

SEMI-NONPARAMETRIC VARYING COEFFICIENTS MODELS FOR IMAGING GENETICS

Ting Li¹, Yang Yu², Xiao Wang³, J.S. Marron² and Hongtu Zhu^{*2}

¹*Shanghai University of Finance and Economics,*

²*University of North Carolina at Chapel Hill and* ³*Purdue University*

Abstract: Motivated by imaging genetics, this paper introduces a semi-nonparametric varying coefficients modeling framework to reveal varying associations between genetic markers and imaging responses. We aim to conduct a comprehensive theoretical analysis of estimation and inference procedures applicable to these models. By employing the kernel machine method, we estimate unknown varying coefficient functions and derive their representer theorem. We also establish the theoretical properties of these estimated functions, including their rate of convergence, Bahadur representation, point-wise limit distributions, and confidence intervals. Additionally, we propose test statistics under a linear mixed effects model framework to assess the significance of all varying coefficients, taking into account within-subject dependence. The efficacy of our proposed methodology is demonstrated through simulation studies and an application to data from the Alzheimer's Disease Neuroimaging Initiative study.

Key words and phrases: Imaging genetics, kernel ridge regression, linear mixed effects model, reproducing kernel Hilbert space.

1. Introduction

Advancements in modern imaging and genetic techniques have facilitated large-scale neuroimaging genetics studies aimed at understanding the genetic foundations of human brain structure and function. Prominent examples include the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Mueller et al., 2005) and the UK Biobank (UKB) study (Miller et al., 2016). Imaging traits are increasingly used as biomarkers for diagnosis and prognosis, as well as endophenotypes to identify genetic markers linked to various brain-related disorders (Elliott et al., 2018; Zhao et al., 2021). These traits provide critical insights into the biological pathways that connect genetics with imaging characteristics and brain-related disorders which are confounded with health factors such as diet and alcohol. Nonetheless, the combined analysis of imaging and genetic data poses significant challenges to current statistical methods due to their high dimensionality and intricate spatio-temporal structures (Zhu, Li and Zhao, 2023).

A compelling data example that motivates our proposed method is derived

*Corresponding author. E-mail: htzhu@email.unc.edu