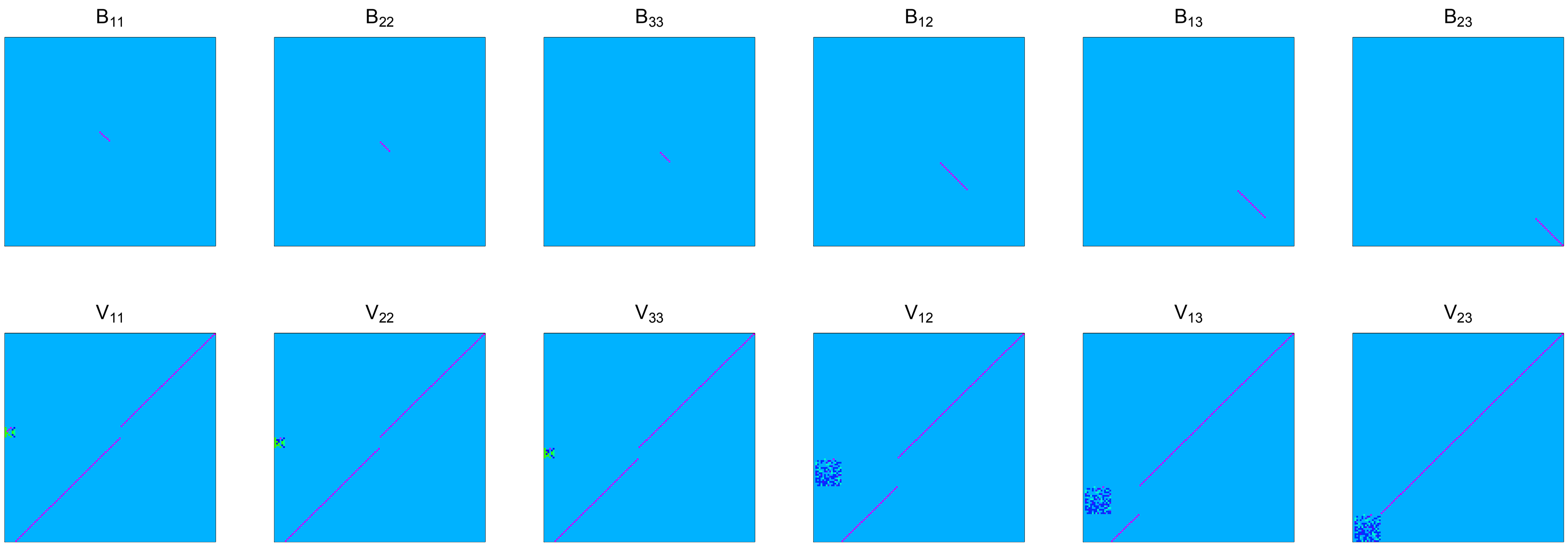


Figure 9.6: Heat maps for the matrices of the eigenvectors associated with random source matrices $B_{pp'}$ matrices (top) and matrices of associated eigenvectors of $\boldsymbol{V}_{pp'}$ (bottom) when $p, p' \in \{1, \ldots, r\}$, $n = 3$ and $n_0 = 4$. The $N_{pp'} \times N_{pp'}$ patchworks correspond to eigenvectors with non-zero elements. The remaining columns correspond to vectors with one non-zero element (in red), and are associated with zero-valued eigenvalues.

This decomposition is favorable in that studying the relevant parts of the eigen-decomposition of $\mathbf{\Sigma}_c$ is equivalent to studying the eigen-decomposition of the relevant parts of $\mathbf{\Sigma}_c$. That is,

$$\begin{aligned}\widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii}\left[\sum_{l=1}^{N_c} \boldsymbol{v}_{ii_l}\boldsymbol{v}_{ii_l}^{'}\right]\boldsymbol{B}_{ii}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_{ii}\left[\sum_{l=1}^{N_{ii}} \boldsymbol{v}_{ii_l}^{*}\boldsymbol{v}_{ii_l}^{*'}\right]\widetilde{\boldsymbol{s}}_{ii} \\ \widetilde{\boldsymbol{s}}_c \boldsymbol{B}_{ii'}\left[\sum_{l=1}^{N_c} \boldsymbol{v}_{ii'_l}\boldsymbol{v}_{ii'_l}^{'}\right]\boldsymbol{B}_{ii'}\widetilde{\boldsymbol{s}}_c &= \widetilde{\boldsymbol{s}}_{ii'}\left[\sum_{l=1}^{N_{ii'}} \boldsymbol{v}_{ii'_l}^{*}\boldsymbol{v}_{ii'_l}^{*'}\right]\widetilde{\boldsymbol{s}}_{ii'}\end{aligned}$$

where $\boldsymbol{v}_{ii_l}^{*}$ are the eigenvectors of the $N_{ii} \times N_{ii}$ matrix formed by considering the non-zero rows of the columns associated with the non-zero eigenvalues of $\boldsymbol{v}_{ii}$, and $\boldsymbol{v}_{ii'_l}^{*}$ are the eigenvectors of the $N_{ii'} \times N_{ii'}$ matrix formed by considering the non-zero portions of the columns associated with the non-zero eigenvalues of $\boldsymbol{v}_{ii'}$ This is equivalent to considering only the indices of $\mathbf{\Sigma}_c$ that correspond to each source comparison. That is, $\boldsymbol{v}_{ii_l}^{*}$ are the eigenvectors of

$$\Sigma_{ii} := \boldsymbol{P}_{ii}\boldsymbol{P}_{ii}'\sigma_a^2 + \boldsymbol{Q}_{ii}\boldsymbol{Q}_{ii}'\sigma_b^2 + \boldsymbol{R}_{ii}\boldsymbol{R}_{ii}'\sigma_c^2 + \boldsymbol{T}_{ii}\boldsymbol{T}_{ii}'\sigma_d^2 + \boldsymbol{I}_{ii}\sigma_e^2, \tag{9.14}$$

the matrix formed by considering the rows of $\mathbf{\Sigma}_c$ associated with some within-source comparison, and $\boldsymbol{v}_{ii'_l}$ are the eigenvectors of

$$\Sigma_{ii'} := \boldsymbol{P}_{ii'}\boldsymbol{P}_{ii'}'\sigma_a^2 + \boldsymbol{Q}_{ii'}\boldsymbol{Q}_{ii'}'\sigma_b^2 + \boldsymbol{R}_{ii'}\boldsymbol{R}_{ii'}'\sigma_c^2 + \boldsymbol{T}_{ii'}\boldsymbol{T}_{ii'}'\sigma_d^2 + \boldsymbol{I}_{ii'}\sigma_e^2, \tag{9.15}$$

the matrix formed by considering the rows of $\mathbf{\Sigma}_c$ associated with some between-source comparison. In addition, we have that $\widetilde{\boldsymbol{s}}_{ii} = (\mathbf{\Delta}_{ii}(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}))$, where $\mathbf{\Delta}_{ii}$ is the $N_{ii} \times N_{ii}$ portion of $\mathbf{\Delta}_c$ that considers σ_{ii}, and $\widetilde{\boldsymbol{s}}_{ii'} = \left(\mathbf{\Delta}_{ii'}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right)$, where $\mathbf{\Delta}_{ii'}$ is the $N_{ii'} \times N_{ii'}$ portion of $\mathbf{\Delta}_c$ that considers $\sigma_{ii'}$.

$\boldsymbol{\Sigma}_{kk}$		
Eigenvalue (ν_{kk_l})	Multiplicity ($m_{\nu_{kk_l}}$)	Eigenvectors ($\boldsymbol{v}_{kk_l}$)
$2\,(n_0-1)\,\sigma_a^2+\binom{n_0}{2}\sigma_c^2+2\,(2n_0-2)\,\sigma_d^2+\sigma_e^2$	1	$\boldsymbol{v}_{kk_1} := \frac{\mathbf{1}_{N_{kk}}}{\sqrt{N_{kk}}}$
$(n_0-2)\,\sigma_a^2+2\,(n_0-2)\,\sigma_d^2+\sigma_e^2$	n_0-1	$\boldsymbol{v}_{kk_l}$ such that $\boldsymbol{V}_{kk}\boldsymbol{v}_{kk_l}=\nu_{kk_2}\boldsymbol{v}_{kk_l}$
σ_e^2	$\binom{n_0}{2}-n_0$	$\boldsymbol{v}_{ii_l}$ such that $\boldsymbol{V}_{kk}\boldsymbol{v}_{kk_l}=\nu_{kk_3}\boldsymbol{v}_{kk_l}$
$\boldsymbol{\Sigma}_{kp}$		
Eigenvalue (ν_{kp_l})	Multiplicity ($m_{\nu_{kp_l}}$)	Eigenvectors ($\boldsymbol{v}_{kp_l}$)
$2n_0\sigma_a^2+n_0^2\sigma_c^2+4n_0\sigma_d^2+\sigma_e^2$	1	$\boldsymbol{v}_{kp_1} := \frac{\mathbf{1}_{N_{kp}}}{\sqrt{N_{kp}}}$
$n_0\sigma_a^2+2n_0\sigma_d^2+\sigma_e^2$	$2n_0-2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{kp}\boldsymbol{v}_{kp_l}=\nu_{kp_2}\boldsymbol{v}_{kp_l}$
σ_e^2	$(n_0-1)^2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{kp}\boldsymbol{v}_{kp_l}=\nu_{kp_3}\boldsymbol{v}_{kp_l}$
$\boldsymbol{\Sigma}_{pp}$		
Eigenvalue (ν_{pp_l})	Multiplicity ($m_{\nu_{pp_l}}$)	Eigenvectors ($\boldsymbol{v}_{pp_l}$)
$2(n_0-1)\sigma_a^2+\binom{n_0}{2}\sigma_c^2+2(2n_0-2)\sigma_d^2+\sigma_e^2$	1	$\boldsymbol{v}_{pp_1} := \frac{\mathbf{1}_{N_{pp}}}{\sqrt{N_{pp}}}$
$(n_0-2)\sigma_a^2+2(n_0-2)\sigma_d^2+\sigma_e^2$	n_0-1	$\boldsymbol{v}_{pp_l}$ such that $\boldsymbol{V}_{pp}\boldsymbol{v}_{pp_l}=\nu_{pp_2}\boldsymbol{v}_{pp_l}$
σ_e^2	$\binom{n_0}{2}-n_0$	$\boldsymbol{v}_{ii'_l}$ such that $\boldsymbol{V}_{ii'}\boldsymbol{v}_{ii'_l}=\nu_{ii'_3}\boldsymbol{v}_{ii'_l}$
$\boldsymbol{\Sigma}_{pp'}$		
Eigenvalue ($\nu_{pp'_l}$)	Multiplicity ($m_{\nu_{pp'_l}}$)	Eigenvectors ($\boldsymbol{v}_{pp'_l}$)
$2n_0\sigma_a^2+n_0^2\sigma_c^2+4n_0\sigma_d^2+\sigma_e^2$	1	$\boldsymbol{v}_{pp'_1} := \frac{\mathbf{1}_{N_{pp'}}}{\sqrt{N_{pp'}}}$
$n_0\sigma_a^2+2n_0\sigma_d^2+\sigma_e^2$	$2n_0-2$	$\boldsymbol{v}_{pp'_l}$ such that $\boldsymbol{V}_{pp'}\boldsymbol{v}_{pp'_l}=\nu_{pp'_2}\boldsymbol{v}_{pp'_l}$
σ_e^2	$(n_0-1)^2$	$\boldsymbol{v}_{kp_l}$ such that $\boldsymbol{V}_{pp'}\boldsymbol{v}_{pp'_l}=\nu_{pp'_3}\boldsymbol{v}_{pp'_l}$

Table 9.4: Eigenstructure of design matrices, $\boldsymbol{\Sigma}_{kk}, \boldsymbol{\Sigma}_{pp}, \boldsymbol{\Sigma}_{kp}$ and $\boldsymbol{\Sigma}_{pp'}$ in (9.14) and (9.15)

These results follow from using the $\boldsymbol{B}_{ii}$ and $\boldsymbol{B}_{ii'}$ matrices to activate certain areas of the vector $\widetilde{\boldsymbol{s}}_c$ and the matrices $\boldsymbol{V}_{ii}$, $\boldsymbol{V}_{ii'}$, i.e., introducing the matrices $\boldsymbol{B}_{ii}$ and $\boldsymbol{B}_{ii'}$ allows us to activate the parts of $\widetilde{\boldsymbol{s}}_c$ and $\boldsymbol{\Sigma}_c$ that correspond to the different source comparisons. Rather than consider a sparse N_c vector alongside a sparse $N_c \times N_c$ matrix, we can directly consider the interesting parts of the vector and matrix by considering the associated $N_{ii}-$ or $N_{ii'}-$ dimensional vector, and $N_{ii} \times N_{ii}$ or $N_{ii'} \times N_{ii'}$ dimensional matrix. Thus, we can explicitly define the sums of squares such that

$$\begin{aligned} SS_{ii} \quad = \quad & \left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right)'\left(\boldsymbol{v}_{ii_1}^{*}\boldsymbol{v}_{ii_1}^{*}{}'\right)\left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right) \;+ \\ & \left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right)'\left(\sum_{l=2}^{n_0}\boldsymbol{v}_{ii_l}^{*}\boldsymbol{v}_{ii_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right) \;+ \\ & \left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{22}\mathbf{1}_{N_{ii}}\right)\right)'\left(\sum_{l=n_0+1}^{N_{ii}}\boldsymbol{v}_{ii_l}^{*}\boldsymbol{v}_{ii_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii}-\theta_{ii}\mathbf{1}_{N_{ii}}\right)\right), \end{aligned} \tag{9.16}$$

for within-source comparisons and

$$\begin{aligned} SS_{ii'} &= \left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right)'\left(\boldsymbol{v}_{ii'_1}^{*}\boldsymbol{v}_{ii'_1}^{*}{}'\right)\left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right) + \\ & \left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right)'\left(\sum_{l=2}^{\binom{n_0}{2}-1}\boldsymbol{v}_{ii'_l}^{*}\boldsymbol{v}_{ii'_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right) + \quad (9.17) \\ & \left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{12}\mathbf{1}_{N_{ii'}}\right)\right)'\left(\sum_{l=\binom{n_0}{2}}^{N_{ii'_l}}\boldsymbol{v}_{ii'_l}^{*}\boldsymbol{v}_{ii'_l}^{*}{}'\right)\left(\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right), \end{aligned}$$

for between-source comparisons, where the degrees of freedom for each line of (9.16) and (9.17) are equal to the multiplicities of the associated eigenvalues in Table 9.4, and the total sum of squares remains as in (9.9). It is trivial to show that

$$\sum_{i=1}^{r+1}\left[\sum_{l=1}^{3} m_{\nu_{ii_l}}\right] + \sum_{i=1}^{r}\sum_{i'=i+1}^{r+1}\left[\sum_{l=1}^{3} m_{\nu_{ii'_l}}\right] = N_c.$$

In the following sections, we analyse the three terms that make up the sums of squares defined in (9.16) and (9.17) so that we can write each term without the use of eigenvectors.

9.2.1 ALTERNATIVE REPRESENTATION OF WITHIN-SOURCE SUMS OF SQUARES INVOLVING THE PUTATIVE SOURCE

We begin by studying the individual terms in the within-source sum of squares for the putative source. This sum of squares follows the form given by (9.16), which is composed of three sums of squared terms. The developments for these terms follow those presented in Appendix D. We re-write the first term as

$$\begin{aligned} \left(\boldsymbol{s}_{kk} - \theta_{kk}\mathbf{1}_{N_{kk}}\right)'\boldsymbol{\Delta}_{kk}^{-1'}\boldsymbol{v}_{kk_1}^{*}\boldsymbol{v}_{kk_1}^{*'}\boldsymbol{\Delta}_{kk}\left(\boldsymbol{s}_{kk} - \theta_{kk}\mathbf{1}_{N_{kk}}\right) &= \frac{N_{kk}}{\sigma_{kk}^2}\left(\bar{s}_{kk} - \theta_{kk}\right)^2 \quad (9.18) \\ &:= SS_{M_{kk}}, \end{aligned}$$

where $\bar{s}_{kk}$ is the average score observed for the within-source comparisons from the putative source, k. Recall that $N_{kk} = \binom{n_0}{2}$.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{kk}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{kk_l}^{*}\boldsymbol{v}_{kk_l}^{*'}\right]\boldsymbol{\Delta}_{kk}^{-1}$.

Following the development by [6], we can write this second sum of squares as

$$\boldsymbol{s}'_{kk}\boldsymbol{\Delta}_{kk}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}^*_{kk_l}\boldsymbol{v}^{*'}_{kk_l}\right]\boldsymbol{\Delta}_{kk}^{-1}\boldsymbol{s}_{kk} = \frac{(n_0-1)^2}{\sigma^2_{kk}(n_0-2)}\sum_{j=1}^{n_0}\left(\bar{s}^{(k_j)}_{kk}-\bar{s}_{kk}\right)^2 := SS_{W_{kk}}, \quad (9.19)$$

where $\bar{s}^{(k_j)}_{kk}$ is the mean value of scores that compare object j from the putative source, k, to any other object in the putative source, k, and $\bar{s}_{kk}$ is as in (9.18). The final result, given by (9.19), gives the within-source sum of squares for the putative source model. By considering this term in conjunction with the total sum of squares from the considered model, $SS_{Tot_{kk}} = \frac{1}{\sigma^2_{kk}}\sum_{j=1}^{n_0}\sum_{j'=j+1}^{n_0-1}\left(s_{kj,kj'}-\bar{s}_{kk}\right)^2$ (this is the sum of the last two terms in (9.16)), we can obtain an eigenvector-free estimate of the last term in SS_{kk} by considering $SS_{Tot_{kk}} - SS_{W_{kk}}$ (see Table 9.5). By using the results of Cochran's theorem presented in Section 3.1.2, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	n_0-1	$SS_{W_{kk}}$	$MS_{W_{kk}} = \frac{SS_{W_{kk}}}{n_0-1}$	$(n_0-2)\sigma_a^2 + 2(n_0-2)\sigma_d^2 + \sigma_e^2$
Error	$N_{kk}-n_0$	$SS_{E_{kk}} = SS_{Tot_{kk}} - SS_{W_{kk}}$	$MS_{E_{kk}} = \frac{SS_{E_{kk}}}{N_{kk}-n_0}$	σ_e^2

Table 9.5: ANOVA table corresponding to within-source sums of squares for the putative source, SS_{kk}.

9.2.2 ALTERNATIVE REPRESENTATION OF BETWEEN-SOURCE SUMS OF SQUARES INVOLVING THE PUTATIVE SOURCE AND A RANDOM SOURCE FROM THE POPULATION

Next, we consider the terms in the between-source sums of squares terms of the form given by (9.17) that consider one object from the putative source compared to one object from a random source in the population. The developments for these terms follow those presented in Appendix E. As in the previous sums of squares development, we can rewrite the first term as

$$\left(\boldsymbol{s}_{kp}-\theta_{k\mathscr{P}}\mathbf{1}_{N_{kp}}\right)'\boldsymbol{\Delta}_{kp}^{-1'}\boldsymbol{v}^*_{kp_1}\boldsymbol{v}^{*'}_{kp_1}\boldsymbol{\Delta}_{kp}^{-1}\left(\boldsymbol{s}_{kp}-\theta_{k\mathscr{P}}\mathbf{1}_{N_{kp}}\right) = \frac{N_{kp}}{\sigma^2_{k\mathscr{P}}}\left(\bar{s}_{kp}-\theta_{k\mathscr{P}}\right)^2 := SS_{M_{kp}}, \quad (9.20)$$

where $\bar{s}_{kp}$ is the average score observed for between-source comparisons that consider an object between the putative source, k, and from a random source from the population, $p \in \{1, \ldots, r\}$. Recall that $N_{kp} = n_0^2$.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{kp}^{-1} \left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}_{kp_l}^{*} \boldsymbol{v}_{kp_l}^{*\prime}\right] \boldsymbol{\Delta}_{kp}^{-1}$. As before, we can write this second sum of squares as

$$\begin{aligned} \boldsymbol{s}_{kp}^{\prime} \boldsymbol{\Delta}_{kp}^{-1\prime} \left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}_{kp_l}^{*} \boldsymbol{v}_{kp_l}^{*\prime}\right] \boldsymbol{\Delta}_{kp}^{-1} \boldsymbol{s}_{kp} &= \frac{n_0^2}{\sigma_{k\mathscr{P}}^2 n_0} \left(\sum_{j=1}^{n_0} \left(\bar{s}_{kp}^{(k_j p)} - \bar{s}_{kp} \right)^2 + \sum_{j'=1}^{n_0} \left(\bar{s}_{kp}^{(kp_{j'})} - \bar{s}_{kp} \right)^2 \right) \\ &= SS_{W_{kp}}, \end{aligned} \tag{9.21}$$

where $\bar{s}_{kp}^{(k_j)}$, $p \in \{1, \ldots, r\}$, $j \in \{1, \ldots, n_0\}$, is the mean value of scores that compare object j in source k to any object in source p, $\bar{s}_{kp}^{(p_{j'})}$, $p \in \{1, \ldots, r\}$, $j' \in \{1, \ldots, n_0\}$ is the mean value of scores that compare object j' in source p to any object in source k, and $\bar{s}_{kp}$ is as in (9.20). The final result given by (9.21), gives the within-source sum of squares for the between-source k, p comparisons. By considering this term in conjunction with the total sum of squares for the considered model, $SS_{Tot_{kp}} = \frac{1}{\sigma_{k\mathscr{P}}^2} \sum_{j=1}^{n_0-1} \sum_{j'=1}^{n_0} \left(s_{kj,pj'} - \bar{s}_{kp} \right)^2$, we can define an eigenvector-free estimate of the last term in SS_{kp} by considering $SS_{Tot_{kp}} - SS_{W_{kp}}$. By using the results of Cochran's theorem presented in Section 3.1.2, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	$2n_0 - 2$	$SS_{W_{kp}}$	$MS_{W_{kp}} = \frac{SS_{W_{kp}}}{2n_0-2}$	$n_0\sigma_a^2 + 2n_0\sigma_d^2 + \sigma_e^2$
Error	$(n_0 - 1)^2$	$SS_{E_{kp}} = SS_{Tot_{kp}} - SS_{W_{kp}}$	$MS_{E_{kp}} = \frac{SS_{E_{kp}}}{(n_0-1)^2}$	σ_e^2

Table 9.6: ANOVA table corresponding to within-source sums of squares for the putative source and random sources from the population, SS_{kp}.

9.2.3 ALTERNATIVE REPRESENTATION OF WITHIN-SOURCE SUMS OF SQUARES INVOLVING RANDOM SOURCES FROM THE POPULATION

We now move to study the individual terms in the within-source sum of squares for random sources from the population. This sum of squares follows the form given by (9.16), which is composed of three sums of squared terms. The developments for these terms follow those presented in Appendix D. We re-write the first term as

$$\begin{aligned}\left(\boldsymbol{s}_{pp} - \theta_{\mathscr{P}\mathscr{P}}\mathbf{1}_{N_{pp}}\right)' \boldsymbol{\Delta}_{pp}^{-1'} \boldsymbol{v}_{pp_1}^{*} \boldsymbol{v}_{pp_1}^{*'} \boldsymbol{\Delta}_{pp} \left(\boldsymbol{s}_{pp} - \theta_{\mathscr{P}\mathscr{P}}\mathbf{1}_{N_{pp}}\right) &= \frac{N_{pp}}{\sigma^2_{\mathscr{P}\mathscr{P}}} \left(\bar{s}_{pp} - \theta_{\mathscr{P}\mathscr{P}}\right)^2 \quad (9.22)\\ &:= SS_{M_{pp}},\end{aligned}$$

where $\bar{s}_{pp}$ is the average score observed for the within-source comparisons from random sources from the population, $p \in \{1, \ldots, r\}$. Recall that $N_{pp} = \binom{n_0}{2}$.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{pp}^{-1'} \left[\sum_{l=2}^{n_0} \boldsymbol{v}_{pp_l}^{*} \boldsymbol{v}_{pp_l}^{*'}\right] \boldsymbol{\Delta}_{pp}^{-1}$. Following the development above, we can write this second sum of squares as

$$\begin{aligned}\boldsymbol{s}_{pp}' \boldsymbol{\Delta}_{pp}^{-1'} \left[\sum_{l=2}^{n_0} \boldsymbol{v}_{pp_l}^{*} \boldsymbol{v}_{pp_l}^{*'}\right] \boldsymbol{\Delta}_{pp}^{-1} \boldsymbol{s}_{pp} &= \frac{(n_0 - 1)^2}{\sigma^2_{\mathscr{P}\mathscr{P}}(n_0 - 2)} \sum_{j=1}^{n_0} \left(\bar{s}_{pp}^{(p_j)} - \bar{s}_{pp}\right)^2 \quad (9.23)\\ &:= SS_{W_{pp}},\end{aligned}$$

where $\bar{s}_{pp}^{(p_j)}$ is the mean value of scores that compare object j from the considered random source, p, to any other object from the considered random source, p, and $\bar{s}_{pp}$ is as in (9.22). The final result, given by (9.23), gives the within-source sum of squares for the random sources model. By considering this term in conjunction with the total sum of squares from the considered model, $SS_{Tot_{pp}} = \frac{1}{\sigma^2_{\mathscr{P}\mathscr{P}}} \sum_{j=1}^{n_0} \sum_{j'=j+1}^{n_0-1} \left(s_{pj,pj'} - \bar{s}_{pp}\right)^2$ (this is the sum of the last two terms in (9.16)), we can obtain an eigenvector-free estimate of the last term in SS_{pp} by considering $SS_{Tot_{pp}} - SS_{W_{pp}}$ (see Table 9.7). By using the results of Cochran's theorem presented in Section 3.1.2, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	$n_0 - 1$	$SS_{W_{pp}}$	$MS_{W_{pp}} = \frac{SS_{W_{pp}}}{n_0 - 1}$	$(n_0 - 2)\sigma_a^2 + 2(n_0 - 2)\sigma_d^2 + \sigma_e^2$
Error	$N_{pp} - n_0$	$SS_{E_{pp}} = SS_{Tot_{pp}} - SS_{W_{pp}}$	$MS_{E_{pp}} = \frac{SS_{E_{pp}}}{N_{pp} - n_0}$	σ_e^2

Table 9.7: ANOVA table corresponding to within-source sums of squares for the putative source, SS_{pp}.

9.2.4 ALTERNATIVE REPRESENTATION OF BETWEEN-SOURCE SUMS OF SQUARES INVOLVING RANDOM SOURCES FROM THE POPULATION

Next, we consider the terms in the between-source sums of squares terms of the form given by (9.17) random sources in the population. The developments for these terms follow those presented in Appendix E. As above, we can rewrite the first term as

$$\left(\boldsymbol{s}_{pp'} - \theta_{\mathscr{P}\mathscr{P}'}\mathbf{1}_{N_{pp'}}\right)' \boldsymbol{\Delta}_{pp'}^{-1'} \boldsymbol{v}_{pp'_1}^{*} \boldsymbol{v}_{pp'_1}^{*'} \boldsymbol{\Delta}_{pp'}^{-1} \left(\boldsymbol{s}_{pp'} - \theta_{\mathscr{P}\mathscr{P}'}\mathbf{1}_{N_{pp'}}\right) = \frac{N_{pp'}}{\sigma^2_{\mathscr{P}\mathscr{P}'}} \left(\bar{s}_{pp'} - \theta_{\mathscr{P}\mathscr{P}'}\right)^2 \quad (9.24)$$

$$:= SS_{M_{pp'}}, \quad (9.25)$$

where $\bar{s}_{pp'}$ is the average score observed for between-source comparisons that consider objects from random sources from the population, $p \neq p' \in \{1, \ldots, r\}$. Recall that $N_{pp'} = n_0^2$.

Next, we consider the structure of the sum given by $\boldsymbol{\Delta}_{pp'}^{-1}\left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}_{pp'_l}^{*} \boldsymbol{v}_{pp'_l}^{*'}\right] \boldsymbol{\Delta}_{pp'}^{-1}$. As before, we can write this second sum of squares as

$$\boldsymbol{s}'_{pp'} \boldsymbol{\Delta}_{pp'}^{-1'} \left[\sum_{l=2}^{2n_0-1} \boldsymbol{v}_{pp'_l}^{*} \boldsymbol{v}_{pp'_l}^{*'}\right] \boldsymbol{\Delta}_{pp'}^{-1} \boldsymbol{s}_{pp'} = \frac{n_0^2}{\sigma^2_{\mathscr{P}\mathscr{P}'} n_0} \left(\sum_{j'=1}^{n_0} \left(\bar{s}_{pp'}^{(pp'_{j'})} - \bar{s}_{pp'}\right)^2 + \sum_{j=1}^{n_0} \left(\bar{s}_{pp'}^{(p_j p')} - \bar{s}_{pp'}\right)^2 \right) \quad (9.26)$$

$$= SS_{W_{kp}},$$

where $\bar{s}_{pp'}^{(p_j)}$, $p, p' \in \{1, \ldots, r\}$, $j \in \{1, \ldots, n_0\}$ is the mean value of scores that compare object j in source p to any object in source p', $\bar{s}_{pp'}^{(p'_{j'})}$, $j' \in \{1, \ldots, n_0\}$ is the mean value of scores that compare object j' in source p' to any object in source p, and $\bar{s}_{pp'}$ is as in (9.24). The final result given by (9.26), gives the within-source sum of squares for the between-source p, p' comparisons. By considering this term in conjunction with the total sum of squares for the considered model, $SS_{Tot_{pp'}} = \frac{1}{\sigma^2_{\mathscr{P}\mathscr{P}'}} \sum_{j=1}^{n_0-1} \sum_{j'=1}^{n_0} \left(s_{pj,p'j'} - \bar{s}_{pp'}\right)^2$, we can define an eigenvector-free estimate of the last term in $SS_{pp'}$ by considering $SS_{Tot_{pp'}} - SS_{W_{pp'}}$. By using the results of Cochran's theorem presented in Section 3.1.2, we obtain the following results.

Source of Variance	df	SS	MS	E(MS)
Within Source	$2n_0-2$	$SS_{W_{pp'}}$	$MS_{W_{pp'}}=\frac{SS_{W_{pp'}}}{2n_0-2}$	$n_0\sigma_a^2+2n_0\sigma_d^2+\sigma_e^2$
Error	$(n_0-1)^2$	$SS_{E_{pp'}}=SS_{Tot_{pp'}}-SS_{W_{pp'}}$	$MS_{E_{pp'}}=\frac{SS_{E_{pp'}}}{(n_0-1)^2}$	σ_e^2

Table 9.8: ANOVA table corresponding to within-source sums of squares for random sources from the population, $SS_{pp'}$.

9.3 PARAMETER ESTIMATION

At this point, we would like to use the results presented in Sections 9.2.1, 9.2.2, 9.2.3, and 9.2.4 to estimate the parameters of our model; however, given the dependencies that exist between the various parameters, we must resort to sampling methods to obtain posterior samples of the model parameters. In particular, we use a Gibbs sampler with a Metropolis-Hastings step to study the distributions of our various parameters [14, 44]. Before defining the Gibbs sampler, we must first assign posterior distributions to the model parameters (developments for the posterior distributions of $\theta_{ii'}$ and $\sigma_{ii'}$ follow those presented in Appendices F.2 and F.3.

We begin by assigning posterior distributions for the variance terms, σ_a^2, σ_b^2, σ_c^2, σ_d^2, and σ_e^2. Given the particular dependency that exists between these parameters, we can define a posterior distribution to study these variance terms simultaneously. We begin by collecting all mean sums of squares terms defined in Tables 9.5, 9.6 9.7, and 9.8 to capitalize on all information related to the values of σ_a^2, σ_b^2, σ_c^2, σ_d^2, and σ_e^2. We have that

$$\frac{SS_{M_{kk}}}{2(n_0-1)\sigma_a^2+\binom{n_0}{2}\sigma_c^2+4(n_0-1)\sigma_d^2+\sigma_e^2}\sim\chi^2_{df=1}\quad\frac{SS_{W_{kk}}}{(n_0-2)\sigma_a^2+2(n_0-2)\sigma_d^2+\sigma_e^2}\sim\chi^2_{df=n_0-1}\quad\frac{SS_{E_{kk}}}{\sigma_e^2}\sim\chi^2_{df=N_{kk}-n_0}$$

for the sum of squares that consider comparisons that occur within the putative source,

$$\frac{SS_{M_{kp}}}{2n_0\sigma_a^2+n_0^2\sigma_c^2+4n_0\sigma_d^2+\sigma_e^2}\sim\chi^2_{df=1}\quad\frac{SS_{W_{kp}}}{n_0\sigma_a^2+2n_0\sigma_d^2+\sigma_e^2}\sim\chi^2_{df=2n_0-2}\quad\frac{SS_{E_{kp}}}{\sigma_e^2}\sim\chi^2_{df=(n_0-1)^2}$$

for the r sums of squares terms that consider comparisons that occur between an object from the putative source and an object from a random source from the population,

$$\frac{SS_{M_{pp}}}{2(n_0-1)\sigma_a^2+\binom{n_0}{2}\sigma_c^2+4(n_0-1)\sigma_d^2+\sigma_e^2} \sim \chi^2_{df=1} \quad \frac{SS_{W_{pp}}}{(n_0-2)\sigma_a^2+2(n_0-2)\sigma_d^2+\sigma_e^2} \sim \chi^2_{df=n_0-1} \quad \frac{SS_{E_{pp}}}{\sigma_e^2} \sim \chi^2_{df=N_{pp}-n_0}$$

for the r sums of squares terms that consider comparisons that occur within the same random source from the population, and

$$\frac{SS_{M_{pp'}}}{2n_0\sigma_a^2+n_0^2\sigma_c^2+4n_0\sigma_d^2+\sigma_e^2} \sim \chi^2_{df=1} \quad \frac{SS_{W_{pp'}}}{n_0\sigma_a^2+2n_0\sigma_d^2+\sigma_e^2} \sim \chi^2_{df=2n_0-2} \quad \frac{SS_{E_{pp'}}}{\sigma_e^2} \sim \chi^2_{df=(n_0-1)^2}$$

for the $\binom{r}{2}$ sums of squares terms that consider comparisons that occur between two random sources from the population. Note that the sums of squares terms for the two within-source scenarios correspond, as do the sums of squares terms for the two between-source scenarios. Using this information, we can define

$$\begin{aligned} MS_\sigma &= \sum_{i\in\{1,\dots,r,k\}} \left(\frac{SS_{M_{ii}}}{C_1} + \frac{SS_{W_{ii}}}{C2} + \frac{SS_{E_{ii}}}{C_3}\right) \\ &\quad + \sum_{i\in\{1,\dots,r\}} \sum_{i'\in\{2,\dots,r,k\}} \left(\frac{SS_{M_{ii'}}}{C_4} + \frac{SS_{W_{ii'}}}{C_5} + \frac{SS_{E_{ii'}}}{C_3}\right) \\ &\sim \chi^2_{df=\binom{(r+1)n_0}{2}-\binom{r+1}{2}-(r+1)} \end{aligned} \tag{9.27}$$

where

$$C_1 = 2(n_0-1)\sigma_a^2 + \binom{n_0}{2}\sigma_c^2 + 4(n_0-1)\sigma_d^2 + \sigma_e^2 \qquad C_2 = (n_0-2)\sigma_a^2 + 2(n_0-2)\sigma_d^2 + \sigma_e^2 \qquad C_3 = \sigma_e^2$$

$$C_4 = 2n_0\sigma_a^2 + n_0^2\sigma_c^2 + 4n_0\sigma_d^2 + \sigma_e^2 \qquad C_5 = n_0\sigma_a^2 + 2n_0\sigma_d^2 + \sigma_e^2.$$

We can now define a posterior distribution for the variance terms, σ_a^2, σ_b^2, σ_c^2, σ_d^2, and σ_e^2, by considering a χ^2 likelihood for the MS_σ term, and assuming a Dirichlet prior (since we have the constraint that $2\sigma_a^2 + \sigma_c^2 + 4\sigma_d^2 + \sigma_e^2 = 2\sigma_a^2 + 2\sigma_b^2 + 4\sigma_d^2 + \sigma_e^2 = 1$, with $\sigma_c^2 = 2\sigma_b^2$), such that

$$\begin{aligned} \pi\left(\{\sigma_a^2,\sigma_b^2,\sigma_d^2,\sigma_e^2\}|MS_\sigma,\{\theta_{ii'}\}_{ii'},\{\sigma_{ii'}\}_{ii'},\boldsymbol{s},\boldsymbol{\alpha}\right) &\propto \chi^2\left(MS_\sigma|\{\sigma_a^2,\sigma_b^2,\sigma_d^2,\sigma_e^2\},\{\theta_{ii'}\}_{ii'},\{\sigma_{ii'}\}_{ii'},\boldsymbol{s},\alpha_\sigma,\beta_\sigma\right) \\ &\quad \times \mathcal{D}\left(\{\sigma_a^2,\sigma_b^2,\sigma_d^2,\sigma_e^2\}|\boldsymbol{\alpha}\right), \end{aligned} \tag{9.28}$$

where the dependence of MS_σ on $\{\theta_{ii'}\}_{ii'}, \{\sigma_{ii'}\}_{ii'}$, $ii' \in \{kk, k\mathscr{P}, \mathscr{P}\mathscr{P}, \mathscr{P}\mathscr{P}'\}$ in (9.28) results from the construction of the MS_σ term as a sum of the various sums of squares terms defined in Section 9.2.

Next, we move to assign the posterior distributions for the mean terms. However, this model requires proceeding via a different route than that considered for the two-class and multi-class models. While the posterior distributions take the same form, we consider a fixed number of terms. In the two-class and multi-class scenarios, the number of mean and standard deviation terms depended on the number of sources being considered. For this model, rather than consider $(r+1)\binom{n_0}{2} + \binom{r+1}{2}n_0^2$ mean terms, we consider four mean and standard deviation terms related to the four varieties of source-comparisons that we encounter in this model (i.e., within putative source comparisons, between putative and random sources from the population, within random sources from the population, and between random from the population), regardless of the number of random sources considered.

Thus, we assign the posterior distributions for each of the four mean parameters, $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$, by considering a multivariate normal likelihood over the scores that consider the particular source comparison, and assuming a Normal prior with mean $\phi_{ii'}$ and variance $\omega_{ii'}$, $ii' \in \{kk, k\mathscr{P}, \mathscr{P}\mathscr{P}, \mathscr{P}\mathscr{P}'\}$ such that

$$\begin{aligned}\pi\left(\theta_{ii'}|\boldsymbol{s}_{ii'}, \sigma_{ii'}, \sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2, \phi_{ii'}, \omega_{ii'}\right) \quad &\propto \quad \mathcal{MVN}\left(\boldsymbol{s}_{ii'}|\theta_{ii'}, \sigma_{ii'}, \sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2, \phi_{ii'}., \omega_{ii'}\right) \\ &\times \quad \mathcal{N}\left(\theta_{ii'}|\phi_{ii'}, \omega_{ii'}\right) \qquad (9.29)\end{aligned}$$

where the resulting posterior distribution is Normally distributed. The parameters of the posterior distribution of $\theta_{ii'}$ are given by

$$\mu_{ii'_p} = \frac{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \phi_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1} \qquad \sigma^2_{ii'_p} = \frac{\omega_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1},$$

where the terms $N_{ii'}$, $\mathbf{1}_{N_{ii'}}$, $\boldsymbol{s}_{ii'}$ and $\boldsymbol{\Sigma}_{ii'}$ have dimension $\binom{n_0}{2}$ when we consider the vector of scores that compare two objects from the putative source (i.e., $ii' = kk$), have dimension rn_0^2 when we consider the vector of scores that compare an object from the putative source to an object from the population (i.e., $ii' = k\mathscr{P}$), have dimension $r\binom{n_0}{2}$ when we consider the vector of scores that consider two objects from the same random source in the population (i.e., $ii' = \mathscr{P}\mathscr{P}$), and have dimension $\binom{r}{2}n_0^2$ when we consider the vector of scores that compare two objects from different random sources in the population (i.e., $ii' = \mathscr{P}\mathscr{P}'$).

As in the case for the mean terms, we find the posterior distributions for each of the four variance parameters $\sigma_{ii'} \in \{\sigma_{kk}, \sigma_{k\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}'}\}$, by considering a Multivariate Normal likelihood over the scores that consider the particular source comparison, and assuming an Inverse-Gamma prior such that

$$\begin{aligned} \pi\left(\sigma_{ii'}^2|\boldsymbol{s}_{ii'}, \theta_{ii'}, \sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2, \alpha_{ii'}, \beta_{ii'}\right) \quad &\propto \quad \mathcal{MVN}\left(\boldsymbol{s}_{ii'}|\sigma_{ii'}^2, \theta_{ii'}, \sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2, \alpha_{ii'}, \beta_{ii'}\right) \\ &\times \quad \mathcal{IG}\left(\sigma_{ii'}^2|\alpha_{ii'}, \beta_{ii'}\right) \end{aligned} \tag{9.30}$$

where the resulting posterior distribution follows an Inverse Gamma distribution. The parameters of the posterior distribution of $\sigma_{ii'}^2$ are given by

$$\alpha_{ii'_p} = \frac{N_{ii'}}{2} + \alpha_{ii'} \qquad \beta_{ii'_p} = \frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Sigma}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right) + \beta_{ii'},$$

where, again, the terms $N_{ii'}$, $\mathbf{1}_{N_{ii'}}$, $\boldsymbol{s}_{ii'}$ and $\boldsymbol{\Sigma}_{ii'}$ have dimension $\binom{n_0}{2}$ when we consider the vector of scores that compare two objects from the putative source (i.e., $ii' = kk$), have dimension rn_0^2 when we consider the vector of scores that compare an object from the putative source to an object from the population (i.e., $ii' = k\mathscr{P}$), have dimension $r\binom{n_0}{2}$ when we consider the vector of scores that consider two objects from the same random source in the population (i.e., $ii' = \mathscr{P}\mathscr{P}$), and have dimension $\binom{r}{2}n_0^2$ when we consider the vector of scores that compare two objects from different random sources in the population (i.e., $ii' = \mathscr{P}\mathscr{P}'$).

The posterior distribution given by (9.30) provides us with samples from the posterior distribution of the variance term, $\sigma_{ii'}^2$. We, however, are interested in the standard deviation term, $\sigma_{ii'}$, and so we simply take the square root of the sampled variance terms to obtain samples of the standard deviations. The resulting inference is not affected.

We note that each of the distributions described in (9.28), (9.29), and (9.30) depends on the value of at least one other parameter; therefore, we must rely on sampling techniques to study the distributions of the model parameters. We construct the following Gibbs sampler, described in Algorithm 5.

Algorithm 5: Gibbs sampler for generating posterior samples from the distributions of the model parameters

---

Data: Initial values for all parameters at $t = 0$; values for hyperparameters
Result: Posterior samples for all parameters
for $t \in 1 : T$ iterations **do**
  1. Draw $\{\sigma_a^{2(t)}, \sigma_b^{2(t)}, \sigma_d^{2(t)}, \sigma_e^{2(t)}\}|MS_e^{(t-1)}, \boldsymbol{\sigma}^{(t-1)}, \boldsymbol{s}, \boldsymbol{\alpha}$ from the distribution defined in (9.28);
  2. Calcualte $\sigma_c^{2(t)} = 2\sigma_b^{2(t)}$;
  for $ii' \in \{kk, k\mathscr{P}, \mathscr{P}\mathscr{P}, \mathscr{P}\mathscr{P}'\}$ source comparisons **do**
    3. Draw $\theta_{ii'}^{(t)}|\boldsymbol{s}_{ii'}, \sigma_{ii'}^{(t-1)}, \sigma_a^{2(t)}, \sigma_b^{2(t)}, \sigma_d^{2(t)}, \sigma_e^{2(t)}, \phi_{ii'}, \omega_{ii'}$ from the distribution defined in (9.29);
    4. Draw $\sigma_{ii'}^{2(t)}|\boldsymbol{s}_{ii'}, \theta_{ii'}^{(t)}, \sigma_a^{2(t)}, \sigma_b^{2(t)}, \sigma_d^{2(t)}, \sigma_e^{2(t)}, \alpha_{ii'}, \beta_{ii'}$ from the distribution defined in (9.30);
  end
end

Algorithm 6: Metropolis-Hastings algorithm for obtaining a sample of $\widetilde{\boldsymbol{\sigma}}^2 = \{\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2\}$ from the posterior distribution of $\{\sigma_a^2, \sigma_b^2, \sigma_d^2, \sigma_e^2\}$

Data: Current value of $\{\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2\}^{(t-1)}$; value for hyperparameter $\boldsymbol{\alpha}$
Result: Posterior sample of $\{\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2\}$
1. Define the vector of $\widetilde{\boldsymbol{\sigma}}^{2(t-1)} := (2\widetilde{\sigma}_a^{2(t-1)}, 2\widetilde{\sigma}_b^{2(t-1)}, 4\widetilde{\sigma}_d^{2(t-1)}, \widetilde{\sigma}_e^{2(t-1)})$ in terms of the current sigma values;
2. Sample a candidate value, $\widetilde{\boldsymbol{\sigma}}^{2*} \sim \mathcal{D}\left(\widetilde{\boldsymbol{\sigma}}^{2(t-1)}\right)$;
3. Calculate the value of MS_σ^* using the candidate value $\widetilde{\boldsymbol{\sigma}}^{2*}$ using (9.27);
4. Calculate the value of $MS_\sigma^{(t-1)}$ using the current value $\widetilde{\boldsymbol{\sigma}}^{2(t-2)}$ using (9.27);
5. Evaluate the posterior density of $\widetilde{\boldsymbol{\sigma}}^{2*}$, f^*, using the hyperparameter $\boldsymbol{\alpha}$, the value of MS_σ^*, and (9.28);
6. Evaluate the posterior density of $\widetilde{\boldsymbol{\sigma}}^{2(t-1)}$, $f^{(t-1)}$, using the hyperparameter $\boldsymbol{\alpha}$, the value of $MS_\sigma^{(t-1)}$, and (9.28);
7. Calculate the probability of acceptance, $p_{acc} = \frac{f^*}{f^{(t-1)}} \frac{\mathcal{D}\left(\widetilde{\boldsymbol{\sigma}}^{2(t-1)}|\widetilde{\boldsymbol{\sigma}}^{2*}\right)}{\mathcal{D}\left(\widetilde{\boldsymbol{\sigma}}^{2*}|\widetilde{\boldsymbol{\sigma}}^{2(t-1)}\right)}$
8. Generate a random probability, $p^* \sim \mathcal{U}(0, 1)$;
9. If $p_{acc} \geq p^*$, then define $\widetilde{\boldsymbol{\sigma}}^{2(t)} := \widetilde{\boldsymbol{\sigma}}^{2*}$; otherwise define $\widetilde{\boldsymbol{\sigma}}^{2(t)} := \widetilde{\boldsymbol{\sigma}}^{2(t-1)}$;
10. Return the vector $\{\widetilde{\sigma}_1^{2(t)}/2, \widetilde{\sigma}_2^{2(t)}/2, \widetilde{\sigma}_2^{2(t)}, \widetilde{\sigma}_3^{2(t)}/4, \widetilde{\sigma}_4^{2(t)}\}$ as the current values of $\{\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2\}$.

Because the posterior distribution defined for $\{\sigma_a^2, \sigma_b^2, \sigma_d^2, \sigma_e^2\}$ in (9.28) is not readily available, we cannot directly sample from this distribution. As a result, the first step in Algorithm 5 is not so straightforward - indeed, obtaining a sample from the posterior distribution of $\{\sigma_a^2, \sigma_b^2, \sigma_d^2, \sigma_e^2\}$ requires introducing a Metropolis-Hastings algorithm [14, 44]. This procedure is summarized in Algorithm 6.

Now that we have identified a method for obtaining samples of the parameters used to define $\boldsymbol{\theta}$, $\boldsymbol{\Delta}$, and $\{\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2, \sigma_e^2\}$, we can assign a Bayes factor to determine if it is more likely that the set of trace objects originated from the putative source, or it if is more likely that the set of trace objects originated from some random source in the population of potential sources. In particular, we define

$$\Lambda = \frac{\int_{\boldsymbol{\Omega}_p} \ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_p, \boldsymbol{s}_c\right) d\Pi\left(\boldsymbol{\Omega}_p|\boldsymbol{s}_c\right)}{\int_{\boldsymbol{\Omega}_d} \ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_d, \boldsymbol{s}_c\right) d\Pi\left(\boldsymbol{\Omega}_d|\boldsymbol{s}_c\right)} \tag{9.31}$$

$$\approx \frac{\frac{1}{T}\sum_{t=1}^{T} \ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_p^{(t)}, \boldsymbol{s}_c\right)}{\frac{1}{T}\sum_{t'=1}^{T} \ell\left(\boldsymbol{s}_t|\boldsymbol{\Omega}_d, \boldsymbol{s}_c\right)} \tag{9.32}$$

where $\boldsymbol{s}_t$ is the vector of scores that consider at least one trace object, the subscripts on $\boldsymbol{\Omega}$ correspond to the model being considered, and $\boldsymbol{\Omega}^{(t)}$ are posterior samples of the parameters obtained using Algorithm 5. We consider the conditional posterior probability of the scores that consider objects of unknown origin from the set $\boldsymbol{X}_u$, $\boldsymbol{s}_t$, rather than the joint posterior probability of $\boldsymbol{s}$, so as not to recycle the information contained in the vector of scores $\boldsymbol{s}_c$, which are used to sample the parameter values.

9.4 RECOVERING THE MODEL PARAMETERS

In this section, we move to assess the performance of the proposed model. We look at simulated data with known parameters to determine the ability of the model to recover the model parameters as the number of random sources sampled from the population, r, and the number of observations per source, n_0, increase. To determine the combination of r and n_0 that allows the model to appropriately estimate the parameters, we consider a scenario in which we sample scores directly from the proposed distribution given by (9.8). We fix $\theta_{kk} = 10.9$, $\theta_{k\mathscr{P}} = 16.5$, $\theta_{\mathscr{P}\mathscr{P}} = 13.0$, and $\theta_{\mathscr{P}\mathscr{P}'} = 9.7$ for the mean terms, and $\sigma_{kk} = 2.6$, $\sigma_{k\mathscr{P}} = 3.4$, $\sigma_{\mathscr{P}\mathscr{P}} = 2.8$, and $\sigma_{\mathscr{P}\mathscr{P}'} = 4.4$ for the standard deviation terms.

To gauge the values of r and n_0 that allow for obtaining a reliable estimate of the parameters, we consider $r \in \{5, 10, 15\}$, and $n_0 \in \{4, 5, 6, 7\}$. Given $\binom{(r+)n_0}{2}$-dimensional vectors of scores, $\boldsymbol{s}_c$, we can use Algorithms 5 and 6 to sample from the posterior distributions of the parameters, and determine at which point we begin to see stabile results. Figures 9.7, 9.8, 9.9, and 9.10 depict the posterior distributions of the different mean and standard deviation parameters when $r = 5, 10$, and 15, respectively. Tables 9.9, 9.10, 9.11, and 9.12 summarize the results of Figures 9.7, 9.8, 9.9, and 9.10.

Parameter	Sample Mean	Posterior Mean	Posterior Median	Posterior Mode
$n_0 = 4$				
θ_{kk}	9.424	9.430	9.469	10.884
$\theta_{k\mathscr{P}}$	15.111	15.046	15.012	15.328
$\theta_{\mathscr{P}\mathscr{P}}$	11.292	11.317	11.364	10.927
$\theta_{\mathscr{P}\mathscr{P}'}$	7.533	7.571	7.539	6.963
$n_0 = 5$				
θ_{kk}	10.302	10.306	10.304	10.256
$\theta_{k\mathscr{P}}$	15.965	15.976	15.976	15.909
$\theta_{\mathscr{P}\mathscr{P}}$	13.043	13.076	13.054	12.94
$\theta_{\mathscr{P}\mathscr{P}'}$	9.437	9.452	9.447	8.373
$n_0 = 6$				
θ_{kk}	9.787	9.815	9.820	9.624
$\theta_{k\mathscr{P}}$	17.666	17.636	17.658	17.970
$\theta_{\mathscr{P}\mathscr{P}}$	13.772	13.789	13.788	14.152
$\theta_{\mathscr{P}\mathscr{P}'}$	11.352	11.318	11.321	11.090
$n_0 = 7$				
θ_{kk}	8.352	8.353	8.386	7.800
$\theta_{k\mathscr{P}}$	15.168	15.191	15.224	15.391
$\theta_{\mathscr{P}\mathscr{P}}$	12.555	12.568	12.580	12.761
$\theta_{\mathscr{P}\mathscr{P}'}$	9.281	9.291	9.292	9.028

Table 9.9: Point estimates for posterior distributions of $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$, when $r = 5$. The mean parameters are fixed such that $\theta_{kk} = 10.9$, $\theta_{k\mathscr{P}} = 16.5$, $\theta_{\mathscr{P}\mathscr{P}} = 13.0$, and $\theta_{\mathscr{P}\mathscr{P}'} = 9.7$.

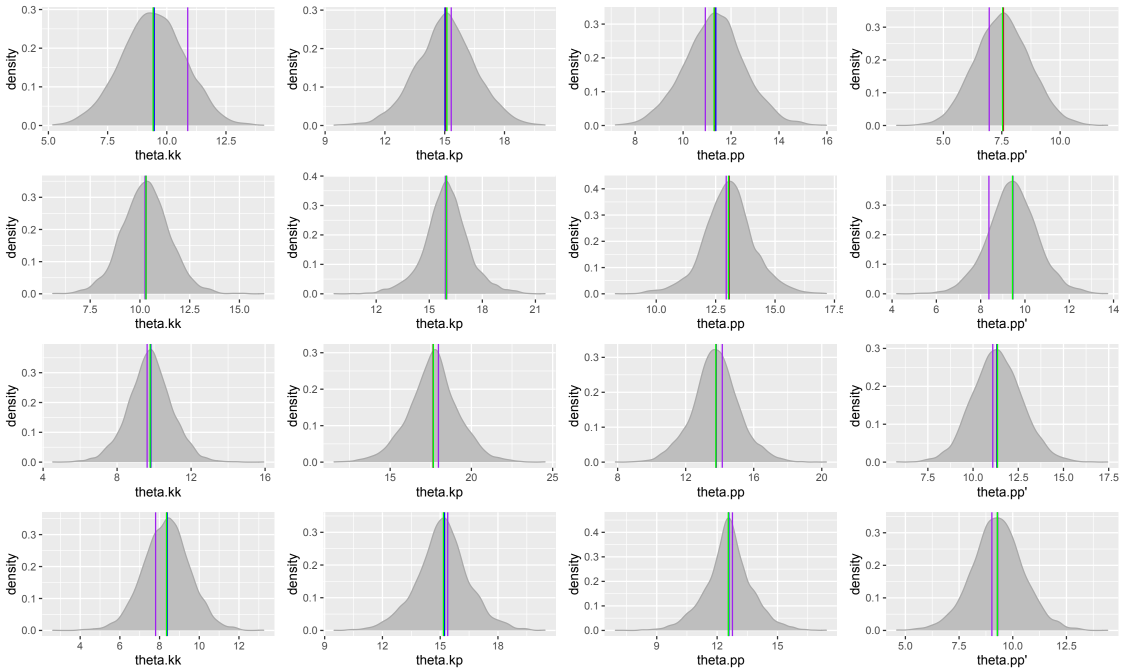


Figure 9.7: Posterior distributions of $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$, for the population-based model when $r = 5$ random sources from the population and $n_0 = 4$ (first row), $n_0 = 5$ (second row), $n_0 = 6$ (third row) and $n_0 = 7$ (fourth row). Red vertical lines correspond to the mean of the posterior distributions obtained using Algorithms 5 and 6; blue vertical lines correspond to the medians of the posterior distributions; purple vertical lines correspond to the mode of the posterior distributions; green vertical lines correspond to the sample means of the different source combinations that exist within the sampled vectors of scores.

Parameter	Sample Standard Deviation	Posterior Mean	Posterior Median	Posterior Mode
$n_0 = 4$				
σ_{kk}	2.832	8.453	7.988	8.306
$\sigma_{k\mathscr{P}}$	3.491	7.874	7.557	7.014
$\sigma_{\mathscr{P}\mathscr{P}}$	2.437	6.499	6.201	6.297
$\sigma_{\mathscr{P}\mathscr{P}'}$	3.317	9.804	9.328	9.944
$n_0 = 5$				
σ_{kk}	1.601	3.215	2.985	2.835
$\sigma_{k\mathscr{P}}$	3.417	5.625	5.985	2.383
$\sigma_{\mathscr{P}\mathscr{P}}$	3.115	4.028	3.790	3.366
$\sigma_{\mathscr{P}\mathscr{P}'}$	4.159	5.805	5.739	2.884
$n_0 = 6$				
σ_{kk}	1.023	3.412	3.752	4.275
$\sigma_{k\mathscr{P}}$	2.981	9.094	11.207	11.097
$\sigma_{\mathscr{P}\mathscr{P}}$	2.799	7.524	9.056	7.670
$\sigma_{\mathscr{P}\mathscr{P}'}$	4.163	11.645	14.444	15.717
$n_0 = 7$				
σ_{kk}	1.788	5.019	4.532	1.551
$\sigma_{k\mathscr{P}}$	2.866	7.472	6.739	4.407
$\sigma_{\mathscr{P}\mathscr{P}}$	1.834	4.942	4.454	4.193
$\sigma_{\mathscr{P}\mathscr{P}'}$	3.324	8.427	7.519	2.548

Table 9.10: Point estimates for posterior distributions of $\sigma_{ii'} \in \{\sigma_{kk}, \sigma_{k\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}'}\}$, when $r = 5$. The standard deviation parameters are fixed such that $\sigma_{kk} = 2.6$, $\sigma_{k\mathscr{P}} = 3.4$, $\sigma_{\mathscr{P}\mathscr{P}} = 2.8$, and $\sigma_{\mathscr{P}\mathscr{P}'} = 4.4$.

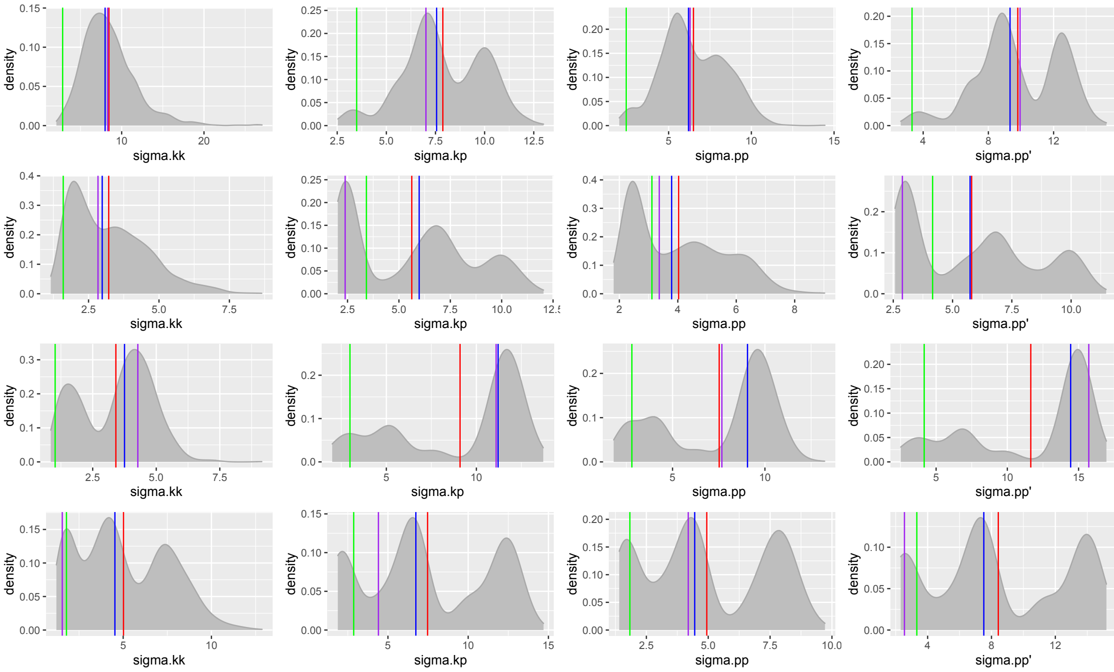


Figure 9.8: Posterior distributions of $\sigma_{ii'} \in \{\sigma_{kk}, \sigma_{k\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}'}\}$, for the population-based model when $r = 5$ random sources from the population and $n_0 = 4$ (first row), $n_0 = 5$ (second row), $n_0 = 6$ (third row) and $n_0 = 7$ (fourth row). Red vertical lines correspond to the mean of the posterior distributions obtained using Algorithms 5 and 6; blue vertical lines correspond to the medians of the posterior distributions; purple vertical lines correspond to the mode of the posterior distributions; green vertical lines correspond to the sample standard deviations of the different source combinations that exist within the sampled vectors of scores.

Parameter	Sample Mean	Posterior Mean	Posterior Median	Posterior Mode
$n_0 = 4$				
θ_{kk}	12.082	12.073	12.086	13.108
$\theta_{k\mathscr{P}}$	15.779	15.773	15.785	16.033
$\theta_{\mathscr{P}\mathscr{P}}$	12.812	12.814	12.810	12.857
$\theta_{\mathscr{P}\mathscr{P}'}$	8.990	8.964	8.972	8.910
$n_0 = 5$				
θ_{kk}	7.894	7.907	7.907	8.155
$\theta_{k\mathscr{P}}$	15.512	15.528	15.525	15.455
$\theta_{\mathscr{P}\mathscr{P}}$	13.196	13.195	13.185	13.158
$\theta_{\mathscr{P}\mathscr{P}'}$	9.686	9.678	9.696	10.035
$n_0 = 6$				
θ_{kk}	13.309	13.323	13.332	13.233
$\theta_{k\mathscr{P}}$	17.892	17.897	17.888	18.067
$\theta_{\mathscr{P}\mathscr{P}}$	13.096	13.103	13.104	13.054
$\theta_{\mathscr{P}\mathscr{P}'}$	10.418	10.430	10.431	10.938
$n_0 = 7$				
θ_{kk}	10.894	10.899	10.902	10.671
$\theta_{k\mathscr{P}}$	16.133	16.130	16.128	15.993
$\theta_{\mathscr{P}\mathscr{P}}$	12.642	12.640	12.642	12.844
$\theta_{\mathscr{P}\mathscr{P}'}$	9.322	9.317	9.329	9.783

Table 9.11: Point estimates for posterior distributions of $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$, when $r = 10$. The mean parameters are fixed such that $\theta_{kk} = 10.9$, $\theta_{k\mathscr{P}} = 16.5$, $\theta_{\mathscr{P}\mathscr{P}} = 13.0$, and $\theta_{\mathscr{P}\mathscr{P}'} = 9.7$.

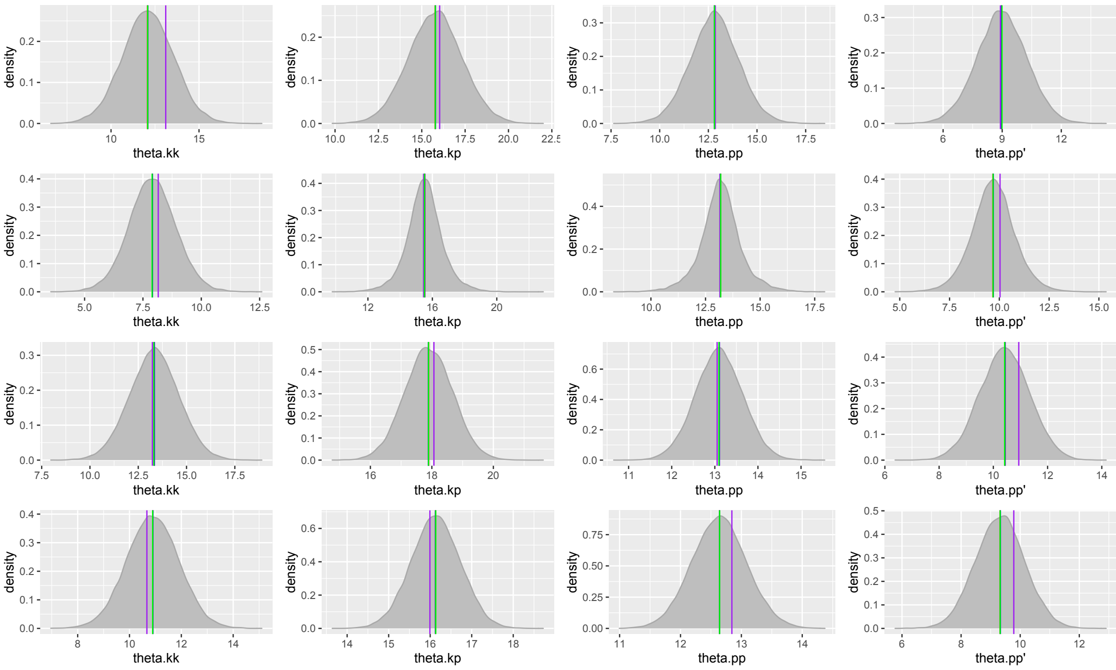


Figure 9.9: Posterior distributions of $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$, for the population-based model when $r = 10$ random sources from the population and $n_0 = 4$ (first row), $n_0 = 5$ (second row), $n_0 = 6$ (third row) and $n_0 = 7$ (fourth row). Red vertical lines correspond to the mean of the posterior distributions obtained using Algorithms 5 and 6; blue vertical lines correspond to the medians of the posterior distributions; purple vertical lines correspond to the mode of the posterior distributions; green vertical lines correspond to the sample means of the different source combinations that exist within the sampled vectors of scores.

Parameter	Sample Standard Deviation	Posterior Mean	Posterior Median	Posterior Mode
$n_0 = 4$				
σ_{kk}	1.607	5.434	4.811	3.855
$\sigma_{k\mathscr{P}}$	3.367	11.253	10.314	10.764
$\sigma_{\mathscr{P}\mathscr{P}}$	3.66	7.293	6.233	5.783
$\sigma_{\mathscr{P}\mathscr{P}'}$	4.744	12.796	11.687	11.690
$n_0 = 5$				
σ_{kk}	1.211	2.500	2.147	1.295
$\sigma_{k\mathscr{P}}$	3.270	5.626	5.302	2.710
$\sigma_{\mathscr{P}\mathscr{P}}$	2.351	4.417	3.905	2.365
$\sigma_{\mathscr{P}\mathscr{P}'}$	3.876	6.668	6.573	3.215
$n_0 = 6$				
σ_{kk}	2.510	2.880	2.817	2.518
$\sigma_{k\mathscr{P}}$	3.292	3.118	3.115	3.129
$\sigma_{\mathscr{P}\mathscr{P}}$	2.290	2.529	2.521	2.508
$\sigma_{\mathscr{P}\mathscr{P}'}$	4.059	4.200	4.198	4.171
$n_0 = 7$				
σ_{kk}	1.774	1.949	1.916	1.814
$\sigma_{k\mathscr{P}}$	2.921	2.415	2.414	2.387
$\sigma_{\mathscr{P}\mathscr{P}}$	2.429	2.058	2.054	2.031
$\sigma_{\mathscr{P}\mathscr{P}'}$	3.851	3.552	3.551	3.557

Table 9.12: Point estimates for posterior distributions of $\sigma_{ii'} \in \{\sigma_{kk}, \sigma_{k\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}'}\}$, when $r = 10$. The standard deviation parameters are fixed such that $\sigma_{kk} = 2.6$, $\sigma_{k\mathscr{P}} = 3.4$, $\sigma_{\mathscr{P}\mathscr{P}} = 2.8$, and $\sigma_{\mathscr{P}\mathscr{P}'} = 4.4$.

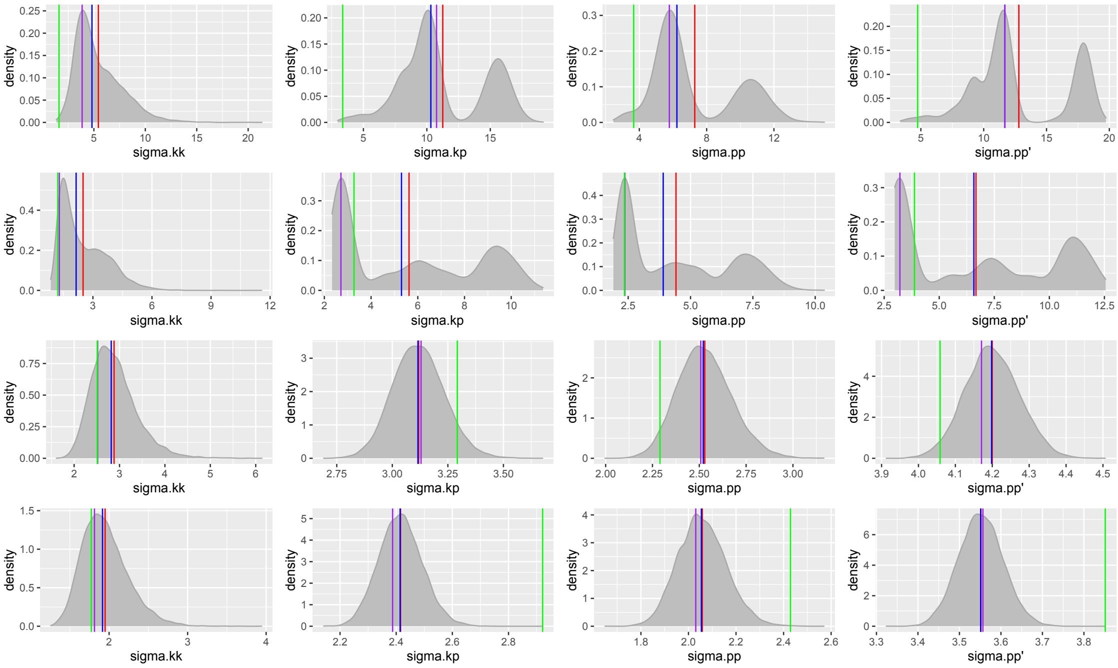


Figure 9.10: Posterior distributions of $\sigma_{ii'} \in \{\sigma_{kk}, \sigma_{k\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}'}\}$, for the population-based model when $r = 10$ random sources from the population and $n_0 = 4$ (first row), $n_0 = 5$ (second row), $n_0 = 6$ (third row) and $n_0 = 7$ (fourth row). Red vertical lines correspond to the mean of the posterior distributions obtained using Algorithms 5 and 6; blue vertical lines correspond to the medians of the posterior distributions; purple vertical lines correspond to the mode of the posterior distributions; green vertical lines correspond to the sample standard deviations of the different source combinations that exist within the sampled vectors of scores.

Parameter	Sample Mean	Posterior Mean	Posterior Median	Posterior Mode
$n_0 = 4$				
θ_{kk}	14.804	14.812	14.840	14.878
$\theta_{k\mathscr{P}}$	18.532	18.555	18.507	18.373
$\theta_{\mathscr{P}\mathscr{P}}$	12.833	12.863	12.842	12.751
$\theta_{\mathscr{P}\mathscr{P}'}$	9.322	9.2667	9.272	9.924
$n_0 = 5$				
θ_{kk}	9.343	9.336	9.338	9.274
$\theta_{k\mathscr{P}}$	16.008	16.004	15.997	16.058
$\theta_{\mathscr{P}\mathscr{P}}$	13.276	13.286	13.286	13.319
$\theta_{\mathscr{P}\mathscr{P}'}$	9.609	9.604	9.604	9.484
$n_0 = 6$				
θ_{kk}	10.618	10.614	10.612	10.741
$\theta_{k\mathscr{P}}$	15.366	15.363	15.358	15.345
$\theta_{\mathscr{P}\mathscr{P}}$	12.070	12.066	12.064	12.021
$\theta_{\mathscr{P}\mathscr{P}'}$	8.992	8.990	8.990	9.066
$n_0 = 7$				
θ_{kk}	14.096	14.081	14.073	13.921
$\theta_{k\mathscr{P}}$	17.365	17.359	17.359	17.187
$\theta_{\mathscr{P}\mathscr{P}}$	12.339	12.338	12.342	12.491
$\theta_{\mathscr{P}\mathscr{P}'}$	8.809	8.808	8.809	8.978

Table 9.13: Point estimates for posterior distributions of $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$, when $r = 15$. The mean parameters are fixed such that $\theta_{kk} = 10.9$, $\theta_{k\mathscr{P}} = 16.5$, $\theta_{\mathscr{P}\mathscr{P}} = 13.0$, and $\theta_{\mathscr{P}\mathscr{P}'} = 9.7$.

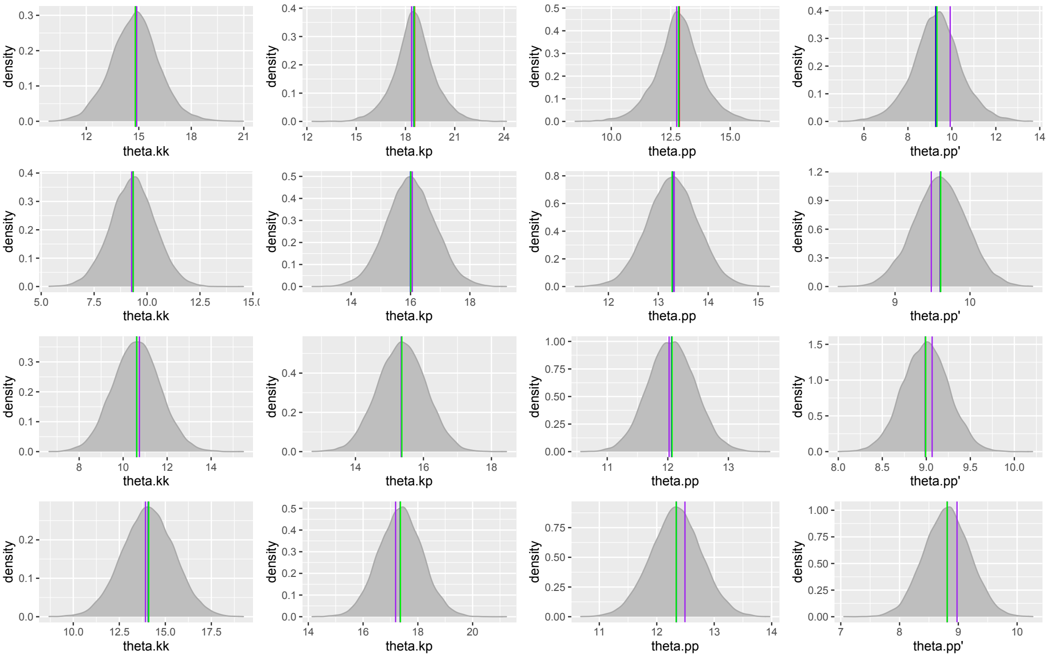


Figure 9.11: Posterior distributions of $\theta_{ii'} \in \{\theta_{kk}, \theta_{k\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}}, \theta_{\mathscr{P}\mathscr{P}'}\}$, for the population-based model when $r = 15$ random sources from the population and $n_0 = 4$ (first row), $n_0 = 5$ (second row), $n_0 = 6$ (third row) and $n_0 = 7$ (fourth row). Red vertical lines correspond to the mean of the posterior distributions obtained using Algorithms 5 and 6; blue vertical lines correspond to the medians of the posterior distributions; purple vertical lines correspond to the mode of the posterior distributions; green vertical lines correspond to the sample means of the different source combinations that exist within the sampled vectors of scores.

Parameter	Sample Standard Deviation	Posterior Mean	Posterior Median	Posterior Mode
$n_0 = 4$				
σ_{kk}	1.352	3.061	2.942	3.061
$\sigma_{k\mathscr{P}}$	3.252	6.999	8.382	7.465
$\sigma_{\mathscr{P}\mathscr{P}}$	3.053	5.377	6.083	5.627
$\sigma_{\mathscr{P}\mathscr{P}'}$	4.573	8.048	9.874	8.575
$n_0 = 5$				
σ_{kk}	2.189	2.479	2.409	2.344
$\sigma_{k\mathscr{P}}$	5.366	3.850	3.844	3.880
$\sigma_{\mathscr{P}\mathscr{P}}$	3.563	2.980	2.980	2.994
$\sigma_{\mathscr{P}\mathscr{P}'}$	2.582	1.779	1.780	1.783
$n_0 = 6$				
σ_{kk}	2.785	2.759	2.685	2.758
$\sigma_{k\mathscr{P}}$	5.632	3.548	3.542	3.591
$\sigma_{\mathscr{P}\mathscr{P}}$	4.036	2.668	2.665	2.648
$\sigma_{\mathscr{P}\mathscr{P}'}$	3.007	1.651	1.651	1.654
$n_0 = 7$				
σ_{kk}	4.190	4.041	3.986	3.798
$\sigma_{k\mathscr{P}}$	5.975	3.899	3.894	3.924
$\sigma_{\mathscr{P}\mathscr{P}}$	3.482	2.687	2.692	2.736
$\sigma_{\mathscr{P}\mathscr{P}'}$	2.727	1.963	1.962	1.958

Table 9.14: Point estimates for posterior distributions of $\sigma_{ii'} \in \{\sigma_{kk}, \sigma_{k\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}'}\}$, when $r = 15$. The standard deviation parameters are fixed such that $\sigma_{kk} = 2.6$, $\sigma_{k\mathscr{P}} = 3.4$, $\sigma_{\mathscr{P}\mathscr{P}} = 2.8$, and $\sigma_{\mathscr{P}\mathscr{P}'} = 4.4$.

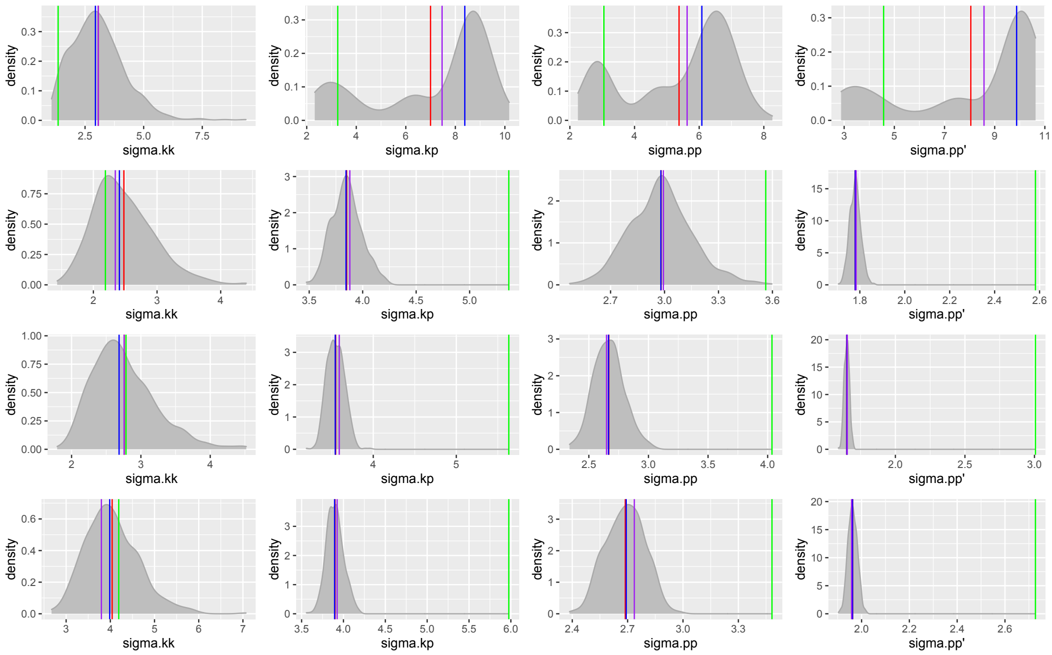


Figure 9.12: Posterior distributions of $\sigma_{ii'} \in \{\sigma_{kk}, \sigma_{k\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}}, \sigma_{\mathscr{P}\mathscr{P}'}\}$, for the population-based model when $r = 15$ random sources from the population and $n_0 = 4$ (first row), $n_0 = 5$ (second row), $n_0 = 6$ (third row) and $n_0 = 7$ (fourth row). Red vertical lines correspond to the mean of the posterior distributions obtained using Algorithms 5 and 6; blue vertical lines correspond to the medians of the posterior distributions; purple vertical lines correspond to the mode of the posterior distributions; green vertical lines correspond to the means (for $\theta_{ii'}$) and standard deviations (for $\sigma_{ii'}$) of the different source combinations that exist within the sampled vectors of scores.

There are three implications of Figures 9.7 through 9.12 and Tables 9.9 through 9.14. First, we see that we obtain better estimates of the parameter values (as given by the posterior means, medians and modes) as we increase r and n_0. Second, we see that the posterior approximations for $\theta_{ii'}$ are better representations than they are for $\sigma_{ii'}$. In particular, there appears to be a lack of convergence whenever $n_0 < 6$ control objects for each of $n \leq 10$ random sources (although the convergence does seem to improve as we move from considering $r = 5$ random sources from the population to considering $r = 10$ random sources from the population). Finally, we note that the posterior samples obtained when $r = 5$ are not stable since the resulting chains do not converge towards a single distribution. Instead, they appear to explore a mixture of distributions. We see this same peculiarity when $r = 10$ random sources from the population and $n_0 = 4$ or $n_0 = 5$ control objects per source, and when $r = 15$ random sources from the population and $n_0 = 4$ control objects per source (note, though, that the phenomenon becomes less extreme as we move to consider $r = 10$ or $r = 15$ random sources from the population). However, when $r = 10$ or $r = 15$, we see that the chains and resulting posterior samples begin to stabilize. This indicates that, given the current status of the sampler, an examiner should consider no less than $r = 10$ random sources from the population alongside at least $n_0 = 6$ control objects per source. Note that there exist several methods for stabilizing the resulting chains. For example, considering a more informative prior can help to stabilize chains in this type of scenario [44].

9.5 ASSESSING NORMALITY ASSUMPTIONS

The score model presented in (9.1) relies on the assumption that the vector of scores $\boldsymbol{s}$ is normally distributed. This assumption can be met by defining an appropriate kernel, or by increasing the intrinsic dimension of the original objects [5, 53] . In this section, we present the results of some simulations that demonstrate that the assumption of Multivariate Normality is reasonable, so long as the dimension of the objects in the original space is sufficiently large. In these simulations, we sample 2500 sets of functional objects described by p B-spline basis functions, $p \in \{5, 50, 500\}$. In these simulations, we sample the coefficients of the basis functions from a Dirichlet distribution. This distribution is chosen to demonstrate that the distribution of the original objects does not impact the convergence of the scores to a Multivariate Normal distribution. This process is outlined in Algorithm 7. After sampling sets of objects, the scores are calculated using the exponential of a squared Euclidean kernel. Given these 2500 sets of $N = \binom{(r+1)n_0+n_u}{2}$-dimensional scores, we are able to compute the associated empirical covariance matrix, and project the scores into their eigenspace.

We proceed by considering the within-source comparisons separate of the between-source comparisons. In the case of the within-source comparisons, we ensure that the objects sampled in each iteration are from the same source by first sampling a single p-dimensional flat Dirichlet object, $\boldsymbol{\alpha}$, using $\boldsymbol{\Lambda} = \{\mathbf{1}_p\}$. We then sample N objects from a p-dimensional Dirichlet distribution whose parameter is $\boldsymbol{\alpha}$, multiplied by $c = 1000$ to ensure that the three objects are very similar, and thus have originated from the same

Algorithm 7: Generating scores for sets of objects

Data: A kernel function, κ; a set of p basis functions, $\{\beta_i(t)\}_{i=1}^p$, over some interval; $[a,b]$, a multivariate distribution function, F and associated sets of parameters, $\boldsymbol{\Lambda}, \boldsymbol{\Omega}$

Result: An matrix of scores, where each row corresponds to an $N = \binom{(r+1)n_0+n_u}{2}$-dimensional vector of scores given the simulated data

for n iterations **do**

1. Sample a mean vector, $\boldsymbol{\alpha} \sim F(\boldsymbol{\Lambda})$, $\boldsymbol{\alpha} \in \mathbb{R}^p$, where Λ is the set of parameters for F;
2. Sample a matrix, $\boldsymbol{B}$, of coefficients such that each row vector, $\boldsymbol{b}_i \sim F(\boldsymbol{\Omega})$, $i \in \{1,2,3\}$, $c \in \mathbb{R}$, $\boldsymbol{\alpha} \subset \boldsymbol{\Omega}$, represents the coefficients of a single object;
3. Calculate the function values $x_{ij} = \sum_{j=1}^p b_{ij}\beta_j(t)$, $\forall i \in \{1,\ldots,N\}$, $\forall t \in [a,b]$;
4. Calculate the vector of scores, $\boldsymbol{s}$;

end

source. In this case, where we sample our coefficients from a Dirichlet distribution, the parameter $\boldsymbol{\Omega} = \{1000\boldsymbol{\alpha}\}$. In the case where we consider that all objects originate from different sources, we proceed using the same steps as in the within-source scenario described above. However, rather than multiply the parameter α by $c = 1000$, we simply use the vector α ($c = 1$) as our parameter for the second sample from a p-dimensional Dirichlet distribution, such that $\boldsymbol{\Omega} = \{\boldsymbol{\alpha}\}$. Keeping the parameters "small" allows for enough variation in the sampled objects such that they originate from different sources. Upon obtaining the functional objects, we proceed as in the above algorithm.

Figures 9.13 and 9.14 demonstrate that, as the dimension of the objects increases, we see a tendency of the distributions of the resulting scores to become spherical or ellipsoidal, which gives us an idea of whether or not the vector of scores follows a Multivariate Normal distribution. In either scenario, when we consider the first three eigen-dimensions, we see that the clouds of points begin to appear spherical or ellipsoidal when $p = 50$. In addition, we note that the projections of the within-source scores are more spherical in appearance than those of the between-source scores.

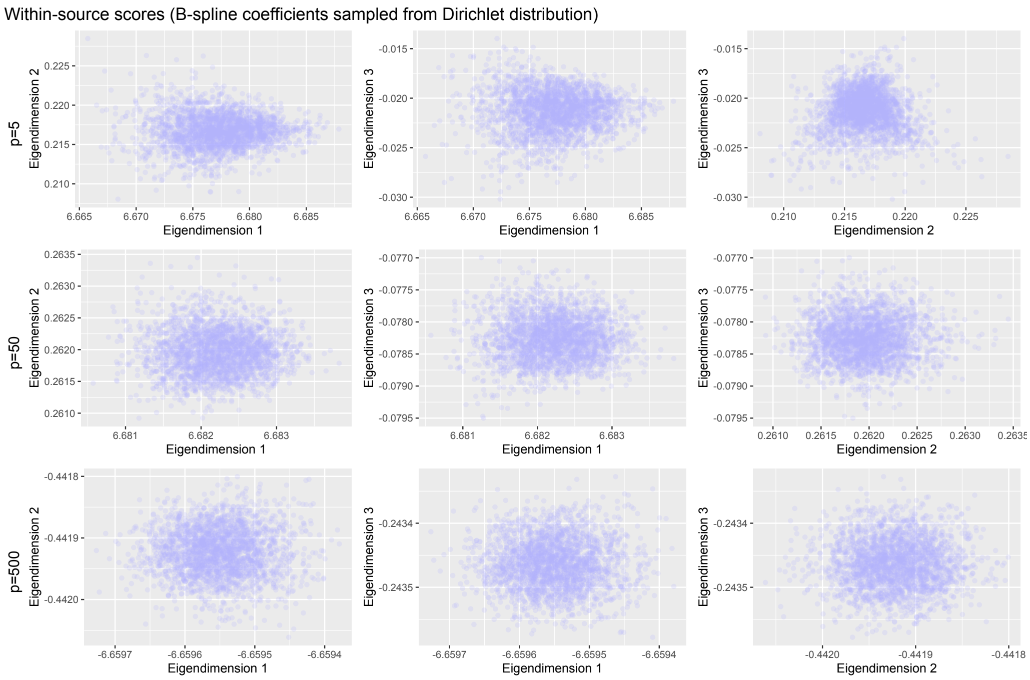


Figure 9.13: Projection of N-dimensional vectors of scores obtained from 2500 sets of within-source objects in the space defined by their respective spectra decompositions of their covariance matrices. Objects correspond to spectra obtained from linear combinations of B-spline bases whose coefficients were sampled from Dirichlet distributions.

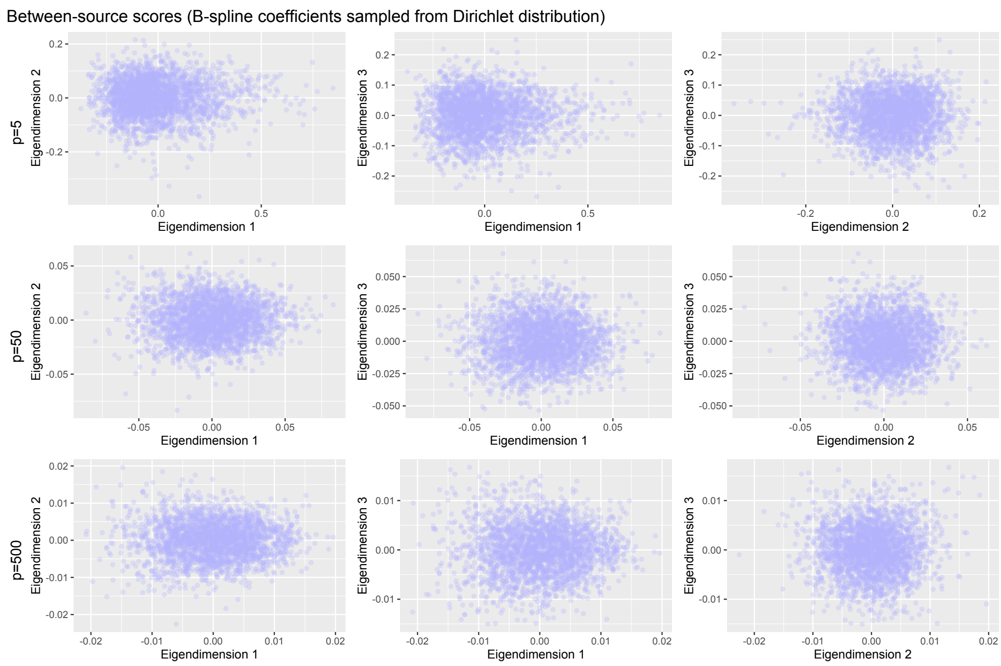


Figure 9.14: Projection of N-dimensional vectors of scores obtained from 2500 sets of between-source objects in the space defined by their respective spectra decompositions of their covariance matrices. Objects correspond to spectra obtained from linear combinations of B-spline bases whose coefficients were sampled from Dirichlet distributions.

Chapter 10

IMPLEMENTING THE POPULATION-BASED MODEL SELECTION ALGORITHM

In this chapter, we apply the proposed model to a set of Fourier-Transform Infrared (FTIR) spectra of paint chips from cans of common household paint. The paint chips considered in this example come from 166 different paint cans. For each paint source, we observe seven replicates, each of which corresponds to a new, distinct observation, and is not a repeated measurement of a single paint chip. That is, the seven replicates correspond to seven exchangeable FTIR spectra. Each spectra represents the absorbance of the paint material for a range of wavelengths (from approximately 550 cm^{-1} to approximately 4,000cm^{-1}), and is captured by an approximately 7,000-dimensional vector[1].

Since we observe only seven spectra per source, we treat the spectra as functional data and express each as a linear combination of 300 B-spline bases for the purpose of this experiment. We assume that the vectors of basis coefficients are $i.i.d.$ Multivariate Normal, and we use the sample mean and covariance matrix of the coefficients for the seven spectra as point estimates for the parameters of their distribution. This strategy is fit-for-purpose in the context of this example, and allows us to "re-sample" new spectra from a considered source to study the performance of our model. Figure 10.1 presents seven observed spectra overlaid with seven simulated spectra from the same can of paint, and indicates that this approach is reasonable.

[1]The set of FTIR spectra proves to be more felicitous dataset for the considered model, given the greater number of sources that constitute the population of potential sources. Given that the MNIST handwritten digit data considers only 10 potential sources, it is not an appropriate dataset for considering the performance of the considered population-based model.

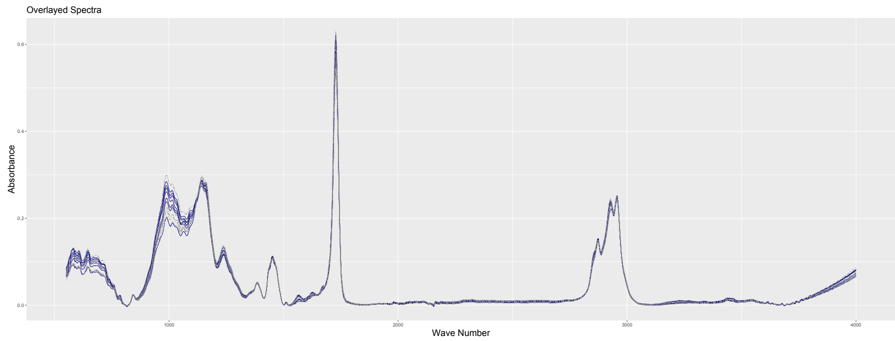


Figure 10.1: Seven observed replicates of FTIR spectra (solid dark blue lines) overlaid with seven generated replicates of pseudo-spectra (dashed light gray lines) from paint can #5 (out of 166 paint cans)

Evaluating the performance of the model in this scenario requires defining a kernel function for comparing two FTIR spectra. The kernel function used in this experiment measures the dissimilarity between two spectra $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{i'j'}$ by considering their cross-correlation over a range of lags, $\tau = -k, \ldots, k$, and the Euclidean norm of their difference. Specifically, we define our kernel as

$$\kappa\left(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}\right) = \log\left[C\left(||\boldsymbol{\omega} \circ \boldsymbol{x}_{ij} - \boldsymbol{\omega} \circ \boldsymbol{x}_{i'j'}||\right)\left(||\mathbf{1}_{2k+1}|| - ||\boldsymbol{r}_{\tau}||\right)\right]. \tag{10.1}$$

There are three components to the kernel defined in (10.1): we define the constant, C, to help satisfy the normality assumption for the resulting vector of scores; we consider the Euclidean norm of the difference between the vectors $\boldsymbol{\omega} \circ \boldsymbol{x}_{ij}$ and $\boldsymbol{\omega} \circ \boldsymbol{x}_{i'j'}$, where $\boldsymbol{\omega}$ is a vector of binary weights that indicates which positions of the spectra are considered in the calculation, and $\circ$ is the Schur product; we convert the $(2k + 1)$-dimensional vector of cross-correlations between spectra $\boldsymbol{\omega} \circ \boldsymbol{x}_{ij}$ and $\boldsymbol{\omega} \circ \boldsymbol{x}_{i'j'}$ into a distance metric by considering the displacement of its norm from the norm of the $(2k+1)$-dimensional one vector. Considering $\boldsymbol{\omega} \circ \boldsymbol{x}_{ij}$, rather than $\boldsymbol{x}_{ij}$, allows us to employ a filtering process such that only the interesting areas of $\boldsymbol{x}_{ij}$ are considered in the comparison process. That is, uninformative (i.e., flat) areas that exist between pairs of spectra are filtered out so as to better discriminate between pairs of spectra (see Figure 10.3 for an example of the results of the filtering process). Finally, the function satisfies Mercer's conditions, and so is a valid kernel function with an inner-product representation in some separable high-dimensional feature space.

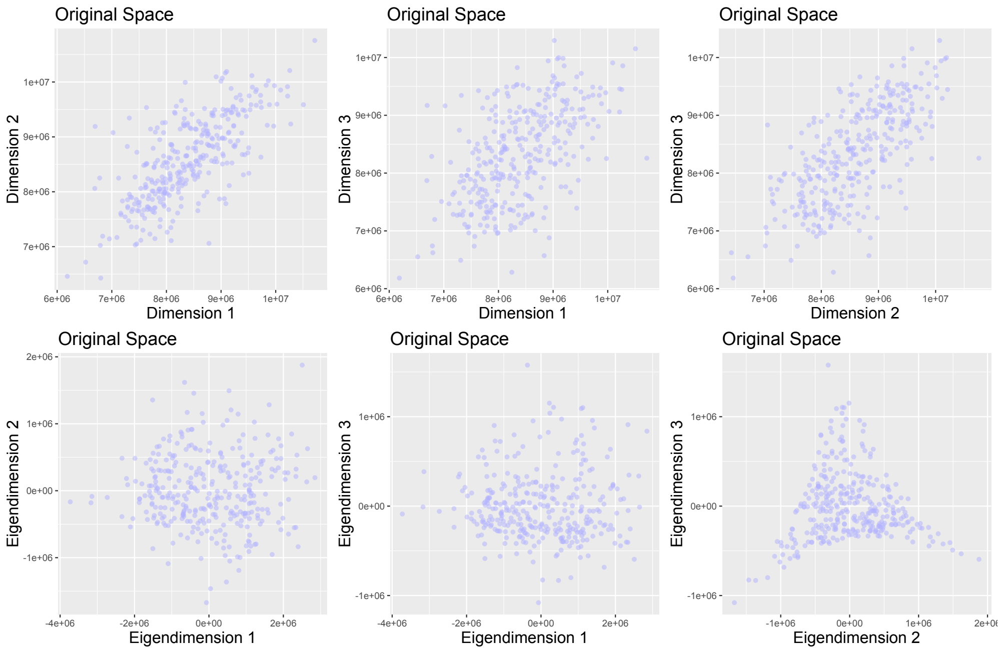


Figure 10.2: Original and projected distributions of scores obtained from comparing FTIR spectra of paint. Top Row: 3-dimensional vectors of scores obtained from 332 triplets of objects originating from the same source in the original space. Bottom Row: Projection of 3-dimensional vectors of scores obtained from 332 triplets of objects originating from the same source in the space defined by the spectral decomposition of their covariance matrix.

Figure 10.2 (top row) portrays the marginal distributions of the scores in their original space. By expressing the original vectors of scores as a function of the space defined by the eigenvectors of their sample covariance matrix, we can observe the marginal distributions of the score vectors along orthogonal axes, and better determine if the marginal distributions follow a Normal distribution. Figure 10.2 (bottom row) shows that, although the data is approximately spherical in the first two dimensions of the eigenspace, there is a rather significant departure from normality when eigendimensions 2 and 3 are plotted against one another. However, given the results presented below, we purport that this deviation from multivariate normality does not affect the ability of the model to correctly classify and differentiate spectra, and thus testifies to the robustness of the model: despite the lack of normality, the model remains able to correctly associate and differentiate spectra originating from the same and different sources, respectively.

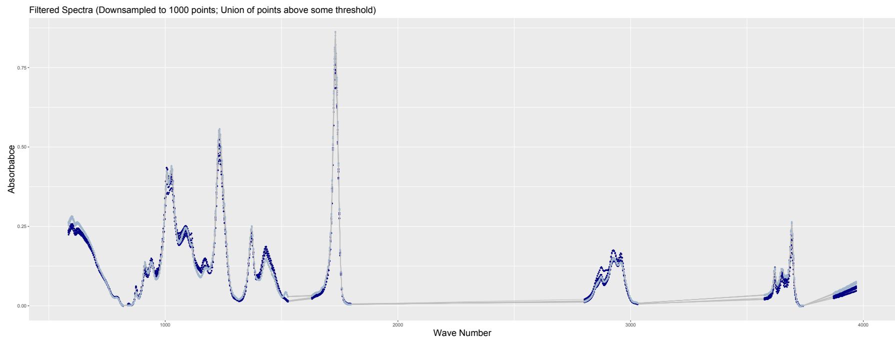


Figure 10.3: Spectra from paint can #19 (light blue) overlaid with spectra from paint can #34 (dark blue). Paint can #34 is the most similar to paint can #19 out of the 166 potential cans of paint, as determined by the kernel function defined in (10.1). Regions in light grey correspond to the areas that are considered to be uninformative in the discrimination process, and so these time stamps are not considered in calculating the score returned by the kernel function. By considering only the areas in blue, we can better discriminate between the two sources.

To assess the performance of the model in the forensic context, we again consider a series of simulations in which we consider a putative source alongside $r = 10$ random sources from the population of potential sources. We consider $n_0 = 6$ control objects per source, and $n_u = 3$ trace objects. We consider this combination of r and n_0 since this is when we begin to see some stability of the distributions resulting from our sampling process (see Figures 9.9 and 9.10 above).

To determine if the rarity of the putative source affects the sensitivity of the model, we consider three instances in which the putative source has a low random match probability, indicating that the source has characteristics which make it rare in the population (we consider that paint cans #37, #77, and #160 are rare) and three instances in which the putative source has a high random match probability, indicating that the source has characteristics that are unremarkable in the population (we consider that paint cans #18, #47, and #85 are unremarkable). The rarity of the sources was determined by considering the random match probability associated with each of the 166 potential sources that make up the population. Figure 4 in [7] is reproduced as Figure 10.4 below, and characterizes the distributions of the random match probabilities associated with each of the 166 paint cans that make up the population of potential sources [2]. See details of full experiment in [7]).

[2]Note that Ausdemore et al. use N to describe the number of control objects, while this

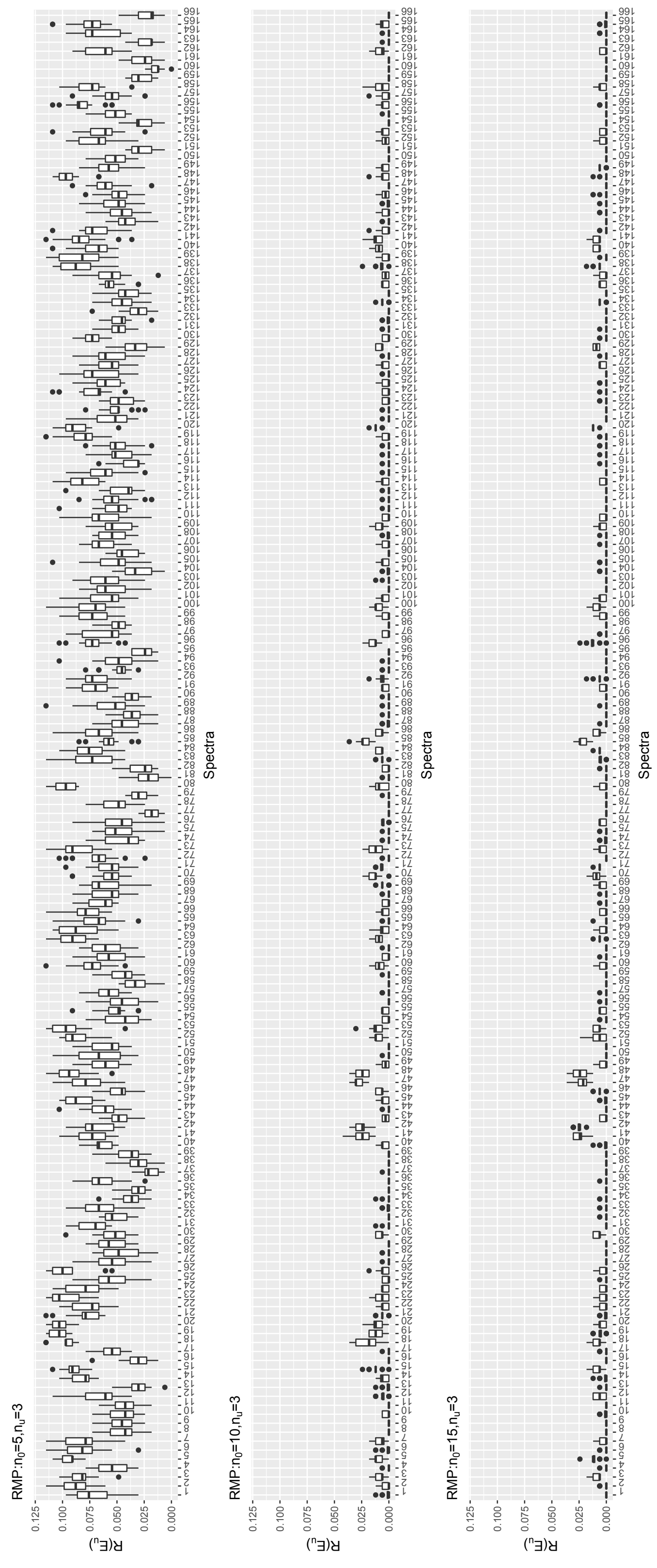


Figure 10.4: Distribution of random match probabilities associated with a population of 166 paint cans making up the population of potential sources when we consider 5 (top), 10 (middle), and 15 (bottom) control objects per source. See details of full experiment in [7].

We consider 250 iterations of the experiment when $\mathcal{H}_p$ is true (that is, the putative source is the true source of the set of trace objects), and when $\mathcal{H}_d$ is true (that is,

dissertation uses n_0 to describe the number of control objects.

some other random source from the population of potential sources is the true source of the set of trace objects). For this set of experiments, when we consider that $\mathcal{H}_d$ is true, we consider two scenarios: we consider a set of simulations in which the set of trace objects truly originates from can #9, since this paint can is considered to be rare in the population of 166 paint cans, and we consider a set of simulations in which the set of trace objects truly originates from can #41, since this paint can is considered to be unremarkable in the population of 166 paint cans. We choose to consider a rare source as the true source under $\mathcal{H}_d$, since it more likely to be poorly described when $r = 10$ random sources are used to describe the population of potential sources, and we choose to consider an unremarkable source as the true source under $\mathcal{H}_d$, since this source should be well described when $r = 10$ random sources are used to describe the population of potential sources.

		Proposed Model		
		Paint Can #18	Random Source	Total
True Source	Paint Can #18	250	0	250
	Paint Can #9	3	247	250
	Total	253	247	500

Table 10.1: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #18. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #18 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #18	Random Source	Total
True Source	Paint Can #18	250	0	250
	Paint Can #41	14	236	250
	Total	264	236	500

Table 10.2: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #18. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #18 and 250 sets of $n_u = 3$ trace objects that truly originate from an unremarkable source (paint can #41) in the population.

		Proposed Model		
		Paint Can #37	Random Source	Total
True Source	Paint Can #37	250	0	250
	Paint Can #9	2	248	250
	Total	252	248	500

Table 10.3: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #37. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #37 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #37	Random Source	Total
True Source	Paint Can #37	250	0	250
	Paint Can #41	10	240	250
	Total	260	240	500

Table 10.4: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #37. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #37 and 250 sets of $n_u = 3$ trace objects that truly originate from an unremarkable source (paint can #41) in the population.

		Proposed Model		
		Paint Can #47	Random Source	Total
True Source	Paint Can #47	250	0	250
	Paint Can #9	4	246	250
	Total	254	246	500

Table 10.5: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #47. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #47 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #47	Random Source	Total
True Source	Paint Can #47	250	0	250
	Paint Can #41	11	239	250
	Total	261	239	500

Table 10.6: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #47. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #47 and 250 sets of $n_u = 3$ trace objects that truly originate from an unremarkable source (paint can #41) in the population.

		Proposed Model		
		Paint Can #77	Random Source	Total
True Source	Paint Can #77	250	0	250
	Paint Can #9	11	239	250
	Total	261	239	500

Table 10.7: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #77. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #77 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #77	Random Source	Total
True Source	Paint Can #77	250	0	250
	Paint Can #41	9	241	250
	Total	259	241	500

Table 10.8: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #77. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #77 and 250 sets of $n_u = 3$ trace objects that truly originate from an unremarkable source (paint can #41) in the population.

		Proposed Model		
		Paint Can #85	Random Source	Total
True Source	Paint Can #85	250	0	250
	Paint Can #9	10	240	250
	Total	260	240	500

Table 10.9: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #85. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #85 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #85	Random Source	Total
True Source	Paint Can #85	250	0	250
	Paint Can #41	9	241	250
	Total	259	241	500

Table 10.10: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #85. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #85 and 250 sets of $n_u = 3$ trace objects that truly originate from an unremarkable source (paint can #41) in the population.

		Proposed Model		
		Paint Can #160	Random Source	Total
True Source	Paint Can #160	250	0	250
	Paint Can #9	12	238	250
	Total	262	238	500

Table 10.11: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #160. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #160 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #160	Random Source	Total
True Source	Paint Can #160	250	0	250
	Paint Can #41	9	241	250
	Total	259	241	500

Table 10.12: Confusion matrix for classification of FTIR spectra when $r = 10$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #160. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #160 and 250 sets of $n_u = 3$ trace objects that truly originate from an unremarkable source (paint can #41) in the population.

Tables 10.1 through 10.12 summarize the results of our experiment for the given kernel function defined in (10.1). These results indicate that, overall, the model is better at distinguishing between the putative source and the population of potential sources when the putative source is the true source of the trace objects. In addition, we note that we have a higher misclassification rate, in general, when the putative source is rare in the population of potential sources. Finally, we note that we have a similar performance, regardless if the true source of the set of trace objects is a rare random source from the population of potential sources, or if it is an unremarkable random source from the population of potential sources. Overall, the results of Tables 10.1 through 10.12 indicate that, for the given kernel, the model performs quite well, despite the apparent departure from normality, indicated in Figure 10.2.

Finally, we perform a test where we consider $r = 5$ random sources from the population of potential sources alongside $n_0 = 6$ control objects per source to determine the ultimate performance of the model given unstable samples of the parameters (see Figure 9.8 in Section 9.4). For this experiment, we consider that, when $\mathcal{H}_d$ is true, the source of the set of trace objects is a rare source from the population of potential sources (paint can #9). Tables 10.13 through 10.18 below indicate that, even when the Gibbs sampler does not necessarily return stable samples of the parameters, the performance of the model is not affected. This issue is addressed in Chapter 11.

		Proposed Model		
		Paint Can #18	Random Source	Total
True Source	Paint Can #160	250	0	250
	Paint Can #9	6	244	250
	Total	256	244	500

Table 10.13: Confusion matrix for classification of FTIR spectra when $r = 5$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #18. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #18 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #37	Random Source	Total
True Source	Paint Can #37	250	0	250
	Paint Can #9	4	246	250
	Total	254	246	500

Table 10.14: Confusion matrix for classification of FTIR spectra when $r = 5$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #37. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #37 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #47	Random Source	Total
True Source	Paint Can #47	250	0	250
	Paint Can #9	2	248	250
	Total	252	248	500

Table 10.15: Confusion matrix for classification of FTIR spectra when $r = 5$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #47. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #47 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #77	Random Source	Total
True Source	Paint Can #77	250	0	250
	Paint Can #9	2	248	250
	Total	252	248	500

Table 10.16: Confusion matrix for classification of FTIR spectra when $r = 5$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #77. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #77 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #85	Random Source	Total
True Source	Paint Can #85	250	0	250
	Paint Can #9	4	246	250
	Total	254	246	500

Table 10.17: Confusion matrix for classification of FTIR spectra when $r = 5$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #85. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #85 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

		Proposed Model		
		Paint Can #160	Random Source	Total
True Source	Paint Can #160	250	0	250
	Paint Can #9	6	244	250
	Total	256	244	500

Table 10.18: Confusion matrix for classification of FTIR spectra when $r = 5$ and $n_0 = 6$ under the null hypothesis that the set of n_u trace objects truly originate from paint can #160. We consider 250 sets of $n_u = 3$ trace objects that truly originate from paint can #160 and 250 sets of $n_u = 3$ trace objects that truly originate from a rare source (paint can #9) in the population.

Chapter 11

EVALUATING THE POPULATION-BASED MODEL SELECTION ALGORITHM

In this part, we considered the development for a population-based model selection algorithm that allows for making inference on the source of a set of test objects. In this scenario, the test objects may have originated from a considered putative source, or from some other random source in a population of potential sources. This method is novel in that it allows for classifying the complete set of n_u trace objects at once, rather than classifying each object in turn. This method relies on a kernel function, which allows for considering virtually any set of high-dimensional, complex, heterogeneous data as a single vector of real-valued scores between observations by merely modifying the kernel to accommodate the considered data. In addition, our method is particularly well-suited for scenarios in which a limited number of observations are available for consideration, as is oftentimes the case in forensic scenarios.

To evaluate the model under various conditions, we conducted a series of simulations, described in Chapters 9 and 10 above. These simulations indicated that the sampler used to study the model parameters (in particular, the variance terms) has the best performance when at least $r = 10$ random sources consisting of at least $n_0 = 6$ control objects are used to describe the population of potential sources. When the model was applied to a real forensic dataset, we saw that the algorithm had an average correct classification rate of 98.60%, and that the algorithm was particularly apt at correctly classifying the trace objects when they truly originated from the considered putative source.

This being said, the current Gibbs sampler appears to be dependent on the Metropolis-Hastings step for the set of variance terms, $\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2$, and σ_e^2 defined in 9.8. A study of the Gibbs sampler indicates that this Metropolis-Hastings step is the most delicate part of the model. That is, if the Metropolis-Hastings step starts at a random point in which one of the positions of the 5-dimensional parameter vector is too small, the sampler gets "stuck", and the entire sampler is offset.

This issue was first noticed in the original development of the Gibbs sampler and Metropolis-Hastings step. Initial attempts to correct this issue consisted of modifying the distribution from which the random starting point was sampled, and also included adjusting the proposal distribution from which the next point in the Metropolis-Hastings step is sampled. While these adjustments did bolster the performance of the model and the stability of the resulting distributions of the posterior samples, there still appear to be instances (albeit rare instances) in which the sampler gets stuck. Future work would focus on stabilizing this step to readily ensure that values of the variance terms do not get too small.

That being said, these issues do not appear to affect the model in terms of distinguishing between $\mathcal{H}_p$ and $\mathcal{H}_d$. That is, the discrepancy in the sampler is consistent between the numerator and the denominator of the Bayes factor in 9.32, and so one model is not "favored" over the other as a result of the discrepancy, and the model is still able to accurately determine whether or not the trace objects came from the putative source or from another random source in the population of potential sources.

In addition, as in the previous two models described in Parts II and III, the computational cost increases as n_0, and n_u increases, and increases drastically as r increases.

Finally, we note that the model is robust to both departures from multivariate normality and instabilities in the posterior samples obtained during the sampling process. That is, should the vector of scores violate the assumption of multivariate normality or should the Gibbs sampler return an unstable chain, the overall performance of the model is not affected. The model is still able to accurately distinguish between $\mathcal{H}_p$ and

$\mathcal{H}_d$, as can be seen in Chapter 10.

Overall, the performance of this model indicates that the model works well for determining whether or not a set of trace objects originates from a given putative source, or from a random source in the population of potential sources.

APPENDICES

A CALCULATIONS FOR $\boldsymbol{SS_{11}}$ BRACKETED TERMS

A.1 FULL DEVELOPMENT OF (3.21)

$$
\begin{aligned}
\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)'\boldsymbol{\Delta}_{11}^{-1'}\boldsymbol{v}^*_{11_1}\boldsymbol{v}^{*\,\prime}_{11_1}\boldsymbol{\Delta}_{11}^{-1}\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right) &= \left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)'\boldsymbol{\Delta}_{11}^{-1'}\boldsymbol{v}^*_{11_1}\left(\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)'\boldsymbol{\Delta}^{-1'}\boldsymbol{v}^*_{11_1}\right)' \\
&= \left[\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)'\boldsymbol{\Delta}_{11}^{-1'}\boldsymbol{v}^*_{11_1}\right]^2 \\
&= \left[\left(\boldsymbol{s}_{11}-\theta_{11}\mathbf{1}_{N_{11}}\right)'\sigma_{11}^{-1}\boldsymbol{v}^*_{11_1}\right]^2 \\
&= \left[\sigma_{11}^{-1}\boldsymbol{s}_{11}\boldsymbol{v}^*_{11_1}-\sigma_{11}^{-1}\theta_{11}\mathbf{1}'_{N_{11}}\boldsymbol{v}^*_{11_1}\right]^2 \\
&= \sigma_{11}^{-2}\left[\left(\boldsymbol{s}'_{11}\frac{\mathbf{1}_{N_{11}}}{\sqrt{N_{11}}}\right)^2-2\left(\boldsymbol{s}'_{11}\frac{\mathbf{1}_{N_{11}}}{\sqrt{N_{11}}}\right)\left(\theta_{11}\mathbf{1}'_{N_{11}}\frac{\mathbf{1}_{N_11}}{\sqrt{N_{11}}}\right)\right. \\
&\quad \left.+\left(\theta_{11}\mathbf{1}'_{N_{11}}\frac{\mathbf{1}_{N_11}}{\sqrt{N_{11}}}\right)^2\right] \\
&= \sigma_{11}^{-2}\left[\frac{\left(\sum_{i=1}^{N_{11}}s_{11_i}\right)^2}{N_{11}}-2\theta_{11}\sum_{i=1}^{N_{11}}s_{11_i}+N_{11}\theta_{11}^2\right] \\
&= \frac{N_{11}}{\sigma_{11}^2}\left[\left(\frac{\sum_{i=1}^{N_{11}}s_{11_i}}{N_{11}}\right)^2-2\theta_{11}\frac{\sum_{i=1}^{N_{11}}s_{11_i}}{N_{11}}+\theta_{11}^2\right] \\
&= \frac{N_{11}}{\sigma_{11}^2}\left[\bar{s}_{11}^2-2\theta_{11}\bar{s}_{11}+\theta_{11}^2\right] \\
&= \frac{N_{11}}{\sigma_{11}^2}\left[\bar{s}_{11}-\theta_{11}\right]^2
\end{aligned}
$$

A.2 FULL DEVELOPMENT OF (3.22)

$$
\begin{aligned}
\boldsymbol{\Delta}_{11}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}^*_{11_l}\boldsymbol{v}^{*\,\prime}_{11_l}\right]\boldsymbol{\Delta}_{11}^{-1} &= \frac{1}{n_0-2}\left[\sigma_{11}^{-2}\left(\boldsymbol{P}_{11}\boldsymbol{P}'_{11}-2(n_0-1)\boldsymbol{v}^*_{11_l}\boldsymbol{v}^{*\,\prime}_{11_l}\right)\right] \\
&= \frac{1}{\sigma_{11}^2(n_0-2)}\left[\boldsymbol{P}_{11}\boldsymbol{P}'_{11}-2(n_0-1)\frac{\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}}{N_{11}}\right] \\
&= \frac{1}{\sigma_{11}^2(n_0-2)}\left[\boldsymbol{P}_{11}\boldsymbol{P}'_{11}-2(n_0-1)\frac{\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}}{\frac{n_0(n_0-1)}{2}}\right] \\
&= \frac{1}{\sigma_{11}^2(n_0-2)}\left[\boldsymbol{P}_{11}\boldsymbol{P}'_{11}-4(n_0-1)\frac{\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}}{n_0(n_0-1)}\right] \\
&= \frac{1}{\sigma_{11}^2(n_0-2)}\frac{(n_0-1)^2}{(n_0-1)^2}\left[\boldsymbol{P}_{11}\boldsymbol{P}'_{11}-\frac{4}{n_0}\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}\right] \\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{11}\boldsymbol{P}'_{11}}{(n_0-1)^2}-\frac{4}{n_0(n_0-1)^2}\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}\right] \\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{11}\boldsymbol{P}'_{11}}{(n_0-1)^2}+\left(\frac{4n_0}{n_0^2(n_0-1)^2}-\frac{8n_0}{n_0^2(n_0-1)^2}\right)\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}\right] \\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{11}\boldsymbol{P}'_{11}}{(n_0-1)^2}+\frac{4n_0}{n_0^2(n_0-1)^2}\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}-\frac{8}{n_0(n_0-1)^2}\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}\right] \\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{11}\boldsymbol{P}'_{11}}{(n_0-1)^2}+\frac{n_0}{N_{11}^2}\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}-\frac{8}{n_0(n_0-1)^2}\mathbf{1}_{N_{11}}\mathbf{1}'_{N_{11}}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{11}\boldsymbol{P}_{11}'}{(n_0-1)^2}+\frac{n_0}{N_{11}^2}\mathbf{1}_{N_{11}}\mathbf{1}_{N_{11}}'-\frac{4}{N_{11}(n_0-1)}\mathbf{1}_{N_{11}}\mathbf{1}_{N_{11}}'\right]\\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{11}\boldsymbol{P}_{11}'}{(n_0-1)^2}-\frac{4}{N_{11}(n_0-1)}\mathbf{1}_{N_{11}}\mathbf{1}_{N_{11}}'+\frac{n_0}{N_{11}^2}\mathbf{1}_{N_{11}}\mathbf{1}_{N_{11}}'\right]\\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\frac{1}{n_0-1}\boldsymbol{P}_{11}-\frac{1}{N_{11}}\mathbf{1}_{N_{11}}\mathbf{1}_{n_0}'\right]\left[\frac{1}{n_0-1}\boldsymbol{P}_{11}'-\frac{1}{N_{11}}\mathbf{1}_{n_0}\mathbf{1}_{N_{11}}'\right]\\
\Longrightarrow \boldsymbol{s}_{11}'\boldsymbol{\Delta}_{11}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{11_l}^*\boldsymbol{v}_{11_l}^{*\prime}\right]\boldsymbol{\Delta}_{11}^{-1}\boldsymbol{s}_{11} &= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\boldsymbol{s}_{11}'\left[\frac{1}{n_0-1}\boldsymbol{P}_{11}-\frac{1}{N_{11}}\mathbf{1}_{N_{11}}\mathbf{1}_{n_0}'\right]\left[\frac{1}{n_0-1}\boldsymbol{P}_{11}'-\frac{1}{N_{11}}\mathbf{1}_{n_0}\mathbf{1}_{N_{11}}'\right]\boldsymbol{s}_{11}\\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\sum_{j=1}^{n_0}\left(\boldsymbol{s}_{11}'\left[\frac{1}{n_0-1}\boldsymbol{P}_{11}-\frac{1}{N_{11}}\mathbf{1}_{N_{11}}\mathbf{1}_{n_0}'\right]\right)_j^2\\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\sum_{j=1}^{n_0}\left(\overline{s}_{11}^{(1j)}-\overline{s}_{11}\right)^2
\end{aligned}
$$

The following two identities are used to move through the final two equalities:

1. $\frac{1}{n_0-1}\boldsymbol{P}_{11}'\boldsymbol{s}_{11}=\left(\overline{s}_{11}^{(11)},\ldots,\overline{s}_{11}^{(1n_0)}\right)'$, where $\overline{s}_{11}^{(1j)}$, $j\in\{1,\ldots,n_0\}$ is the mean value of scores that compare object j in source 1 to any other object in source 1;

2. $\frac{1}{N_{11}}\mathbf{1}_{n_0}\mathbf{1}_{N_{11}}'\boldsymbol{s}_{11}=\overline{\boldsymbol{s}}_{11}$, where $\overline{\boldsymbol{s}}_{11}$ is the n_0-dimensional vector of $\overline{s}_{11}$.

Together, these imply that

$$
\frac{1}{n_0-1}\boldsymbol{P}_{11}'\boldsymbol{s}_{11}-\frac{1}{N_{11}}\mathbf{1}_{n_0}\mathbf{1}_{N_{11}}'\boldsymbol{s}_{11}=\left(\overline{s}^{(11)}-\overline{s}_{11},\ldots,\overline{s}^{(1n_0)}-\overline{s}_{11}\right)'.
$$

B CALCULATIONS FOR SS_{22} BRACKETED TERMS

B.1 FULL DEVELOPMENT OF (3.23)

$$
\begin{aligned}
\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)'\boldsymbol{\Delta}_{22}^{-1'}\boldsymbol{v}_{22_1}^*\boldsymbol{v}_{22_1}^{*\prime}\boldsymbol{\Delta}_{22}^{-1}\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right) &= \left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)'\boldsymbol{\Delta}_{22}^{-1'}\boldsymbol{v}_{22_1}^*\left(\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)'\boldsymbol{\Delta}^{-1'}\boldsymbol{v}_{22_1}^*\right)'\\
&= \left[\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)'\boldsymbol{\Delta}_{22}^{-1'}\boldsymbol{v}_{22_1}^*\right]^2\\
&= \left[\left(\boldsymbol{s}_{22}-\theta_{22}\mathbf{1}_{N_{22}}\right)'\sigma_{22}^{-1}\boldsymbol{v}_{22_1}^*\right]^2\\
&= \left[\sigma_{22}^{-1}\boldsymbol{s}_{22}\boldsymbol{v}_{22_1}^*-\sigma_{22}^{-1}\theta_{22}\mathbf{1}_{N_{22}}'\boldsymbol{v}_{22_1}^*\right]^2\\
&= \sigma_{22}^{-2}\left[\left(\boldsymbol{s}_{22}'\frac{\mathbf{1}_{N_{22}}}{\sqrt{N_{22}}}\right)^2-2\left(\boldsymbol{s}_{22}'\frac{\mathbf{1}_{N_{22}}}{\sqrt{N_{22}}}\right)\left(\theta_{22}\mathbf{1}_{N_{22}}'\frac{\mathbf{1}_{N_22}}{\sqrt{N_{22}}}\right)\right.\\
&\quad\left.+\left(\theta_{22}\mathbf{1}_{N_{22}}'\frac{\mathbf{1}_{N_22}}{\sqrt{N_{22}}}\right)^2\right]\\
&= \sigma_{22}^{-2}\left[\frac{\left(\sum_{i=1}^{N_{22}}s_{22_i}\right)^2}{N_{22}}-2\theta_{22}\sum_{i=1}^{N_{22}}s_{22_i}+N_{22}\theta_{22}^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{N_{22}}{\sigma_{22}^2}\left[\left(\frac{\sum_{i=1}^{N_{22}} s_{22_i}}{N_{22}}\right)^2 - 2\theta_{22}\frac{\sum_{i=1}^{N_{22}} s_{22_i}}{N_{22}} + \theta_{22}^2\right] \\
&= \frac{N_{22}}{\sigma_{22}^2}\left[\bar{s}_{22}^2 - 2\theta_{22}\bar{s}_{22} + \theta_{22}^2\right] \\
&= \frac{N_{22}}{\sigma_{22}^2}\left[\bar{s}_{22} - \theta_{22}\right]^2
\end{aligned}
$$

B.2 FULL DEVELOPMENT OF (3.24)

$$
\begin{aligned}
\boldsymbol{s}_{22}'\boldsymbol{\Delta}_{22}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{22_l}^*\boldsymbol{v}_{22_l}^{*\,\prime}\right]\boldsymbol{\Delta}_{22}^{-1}\boldsymbol{s}_{22} &= \frac{1}{n_0-2}\left[\sigma_{22}^{-2}\left(\boldsymbol{P}_{22}\boldsymbol{P}_{22}' - 2(n_0-1)\boldsymbol{v}_{22_1}^*\boldsymbol{v}_{22_1}^{*\,\prime}\right)\right] \\
&= \frac{1}{\sigma_{22}^2(n_0-2)}\left[\boldsymbol{P}_{22}\boldsymbol{P}_{22}' - 2(n_0-1)\frac{\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'}{N_{22}}\right] \\
&= \frac{1}{\sigma_{22}^2(n_0-2)}\left[\boldsymbol{P}_{22}\boldsymbol{P}_{22}' - 2(n_0-1)\frac{\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'}{\frac{n_0(n_0-1)}{2}}\right] \\
&= \frac{1}{\sigma_{22}^2(n_0-2)}\left[\boldsymbol{P}_{22}\boldsymbol{P}_{22}' - 4(n_0-1)\frac{\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'}{n_0(n_0-1)}\right] \\
&= \frac{1}{\sigma_{22}^2(n_0-2)}\frac{(n_0-1)^2}{(n_0-1)^2}\left[\boldsymbol{P}_{22}\boldsymbol{P}_{22}' - \frac{4}{n_0}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{22}\boldsymbol{P}_{22}'}{(n_0-1)^2} - \frac{4}{n_0(n_0-1)^2}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{22}\boldsymbol{P}_{22}'}{(n_0-1)^2} + \left(\frac{4n_0}{n_0^2(n_0-1)^2} - \frac{8n_0}{n_0^2(n_0-1)^2}\right)\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{22}\boldsymbol{P}_{22}'}{(n_0-1)^2} + \frac{4n_0}{n_0^2(n_0-1)^2}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}' - \frac{8}{n_0(n_0-1)^2}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{22}\boldsymbol{P}_{22}'}{(n_0-1)^2} + \frac{n_0}{N_{22}^2}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}' - \frac{8}{n_0(n_0-1)^2}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{22}\boldsymbol{P}_{22}'}{(n_0-1)^2} + \frac{n_0}{N_{22}^2}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}' - \frac{4}{N_{22}(n_0-1)}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{22}\boldsymbol{P}_{22}'}{(n_0-1)^2} - \frac{4}{N_{22}(n_0-1)}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}' + \frac{n_0}{N_{22}^2}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{N_{22}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\left[\frac{1}{n_0-1}\boldsymbol{P}_{22} - \frac{1}{N_{22}}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{n_0}'\right]\left[\frac{1}{n_0-1}\boldsymbol{P}_{22}' - \frac{1}{N_{22}}\boldsymbol{1}_{n_0}\boldsymbol{1}_{N_{22}}'\right] \\
\Longrightarrow \boldsymbol{s}_{22}'\boldsymbol{\Delta}_{22}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{22_l}^*\boldsymbol{v}_{22_l}^{*\,\prime}\right]\boldsymbol{\Delta}_{22}^{-1}\boldsymbol{s}_{22} &= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\boldsymbol{s}_{22}'\left[\frac{1}{n_0-1}\boldsymbol{P}_{22} - \frac{1}{N_{22}}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{n_0}'\right]\left[\frac{1}{n_0-1}\boldsymbol{P}_{22}' - \frac{1}{N_{22}}\boldsymbol{1}_{n_0}\boldsymbol{1}_{N_{22}}'\right]\boldsymbol{s}_{22} \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\sum_{j=1}^{n_0}\left(\boldsymbol{s}_{22}'\left[\frac{1}{n_0-1}\boldsymbol{P}_{22} - \frac{1}{N_{22}}\boldsymbol{1}_{N_{22}}\boldsymbol{1}_{n_0}'\right]\right)_j^2 \\
&= \frac{(n_0-1)^2}{\sigma_{22}^2(n_0-2)}\sum_{j=1}^{n_0}\left(\bar{s}^{(2j)} - \bar{s}_{22}\right)^2
\end{aligned}
$$

To reach the final equality, we make use of the two identities expressed in Appendix A.

C CALCULATIONS FOR SS_{12} BRACKETED TERMS

C.1 FULL DEVELOPMENT OF (3.25)

$$
\begin{aligned}
\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)'\boldsymbol{\Delta}_{12}^{-1'}\boldsymbol{v}_{12_1}^{*}\boldsymbol{v}_{12_1}^{*}{}'\boldsymbol{\Delta}_{12}^{-1}\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right) &= \left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)'\boldsymbol{\Delta}_{12}^{-1'}\boldsymbol{v}_{12_1}^{*}\left(\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)'\boldsymbol{\Delta}^{-1'}\boldsymbol{v}_{12_1}^{*}\right)' \\
&= \left[\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)'\boldsymbol{\Delta}_{12}^{-1'}\boldsymbol{v}_{12_1}^{*}\right]^2 \\
&= \left[\left(\boldsymbol{s}_{12}-\theta_{12}\mathbf{1}_{N_{12}}\right)'\sigma_{12}^{-1}\boldsymbol{v}_{12_1}^{*}\right]^2 \\
&= \left[\sigma_{12}^{-1}\boldsymbol{s}_{12}\boldsymbol{v}_{12_1}^{*}-\sigma_{12}^{-1}\theta_{12}\mathbf{1}_{N_{12}}'\boldsymbol{v}_{12_1}^{*}\right]^2 \\
&= \sigma_{12}^{-2}\left[\left(\boldsymbol{s}_{12}'\frac{\mathbf{1}_{N_{12}}}{\sqrt{N_{12}}}\right)^2-2\left(\boldsymbol{s}_{12}'\frac{\mathbf{1}_{N_{12}}}{\sqrt{N_{12}}}\right)\left(\theta_{12}\mathbf{1}_{N_{12}}'\frac{\mathbf{1}_{N_12}}{\sqrt{N_{12}}}\right)\right. \\
&\quad \left.+\left(\theta_{12}\mathbf{1}_{N_{12}}'\frac{\mathbf{1}_{N_12}}{\sqrt{N_{12}}}\right)^2\right] \\
&= \sigma_{12}^{-2}\left[\frac{\left(\sum_{i=1}^{N_{12}}s_{12_i}\right)^2}{N_{12}}-2\theta_{12}\sum_{i=1}^{N_{12}}s_{12_i}+N_{12}\theta_{12}^2\right] \\
&= \frac{N_{12}}{\sigma_{12}^2}\left[\left(\frac{\sum_{i=1}^{N_{12}}s_{12_i}}{N_{12}}\right)^2-2\theta_{12}\frac{\sum_{i=1}^{N_{12}}s_{12_i}}{N_{12}}+\theta_{12}^2\right] \\
&= \frac{N_{12}}{\sigma_{12}^2}\left[\overline{s}_{12}^2-2\theta_{12}\overline{s}_{12}+\theta_{12}^2\right] \\
&= \frac{N_{12}}{\sigma_{12}^2}\left[\overline{s}_{12}-\theta_{12}\right]^2
\end{aligned}
$$

C.2 FULL DEVELOPMENT OF (3.26)

$$
\begin{aligned}
\boldsymbol{\Delta}_{12}^{-1'}\left[\sum_{l=2}^{2n_0-1}\boldsymbol{v}_{12_l}^{*}\boldsymbol{v}_{12_l}^{*}{}'\right]\boldsymbol{\Delta}_{12}^{-1} &= \frac{1}{\sigma_{12}^2 n_0}\left[\boldsymbol{P}_{12}\boldsymbol{P}_{12}'-nn_0\boldsymbol{v}_{12_1}^{*}\boldsymbol{v}_{12_1}^{*}{}'\right] \\
&= \frac{1}{\sigma_{12}^2 n_0}\left[\boldsymbol{P}_{12}\boldsymbol{P}_{12}'-2n_0\frac{\mathbf{1}_{N_{12}}}{\sqrt{n_0^2}}\frac{\mathbf{1}_{N_{12}}'}{\sqrt{n_0^2}}\right] \\
&= \frac{1}{\sigma_{12}^2 n_0}\left[\boldsymbol{P}_{12}\boldsymbol{P}_{12}'-\frac{2n_0}{n_0^2}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'\right] \\
&= \frac{1}{\sigma_{12}^2 n_0}\frac{n_0^2}{n_0^2}\left[\boldsymbol{P}_{12}\boldsymbol{P}_{12}'-\frac{2}{n_0}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'\right] \\
&= \frac{n_0^2}{\sigma_{12}^2 n_0}\frac{1}{n_0^2}\left[\boldsymbol{P}_{12}\boldsymbol{P}_{12}'-\frac{2}{n_0}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'\right] \\
&= \frac{n_0^2}{\sigma_{12}^2 n_0}\left[\frac{\boldsymbol{P}_{12}\boldsymbol{P}_{12}'}{n_0^2}-\frac{2}{n_0^3}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'\right] \\
&= \frac{n_0^2}{\sigma_{12}^2 n_0}\left[\frac{\boldsymbol{P}_{12}\boldsymbol{P}_{12}'}{n_0^2}+\left(\frac{2}{n_0^3}-\frac{4}{n_0^3}\right)\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'\right] \\
&= \frac{n_0^2}{\sigma_{12}^2 n_0}\left[\frac{\boldsymbol{P}_{12}\boldsymbol{P}_{12}'}{n_0^2}+\frac{2n_0}{n_0^4}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'-\frac{4n_0}{n_0^4}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'\right] \\
&= \frac{n_0^2}{\sigma_{12}^2 n_0}\left[\frac{\boldsymbol{P}_{12}\boldsymbol{P}_{12}'}{n_0^2}-\frac{4n_0}{n_0^4}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'+\frac{2n_0}{n_0^4}\mathbf{1}_{N_{12}}\mathbf{1}_{N_{12}}'\right] \\
&= \frac{n_0^2}{\sigma_{12} n_0}\left[\frac{1}{n_0}\boldsymbol{P}_{12}-\frac{1}{n_0^2}\mathbf{1}_{N_{12}}\mathbf{1}_{2n_0}'\right]\left[\frac{1}{n_0}\boldsymbol{P}_{12}'-\frac{1}{n_0^2}\mathbf{1}_{2n_0}\mathbf{1}_{N_{12}}'\right]
\end{aligned}
$$

$$\Longrightarrow \boldsymbol{s}'_{12}\boldsymbol{\Delta}_{12}^{-1'}\left[\sum_{l=2}^{2n_0-1}\boldsymbol{v}^*_{12_l}\boldsymbol{v}^{*\prime}_{12_l}\right]\boldsymbol{\Delta}_{12}^{-1}\boldsymbol{s}_{12} = \frac{n_0^2}{\sigma_{12}^2 n_0}\boldsymbol{s}'_{12}\left[\frac{1}{n_0}\boldsymbol{P}_{12} - \frac{1}{n_0^2}\mathbf{1}_{N_{12}}\mathbf{1}'_{2n_0}\right]\left[\frac{1}{n_0}\boldsymbol{P}'_{12} - \frac{1}{n_0^2}\mathbf{1}_{2n_0}\mathbf{1}'_{N_{12}}\right]\boldsymbol{s}_{12}$$

$$= \frac{n_0^2}{\sigma_{12}^2 n_0}\sum_{j=1}\left(\boldsymbol{s}'_{12}\left[\frac{1}{n_0}\boldsymbol{P}_{12} - \frac{1}{n_0^2}\mathbf{1}_{N_{12}}\mathbf{1}'_{2n_0}\right]\right)_j^2$$

$$= \frac{n_0^2}{\sigma_{12}^2 n_0}\sum_{i=1}^{2}\sum_{j=2}^{n_0}\left(\overline{s}_{12}^{(ij)} - \overline{s}_{12}\right)^2$$

To reach the final equality, we make use of the two identities expressed in Appendix A.

D GENERALIZATION OF DEVELOPMENT FOR WITHIN-SOURCE SUMS OF SQUARES

D.1 FULL DEVELOPMENT OF (6.18), (9.18) AND (9.22)

$$\left(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}\right)'\boldsymbol{\Delta}_{ii}^{-1'}\boldsymbol{v}^*_{ii_1}\boldsymbol{v}^{*\prime}_{ii_1}\boldsymbol{\Delta}_{ii}^{-1}\left(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}\right) = \left(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}\right)'\boldsymbol{\Delta}_{ii}^{-1'}\boldsymbol{v}^*_{ii_1}\left(\left(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}\right)'\boldsymbol{\Delta}^{-1'}\boldsymbol{v}^*_{ii_1}\right)'$$

$$= \left[\left(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}\right)'\boldsymbol{\Delta}_{ii}^{-1'}\boldsymbol{v}^*_{ii_1}\right]^2$$

$$= \left[\left(\boldsymbol{s}_{ii} - \theta_{ii}\mathbf{1}_{N_{ii}}\right)'\sigma_{ii}^{-1}\boldsymbol{v}^*_{ii_1}\right]^2$$

$$= \left[\sigma_{ii}^{-1}\boldsymbol{s}_{ii}\boldsymbol{v}^*x_{ii_1} - \sigma_{ii}^{-1}\theta_{ii}\mathbf{1}'_{N_{ii}}\boldsymbol{v}^*_{ii_1}\right]^2$$

$$= \sigma_{ii}^{-2}\left[\left(\boldsymbol{s}'_{ii}\frac{\mathbf{1}_{N_{ii}}}{\sqrt{N_{ii}}}\right)^2 - 2\left(\boldsymbol{s}'_{ii}\frac{\mathbf{1}_{N_{ii}}}{\sqrt{N_{ii}}}\right)\left(\theta_{ii}\mathbf{1}'_{N_{ii}}\frac{\mathbf{1}_{N_{ii}}}{\sqrt{N_{ii}}}\right) + \left(\theta_{ii}\mathbf{1}'_{N_{ii}}\frac{\mathbf{1}_{N_{ii}}}{\sqrt{N_{ii}}}\right)^2\right]$$

$$= \sigma_{ii}^{-2}\left[\frac{\left(\sum_{l=1}^{N_{ii}} s_{ii_l}\right)^2}{N_{ii}} - 2\theta_{ii}\sum_{l=1}^{N_{ii}} s_{ii_l} + N_{ii}\theta_{ii}^2\right]$$

$$= \frac{N_{ii}}{\sigma_{ii}^2}\left[\left(\frac{\sum_{l=1}^{N_{ii}} s_{ii_l}}{N_{ii}}\right)^2 - 2\theta_{ii}\frac{\sum_{l=1}^{N_{ii}} s_{ii_l}}{N_{ii}} + \theta_{ii}^2\right]$$

$$= \frac{N_{ii}}{\sigma_{ii}^2}\left[\overline{s}_{ii}^2 - 2\theta_{ii}\overline{s}_{ii} + \theta_{ii}^2\right]$$

$$= \frac{N_{ii}}{\sigma_{ii}^2}\left[\overline{s}_{ii} - \theta_{ii}\right]^2$$

D.2 FULL DEVELOPMENT OF (6.19), (9.19) AND (9.23)

$$\boldsymbol{\Delta}_{ii}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}^*_{ii_l}\boldsymbol{v}^{*\prime}_{ii_l}\right]\boldsymbol{\Delta}_{ii}^{-1} = \frac{1}{n_0 - 2}\left[\sigma_{ii}^{-2}\left(\boldsymbol{P}_{ii}\boldsymbol{P}'_{ii} - 2(n_0-1)\boldsymbol{v}^*_{ii_1}\boldsymbol{v}^{*\prime}_{ii_1}\right)\right]$$

$$= \frac{1}{\sigma_{ii}^2(n_0-2)}\left[\boldsymbol{P}_{ii}\boldsymbol{P}'_{ii} - 2(n_0-1)\frac{\mathbf{1}_{N_{ii}}\mathbf{1}'_{N_{ii}}}{N_{ii}}\right]$$

$$= \frac{1}{\sigma_{ii}^2(n_0-2)}\left[\boldsymbol{P}_{ii}\boldsymbol{P}'_{ii} - 2(n_0-1)\frac{\mathbf{1}_{N_{ii}}\mathbf{1}'_{N_{ii}}}{\frac{n_0(n_0-1)}{2}}\right]$$

$$= \frac{1}{\sigma_{ii}^2(n_0-2)}\left[\boldsymbol{P}_{ii}\boldsymbol{P}'_{ii} - 4(n_0-1)\frac{\mathbf{1}_{N_{ii}}\mathbf{1}'_{N_{ii}}}{n_0(n_0-1)}\right]$$

$$
\begin{aligned}
&= \frac{1}{\sigma_{ii}^2(n_0-2)}\frac{(n_0-1)^2}{(n_0-1)^2}\left[\boldsymbol{P}_{ii}\boldsymbol{P}_{ii}' - \frac{4}{n_0}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{ii}\boldsymbol{P}_{ii}'}{(n_0-1)^2} - \frac{4}{n_0(n_0-1)^2}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{ii}\boldsymbol{P}_{ii}'}{(n_0-1)^2} + \left(\frac{4n_0}{n_0^2(n_0-1)^2} - \frac{8n_0}{n_0^2(n_0-1)^2}\right)\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{ii}\boldsymbol{P}_{ii}'}{(n_0-1)^2} + \frac{4n_0}{n_0^2(n_0-1)^2}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}' - \frac{8}{n_0(n_0-1)^2}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{11}^2(n_0-2)}\left[\boldsymbol{P}_{ii}\boldsymbol{P}_{ii}' + \frac{n_0}{N_{ii}^2}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}' - \frac{8}{n_0(n_0-1)^2}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\left[\frac{\boldsymbol{P}_{ii}\boldsymbol{P}_{ii}'}{(n_0-1)^2} - \frac{4}{N_{ii}(n_0-1)}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}' + \frac{n_0}{N_{ii}^2}\mathbf{1}_{N_{ii}}\mathbf{1}_{N_{ii}}'\right] \\
&= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\left[\frac{1}{n_0-1}\boldsymbol{P}_{ii} - \frac{1}{N_{ii}}\mathbf{1}_{N_{ii}}\mathbf{1}_{n_0}'\right]\left[\frac{1}{n_0-1}\boldsymbol{P}_{ii}' - \frac{1}{N_{ii}}\mathbf{1}_{n_0}\mathbf{1}_{N_{ii}}'\right]
\end{aligned}
$$

$$
\begin{aligned}
\Longrightarrow \boldsymbol{s}_{ii}'\boldsymbol{\Delta}_{ii}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}_{ii_l}^*\boldsymbol{v}_{ii_l}^{*\,\prime}\right]\boldsymbol{\Delta}_{ii}^{-1}\boldsymbol{s}_{ii} &= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\boldsymbol{s}_{ii}'\left[\frac{1}{n_0-1}\boldsymbol{P}_{ii} - \frac{1}{N_{ii}}\mathbf{1}_{N_{ii}}\mathbf{1}_{n_0}'\right]\left[\frac{1}{n_0-1}\boldsymbol{P}_{ii}' - \frac{1}{N_{ii}}\mathbf{1}_{n_0}\mathbf{1}_{N_{ii}}'\right]\boldsymbol{s}_{ii} \\
&= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\sum_{j=1}^{n_0}\left(\boldsymbol{s}_{ii}'\left[\frac{1}{n_0-1}\boldsymbol{P}_{ii} - \frac{1}{N_{ii}}\mathbf{1}_{N_{ii}}\mathbf{1}_{n_0}'\right]\right)_j^2 \\
&= \frac{(n_0-1)^2}{\sigma_{ii}^2(n_0-2)}\sum_{j=1}^{n_0}\left(\overline{s}^{(ij)} - \overline{s}_{ii}\right)^2
\end{aligned}
$$

E GENERALIZATION OF DEVELOPMENT FOR BETWEEN-SOURCE SUMS OF SQUARES

E.1 FULL DEVELOPMENT OF (6.20), (9.20), AND (9.24)

$$
\begin{aligned}
\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Delta}_{ii'}^{-1'}\boldsymbol{v}_{ii_1'}^*\boldsymbol{v}_{ii_1'}^{*\,\prime}\boldsymbol{\Delta}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right) &= \left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Delta}_{ii'}^{-1'}\boldsymbol{v}_{ii_1'}^*\left(\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Delta}^{-1'}\boldsymbol{v}_{ii_1'}^*\right)' \\
&= \left[\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Delta}_{ii'}^{-1'}\boldsymbol{v}_{ii_1'}^*\right]^2 \\
&= \left[\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\sigma_{ii'}^{-1}\boldsymbol{v}_{ii_1'}^*\right]^2 \\
&= \left[\sigma_{ii'}^{-1}\boldsymbol{s}_{ii'}\boldsymbol{v}_{ii_1'} - \sigma_{ii'}^{-1}\theta_{ii'}\mathbf{1}_{N_{ii'}}'\boldsymbol{v}_{ii_1'}^*\right]^2 \\
&= \sigma_{ii'}^{-2}\left[\left(\boldsymbol{s}_{ii'}'\frac{\mathbf{1}_{N_{ii'}}}{\sqrt{N_{ii'}}}\right)^2 - 2\left(\boldsymbol{s}_{ii'}'\frac{\mathbf{1}_{N_{ii'}}}{\sqrt{N_{ii'}}}\right)\left(\theta_{ii'}\mathbf{1}_{N_{ii'}}'\frac{\mathbf{1}_{N_{ii'}}}{\sqrt{N_{ii'}}}\right)\right. \\
&\quad \left. + \left(\theta_{ii'}\mathbf{1}_{N_{ii'}}'\frac{\mathbf{1}_{N_{ii'}}}{\sqrt{N_{ii'}}}\right)^2\right] \\
&= \sigma_{ii'}^{-2}\left[\frac{\left(\sum_{l=1}^{N_{ii'}} s_{ii_l'}\right)^2}{N_{ii'}} - 2\theta_{ii'}\sum_{l=1}^{N_{ii'}} s_{ii_l'} + N_{ii'}\theta_{ii'}^2\right] \\
&= \frac{N_{ii'}}{\sigma_{ii'}^2}\left[\left(\frac{\sum_{l=1}^{N_{ii'}} s_{ii_l'}}{N_{ii'}}\right)^2 - 2\theta_{ii'}\frac{\sum_{l=1}^{N_{ii'}} s_{ii_l'}}{N_{ii'}} + \theta_{ii'}^2\right] \\
&= \frac{N_{ii'}}{\sigma_{ii'}^2}\left[\overline{s}_{ii'}^2 - 2\theta_{ii'}\overline{s}_{ii'} + \theta_{ii'}^2\right] \\
&= \frac{N_{ii'}}{\sigma_{ii'}^2}\left[\overline{s}_{ii'} - \theta_{ii'}\right]^2
\end{aligned}
$$

E.2 FULL DEVELOPMENT OF (6.21), (9.21), AND (9.26)

$$
\begin{aligned}
\boldsymbol{\Delta}_{ii'}^{-1'}\left[\sum_{l=2}^{2n_0-1}\boldsymbol{v}^*_{ii'_l}\boldsymbol{v}^{*\prime}_{ii'_l}\right]\boldsymbol{\Delta}_{ii'}^{-1} &= \frac{1}{n_0}\left[\sigma_{ii'}^{-2}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'} - 2n_0\boldsymbol{v}^*_{ii'_l}\boldsymbol{v}^{*\prime}_{ii'_l}\right)\right] \\
&= \frac{1}{\sigma_{ii'}^2 n_0}\left[\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'} - 2n_0\frac{\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}}{N_{ii'}}\right] \\
&= \frac{1}{\sigma_{ii'}^2 n_0}\left[\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'} - \frac{2n_0}{n_0^2}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{1}{\sigma_{ii'}^2 n_0}\frac{n_0^2}{n_0^2}\left[\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'} - \frac{2}{n_0}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{n_0^2}{\sigma_{ii'}^2 n_0}\frac{1}{n_0^2}\left[\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'} - \frac{2}{n_0}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{n_0^2}{\sigma_{12}^2 n_0}\left[\frac{\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}}{n_0^2} - \frac{2}{n_0^3}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{n_0^2}{\sigma_{ii'}^2 n_0}\left[\frac{\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}}{n_0^2} + \left(\frac{2}{n_0^3} - \frac{4}{n_0^3}\right)\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{n_0^2}{\sigma_{ii'}^2 n_0}\left[\frac{\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}}{n_0^2} + \frac{2n_0}{n_0^4}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}} - \frac{4n_0}{n_0^4}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{n_0^2}{\sigma_{ii'}^2 n_0}\left[\frac{\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}}{n_0^2} - \frac{4n_0}{n_0^4}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}} + \frac{2n_0}{n_0^4}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{n_0^2}{\sigma_{ii'}^2 n_0}\left[\frac{\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}}{n_0^2} - \frac{4}{n_0^3}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}} + \frac{2n_0}{n_0^4}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{N_{ii'}}\right] \\
&= \frac{n_0^2}{\sigma_{ii'} n_0}\left[\frac{1}{n_0}\boldsymbol{P}_{ii'} - \frac{1}{n_0^2}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{2n_0}\right]\left[\frac{1}{n_0}\boldsymbol{P}'_{ii'} - \frac{1}{n_0^2}\mathbf{1}_{2n_0}\mathbf{1}'_{N_{ii'}}\right] \\
\Longrightarrow \boldsymbol{s}'_{ii'}\boldsymbol{\Delta}_{ii'}^{-1'}\left[\sum_{l=2}^{n_0}\boldsymbol{v}^*_{ii'_l}\boldsymbol{v}^{*\prime}_{ii'_l}\right]\boldsymbol{\Delta}_{ii'}^{-1}\boldsymbol{s}_{ii'} &= \frac{n_0^2}{\sigma_{12}^2 n_0}\boldsymbol{s}'_{ii'}\left[\frac{1}{n_0}\boldsymbol{P}_{ii'} - \frac{1}{n_0^2}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{2n_0}\right]\left[\frac{1}{n_0}\boldsymbol{P}'_{ii'} - \frac{1}{n_0^2}\mathbf{1}_{2n_0}\mathbf{1}'_{N_{ii'}}\right]\boldsymbol{s}_{ii'} \\
&= \frac{n_0^2}{\sigma_{ii'}^2 n_0}\sum_{j=1}\left(\boldsymbol{s}'_{ii'}\left[\frac{1}{n_0}\boldsymbol{P}_{ii'} - \frac{1}{N_{ii'}}\mathbf{1}_{N_{ii'}}\mathbf{1}'_{2n_0}\right]\right)_j^2 \\
&= \frac{n_0^2}{\sigma_{ii'}^2 n_0}\left(\sum_{j=1}^{n_0}\left(\overline{s}^{(ij_{i'})} - \overline{s}_{ii'}\right)^2 + \sum_{j=1}^{n_0}\left(\overline{s}^{(i'j_i)} - \overline{s}_{ii'}\right)^2\right)
\end{aligned}
$$

F GENERALIZATION OF ASSIGNING DISTRIBUTIONS TO PARAMETERS

F.1 POSTERIOR DISTRIBUTION FOR σ_e^2 FOR THE TWO- AND MULTI-CLASS MODELS

We begin by defining MS_e as a function of all available information pertaining to σ_e^2 in Tables 6.5 through 6.6:

$$
MS_e = \sum_{i=1}^{n}\left(\frac{SS_{W_{ii}}}{C_2} + \frac{SS_{E_{ii}}}{C_3}\right) + \sum_{i=1}^{n-1}\sum_{i'=2}^{n}\left(\frac{SS_{W_{ii'}}}{C_1} + \frac{SS_{E_{ii'}}}{C_3}\right).
$$

For ease of notation, we define

$$C_1 = n_0 \left(\frac{1-\sigma_e^2}{2} \right) + \sigma_e^2 \qquad C_2 = (n_0 - 2) \left(\frac{1-\sigma_e^2}{2} \right) + \sigma_e^2 \qquad C_3 = \sigma_e^2,$$

such that

$$\begin{aligned} MS_e &= \frac{C_1}{C_1}\frac{C_3}{C_3}\sum_{i=1}^{n}\left(\frac{SS_{W_{ii}}}{C_2}\right) + \frac{C_2}{C_2}\frac{C_3}{C_3}\sum_{i=1}^{n-1}\sum_{i'=2}^{n}\left(\frac{SS_{W_{ii'}}}{C_1}\right) + \frac{C_1}{C_1}\frac{C_2}{C_2}\left(\sum_{i=1}^{n}\left(\frac{SS_{E_{ii}}}{C_3}\right) + \sum_{i=1}^{n-1}\sum_{i'=2}^{n}\left(\frac{SS_{E_{ii'}}}{C_3}\right)\right) \\ &= \frac{C_1C_3\left(\sum_{i=1}^{n} SS_{W_{ii}}\right) + C_2C_3\left(\sum_{i=1}^{n-1}\sum_{i=2}^{n} SS_{W_{ii'}}\right) + C_1C_2\left(\sum_{i=1}^{n} SS_{E_{ii}} + \sum_{i=1}^{n-1}\sum_{i'=2}^{n} SS_{E_{ii'}}\right)}{C_1C_2C_3} \\ &\sim \chi^2_{df=\binom{nn_0}{2}-\binom{n}{2}-n}. \end{aligned}$$

We obtain the posterior distribution of σ_e^2, $\pi(\sigma_e^2|MS_e)$ by considering the above distribution of MS_e, $f(MS_e|\sigma_e^2)$, in conjunction with a Beta prior on σ_e^2, with hyper-parameters α_e, β_e, such that

$$\begin{aligned} \pi(\sigma_e^2|MS_e) &\propto f(MS_e|\sigma_e^2)\pi(\sigma_e^2) \\ &= \chi^2(MS_e|\sigma_e^2)\mathcal{B}(\sigma_e^2|\alpha_e, \beta_e). \end{aligned}$$

F.2 POSTERIOR DISTRIBUTION FOR $\theta_{ii'}$ FOR THE TWO- AND MULTI-CLASS MODELS

We obtain the posterior distribution of $\theta_{ii'}$, $\pi(\theta_{ii'}|\boldsymbol{s}_{ii'}, \sigma_{ii'}, \sigma_a^2, \sigma_e^2)$, for $i = i'$ and $i \neq i'$, by considering a Multivariate Normal likelihood for our vector of scores, $\boldsymbol{s}_{ii'}$, and a Normal prior on $\theta_{ii'}$, with hyper-parameters $\phi_{ii'}, \omega_{ii'}$. We define $\theta_{ii'}$ and $\boldsymbol{\Delta}_{ii'}$ as the parameters associated with $\boldsymbol{s}_{ii'}$, $\boldsymbol{P}_{ii'}$ and $\boldsymbol{I}_{ii'}$ as design matrices corresponding to these source comparisons, and $N_{ii'}$ as the cardinality of $\boldsymbol{s}_{ii'}$. We begin by defining

$$
\begin{aligned}
\pi\left(\theta_{ii'}|\boldsymbol{s}_{ii'},\sigma_a^2,\sigma_e^2\right) &\propto f\left(\boldsymbol{s}_{ii'}|\theta_{ii'},\sigma_{ii'},\sigma_a^2,\sigma_e^2,\phi_{ii'},\omega_{ii'}\right)\pi\left(\theta_{ii'}|\phi_{ii'},\omega_{ii'}\right)\\
&= \mathcal{MVN}\left(\boldsymbol{s}_{ii'}|\theta_{ii'},\sigma_{ii'}.\sigma_a^2,\sigma_e^2,\phi_{ii'},\omega_{ii'}\right)\mathcal{N}\left(\theta_{ii'}|\phi_{ii'},\omega_{ii'}\right)\\
&= (2\pi)^{-\frac{N_{ii'}}{2}}|\boldsymbol{\Delta}_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}_{ii'}'\sigma_a^2+\boldsymbol{I}_{ii'}\sigma_e^2\right)\boldsymbol{\Delta}_{ii'}'|^{-\frac{1}{2}}\\
&\quad\times\exp\left\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'}-\theta_{kk'}\mathbf{1}_{N_{ii'}}\right)'\left[\boldsymbol{\Delta}_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}_{ii'}'\sigma_a^2+\boldsymbol{I}_{ii'}\sigma_e^2\right)\boldsymbol{\Delta}_{ii'}\right]^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right\}\\
&\quad\times\frac{1}{\sqrt{2\pi\omega_{ii'}}}\exp\left\{-\frac{1}{2\omega_{ii'}}\left(\theta_{ii'}-\phi_{ii'}\right)^2\right\}.
\end{aligned}
$$

From this point on, we define $\boldsymbol{\Sigma}_{ii'}:=\boldsymbol{P}_{ii'}\boldsymbol{P}_{ii'}'\sigma_a^2+\boldsymbol{I}_{ii'}\sigma_e^2$ for ease of notation, and so

$$
\begin{aligned}
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Sigma}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right\}\times\exp\left\{-\frac{1}{2\omega_{ii'}}\left(\theta_{ii'}-\phi_{ii'}\right)^2\right\}\\
&= \exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Sigma}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)+\frac{1}{\omega_{ii'}}\left(\theta_{ii'}-\phi_{ii'}\right)^2\right]\right\}\\
&= \exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{s}_{ii'}'\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'}-\boldsymbol{s}_{ii'}'\boldsymbol{\Sigma}_{ii'}^{-1}\theta_{ii'}\mathbf{1}_{N_{ii'}}-\theta_{ii'}\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'}+\theta_{ii'}\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\theta_{ii'}\mathbf{1}_{N_{ii'}}\\
&\quad+\frac{1}{\omega_{ii'}}\theta_{ii'}^2-\frac{2}{\omega_{ii'}}\theta_{ii'}\phi_{ii'}+\frac{1}{\omega_{ii'}}\phi_{ii'}^2\Bigg]\Bigg\}\\
&= \exp\left\{-\frac{1}{2}\left[\left(\theta_{ii'}\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}+\frac{1}{\omega_{ii'}}\theta_{ii}^2\right)-2\left(\theta_{ii'}\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'}+\frac{1}{\omega_{ii'}}\theta_{ii'}\phi_{ii'}\right)+C\right]\right\}\\
&= \exp\left\{-\frac{1}{2}\left[\theta_{ii'}^2\left(\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}+\frac{1}{\omega_{ii'}}\right)-2\theta_{ii'}\left(\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'}+\frac{1}{\omega_{ii'}}\phi_{ii'}\right)+C\right]\right\}.
\end{aligned}
$$

At this point, we define

$$
\begin{aligned}
\frac{1}{\sigma_{ii_p'}^2}=\frac{1}{\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}+\frac{1}{\omega_{ii'}}}\Longrightarrow\sigma_{ii_p'}^2 &= \left(\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}+\omega_{ii'}^{-1}\right)^{-1}\\
&= \frac{\omega_{ii'}}{\left(\mathbf{1}_{N_{ii'}}'\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\right)\omega_{ii'}+1},
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\mu_{ii'_p}}{\sigma^2_{ii'_p}} = \mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'} \Longrightarrow \mu_{ii'_p} &= \left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}\right)\sigma^2_{ii'_p} \\
&= \left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}\right) \times \frac{\omega_{ii'}}{\left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\mathbf{1}_{N_{ii'}}\right)\omega_{ii'} + 1} \\
&= \left[\frac{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1}\right]\omega_{ii'} \\
&= \frac{\mathbf{1}_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\boldsymbol{s}_{ii'}\omega_{ii'} + \phi_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}^{-1}_{ii'}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1}
\end{aligned}
$$

And so $\pi\left(\theta_{ii'}|\boldsymbol{s}_{ii'}, \sigma^2_a, \sigma^2_e\right) \sim N(\mu_{ii'_p}, \sigma^2_{ii'_p})$, with $\mu_{ii'_p}$ and $\sigma^2_{ii'_p}$ defined as above.

F.3 POSTERIOR DISTRIBUTION FOR $\sigma^2_{ii'}$ FOR THE TWO- AND MULTI-CLASS MODELS

We obtain the posterior distribution of $\sigma^2_{ii'}$, $\pi(\sigma^2_{ii'}|\boldsymbol{s}_{ii'}, \theta_{ii'}, \sigma^2_a, \sigma^2_e)$, by considering a Multivariate Normal likelihood for our vector of scores, $\boldsymbol{s}_{ii'}$, and an Inverse-Gamma prior on $\sigma^2_{ii'}$, with hyper-parameters $\alpha_{ii'}, \beta_{ii'}$. As in Appendix F.2, the subscript ii' indicates the type of relationship we are considering between our scores. We begin by defining

$$
\begin{aligned}
\pi(\sigma^2_{ii'}|\boldsymbol{s}_{ii'}, \theta_{ii'}, \sigma^2_a, \sigma^2_e) &\propto f(\boldsymbol{s}_{ii'}|\sigma^2_{ii'}, \theta_{ii'}, \sigma^2_a, \sigma^2_e, \alpha_{ii'}, \beta_{ii'})\pi(\sigma^2_{ii'}|\alpha_{ii'}, \beta_{ii'}) \\
&= \mathcal{MVN}(\boldsymbol{s}_{ii'}|\sigma^2_{ii'}, \theta_{ii'}, \sigma^2_a, \sigma^2_e, \alpha_{ii'}, \beta_{ii'})\mathcal{IG}(\sigma^2_{ii'}|\alpha_{ii'}, \beta_{ii'}) \\
&= (2\pi)^{-\frac{N}{2}}|\boldsymbol{\Delta}_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{I}_{ii'}\sigma^2_e\right)\boldsymbol{\Delta}'_{kk'}|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{kk'}}\right)'\left[\boldsymbol{\Delta}_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{I}\sigma^2_e\right)\boldsymbol{\Delta}'_{ii'}\right]^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right\} \\
&\quad \times \frac{\beta^{\alpha_{ii'}}_{ii'}}{\Gamma(\alpha_{ii'})}\sigma^{2^{-(\alpha_{ii'}+1)}}_{ii'} \exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
&\propto |\sigma^2_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{I}_{ii'}\sigma^2_e\right)|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\left[\sigma^2_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{I}_{ii'}\sigma^2_e\right)\right]^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right\} \\
&\quad \times \sigma^{2^{-(\alpha_{ii'}+1)}}_{ii'} \exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
&= \left(\sigma^{2^{N_{ii'}}}_{ii'}|\boldsymbol{P}_{ii'}\boldsymbol{P}_{ii'}\sigma^2_a + \boldsymbol{I}_{ii'}\sigma^2_e|\right)^{-\frac{1}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\left[\sigma^2_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{I}_{ii'}\sigma^2_e\right)\right]^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
& \times\, \sigma_{ii'}^{2^{-(\alpha_{ii'}+1)}} \exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
\propto\;& \sigma_{ii'}^{2^{-\frac{N_{ii'}}{2}}} \exp\left\{-\frac{1}{2\sigma^2_{ii'}}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{ii'}\right)'\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a+\boldsymbol{I}_{ii'}\sigma^2_e\right)^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right\} \\
& \times\, \sigma_{ii'}^{2^{-\alpha_{ii'}-1}} \exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
=\;& \sigma_{ii'}^{2^{-\left(\frac{N_{ii'}}{2}+\alpha_{ii'}\right)-1}} \\
& \times \exp\left\{-\frac{1}{\sigma^2_{ii'}}\left[\frac{1}{2}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a+\boldsymbol{I}_{ii'}\sigma^2_e\right)^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)+\beta_{ii'}\right]\right\} \\
\sim\;& \mathcal{IG}\left(\alpha_{ii'_p},\beta_{ii'_p}\right)
\end{aligned}
$$

where

$$\alpha_{kk'_p} := \frac{N_{ii'}}{2}+\alpha_{ii'} \quad \text{and} \quad \beta_{ii'_p} := \frac{1}{2}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a+\boldsymbol{I}_{ii'}\sigma^2_e\right)^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)+\beta_{ii'}.$$

G GENERALIZATION OF ASSIGNING DISTRIBUTIONS TO PARAMETERS FOR THE POPULATION-BASED MODEL

G.1 POSTERIOR DISTRIBUTION FOR $\theta_{ii'}$ FOR THE POPULATION-BASED MODEL

We obtain the posterior distribution of $\theta_{ii'}$, $\pi(\theta_{ii'}|\boldsymbol{s}_{ii'},\sigma_{ii'},\widetilde{\boldsymbol{\sigma}}^2)$ for $ii'\in\{kk,kp,pp,pp'\}$ by considering a Multivariate Normal likelihood for our vector of scores, $\boldsymbol{s}_{ii'}$, and a Normal prior on $\theta_{ii'}$, with hyper-parameters $\phi_{ii'},\omega_{ii'}$. We define $\theta_{ii'}$ and $\boldsymbol{\Delta}_{ii'}$ as the parameters associated with $\boldsymbol{s}_{ii'}$, $\boldsymbol{P}_{ii'}$, $\boldsymbol{Q}_{ii'}$, $\boldsymbol{R}_{ii'}$, $\boldsymbol{T}_{ii'}$, and $\boldsymbol{I}_{ii'}$ as design matrices corresponding to these source comparisons, and $N_{ii'}$ as the cardinality of $\boldsymbol{s}_{ii'}$. We begin by defining

$$
\begin{aligned}
\pi\left(\theta_{ii'}|\boldsymbol{s}_{ii'},\widetilde{\boldsymbol{\sigma}}^2\right) \;\propto\;& f\left(\boldsymbol{s}_{ii'}|\theta_{ii'},\sigma_{ii'},\widetilde{\boldsymbol{\sigma}}^2,\phi_{ii'},\omega_{ii'}\right)\pi\left(\theta_{ii'}|\phi_{ii'},\omega_{ii'}\right) \\
=\;& \mathcal{MVN}\left(\boldsymbol{s}_{ii'}|\theta_{ii'},\sigma_{ii'},\widetilde{\boldsymbol{\sigma}}^2,\phi_{ii'},\omega_{ii'}\right)\mathcal{N}\left(\theta_{ii'}|\phi_{ii'},\omega_{ii'}\right) \\
=\;& (2\pi)^{-\frac{N_{ii'}}{2}}|\boldsymbol{\Delta}_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a+\boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b+\boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c+\boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d+\boldsymbol{I}_{ii'}\sigma^2_e\right)\boldsymbol{\Delta}'_{ii'}|^{-\frac{1}{2}} \\
& \times\exp\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'}-\theta_{kk'}\mathbf{1}_{N_{ii'}}\right)'[\boldsymbol{\Delta}_{ii'}(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a+\boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b+\boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c+\boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d \\
& +\boldsymbol{I}_{ii'}\sigma^2_e)\boldsymbol{\Delta}_{ii'}]^{-1}\left(\boldsymbol{s}_{ii'}-\theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\}\times\frac{1}{\sqrt{2\pi\omega_{ii'}}}\exp\left\{-\frac{1}{2\omega_{ii'}}\left(\theta_{ii'}-\phi_{ii'}\right)^2\right\}.
\end{aligned}
$$

From this point on, we define $\boldsymbol{\Sigma}_{ii'} := \boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma_a^2 + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma_b^2 + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma_c^2 + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma_d^2 + \boldsymbol{I}_{ii'}\sigma_e^2$ for ease of notation, and so

$$
\begin{aligned}
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Sigma}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)\right\} \times \exp\left\{-\frac{1}{2\omega_{ii'}}\left(\theta_{ii'} - \phi_{ii'}\right)^2\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right)'\boldsymbol{\Sigma}_{ii'}^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\mathbf{1}_{N_{ii'}}\right) + \frac{1}{\omega_{ii'}}\left(\theta_{ii'} - \phi_{ii'}\right)^2\right]\right\} \\
&= \exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{s}'_{ii'}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} - \boldsymbol{s}'_{ii'}\boldsymbol{\Sigma}_{ii'}^{-1}\theta_{ii'}\mathbf{1}_{N_{ii'}} - \theta_{ii'}\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \theta_{ii'}\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\theta_{ii'}\mathbf{1}_{N_{ii'}} \\
&\quad + \frac{1}{\omega_{ii'}}\theta_{ii'}^2 - \frac{2}{\omega_{ii'}}\theta_{ii'}\phi_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}^2\Bigg]\Bigg\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\theta_{ii'}\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}} + \frac{1}{\omega_{ii'}}\theta_{ii}^2\right) - 2\left(\theta_{ii'}\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\theta_{ii'}\phi_{ii'}\right) + C\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\theta_{ii'}^2\left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}} + \frac{1}{\omega_{ii'}}\right) - 2\theta_{ii'}\left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}\right) + C\right]\right\}.
\end{aligned}
$$

At this point, we define

$$
\begin{aligned}
\frac{1}{\sigma_{ii'_p}^2} = \frac{1}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}} + \frac{1}{\omega_{ii'}}} \Longrightarrow \sigma_{ii'_p}^2 &= \left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}} + \omega_{ii'}^{-1}\right)^{-1} \\
&= \frac{\omega_{ii'}}{\left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\right)\omega_{ii'} + 1},
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\mu_{ii'_p}}{\sigma_{ii'_p}^2} = \mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'} \Longrightarrow \mu_{ii'_p} &= \left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}\right)\sigma_{ii'_p}^2 \\
&= \left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}\right) \times \frac{\omega_{ii'}}{\left(\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\right)\omega_{ii'} + 1} \\
&= \left[\frac{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'} + \frac{1}{\omega_{ii'}}\phi_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1}\right]\omega_{ii'} \\
&= \frac{\mathbf{1}_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{s}_{ii'}\omega_{ii'} + \phi_{ii'}}{\mathbf{1}'_{N_{ii'}}\boldsymbol{\Sigma}_{ii'}^{-1}\mathbf{1}_{N_{ii'}}\omega_{ii'} + 1}
\end{aligned}
$$

And so $\pi\left(\theta_{ii'}|\boldsymbol{s}_{ii'}, \sigma_a^2, \sigma_e^2\right) \sim N(\mu_{ii'_p}, \sigma_{ii'_p}^2)$, with $\mu_{ii'_p}$ and $\sigma_{ii'_p}^2$ defined as above.

G.2 POSTERIOR DISTRIBUTION FOR $\sigma_{ii'}^2$ FOR THE POPULATION-BASED MODEL:

We obtain the posterior distribution of $\sigma_{ii'}^2$, $\pi(\sigma_{ii'}^2|\boldsymbol{s}_{ii'}, \theta_{ii'}, \widetilde{\boldsymbol{\sigma}}^2)$, by considering a Multivariate Normal likelihood for our vector of scores, $\boldsymbol{s}_{ii'}$, and an Inverse-Gamma prior on $\sigma_{ii'}^2$, with hyper-parameters $\alpha_{ii'}, \beta_{ii'}$. As in Appendix F.2, the subscript

$ii' \in \{kk, kp, pp, pp'\}$ indicates the type of relationship we are considering between our scores. We begin by defining

$$
\begin{aligned}
\pi(\sigma^2_{ii'}|\boldsymbol{s}_{ii'}, \theta_{ii'}, \widetilde{\boldsymbol{\sigma}}^2) &\propto f(\boldsymbol{s}_{ii'}|\sigma^2_{ii'}, \theta_{ii'}, \widetilde{\boldsymbol{\sigma}}^2, \alpha_{ii'}, \beta_{ii'})\pi(\sigma^2_{ii'}|\alpha_{ii'}, \beta_{ii'}) \\
&= \mathcal{MVN}(\boldsymbol{s}_{ii'}|\sigma^2_{ii'}, \theta_{ii'}, \widetilde{\boldsymbol{\sigma}}^2, \alpha_{ii'}, \beta_{ii'})\mathcal{IG}(\sigma^2_{ii'}|\alpha_{ii'}, \beta_{ii'}) \\
&= (2\pi)^{-\frac{N}{2}}|\boldsymbol{\Delta}_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d + \boldsymbol{I}_{ii'}\sigma^2_e\right)\boldsymbol{\Delta}'_{kk'}|^{-\frac{1}{2}} \\
&\quad \times \exp\big\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{kk'}}\right)'[\boldsymbol{\Delta}_{ii'}(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d \\
&\quad + \boldsymbol{I}_{ii'}\sigma^2_e)\boldsymbol{\Delta}'_{ii'}]^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)\big\} \times \frac{\beta_{ii'}^{\alpha_{ii'}}}{\Gamma(\alpha_{ii'})}\sigma^{2^{-(\alpha_{ii'}+1)}}_{ii'}\exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
&\propto |\sigma^2_{ii'}\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d + \boldsymbol{I}_{ii'}\sigma^2_e\right)|^{-\frac{1}{2}} \\
&\quad \times \exp\big\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)'[\sigma^2_{ii'}(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d \\
&\quad + \boldsymbol{I}_{ii'}\sigma^2_e)]^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)\big\} \times \ \sigma^{2^{-(\alpha_{ii'}+1)}}_{ii'}\exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
&= \left(\sigma^{2^{N_{ii'}}}_{ii'}|\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d + \boldsymbol{I}_{ii'}\sigma^2_e|\right)^{-\frac{1}{2}} \\
&\quad \times \exp\big\{-\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)'[\sigma^2_{ii'}(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d \\
&\quad + \boldsymbol{I}_{ii'}\sigma^2_e)]^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)\big\} \times \ \sigma^{2^{-(\alpha_{ii'}+1)}}_{ii'}\exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
&\propto \sigma^{2^{-\frac{N_{ii'}}{2}}}_{ii'}\exp\big\{-\frac{1}{2\sigma^2_{ii'}}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{ii'}\right)'\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d \right. \\
&\quad \left. + \boldsymbol{I}_{ii'}\sigma^2_e\right)^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)\big\} \times \ \sigma^{2^{-\alpha_{ii'}-1}}_{ii'}\exp\left\{-\frac{\beta_{ii'}}{\sigma^2_{ii'}}\right\} \\
&= \sigma^{2^{-\left(\frac{N_{ii'}}{2}+\alpha_{ii'}\right)-1}}_{ii'} \times \exp\big\{-\frac{1}{\sigma^2_{ii'}}[\frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)'\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c \right. \\
&\quad \left. + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d + \boldsymbol{I}_{ii'}\sigma^2_e\right)^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right) + \beta_{ii'}]\big\} \\
&\sim \mathcal{IG}\left(\alpha_{ii'_p}, \beta_{ii'_p}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_{kk'_p} &:= \frac{N_{ii'}}{2} + \alpha_{ii'} \\
\beta_{ii'_p} &:= \frac{1}{2}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right)'\left(\boldsymbol{P}_{ii'}\boldsymbol{P}'_{ii'}\sigma^2_a + \boldsymbol{Q}_{ii'}\boldsymbol{Q}'_{ii'}\sigma^2_b + \boldsymbol{R}_{ii'}\boldsymbol{R}'_{ii'}\sigma^2_c + \boldsymbol{T}_{ii'}\boldsymbol{T}'_{ii'}\sigma^2_d + \boldsymbol{I}_{ii'}\sigma^2_e\right)^{-1}\left(\boldsymbol{s}_{ii'} - \theta_{ii'}\boldsymbol{1}_{N_{ii'}}\right) + \beta_{ii'}.
\end{aligned}
$$

REFERENCES

[1] C.C. Aggarwal. *Data Mining.* Springer International Publishing, 2015.

[2] C.G.G. Aitken and F. Taroni. *Evaluation of Evidence for Forensic Scientists.* 2^{nd}. Wiley and Sons Ltd, Chichester, 2004, p. 540.

[3] M.A. Aizerman, E.M. Braverman, and L.I. Rozonoer. "Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning". In: *Automation and Remote Control* 26.6 (1964), pp. 821–837.

[4] T.W. Anderson and G. P. H. Styan. *Cochran's Theorem, Rank Additivity, and Tripotent Matrices.* Tech. rep. Department of Statistics, Standford University, 1980.

[5] D. Armstrong. "Development and Properties of Kernel-Based Methods for the Interpretation and Presentation of Forensic Evidence". PhD thesis. South Dakota State University, 2018.

[6] D. Armstrong et al. "Kernel-based methods for source identification using very small particles from carpet fibers". In: *Chemometrics and Intelligent Laboratory Systems* 160 (2017), pp. 99–109.

[7] M. Ausdemore et al. "Two-Stage Approach for the Inference of the Source of High-Dimensional and Complex Chemical Data in Forensic Science". In: *Journal of Chemometrics* 35.1 (2020).

[8] G. Baudat and F. Anouar. "Generalized discriminant analysis using a kernel approach". In: *Neural Computation* 12 (2000), pp. 2385–2404.

[9] C. Berg, J.P.R. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups.* New York: Springer, 1984.

[10] Alphonse Bertillon. "Archives de l'anthropologie criminelle et des sciences pénales". In: vol. 1. 1886. Chap. De l'identification par les signalements anthropométriques, pp. 193–223.

[11] A. Biedermann, S. Bozza, and F. Taroni. "Decision theoretic properties of forensic identification: underlying logic and argumentative implications". In: *Forensic science international* 177 (2008), pp. 120–132.

[12] C.M. Bishop. *Pattern Recognition and Machine Learning.* Springer Texts in Statistics, 2006.

[13] B.E. Boser, I.M. Guyon, and Vladimir N. Vapnik. "A Training Algorithm for Optimal Margin Classifiers". In: *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory.* ACM Press, 1992, pp. 144–152.

[14] S. Brooks et al., eds. *Handbook of Markov Chain Monte Carlo.* Taylor and Francis Group, LLC, 2011.

[15] M. Capiński and E. Kopp. *Measure, Integral and Probability.* Second. Springer, 2008.

[16] W.G. Cochran. "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 30.2 (1934), pp. 178–191.

[17] C. Cortes and Vladimir N. Vapnik. "Support vector networks". In: *Machine Learning* 20 (1995), pp. 273–297.

[18] X. Didelot et al. "Likelihood-free estimation of model evidence". In: *Bayesian Analysis* 6.1 (2011), pp. 49–76.

[19] Alfred Dreyfus. "La révision du procs̀ Dreyfus - Enquête de la Cour de cassation". In: Paris: P.-V. Stock, 1899. Chap. Déposition Bertillon (du 18 janvier, 4 février et 6 février 1899), pp. 482–500.

[20] N.M. Egli-Anthonioz and C. Champod. "Evidence evaluation in fingerprint comparison and automated fingerprint identification systems - modeling between finger variability". In: *Forensic Science International* 235 (2014), pp. 86–101.

[21] N.M. Egli, C. Champod, and P. Margot. "Evidence evaluation in fingerprint comparison and automated fingerprint identification systems - modelling within finger variability". In: *Forensic Science International* 176 (2006), pp. 189–195.

[22] *ENFSI Guideline for Evaluative Reporting in Forensic Science.* http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf. European Network of Forensic Science Institutes (ENFSI). 2016.

[23] I.W. Evett. "Towards a Uniform Framework for Reporting Opinions in Forensic Science Casework". In: *Science & Justice* 38(3) (1998), pp. 1198–202.

[24] T. Evgeniou, M. Pontil, and T. Poggio. *A Unified Framework for Regularization Networks and Support Vector Machines.* Tech. rep. Cambridge, MA: Massachusetts Institute of Technology, 1999.

[25] G. Flake and C. Lawrence. "Efficient SVM regression training with SMO". In: *Machine Learning* (2002).

[26] D. Gantz and C. Saunders. *Quantifying the effects of database size and sample quality on measures of individualization validity and accuracy in forensics.* Tech. rep. Department of Justice, Award No. 2009-DN-BX-K234, 2014.

[27] J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, and J. Ortega-Garcia. "Forensic identification reporting using automatic speaker recognition systems". In: *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)* (2003), pp. 11–93.

[28] J. Gonzalez-Rodriguez et al. "Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems". In: *Forensic Science International* 155 (2005), pp. 126–140.

[29] J. Gonzalez-Rodriguez et al. “Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition”. In: *Computer Speech and Language* 20 (2006), pp. 331–355.

[30] I.J. Good. *Probability and the Weighting of Evidence*. Charles Griffin & Co., London, U.K., 1950.

[31] J. Hendricks, C. Neumann, and C. Saunders. *A ROC-based Approximate Bayesian Computation algorithm for model selection: application to fingerprint comparisons in forensic science*. Tech. rep. 2019.

[32] D. Hilbert. “Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen.” In: *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische* 1 (1904), pp. 49–91.

[33] T.B. Hoffman, B. Scholköpf, and A.J. Smola. “A review of kernel methods in machine learning”. In: *The Annals of Statistics* 36.3 (2008), pp. 1171–1220.

[34] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003, p. 727.

[35] H. Jeffreys. *Theory of Probability*. Third. Oxford University Press, 1961, p. 470.

[36] T. Joachims. “Advances in Kernel Methods - Support Vector Learning”. In: ed. by B. Schölkopt, C.J.C. Burges, and A.J. Smola. MIT Press, 1999. Chap. Making large-scale SVM learning practical.

[37] T. Joachims. “Transductive inference for text classification using support vector machines”. In: *Proceedings of the 16th International Conference on Machine Learning*. Ed. by Morgan Kaufmann. San Francisco, CA, 1999, pp. 200–209.

[38] Alexandros Karatzoglou et al. “kernlab – An S4 Package for Kernel Methods in R”. In: *Journal of Statistical Software* 11.9 (2004), 1:20.

[39] Yann LeCun, Corinna Cortes, and Christopher J.C. Burges. *MNIST handwritten digit database*. http://yann.lecun.com/exdb/mnist/. 2010.

[40] Christina Leslie, Eleazar Eskin, and William Stafford Noble. “The Spectrum Kernel: A String Kernel for SVM Protein Cassification”. In: *Proceedings of the Pacific Symposium on Biocomputing, 2002* (2002), pp. 564–575.

[41] Christina Leslie and Rui Kuang. “Fast kernels for inexact string matching”. In: *Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003*. 2003.

[42] D.V. Lindley. *Understanding Uncertainty*. John Wiley & Sons, Inc., 2006.

[43] William G. Madow. “The Distribution of Quadratic Forms in Non-Central Normal Random Variables”. In: *The Annals of Mathematical Statistics* 11.1 (1940), pp. 100–103.

[44] Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Taylor and Francis Group, LLC, 2016.

[45] J. Mercer. “Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations”. In: *Philosophical Transactions of the Royal Society of London. Series A. Containing Papers of a Mathematical or Phiscal Character* 209 (1909), pp. 415–446.

[46] D Meuwly. “Forensic Individualisation from Biometric Data”. In: *Science and Justice* 46 (2006), pp. 205–213.

[47] D. Meuwly and A. Drygajlo. “Forensic Speaker Recognition based on a Bayesian Framework and Gaussian Mixture Modeling”. In: *Proc. of Odyssey 2001 Speaker Recognition Workshop, Crete (Greece)* (2001).

[48] S. Mika et al. “Fisher discriminant analysis with kernels”. In: *Neural Networks for Signal Processings IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop*. Ed. by Y.-H. Hu et al. 1999, pp. 41–48.

[49] C. Neumann and M. Ausdemore. “Defence Against the Modern Arts: the Curse of Statistics "Score-based likelihood ratios"”. In: (2019).

[50] C. Neumann, I.W. Evett, and J.E. Skerrett. “Quantifying the weight of evidence from a forensic fingerprint comparison: A new paradigm”. In: *Journal of the Royal Statistical Society (Series A)* 175 (2012), pp. 1–26.

[51] C. Neumann and P.M. Margot. “New perspectives in the use of ink evidence in forensic science: Part III: Operational applications and evaluation”. In: *Forensic Science International* 192 (2009), pp. 29–42.

[52] C. Neumann et al. “Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae”. In: *Journal of Forensic Sciences* 52 (2007), pp. 54–64.

[53] Austin O’Brien. “A Kernel Based Approach to Determine Atypicality”. PhD thesis. South Dakota State University, 2017.

[54] T. Ogasawara and M. Takahashi. “Independence of quadratic quantitiesl in a normal system”. In: *J. Sci. Hiroshima University A* 15 (1951), pp. 1–9.

[55] J. Ogawa. “On the independence of statistics of quadratic forms. (In Japanese)”. In: *Notes on the Institue of Statistical Mathematics*. Vol. 2. 1946, pp. 98–111.

[56] J. Ogawa. “On the independence of statistics of quadratic forms. (Translation of Ogawa, 1946)”. In: *Notes on the Institue of Statistical Mathematics*. Vol. 3. 1947, pp. 137–151.

[57] N. Oliver, B. Scholköpf, and A.J. Smola. “Natural Regularization from Generative Models”. In: *Advances in Large Margin Classifiers*. Ed. by A.J. Smola et al. Cambridge, MA: MIT Press, 2000, pp. 51–60.

[58] D.M. Ommen and C.P. Saunders. “Reconciling the Bayes Factor and Likelihood Ratio for Two Non-Nested Model Selection Problems”. In: *https://arxiv.org/pdf/1901.* (2019).

[59] D.M. Ommen, C.P. Saunders, and C. Neumann. “The characterization of Monte Carlo errors for the quantification of the value of forensic evidence”. In: *Journal of Statistical Computation and Simulation* 87.8 (2017), pp. 1608–1643.

[60] J.B. Parker. “A statistical treatment of identification problems”. In: *Journal of FSSoc* 6 (1966), pp. 33–39.

[61] J.B. Parker. “The mathematical evaluation of numerical evidence”. In: *Journal of FSSoc* 7 (1967), pp. 134–144.

[62] J.B. Parker and A. Holford. “Optimum Test Statistics with Particular Reference to a Forensic Science Problem”. In: *Journal of the Royal Statistical Society. Series C* 17.3 (1968), pp. 237–251.

[63] J. Platt. “Advances in Kernel Methods - Support Vector Learning”. In: ed. by B. Schölkopt, C.J.C. Burges, and A.J. Smola. MIT Press, 1999. Chap. Fast training of support vector machines using sequential minimal optimization, pp. 185–208.

[64] Damoulas T. Psorakis I. and M.A. Girolami. “Multiclass relevance vector machines: sparsity and accuracy”. In: *IEEE Transactions on Neural Networks* 21.10 (2010).

[65] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[66] C. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Second. Springer Texts in Statistics, 2007.

[67] Cornuet J. Robert C. and N. Pillai. “Lack of confidence in abc model choice”. In: *Proceedings of the National Academy of Sciences in the United States of America* 108.37 (2011), pp. 5112–5117.

[68] J. Robins and L. Wasserman. “Conditioning, Likelihood, and Coherence: A Review of Some Fundamental Concepts”. In: *Journal of the American Statistical Association* (2000).

[69] F. Rosenblatt. “The perceptron: A probabilistic model for information storage and organization in the brain”. In: *Psychological Review* 65.6 (1958), pp. 386–408.

[70] V. Roth and V. Steinhage. “Nonlinear discriminant analysis using kernel functions”. In: *Advance in Neural Information Processing Systems 12*. Ed. by A. Solla, T.K. Leen, and K.-R. Mu. MIT Press, 2000, pp. 568–574.

[71] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd ed. Pearson Education, Inc., 2010.

[72] L.J. Savage. *The Foundations of Statistics*. 2^{nd}. Dover Publications, Inc. New York., 1972, p. 310.

[73] M Schmidt and H. Gish. “Speaker identification via support vector classifiers”. In: *Proceedings ICASSP’96*. 1996, pp. 105–108.

[74] B. Scholköpf and A.J. Smola. *Learning with Kernels*. The MIT Press, 2002.

[75] B. Scholköpf, A.J. Smola, and Klaus-Robert Müller. *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*. Tech. rep. 44. Bülthoff, Tübingen: Max-Planck-Institut für biologische Kybernetik Arbeitsgruppe, 1996.

[76] B. Scholköpf, A.J. Smola, and Klaus-Robert Müller. “Kernel Principal Component Analysis”. In: *Proceedings of the 7th International Conference on Artificial Neural Networks*. ICANN ’97. Berlin, Heidelberg: Springer-Verlag, 1997, pp. 583–588.

[77] A.J. Smola and B. Scholköpf. “On a Kernel-Based Method for Pattern Recognition, Regression, Approximation, and Operator Inversion”. In: *Algorithmica* 22.1-2 (1998), pp. 211–231.

[78] A.J. Smola, B. Scholköpf, and Klaus-Robert Müller. “The Connection Between Regularization Operators and Support Vector Kernels”. In: *Neural Networks* 11.4 (1998), pp. 637–649.

[79] I. Steinwart. “Support Vector Machines and Universally Consistent”. In: *Journal of Complexity* 18.3 (2002), pp. 768–791.

[80] G.W. Stewart. “Fredholm, Hilbert, Schmidt. Three Fundamental Papers on Integral Equations.” Translated with commentary. 2001.

[81] M.O. Stitson et al. *Support Vector Regression with ANOVA Decomposition Kernels*. Tech. rep. CSD-TR-97-22. Department of Computer Science, Egham, Surrey TW20, 0EX, England: Royal Holloway University of London, 1997.

[82] H. Swofford et al. “A method for the statistical interpretation of friction ridge skin impression evidence: Method development and validation”. In: *Forensic Science International* 287 (2018), pp. 113–126.

[83] F. Taroni, C. Champod, and P. Margot. “Forerunners of Bayesianism in Early Forensic Science”. In: *Jurimetrics* 38 (1998), pp. 183–200.

[84] F. Taroni et al. “Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio”. In: *Law, Probability and Risk* 15.1 (2016), pp. 1–16.

[85] M. Tipping. “Sparse bayesian learning and the relevance vector machine”. In: *Jorurnal of Machine Learning Research* 38 (2001), pp. 183–200.

[86] V. Vapnik and A. Chervonenkis. “A note on one class of perceptrons”. In: *Automation and Remote Control* 25 (1964).

[87] V. Vapnik and A. Chervonenkis. *Theory of Pattern Recognition [in Russian]*. Nauka, Moscow, 1974.

[88] V. Vapnik and A. Lerner. “Pattern recognition using generalized portrait method”. In: *Automation and Remote Control* 24 (1963), pp. 774–780.

[89] Vladimir N. Vapnik. *Statistical Leaning Theory*. John Wiley & Sons, Inc., 1998.

[90] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. 2nd ed. Springer Science + Business Media, 2000.

[91] Jean-Philippe Vert, Koji Tsuda, and Bernhard Schölkopt. “Kernel Methods in Computational Biology”. In: MIT Press, 2004. Chap. A Primer on Kernel Methods, pp. 35–70.