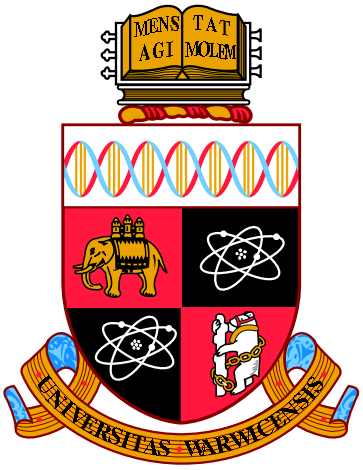


Computationally Efficient Mixed Effect Model for Genetic Analysis of High Dimensional Neuroimaging Data

by

Habib Ganjgahi

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

November 2016

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Thomas E. Nichols for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and personal life. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would also like to thank those at the Department of Statistics and the University of Warwick for always providing help when needed. Also, special thanks go out to all the past and present members of the Neuroimaging group. Without their cheerful attitude, friendship, my time at Warwick would have been a very lonely experience.

Last but not the least, I would like to thank my family: my parents and to my brother for supporting me spiritually throughout writing this thesis and the unconditional love through my life in general.

Declarations

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in these, or any other Universities. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

- The work in Chapter 3 has been published in Neuroimage (Ganjgahi *et al.*, 2015) without revision.
- The work in Chapter 4 is based on a manuscript which is currently under a revision for submission.
- The work in Chapter 5 will shortly be submitted for publication.

Habib Ganjgahi 2016

Abstract

A new research direction in the neuroimaging discipline, so called imaging genetic, has emerged recently concerns describing individual differences in imaging phenotypes using genetic and environmental factors. The large number of voxel- and vertex-wise measurements in imaging genetics studies present a challenge both in terms of computational intensity and the need to account for elevated false positive risk because of the multiple testing problem. There is a gap in existing tools, as standard neuroimaging software cannot perform essential genetic analyses including heritability and association estimations and testings, and yet standard quantitative genetics tools cannot provide essential neuroimaging inferences, like family-wise error corrected voxel- wise or cluster-wise P-values. Moreover, available genetic tools rely on P-values that can be inaccurate with usual parametric inference methods.

In this thesis computationally efficient linear mixed effect model for voxel-wise genetic analyses of high-dimensional imaging phenotypes are developed. Specifically, fast estimation and inference procedures for heritability and association analyses are introduced using orthogonal transformations that dramatically simplify the likelihood and restricted likelihood functions of mixed effect model. We review the family of score, likelihood ratio and Wald tests and propose novel inference methods for fixed and random effect terms in the mixed effect models. To address problems with inaccuracies with the standard results used to find P-values, we propose different permutation schemes to allow semi-parametric inference (parametric likelihood-based estimation, non-parametric sampling distribution). In total, we evaluate different significance tests for heritability and association, with either asymptotic parametric or permutation-based P-value computations. We identify a number of tests that are both computationally efficient and powerful, making them ideal candidates for heritability and genome-wide association studies in the massive data setting.

Chapter 1

Introduction

The discipline of neuroimaging comprises various techniques for recording the structure and function of the living human brain. While some methods use electrical or magnetic signals emitted from the brain through the scalp EEG/MEG (Electroencephalography/Magnetoencephalography), and others rely on injection of radioactive tracers (PET) (Positron Emission Tomography), the most widely used methods are based on MRI (Magnetic Resonance Imaging). MRI facilitates non-invasive 3-dimensional imaging of the brain based on the local magnetic properties of hydrogen atoms. Functional MRI (fMRI) has become an indispensable tool for cognitive psychologists to map the brain regions responsible for basic behaviour and thought. Structural MRI provides high resolution images of the convoluted gray matter, white matter and subcortical structures that construct the brain; neurologists use subtle changes in gray matter to track the progression of Alzheimer's disease and other disorders. And Diffusion Tensor Imaging (DTI) provides unique information on the white matter pathways that connect different parts of cortex.

As with any biomedical measurement, the overall goal of neuroimaging is to understand any differences between populations of subjects and variation within these populations. In the last 5 years there has been considerable interest in explaining such variation with genetic markers, treating the brain imaging measure as a phenotype (see, e.g., Glahn *et al.* (2007)). A phenotype is strictly defined as a *heritable trait* where a trait can be defined as an observable physical or biochemical characteristic of an organism. In this sense, Heritability is the proportion of trait variance that can be explained by genetic sources (a formal definition is provided in Chapter 2). Significant and reproducible heritability has been established for many neuroimaging traits assessing brain structure and function, including, for instance, location and strength of task-related brain activation (Blokland *et al.*, 2008;

Koten *et al.*, 2009; Matthews *et al.*, 2007; Polk *et al.*, 2007), white matter integrity (Kochunov *et al.*, 2014a,b; Jahanshad *et al.*, 2013; Brouwer *et al.*, 2010; Chiang *et al.*, 2009, 2011; Kochunov *et al.*, 2010), cortical and subcortical volumes, cortical thickness and density (Winkler *et al.*, 2010; Rimol *et al.*, 2010; Kochunov *et al.*, 011a,b; Kremen *et al.*, 2010; den Braber *et al.*, 2013). The aim of any heritability study is to find aspects of the human brain structure and function that is under the overall genetic control using the expected genetic similarity among different types of related and unrelated individuals.

There has been growing interest in the field of imaging genetic to move from establishing heritable phenotypes to finding genetic variants that influence brain structure and function in order to better understand the biological basis of neurological and psychiatric illnesses in patients and healthy individuals (Hibar *et al.*, 2015; Stein *et al.*, 2012, 2010a,b; Potkin *et al.*, 2009b,a). Association studies address the effect of genetic variants including genes, single nucleotide polymorphisms (SNPs) or copy number variations (CNVs), on a variety of functional and structural brain imaging phenotypes. In brain imaging, evidence has accumulated over the past decade, showing that certain brain-relevant genes have an influence on brain structure and function. For example, the Alzheimers disease risk apolipoprotein E epsilon-4 allele (ApoE4) associated with reduced grey matter (Burggren *et al.*, 2008; Pievani *et al.*, 2009) and white matter volumes (Hua *et al.*, 2008). However, the proportion of the variance in imaging phenotypes explained by these variants is generally very small, leaving a large proportion of the heritability of imaging phenotypes unaccounted for. Thus there remains intense interest to discover more genes that influence brain structure and function.

Variance component models are the best-practice approach for deriving heritability estimates based on familial data (Almasy and Blangero, 1998; Blangero and Almasy, 1997; Amos, 1994; Hopper and Mathews, 1982), for allowing great flexibility in modeling of genetic additive and dominance effects, as well as common and unique environmental influences. Estimation of parameters typically uses maximum likelihood under the assumption that the additive error follows a multivariate normal distribution. The iterative optimization of the likelihood function requires computationally intensive procedures, that are prone to convergence failures, something particularly problematic when fitting data at every voxel/element; both the computation burden and the algorithmic fragility pose significant problems when applied to 100's of 1,000's of voxels.

Typically a likelihood ratio test (LRT) is used for heritability hypothesis testing. As the null hypothesis value is on the boundary of the parameter space, the

asymptotic distribution of LRT is not χ^2 with 1 degree of freedom (DF), but rather approximately as a $50:50$ mixture of χ^2 distributions with 1 and 0 DF, where a 0 DF χ^2 is a point mass at 0 (Chernoff, 1954; Self and Liang, 1987; Stram and Lee, 1994; Dominicus *et al.*, 2006; Verbeke and Molenberghs, 2003). As with most statistical models, the quantitative genetic models used here are based on an assumption of multivariate Gaussianity, and this assumption is the basis of the estimation and hypothesis testing. However, the heritability test statistic's null distribution may be inaccurate even when Gaussianity is perfectly satisfied, due to the limitations of the $50:50$ χ^2 result (see section 2 for more details). Hence, there is a compelling need for alternative inference procedures that provide valid inference.

The genetic association analysis with the quantitative phenotypes from structural (i.e. brain volume, cortical thickness, white matter integrity) or functional imaging modalities (brain response to particular cognitive task or resting state) at hundred thousand locations in the human brain present statistical challenges including computational intensity, correction for population structure, statistical power and intense multiple comparisons correction.

While genomic data can be used to control for population stratification (formal definition is provided in chapter 2) by including the top principal components as a fixed effect covariates in a linear regression model (Price *et al.*, 2006), usually individuals with close estimated relatedness from identity-by-state (IBS) matrix or different ancestry are excluded from the study sample. This might not be a problem in genetic studies with 4 digits sample sizes, but may make substantial differences in genome-wide association (GWA) studies with neuroimaging phenotypes where sample size is much smaller. Moreover, GWA studies with neuroimaging phenotypes require fitting a marginal model at each point (voxel/element) in the brain leading to a large number of measurements which presents a challenge both in terms of computational intensity and the need to account for elevated false positive risk because of the multiple testing problems regarding both the number of elements and the number of markers being tested.

Although the emergence of large scale neuroimaging consortia like ENIGMA or CHARGE can help to conduct well-powered genetic association studies through meta analysis framework, it is still/yet crucial to use a powerful statistical method at the site level. There is, therefore, a compelling need for a analytical technique that addresses these challenges.

The linear mixed effect model (LMM) efficiency in controlling population structure in the genetic association analysis and possible boost in power inspires using it with high-dimensional imaging phenotypes. However fitting LMM at each

voxel/ROI in the brain is computationally intensive or even intractable at the voxel level while variance component estimation relies on likelihood function optimisation using numerical methods. Despite many analytical techniques being developed to accelerate the GWA with LMM, these advances do not eliminate problems related to numerical optimisation nor multiple testing problem.

In current practice, practitioners usually extract measures from imaging data, either voxel-by-voxel or by regions of interest (ROI), and submit the data, one voxel or ROI at a time, to widely used quantitative genetics software. A limitation of many neuroimaging heritability and gene-finding studies is the reliance on ROI's. While ROI's simplify the analysis by reducing the high dimensional image data to a few numbers, ROI definitions are problematic. Usually a standard atlas specifies the ROI's, but individual differences in brain shape or deficiencies in the atlas can result in ROI's missing the targeted brain structure. This motivates the use of voxel based genetic analyses. The opportunity is that, by working with images as whole (instead of voxel-by-voxel), we can consider spatially informed statistics , like cluster size, that will allow greater power to detect low levels of heritability and adjustable for family wise error rate.

While the statistical methods for estimating heritability on univariate traits are well-established (Almasy and Blangero, 1998; Blangero and Almasy, 1997; Amos, 1994; Hopper and Mathews, 1982), at present there are no established methods to estimate heritability for imaging data and provide the standard neuroimaging inferences, like voxel-wise and cluster-wise P-values, with family-wise error correction for searching the brain for significance. Although random field theory (Worsley *et al.*, 1992; Friston *et al.*, 1994b; Nichols and Hayasaka, 2003) results exist for χ^2 images (Cao, 1999), they are not directly applicable here as the test statistic image can not be expressed as a linear combination of component error fields.

Search for genetic association across the genome at different locations with imaging phenotypes requires intense multiple testing corrections both for number of elements in an image and number of markers being tested. Whether the association analysis is conducted at the reduced search space in the brain i.e., summary measure from a region of interest or voxel level, naive application of bonferroni correction for number of hypothesis testing in the image with usual GWA P-value leads to invalid statistical inference procedure while it ignores complex spatial dependence between elements in the imaging phenotypes. Moreover, parametric null distribution of cluster size (Friston *et al.*, 1994b) or threshold free cluster enhancement (TFCE) statistics (Smith and Nichols, 2009) that are the most common and sensitive inference tools in imaging, could be invalid due to untenable stationary assumption or in

the latter case could be unknown. Familywise error rate (FWE) correction, controlling the chance of one or more false positives across the whole set (family) of tests (Nichols and Hayasaka, 2003) requires the distribution the maximum statistic, can be computed for either voxels/ROI or cluster size with permutation test (Nichols and Holmes, 2002).

The purpose of this work is to propose computationally efficient linear mixed effect model that provides fast methods for heritability and association analyses that are specifically oriented toward brain image data, explicitly accounting for the multiple testing problem. The remainder of this thesis is organised as follows:

- **Background**: This section first provides a review background on heritability, genome-wide association analyses methods that are largely drawn from quantitative genetic field. and then, describes neuroimaging phenotypes, fundamentals of permutation test and multiple-testing corrections following standard neuroimaging spatial statistics. This work is the basis of Ganjgahi *et al.* (2015).

- **Heritability**: In this chapter, we draw on recent results that simplify heritability likelihood computations, converting a correlated data problem into an independent but heteroscedastic one. A suite of non-iterative estimation methods are proposed that are so fast as to be amenable to permutation, and thus allow arbitrary statistics like maximum voxel-wise and cluster-wise statistics, which provide spatial family-wise error control needed in brain imaging. Comprehensive simulation studies are conducted to compare the proposed methods. Real data analysis is also included.

- **Genome-wide association analysis**: In this chapter, two major contributions are introduced to reduce the complexity of LMM in the genetic association specifically with the imaging phenotypes. First, variance component estimation step computational cost is reduced with building more accurate 1-step random effect estimator. Second, complexity of association testing is dramatically decreased with projecting the model and phenotype to a lower dimension space. The score, LRT and Wald statistics performance for hypothesis testing based on the permutation test and parametric framework are compared using simulation studies. Real data analysis is provided for the evaluation purposes.

- **Bivariate genetic modelling**: This chapter introduces a method to accelerate bivariate LMM likelihood function for genetic correlation and association

estimation and testing.

- **Conclusions**: The final part comprises of a presentation of main findings and open problems.

Chapter 2

Background

In this chapter we review basic genetic analyses including heritability and association estimation and testing methods that form the core methodology of quantitative genetics, followed by a brief review of the neuroimaging phenotypes. Next we review the neuroimaging spatial statistic, family wise error correction and use of permutation test to perform valid statistical inference.

2.1 Basic Genetic Concepts

All known living organisms contain long chains of *Deoxyribonucleic acid* (DNA) that encode instructions for cell development and function. Each molecule of DNA can bind to one of four nucleobases, guanine, adenine, thymine, and cytosine, recorded using the letters G, A, T, and C. Each DNA molecule consist of two strands that are made stable by the complementary bonding of the nucleobases (G-C, A-T). DNA is folded into elongated structures called *chromosomes*. The set of chromosomes in a cell makes up its *genome*, and the human genome consists of 23 chromosome pairs, the mother and father contributing each one chromosome to each pair. Sections of DNA that are transcribed into proteins are called *genes*. Different variants of a gene are called *alleles*. A region of a chromosome at which a particular gene is located is called its *locus*.

Modern quantitative genetics is established based on the foundational work of Mendal, and grounded on two fundamental laws.

- The *Law of segregation* states that every individual holds a pair of allele for a particular trait which is inherited randomly from their parents, one from the father (paternal) and the other from the mother (maternal). Each pair of alleles at a locus influences how the individual expresses the corresponding

trait. The influence can be *dominant*, *recessive* or *additive*. Consider a gene with only two possible alleles, denoted "A" and "a". An individual has one paternal allele and one maternal allele and thus has 3 possible *genotypes*, aa, Aa and AA. If a genetic effect only occurs with aA or AA, we say there is a dominant genetic effect of A; "a" is said to be *recessive*. If the effect changes incrementally with the count of an allele (e.g. 0, 1, or 2 copies of A), we say there is an additive genetic effect.

- According to the *Law of independent assortment* different genes are transferred independently from parents to their offspring. This law is also called inheritance law. This law asserts that alleles of different genes get shuffled between parents to form offspring with many different combinations of the genes. This law is a basis for *linkage analysis*, that aims to identify the loci in a chromosome related to a given trait. For two genes that are close together on chromosome this law can be violated, and in such cases the two loci are said to be in *linkage disequilibrium* (Reich *et al.* (2001)).

2.2 Heritability Studies

Genetic and environmental factors as well as measurement error can explain the trait variation of a trait in a population. For instance, some humans are taller or shorter than the others and this variation may be due to genetic factors (e.g. tall parents), environmental factors (e.g. ample nutrition as a child), and random error (e.g. imperfect reproducibility of height measurements). *Heritability* is defined as the proportion of phenotypic variance that can be explained by genetic sources. A genetic effect can be dominant or additive [1]. If the genetic effect for a trait changes uniformly with an allele frequency, we say there is an additive genetic otherwise we call it dominant genetic effect. Two kinds of heritability have been defined, broad sense and narrow sense heritability. Broad sense heritability refers to all genetic variance, both additive and dominant, in the phenotype. However, narrow-sense heritability concerns only the variation explained by additive genetic sources.

2.2.1 Heritability Estimation

Variance component models have been used in behavioural genetics studies since 1980's for linkage analysis and heritability estimation (Almasy and Blangero, 1998;

[1] It can also be recessive, but a trait that is recessive for "a" can equivalently be described as dominant for "A".

Blangero and Almasy, 1997; Amos, 1994; Hopper and Mathews, 1982). In these studies typically a phenotype covariance is decomposed in terms of (1) an additive genetic effect, (2) environmental factors common to the family, like socioeconomic status, and (3) individual environmental factors and measurement error. This model is known as ACE, for Additive genetic variation, Common environment and Error. The 'C' effect is well defined in twin studies, where each twin pair experiences a common upbringing (excluding the rare case of twins separated at birth). In the case of family studies, common environment is harder to define (surely twin pairs have a more similar environment than a pair of siblings different, say, by 10 years); at most the C effect in family studies is defined as a household effect. In practice, even a household effect in family studies is often negligible (Blangero *et al.*, 2000), and so standard practice is to neglect the C term and only consider a AE model (twin studies, though, use ACE by default).

For an AE model the phenotype covariance matrix is decomposed in to two components, one for the additive genetic effect and one for the combination of individual-specific environmental effects and measurement error. We use a linear model for the phenotype Y measured on N individuals,

$$Y = X\beta + g + \epsilon, \tag{2.1}$$

where X is an $N \times p$ matrix of covariates, like age, gender, etc, β is the p-vector of regression coefficients, g is the N-vector of latent additive genetic effects and ϵ is the N-vector of residual errors. Then the trait covariance, $\text{cov}(Y) = \text{cov}(g + \epsilon) = \Sigma$ can be written as

$$\Sigma = 2\sigma_A^2 \Phi + \sigma_E^2 I, \tag{2.2}$$

where Φ is the kinship matrix that captures family resemblance, σ_A^2 and σ_E^2 are the additive genetic and the environmental variance components, respectively, and I is the identity matrix. A kinship matrix codes the relatedness between pairs of individuals; twice the kinship coefficient is the expected proportion of genetic material shared between each pair of individuals (Lange, 2003). In this model narrow sense heritability is $h^2 = \sigma_A^2/\sigma_p^2$, where $\sigma_p^2 = \sigma_A^2 + \sigma_E^2$ is the trait variance. Maximum likelihood is used for parameter estimation with the assumption that the data follows a multivariate normal distribution. One of the advantages of this model is that covariate effects can be incorporated in X, reducing the unexplained phenotypic variance and potentially increasing the power to detect non-zero heritability.

2.2.2 Heritability Hypothesis Testing with Likelihood Ratio Test

Typically a likelihood ratio test (LRT) is used for heritability hypothesis testing. As the null hypothesis value is on the boundary of the parameter space, the asymptotic distribution of LRT is not χ^2 with 1 degree of freedom (DF), but rather approximately as a $50:50$ mixture of χ^2 distributions with 1 and 0 DF, where a 0 DF χ^2 is a point mass at 0 (Chernoff, 1954; Self and Liang, 1987; Stram and Lee, 1994; Dominicus *et al.*, 2006; Verbeke and Molenberghs, 2003). However, this result depends on the assumption of independent and identically distributed (i.i.d.) data (Crainiceanu, 2008; Crainiceanu and Ruppert, 2004c,b,a), which is violated in the heritability problem. It has been shown that 0 values occur at a rate greater than 50%, producing conservative inferences (Blangero *et al.*, 2013; Crainiceanu and Ruppert, 2004c; Shephard, 1993; Shephard and Harvey, 1990).

2.3 Genome-wide Association Studies

Genome-wide association analysis concerns finding genetic variants, usually single nucleotide polymorphism (SNP), that correlates with quantitative traits in a sample of unrelated individuals from a population. A number of issues complicate the interpretation of GWA studies, including cryptic relatedness and population stratification. Cryptic relatedness refers to the presence of unknown genetic relationships between individuals and violates the independence assumption typically made for a sample (Voight and Pritchard, 2005; Weir *et al.*, 2006). Population stratification is where subjects from different populations are included in a study and can lead to false positive associations. These two problems have been studied thoroughly (Pritchard *et al.*, 2000; Cardon and Palmer, 2003; Helgason *et al.*, 2005; Balding, 2006; Price *et al.*, 2010). Even in a carefully design GWA study, it is hard to avoid spurious associations because of population structure, a term that encompasses cryptic/family relatedness and population stratification; in particular it is likely that in studies with large sample sizes some level of population structure are induced within a same population.

Statistical methods have been developed to control for population structure in GWA studies including genomic control (Devlin and Roeder, 1999) and EIGENSTRAT (Price *et al.*, 2006). Population structure can induce weak random associations in a large number of SNPs, even when there are no true associations. In the genomic control approach, the association statistic at each marker is rescaled with an empirical value, the so called genomic factor, that is estimated from the distribution of whole-genome association statistics. For example, for association

with a binary trait the test statistic is a χ^2 with 1-degree-of-freedom; population structure can cause the distribution of (mostly null) test statistics to skew positive; the standard variant of genomic control finds the genomic factor that scales the test statistics so that the empirical median to matches the theoretical median. Value of genomic control close to 1 refers to complete control of confoundings while values greater than 1.2 represents lack of control. This method does not take into account correlation between individual nor allele frequency and individuals. EIGENSTRAT is based on principal component analysis (PCA) and is the most widely used method to control population structure in GWA studies. Here PCA is applied to genomic data to detect structure due to population stratification. By including the top principal components in linear model as a fixed effect term large scale differences in allele frequency between individuals will be discounted.

In practice, these two methods are combined with other steps to minimise GWA artefacts. First, if any information is available about self-reported ethnicity, the sample may be reduced to consider only a single ancestry. It is possible that self-declared ethnicity does not match with true ethnicity, and that substructure exists even within the ethnicity group. To examine this, multidimensional scaling (MDS) of genome-wide average proportion of alleles shared identical by state (IBS) is performed (Purcell *et al.*, 2007). A SNP is called IBS if two or more individuals share the same allele. The sample genetic homogeneity is assessed through visual inspection of the first two MDS components where outlier individuals are excluded. Secondly, subjects are dropped if any close relatives, 1st and 2nd degree relatives are found. Then EIGENSTRAT is used to correct for broad sample structure following genomic control correction (Sabatti *et al.*, 2009; Cho *et al.*, 2009; Ober *et al.*, 2001).

2.3.1 Linear Mixed Model for GWAS

There has been great interest in the field of quantitative genetic to develop sophisticated statistical methods to control confounding factors in GWA studies. Recently, Linear mixed effect models (LMM) have been introduced as an alternative method to linear regression models. (Yu *et al.*, 2006; Kang *et al.*, 2008; Zhang *et al.*, 2010; Kang *et al.*, 2010; Lippert *et al.*, 2011a,b; Zhou and Stephens, 2012; Svishcheva *et al.*, 2012; Pirinen *et al.*, 2013; Listgarten *et al.*, 2013; Widmer *et al.*, 2014; Kadri *et al.*, 2014). In this approach, we assume that phenotypic variation due to differences between latent sub-populations can be explained by genetic similarity covariance

structure. An genetic similarity matrix (GSM) is computed

$$\phi_{i,j} = \frac{1}{M}\sum_{k=1}^{M}\frac{(x_{ik}-2p_k)(x_{jk}-2p_k)}{2p_k(1-p_k)},$$

where $\phi_{i,j}$ is the (i,j) element of GSM; x_{ik} is the minor allele count of the i-th subject's k-th marker, coded as coded as 0, 1 or 2; p_k is frequency of the k-th marker; M is the total number of markers; and is treated as an empirical kinship matrix (see 2Φ above). Consequently, the LMM corrects for population structure by incorporating this GSM component of trait variance in the association statistic. In particular it has been shown that the correction for population structure in GWA studies with LMM could be outstanding (Kadri *et al.*, 2014; Widmer *et al.*, 2014; Yang *et al.*, 2014). Moreover, it has been proposed that LMM could increase GWA power in comparison to conventional linear model (Yang *et al.*, 2014).

However, the LMM is computationally intensive where due to computation of the GSM, variance component parameter estimation, and calculation of the association statistic; specifically, the complexity grows with both the sample size and number of markers. Several approximate or exact methods have been proposed to speed up LMM. In this literature, the term "approximate" is specifically used to refer methods that assume the total polygenic effect is same for all markers under the null hypothesis of no marker effect. This produces an enormous computational saving, as the GSM and related variance component is estimated only once using all markers. In contrast methods that do not make this approximation are referred to as "exact" (Lippert *et al.*, 2011a,b; Widmer *et al.*, 2014), and these exact methods use an LMM with a marker-specific GSM, where the GSM is constructed with the candidate marker and surrounding markers in linkage disequilibrium omitted. It has been shown that exact methods are more powerful than approximate ones because they prevent "proximal contamination", the double fitting of the candidate marker as a fixed covariate and random effect (Lippert *et al.*, 2011a,b).

Several speed-up algorithms have been proposed for exact methods based on using a small sub-set of genetic markers in the GSM. In these approaches, the GSM is constructed using markers with the most significant linear regression P values. The markers are determined either based on the first local minimum of the genomic control factor (Listgarten *et al.*, 2013) or the global maximum of out-of-sample prediction accuracy using the resulting GSM (Lippert *et al.*, 2011b, 2013). Furthermore, these advances can boost the study power by implicitly conditioning only on loci that are relatively likely to be truly associated. However, they might compromise stratification control depending on genetic structure of trait (Yang *et al.*,

2014; Widmer *et al.*, 2014). Regardless of all these efforts, exact methods are still computationally more complex than approximate ones because variance component needs to be estimated for each GSM using numerical methods.

All of the mentioned advances mainly concern accelerating LMM with growing sample sizes and number of markers being tested for univariate trait genome-wide association analysis. However, using LMM for high-dimensional imaging GWA presents enormous challenges in terms of computational intensity.

2.4 LMM Parameter Estimation

Maximum likelihood (ML) can be used for the LMM model (Eqns. (2.1) & (2.2)) parameter estimation with the assumption that the data follows a multivariate normal distribution:

$$\ell_{\mathrm{ML}}(\beta_{\mathrm{ML}}, \Sigma_{\mathrm{ML}}; Y, X) = -\frac{1}{2}\left[N\log(2\pi) + \log(|\Sigma|) + (Y - X\beta)'\Sigma^{-1}(Y - X\beta)\right]. \quad (2.3)$$

One criticism of using likelihood function in LLM to estimate the variance component parameters is that it can produce biased estimates, as it does not account the loss of degrees of freedom due to the fixed effect terms. To overcome this issue Harville (1974) suggested the optimisation of the likelihood function of the residualised data, the so called Restricted maximum likelihood (REML):

$$\begin{aligned}\ell_{\mathrm{REML}}(\Sigma_{\mathrm{REML}}; Y, X) = -\frac{1}{2}[(N-p)\log(2\pi) - \log|X'X| + \log|\Sigma| \\ + \log|X'\Sigma^{-1}X| + Y'PY],\end{aligned} \quad (2.4)$$

where $P = \Sigma^{-1}\left(I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right)$(Patterson and Thompson, 1971; Harville, 1974, 1977). With REML, variance parameters Σ_{REML} are estimated by optimisation of Eqn (2.4) and the fixed effects parameters are estimated using generalized least squares (GLS):

$$\hat{\beta}_{\mathrm{REML}} = (X'\hat{\Sigma}^{-1}_{\mathrm{REML}}X)^{-1}X'\hat{\Sigma}^{-1}_{\mathrm{REML}}Y.$$

A number of numerical methods for optimisation of the ML (Eq, (2.3)) or REML functions have been proposed, including Fisher's scoring (Longford, 1987), Newton-Rophson (Jennrich and Sampson, 1976) and Expectation Maximisation (EM) (Dempster *et al.*, 1977; Laird and Ware, 1982). These approaches are all based on iterative optimisation of the likelihood function, which is computationally intensive and prone to convergence failures; both the computation burden and the

algorithmic fragility pose significant problems when applied to hundreds of thousands of voxels.

2.4.1 Eigensimplification

For large datasets with arbitrary family structure, the computational burden of evaluating of the likelihood functions (Eqn. (2.3) & (2.4)) can be substantial. In particular, the determinant of Σ must be computed, along with a quadratic form of Σ with the residuals. Several algorithms have been proposed to speed up likelihood function optimisation. These advances can be broadly categorized into eigensimplification and covariance matrix reparametrisation followed by eigensimplification.

In the eigensimplification approach, the likelihood function is simplified using the eigenvectors of the kinship matrix. Based on the model in Equation (2.2), the eigenvectors of the phenotypic covariance Σ coincide with those of the kinship matrix. Applying this orthogonal transformation matrix S, which satisfies

$$(2\Phi) = SD_gS',$$

where S is the $N \times N$ matrix of eigenvectors; and $D_g = \text{diag}\{\lambda_{gi}\}$ is a diagonal matrix of the eigenvalues of 2Φ, to Equation (2.1) gives the transformed model

$$S'Y \quad = \quad S'X\beta + S'g + S'\epsilon$$

which we write as

$$Y^* \quad = \quad X^*\beta + g^* + \epsilon^*, \tag{2.5}$$

where Y^* is the transformed data; X^* are the transformed covariates; g^* is the transformed random genetic effect; and ϵ^* is the transformed residual vector. The diagonalising property of the eigenvectors then gives a simplified form for the variance:

$$\text{var}(\epsilon^*) = \Sigma^* = \sigma^2_{\text{A}} D_g + \sigma^2_{\text{E}} I, \tag{2.6}$$

where Σ^* is the variance of the transformed data. Whitening the data with this orthogonal transformation matrix reduces the optimization time substantially because of explicit formula for the determinant and inverse of the covariance matrix in the likelihood function (Blangero *et al.*, 2013).

2.4.2 Eigensimplification combined with profile likelihood

Further improvement in speed can be achieved by covariance matrix reparametrisation based on ratio of residual to genetic variance components accompanying with eigensimplfication. By parametrising the covariance matrix as

$$\Sigma = \sigma_g^2(2\Phi + \delta I),$$

where $\delta = \frac{\sigma_e^2}{\sigma_g^2}$, giving σ_g^2 the role of the traditional scalar variance parameter. Thus given a value of δ, β and σ_g^2 can be solved non-iteratively by generalised least squares as follows:

$$\begin{aligned} \hat{\beta} &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y, \\ \sigma_g^2 &= \frac{(Y - X\hat{\beta})'\Sigma^{-1}(Y - X\hat{\beta})}{N}. \end{aligned}$$

Given $\hat{\beta}$ and $\hat{\sigma}_g^2$, the likelihood has only one unknown (Kang *et al.*, 2008, 2010; Lippert *et al.*, 2011a). However, neither of these advances eliminate iterative optimization nor possible convergence problems.

2.5 Bivariate Genetic Modelling

In many population neuroimaging studies, multiple correlated measurements are taken from individuals. Genetic analysis of two or more quantitative traits can be used to measure the genetic effect on the observed phenotypic correlation. Pleiotropy is the condition when a single gene affects two or more traits simultaneously. When modelling two or more traits simultaneously, pleiotropy requires more variance parameters, specifically genetic and environmental correlations (Falconer and Mackay (1996)).

Bivariate Mixed Effect Model

When a pair of correlated traits are of interest, univariate polygenic model (Eq. (2.1)) can be extended to model the phenotypes jointly:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix},$$

where Y_i, for trait $i = 1, 2$ are the N-vectors of phenotype, X_i, are the fixed effect matrix of covariates, β_i are the p-vectors of regression coefficients, g_i are the N-

vectors of latent additive genetic effect and ϵ_i are the N-vectors of residual errors. When the matrix of covariates is same for both traits $X = X_i$ and there is no missing data for each trait, bivariate polygenic model can be expressed as:

$$Y = (I \otimes X)\beta + g + \epsilon, \tag{2.7}$$

where Y is the $2N$-vector of stacked traits, I is the 2×2 identity matrix, $\otimes$ denotes kronecker product, g and ϵ are the $2N$-vector of latent additive genetic and residual error effects, respectively. The distributional assumptions of the model can then be concisely stated as

$$\begin{aligned} g &\sim N(0, \Sigma_g \otimes (2\Phi)), \\ \epsilon &\sim N(0, \Sigma_e \otimes I_{N\times N}), \end{aligned}$$

such that $\text{cov}(g, e) = 0$, where Σ_g is a 2×2 genetic covariance matrix and Σ_e is a 2×2 residual covariance matrix. In this setting, the trait covariance matrix (Σ) is modelled as the sum of the genetic and environmental covariance components

$$\Sigma = \Sigma_g \otimes (2\Phi) + \Sigma_e \otimes I. \tag{2.8}$$

The variance components and regression coefficients are again estimated by maximising the likelihood or restricted likelihood functions (Almasy *et al.*, 1997; Zhou and Stephens, 2014).

Genetic and Environmental Correlation

Like the genetic and the environmental effects, the genetic and the environmental correlations can not be directly measured, but they can be estimated from the covariance between the two traits.

The phenotype covariance matrix can be decomposed to into the genetic and environmental covariance components as follows:

$$\Sigma_g = \begin{bmatrix} \sigma^2_{g,11} & \sigma_{g,12} \\ \sigma_{g,12} & \sigma^2_{g,22} \end{bmatrix}, \quad \Sigma_e = \begin{bmatrix} \sigma^2_{e,11} & \sigma_{e,12} \\ \sigma_{e,12} & \sigma^2_{e,22} \end{bmatrix},$$

where $\sigma^2_{g,ii}$ and $\sigma^2_{e,ii}$, $i = 1, 2$ are the genetic and environmental variance components, respectively, and $\sigma_{g,ij}$ and $\sigma_{e,ij}$, $i, j = 1, 2$ are the genetic and environmental covariance components, respectively. In this setting, the genetic (ρ_g) and environmental

(ρ_e) correlations can be defined as follows (Almasy *et al.*, 1997)

$$\rho_g = \frac{\sigma_{g,12}}{\sqrt{\sigma^2_{g_1}\sigma^2_{g_2}}}, \quad \rho_e = \frac{\sigma_{e,12}}{\sqrt{\sigma^2_{e_1}\sigma^2_{e_2}}}.$$

2.6 Neuroimaging Phenotypes

Human brain consists of three types of tissue, gray matter, white matter and cerebral spinal fluid (CSF). Fundamental to brain function is the firing of neurons in the gray matter, and information from distant neurons is communicated via pathways in the white matter; CSF is a liquid that fills voids in the brain. Since the invention of MRI, it has been a popular tool to visualize the human brain in vivo. In this section, we describe briefly the ideas and procedures behind an important phenotype in the neuroimaging field which includes summary of data preprocessing, analysis and rational of data gathering.

2.6.1 fMRI

fMRI signal is based on the *BOLD signal*. When human brain involves in a cognitive process, blood velocity increased in an area which is involved in the cognitive process. This phenomena yield to *Heodynamic Responce Function* (HRF). In a fMRI experiment a stimuli is presented to the subjects in a special paradigm with aim of increasing neural activity to its peak and then relaxing it to reach baseline level. Brain areas which their signal change looks like presented paradigm are responsible for cognitive process related to the stimuli. Neural activity associated with cognitive processes occurs in both time and space. In investigating neural activity using fMRI experiment both spatial and temporal properties should be considered. One of the advantage of fMRI experiment is that it has spatial and temporal resolution. Spatial resolution means ability to distinguish change related to stimuli in different areas in the brain and temporal resolution means ability to differentiate change of signal during stimuli presenting time. Some preprocessing steps should be implemented on fMRI data to increase signal to noise ratio and getting them ready for single and group level analysis which are motion correction, slice timing correction, spatial smoothing and registration.

fMRI Data Analysis

When data is preprocessed, GLM model is applied to find brain areas where their signal change looks like the presented stimuli paradigm.

$$Y_k = X\beta_k + \epsilon, \tag{2.9}$$

where $Y_k = (Y_{k_1}, ...Y_{k_N})$ is a measured time series in voxel k, X is a $N \times p$ design matrix which includes stimulus timing where convolved with HRF function, β_k, regression coefficient is a vector of size $p \times 1$ and ϵ is a $N \times 1$ vector of error terms. In the GLM context is assumed $\text{var}(\epsilon) = \sigma^2 I$ to estimate β optimally, however in the fMRI experiment data is correlated ($\text{var}(\epsilon) = \sigma^2 V$). Prewhitening matrix like W where $WVW' = I_N$ is employed to convert V matrix to identity (I_N), then GLM theory can be implemented to estimate the prewhiten model parameters.

$$WY_k = WX\beta_k + W\epsilon, \quad \text{cov}(W\epsilon) = \sigma^2 WVW' = \sigma^2 I$$

Which prewhitening matrix W can be estimated from autoregressive model of order 1 (AR(1)) (Friston *et al.*, 1994a; Woolrich *et al.*, 2001).

After fitting 2.9 to all voxels of subjects, in the next step, statistical inference in population level should be done. Hierarchical random effect modeling (Friston *et al.*, 2002) is deployed to this aim. In this step, separate GLM is fitted to the β_k with a random group error term like first level for all voxels separately.

Consider an experiment with N_k first level sessions and for each session preprocessed fMRI data is a $N \times 1$ vector Y_k , the $N \times P_k$ design matrix is X_k, and β_k is a $P_k \times 1$ vector of parameter estimates ($k = 1, \ldots, N_k$). Also Y_k is assumed to have been prewhitened. An individual GLM is deployed to find first-level parameters to the N_k individual data sets:

$$Y_k = X_k\beta_k + \epsilon_k,$$

where $\epsilon_k \approx N(0, \sigma_k^2 I)$. Note that in the fMRI time series analysis, the first level design matrix, X_k, doesn't need to be same for all k. Using the block diagonal forms:

$$Y_k = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{N_k} \end{bmatrix}, \; X_k = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & X_{N_k} \end{pmatrix}, \; \beta_k = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{N_k} \end{bmatrix}, \; \epsilon_k = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{N_k} \end{bmatrix},$$

The hierarchical model is

$$\begin{aligned} Y_k &= X_k\beta_k + \epsilon_k, \\ \beta_k &= X_G\beta_G + \epsilon_G, \end{aligned}$$

where X_G is the $N_k \times P_G$ second level design matrix, β_G is the $P_G \times 1$ vector of second level parameters and $\epsilon_G \approx N(0, \sigma_G^2 I)$ which σ_G^2 is a random effect variance. Mixed effect model can be applied in this context to estimate the population level parameters:

$$\hat{\beta}_k = X_G\beta_G + \epsilon_G + \epsilon. \tag{2.10}$$

Equation (2.10) contains the within subject and between subject variance components which is estimated by *Ordinary Least Square* (OLS) where $\hat{\beta}$ is consisting of one measurement for each subject (Mumford and Nichols (2006)).

2.6.2 Diffusion Tensor Imaging

Human white matter is a complex system which has an important role in neurological and psychological disease, aging process and cognitive behaviors. These capacities encourage scholars to study this part of brain in detail. *Diffusion Tensor Imaging* (DTI), is an in vivo method which is based on the water molecules diffusion. This diffusive pattern can be captured by an advanced MRI protocol *DWI*. Study of diffusion in human brain is interesting because boundary of a tissue in white matter impose *Anisotropic* diffusion which means it's a marker for white matter microstructure. On the other hand, diffusion in other part of brain like gray matter of CSF is *Isotropic*. *Diffusion Tensor Model* (Basser *et al.* (1994)) is employed to a set of DWI images to model this displacement of molecules and derive diffusion maps. *Fractional Anisotropy* (FA) which represent the white matter integrity and mean *Mean Diffusivity* (MD) are two popular quantity to study the human white matter.

2.6.3 Cortical Thickness and VBM

Regardless of studying gray matter function impact, its structure plays an important role in finding markers for neurodegenerative and psychiatric disease. Accurate and automated measurement of human cerebral cortex thickness (*Cortical thickness*) is one of the popular methods (Fischl and Dale, 2000). This approach, starts with preprocessing of T1 images which mainly includes motion correction, registration to

standard space, intensity normalization with more details and advanced algorithms, after preprocessing the brain is segmented to the gray matter, white matter and CSF, then with the state-of-the-art algorithms white matter surface and then gray matter surface is reconstructed by tessellation. Finally distance between gray and white matter is calculated at vertex of each tessellation which is regarded as cortical thickness.

Voxel Based Morphometry (VBM) is the another way to study human brain cerebral cortex which is based on the gray matter concentration (Ashburner and Friston, 2000). In this method gray matter concentration is compared between groups or correlated with cognitive score or disease severity. To achieve VBM map, from each subject T1 image, gray matter is segmented, then all subjects gray matter is aligned to standard space and finally smoothness is applied to aligned images.

2.7 Multiple Testing Correction

The final step in neuroimaging data analysis is inference, the determination of which voxels are significant and incompatible with the null hypothesis. For the standard mass univariate approach, there are two approaches to inference commonly used, voxel-wise and cluster-wise. Voxel-wise inference, is intuitive: A threshold is applied to the statistic image and each voxel exceeding the threshold is marked, individually as significant. In cluster-wise inference, an arbitrary threshold is applied to the statistic image and "clusters" are formed; in the neuroimaging setting, clusters are contiguous suprathreshold voxels, informally called "blobs". Statistical inference is based on the volume or spatial extent of these clusters; if a cluster has a sufficiently large extent, the set of voxels in the cluster are jointly marked as significant. While cluster-wise inference is generally more sensitive (Friston *et al.*, 1994b), it lacks spatial precision, as inference is on the cluster as a whole, and no individual voxel in the cluster can be identified as the source of the effect.

Whether performing inference on voxels or clusters, naive use of a $\alpha = 5\%$ level leads to many false positives. For example, if there are V voxels in the brain, $\alpha \times V$ false positive voxels are expected to be significant under H_0 with voxel-wise inference; similar problems apply to cluster-wise inference. The solution to the multiple testing problem is to use a more stringent inference procedure that controls a measure of multiple false positives over a family (i.e. an image) of hypotheses. The standard measure of false positives is the family wise error rate (FWE), the chance of *any* false positives occurring.

There is an intrinsic relationship between FWE and the distribution of the

maximal statistic (i.e. the largest statistic value in the brain) (Nichols and Hayasaka, 2003). Under the complete (image-wide) null hypothesis and for a given statistic threshold u, the occurrence of any false positives coincides with the event of the maximal statistic exceeding u. Thus to control the FWE at level α_{FWE}, one can set u equal to the $100(1-\alpha_{FWE})$ percentile of the distribution of the maximal statistic under the null hypothesis. In practice obtaining this maximal distribution for a correlated data is a challenge. However using permutation, it is straightforward to obtain the empirical distribution of the maximum under the null, and thus find the FWE level threshold (Holmes *et al.*, 1996; Nichols and Holmes, 2002). FWE corrected P-values can also be defined by reference to this maximum distribution: the probability of observing a maximum value as or larger than a particular observed statistic. Finally, note that FWE inference applies equally to voxel-wise and cluster-wise inference, in the latter the inference being based on cluster extent instead of signal intensity (Nichols and Holmes, 2002).

2.7.1 Permutation

Permutation is a non parametric method for making statistical inference which needs a few assumption in contrast of parametric inference. This strategy was introduced by Fisher (1935) and it's getting popular among researchers when inexpensive, fast computers have been available. The only assumption about the permutation test is an exchangebility under the null hypothesis. Null hypothesis and the exchangebility together define the permutation strategy. Let $\mathcal{P} = \{P_j\}$, where P_j is a $n \times n$ permutation matrix, also suppose that X be random variable which is distributed according to a probability distribution $P_\theta, \theta \in \Omega$. If $P_j X$ and X have same distribution then data is exchangeable (Good, 2005). Exchangeability and independence are similar concepts but exchangebility is more general. Dependent data can be exchangeable if their permuted joint distribution is maintained under the null after permutation. (Good, 2005).

P-value for Permutation Test

P-value is a measure for assessing a null hypothesis which is calculated in the parametric testing, as $p = P(T \geq T_0 \mid \mathcal{H}_\prime)$, where T is a test statistics under the null hypothesis and related parametric assumptions, T_0 is value of test statistic after running an experiment. However in nonparametric testing like permutation test, according to the rational of it, under the null hypothesis the data are exchangeable and we don't expect to observe permuted data test statistics greater than unper-

muted one so P-value for permutation test can be defined as number of times which permuted test statistics exceeds the unpermuted one.

$$p = \frac{1}{N+1} \sum_{n} I(T^* \geq T_0),$$

where N is the number of permutations, I is an indicator function, T^* is a permuted test statistic and T_0 is the unpermuted data test statistic. Furthermore, permutation tests can provide exact control or approximately exact when there are nuisance variables of false positive risk (Ernst, 2004) .

Permutation test in Neuroimaging

In the neuroimaging, permutation strategy was initiated by Holmes *et al.* (1996) and Nichols and Holmes (2001) validated and provided practical consideration. They showed that based on minimal assumption of permutation theory test, multiple testing problem can be solved easily, specially in the situations that parametric test requirements untenable. Let describe the concept of permutation test in Neuroimaging context with an example. Consider a fMRI group study with two groups of patient and healthy normal subjects. If there isn't any experimental effect (null hypothesis) then labeling of subjects to two groups doesn't matter then with computing new test statistic (relabeling subjects), we would decide to accept or reject the existence of experimental effect at each voxel, if most of the relabeled test statistics are greater or smaller than the original label of subjects. This procedure yields to uncorrected non-parametric p-value for each voxel. In the next step, adjusting for multiple tests is performed based on the signal intensity or spatial extension which is based on the maximum statistic.

Single Threshold Test

The permutation approach can yield the empirical distribution of maximum statistics under the null hypothesis when there isn't any effect entire images in a straightforward way. Instead of calculating test statistics for each voxel separately, Maximum statistic is manipulated in entire search volume. This approach leads to empirical distribution of maximum statistic. When the critical value is manipulated for maximum statistics, it's the critical threshold for a single threshold test over the same search volume, then voxels statistic exceeds this value show experiment effect. Holmes *et al.* (1996) showed that this test has a strong control *experiment-wise* type I error.

Suprathreshold Maximum Cluster Size Test

In this approach, a critical value for maximum suprathreshold cluster size is calculated and each cluster of a size that exceeds this value shows the experiment effect. To reach this aim, in each permutation, maximum cluster size is calculated for relabeled data which yield to empirical distribution of it under the null hypothesis. The Null hypothesis is rejected at level of α if the maximum cluster size of unpermuted data is in the top of 100% of the permutation distribution Nichols and Holmes (2001).

In general this test is more powerful than single threshold approach Friston *et al.* (1994b). However this increasing power comes from reduced localized power. In the parametric maximum cluster size test, pre-defined threshold should be chosen high to meet the RFT requirement, however, in the permutation test it isn't necessary. Low pre-defined threshold impose high power to detect various type of null hypothesis deviations. For instance, a large suprathreshold cluster which is detected from low pre-defined threshold doesn't contain intense focal, in contrast, high threshold one miss lower diffuse signals.

Permutation Test for Heritability Inference

Most models used in science, and in particular the quantitative genetic models of interest here, are based on an assumption of multivariate Gaussianity. These assumptions are the basis of the estimation and test procedures described above. However, our data may exhibit non-Normality, e.g. heavy tails, skew or extreme outliers. Moreover, with the exception of linear models with i.i.d. errors, most inference procedures depend on large sample results that may produce invalid or conservative tests with finite sample sizes. Even with large sample sizes, standard asymptotic results may make simplifying assumptions that are inappropriate for our data (see above, for problems with the 50:50 χ^2 mixture result). Further, some test procedures depend on simply intractable distributions, for example the maximum distribution over an image of statistics, essential for control of the familywise error rate. Hence, there is a compelling need for alternative inference procedures that make fewer assumptions and provide valid P-values.

To our knowledge, there is little work on permutation tests for variance component inference. The typical application of components-of-variance models is not in quantitative genetics, but in an extension of regression models in which data have a hierarchical structure with units nested in clusters, like repeated measures designs. In that setting there are clusters of observations (e.g. observations from a

particular subject) that are dependent. Of the few permutation methods proposed in this setting, they all permute the residuals after removing the fixed-term covariate effects (marginal residuals) between and within clusters while fixing the model structure (keep the number of subjects in each cluster fix). The test procedures differ in the particular test statistics used to make inference on a random effect parameter. Fitzmaurice *et al.* (2007) used the LRT as the statistic, while Lee and Braun (2012) used the sample variance of estimated random effect. It is important to note that both of these statistics are based on optimizing the non linear likelihood function, and thus as permutation procedures they are yet more computationally demanding. The sample variance statistic is faster to compute than the LRT, as the likelihood function needs to be optimized only once per permutation. Even so, any iterative procedure with brain image data, requiring fitting at each of 10,000 to 100,000 voxels, will be quite slow.

Samuh *et al.* (2012) presented a fast permutation test though only for the random intercept model. Under the null hypothesis of no random effect, the problem is cast as a fixed-effects ANOVA model, so exact p-value for random effect hypothesis testing can be calculated by applying the F-statistic for equality of variance among groups in each step. Recently, Drikvandi *et al.* (2013) introduced fast permutation test based on variance least square estimator which is in general applicable to any type of mixed effect models. In their approach, random effects are estimated in each permutation step by applying least square estimator on squared residuals.

Chapter 3

Fast and Powerful Heritability Estimation and Inference

Heritability estimation has become an important tool for imaging genetics studies. The large number of voxel- and vertex-wise measurements in imaging genetics studies present a challenge both in terms of computational intensity and the need to account for elevated false positive risk because of the multiple testing problem. Blangero *et al.* (2013) presented a method to accelerate maximum likelihood estimation by applying an orthonormal data transformation that diagonalises the phenotypic covariance, transforming a correlated heritability model into an independent but heterogeneous variance model. However, this advance doesn't eliminate iterative optimization nor possible convergence problems.

In this chapter, we expanded upon this work to derive fast, non-iterative estimates and test statistics based on the first iteration of Newton's method suitable for voxel-wise heritability analyses. These procedures can be constructed with an auxiliary model based on regressing squared residuals on the kinship matrix eigenvalues. Then the Wald and score hypothesis tests can then be seen as generalized and ordinary explained sum of squares of the auxiliary model. In addition, as the null hypothesis of no heritability corresponds to homogeneous variance of the transformed phenotype, we draw from the statistical literature on tests of heteroscedasticity for a new and completely different test for heritability detection. To address problems with inaccuracies with the standard results used to find P-values, we propose four different permutation schemes to allow semi-parametric inference (parametric likelihood-based estimation, non-parametric sampling distribution). In total, we evaluate 5 different significance tests for heritability, with either asymptotic parametric or permutation-based P-value computations. We identify a number of tests

that are both computationally efficient and powerful, making them ideal candidates for heritability studies in the massive data setting. We illustrate our method on fractional anisotropy measures in 859 subjects from the Genetics of Brain Structure study.

The remainder of this chapter is organized as follows. In the next section we detail the statistical model used and describe each of our proposed methods. The simulation framework used to evaluate the methods, and the real data analysis used for illustration are described in evaluation section. We then present and interpret results, and offer concluding remarks.

3.1 Theory

In this section we detail the statistical models used, introduce our fast heritability estimators and tests, and then propose several permutation strategies for these tests.

3.1.1 Original and Eigensimplified Polygenic Models

At each voxel/element, a polygenic model for the phenotype Y measured on N individuals can be written

$$Y = X\beta + g + \epsilon, \tag{3.1}$$

where X is an $N \times p$ matrix consisting of an intercept and covariates, like age, sex, etc; β is the p-vector of regression coefficients; g is the N-vector of latent (unobserved) additive genetic effect; and ϵ is the N-vector of residual errors. In this study we consider the most common variance components model, with only additive and unique environmental components.

The trait covariance, $\text{var}(Y) = \text{var}(g + \epsilon) = \Sigma$ can be written

$$\Sigma = 2\sigma^2_{\text{A}}\Phi + \sigma^2_{\text{E}}I, \tag{3.2}$$

where Φ is the kinship matrix; σ^2_{A} and σ^2_E are the additive genetic and the environmental variance components, respectively; and I is the identity matrix. The kinship matrix is comprised of kinship coefficients, half the expected proportion of genetic material shared between each pair of individuals (Lange, 2003).

The narrow sense heritability is

$$h^2 = \frac{\sigma^2_{\text{A}}}{\sigma^2_{\text{A}} + \sigma^2_{\text{E}}}. \tag{3.3}$$

Maximum likelihood is used for parameter estimation with the assumption that the data follows a multivariate normal distribution. The log likelihood for the polygenic model (Eqns. (3.1) & (3.2)) is

$$\ell(\beta, \Sigma; Y, X) = -\frac{1}{2} N \log(2\pi) - \frac{1}{2} \log(|\Sigma|) - \frac{1}{2}(Y - X\beta)' \Sigma^{-1} (Y - X\beta). \tag{3.4}$$

For large datasets with arbitrary family structure, the computational burden of evaluating of the likelihood can be substantial. In particular, a quadratic form of the inverse covariance, Σ^{-1}, must be computed, along with the determinant of Σ. We take the approach of Blangero *et al.* (2013), who proposed an orthogonal transformation based on the eigenvectors of the kinship matrix, thus diagonalising the covariance and simplifying the computation of the likelihood (3.4).

The eigensimplified polygenic model is obtained by transforming the data and model with a matrix S, the matrix of eigenvectors of Φ which are the same as the eigenvectors of Σ, Eq. (3.2). Applying this transformation to Equation (3.1) gives the transformed model

$$S'Y = S'X\beta + S'g + S'\epsilon$$

which we write as

$$Y^* = X^*\beta + \epsilon^*, \tag{3.5}$$

where Y^* is the transformed data, X^* are the transformed covariates and ϵ^* is the transformed random component, where ϵ^* now encompasses both the genetic and non-genetic random variation. The diagonalising property of the eigenvectors then gives a simplified form for the variance:

$$\text{var}(\epsilon^*) = \Sigma^* = \sigma^2_{\text{A}} D_{\text{g}} + \sigma^2_{\text{E}} I, \tag{3.6}$$

where Σ^* is the variance of the transformed data and $D_{\text{g}} = \text{diag}\{\lambda_{\text{g}i}\}$ is a diagonal matrix of the eigenvalues of 2Φ.

The log likelihood takes on the exact same form as Equation (3.4) for Y^*, X^*, β and Σ^*, except is much easier to work with since Σ^* is diagonal:

$$\ell(\beta^*, \sigma^*_{\text{A}}, \sigma^*_{\text{E}}; Y^*, X^*) = -\frac{1}{2} N \log(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \log(\sigma^2_{\text{A}} \lambda_{\text{g}i} + \sigma^2_{\text{E}}) - \frac{1}{2} \sum_{i=1}^{N} \frac{\epsilon^{*2}_i}{\sigma^2_{\text{A}} \lambda_{\text{g}i} + \sigma^2_{\text{E}}}.$$

Note that, while S' can be seen as a semi-whitening step, the transformed

model can also be seen as a change of variables, where the variance is reparametrized as $\Sigma = S\Sigma^* S'$. As a reparametrization, the invariance property of maximum likelihood guarantees that the same values of β, σ^2_{A} and σ^2_{E} will optimize both the original and transformed likelihoods.

Use of this transformation has two major benefits. First, optimization time is substantially reduced, as the inverse and determinant of the transformed covariance is now trivial. Second, applying standard statistical inference procedures, including the score and the Wald test, to the eigensimplified polygenic model produces simple algebraic forms that can be harnessed for fast approximations. Both of these speed improvements facilitate the use of permutation tests that avoid asymptotic approximations.

3.1.2 Heritability Estimation and Test Statistics

We segregate the transformed model parameters into fixed β and random $\theta = (\sigma^2_{\mathrm{A}}, \sigma^2_{\mathrm{E}})$ terms, and estimate them by maximizing the likelihood function via iterative numerical methods. Here, we consider Fisher's scoring method because it leads to computationally efficient heritability estimators and associated tests. Scoring method requires the score and expected information matrix of the transformed model, which are

$$S(\beta, \theta) = \begin{bmatrix} X^{*\prime}\Sigma^{*-1}\epsilon^* \\ -\frac{1}{2}\left[U'\Sigma^{*-1}1 - U'\Sigma^{*-2}\epsilon^{*2}\right] \end{bmatrix}, \tag{3.7}$$

and

$$I(\beta, \theta) = \begin{bmatrix} X^{*\prime}\Sigma^{*-1}X^* & 0 \\ 0 & \frac{1}{2}U'\Sigma^{*-2}U \end{bmatrix}, \tag{3.8}$$

respectively, where $U = [\mathbf{1}, \lambda_{\mathrm{g}}]$ is a $N \times 2$ matrix, $\mathbf{1}$ is a $N \times 1$ vector of ones and $\lambda_{\mathrm{g}} = \{\lambda_{\mathrm{g}i}\}$ is a $N \times 1$ vector of kinship matrix eigenvalues. It is useful to write f^* for the vector with elements $f^*_i = \hat{\epsilon}^{*2}_i$, where $\hat{\epsilon}^* = Y^* - X^*\hat{\beta}$ are the transformed model residuals. Fisher's scoring method gives update equations for $\hat{\beta}$ and $\hat{\theta}$ at iteration $j+1$ as:

$$\hat{\beta}_{j+1} = \left(X^{*\prime}(\hat{\Sigma}^*_j)^{-1}X^*\right)^{-1} X^{*\prime}(\hat{\Sigma}^*_j)^{-1}Y^*, \tag{3.9}$$

$$\hat{\theta}_{j+1} = \max\left\{0,\ \left(U'(\hat{\Sigma}^{*2}_j)^{-1}U\right)^{-1} U'(\hat{\Sigma}^{*2}_j)^{-1}f^*_j\right\}, \tag{3.10}$$

where j indexes iteration; the variance parameters θ must be positive, hence the maximum operator. When these updates are iterated until convergence as usual, we denote the estimates with a ML subscript, e.g. $\hat{\beta}_{\text{ML}}$, $\hat{\theta}_{\text{ML}}$ and $\hat{h}^2_{\text{ML}} = \hat{\sigma}^2_{\text{A,ML}} / (\hat{\sigma}^2_{\text{A,ML}} + \hat{\sigma}^2_{\text{E,ML}})$.

To allow for potential improvements on speed, we also consider a one-step estimator. First, observe that since Σ^* is diagonal, (3.9) is the Weighted Least Squares (WLS) regression of Y^* on X^*, and (3.10) is based on the WLS regression of f^* on U. This immediately suggests initial values based on Ordinary Least Squares (OLS),

$$\begin{aligned} \hat{\beta}_{\text{OLS}} &= \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y^*, \\ \hat{\theta}_{\text{OLS}} &= \max\left\{0,\ \left(U'U\right)^{-1} U' f^*_{\text{OLS}}\right\}, \end{aligned} \tag{3.11}$$

where f^*_{OLS} is the square of the OLS residuals

$$\hat{\epsilon}_{\text{OLS}} = Y^* - X^* \hat{\beta}_{\text{OLS}}; \tag{3.12}$$

while not recommended as a final estimate, it also produces $\hat{h}^2_{\text{OLS}} = \hat{\sigma}^2_{\text{A,OLS}} / (\hat{\sigma}^2_{\text{A,OLS}} + \hat{\sigma}^2_{\text{E,OLS}})$. Finally, our proposed one-step estimators are:

$$\begin{aligned} \hat{\beta}_{\text{WLS}} &= \left(X^{*\prime}(\hat{\Sigma}^*_{\text{OLS}})^{-1}X^*\right)^{-1} X^{*\prime}(\hat{\Sigma}^*_{\text{OLS}})^{-1}Y^*, \\ \hat{\theta}_{\text{WLS}} &= \max\left\{0,\ \left(U'(\hat{\Sigma}^{*2}_{\text{OLS}})^{-1}U\right)^{-1} U'(\hat{\Sigma}^{*2}_{\text{OLS}})^{-1} f^*_{\text{OLS}}\right\}, \end{aligned} \tag{3.13}$$

where $\hat{\Sigma}^*_{\text{OLS}}$ is formed by $\hat{\theta}_{\text{OLS}} = (\sigma^2_{\text{A,OLS}}, \sigma^2_{\text{E,OLS}})$, also producing $\hat{h}^2_{\text{WLS}} = \hat{\sigma}^2_{\text{A,WLS}} / (\hat{\sigma}^2_{\text{A,WLS}} + \hat{\sigma}^2_{\text{E,WLS}})$.

Amemiya (1977) showed that such one-step maximum likelihood estimators are asymptotically normal and consistent. Going forward, we will use "ML" to refer to the maximum-likelihood, iterated estimator and "WLS" to refer to this one-step estimator.

3.1.3 Test statistics

In this section we describe Likelihood Ratio Tests (LRTs), Wald Tests, and Score Test for hypothesis tests of nonzero heritability; we also add an additional test based on detecting heterogeneous variance structure to detect heritability. We only consider the transformed model, and tests on $H_0 : \sigma^2_{\text{A}} = 0$ vs. $H_1 : \sigma^2_{\text{A}} > 0$, equivalent to inference for heritability (3.3). Table 3.1 organizes the models and test statistics we consider.

Table 3.1: Comparison of Model and Test Statistic Properties. Usual P-values and CI's (confidence Intervals) refer to the best practice inference tools used with maximum likelihood estimation.

Model Name	Model Expression	Estimation Method	Test statistics			
			LRT	WALD	Score	GQ
Original	$Y = X\beta + g + \epsilon$	ML	*(usual P-values)*	*(usual CI's)*		
Transformed	$Y^* = X^*\beta + \epsilon^*$	WLS, "1 Step" ML, fully converged	$T_{\mathrm{L,WLS}}$ $T_{\mathrm{L,ML}}$	$T_{\mathrm{W,WLS}}$ $T_{\mathrm{W,ML}}$	T_{S}	
Transformed, Split	$Y^*_{\mathrm{A}} = X^*_{\mathrm{A}}\beta_{\mathrm{A}} + \epsilon^*_{\mathrm{A}}$ $Y^*_{\mathrm{B}} = X^*_{\mathrm{B}}\beta_{\mathrm{B}} + \epsilon^*_{\mathrm{B}}$	OLS, "0 Step"				T_{GQ}

Likelihood Ratio Test

The LRT (Neyman and Pearson, 1933) statistic is twice the difference of the log-likelihoods, unrestricted minus H_0-restricted. For ML this requires optimizing the likelihood function twice, once under the null $H_0 : \sigma^2_{\mathrm{A}} = 0$, once under the alternative (though the null model is trivial, equivalent to OLS). We denote the test statistic for this test $T_{\mathrm{L,ML}}$. In addition, LRT can be constructed for the transformed model in terms of the one-step weighted least square (WLS) estimator; we denote this statistic $T_{\mathrm{L,WLS}}$.

Wald Test

The Wald test consists of a quadratic form of the parameter estimate minus its null value, and its inverse asymptotic variance (i.e. expected Fisher's information matrix). Both the estimate and its variance are computed under the full, alternative model.

The Wald test for the ML estimator (Rao, 2008) is

$$\begin{aligned} T_{\mathrm{W,ML}} &= \frac{1}{2}(\hat{\sigma}^2_{\mathrm{A,ML}})^2[C(U'\hat{\Sigma}^{*-2}_{\mathrm{ML}}U)^{-1}C']^{-1} \\ &= \frac{1}{2}\left(N - (\mathbf{1}'\hat{\Sigma}^{*-1}_{\mathrm{ML}}\mathbf{1})^2(\mathbf{1}'\hat{\Sigma}^{*-2}_{\mathrm{ML}}\mathbf{1})^{-1}\right), \end{aligned}$$

where $C = [0\ 1]$ is a contrast row vector, and the latter is a simpler form found in Buse (1984). Iterative optimizations is required for $T_{\mathrm{W,ML}}$, though it can be considerably more amenable to compute than LRT because the likelihood function is optimized only once.

The Wald test for our one-step WLS estimator can be written

$$T_{\mathrm{W,WLS}} = \frac{1}{2}(\hat{\sigma}^2_{\mathrm{A,WLS}})^2[C(U'\hat{\Sigma}^{*-2}_{\mathrm{WLS}}U)^{-1}C']^{-1}$$

$$
\begin{aligned}
&= \frac{1}{2}(\hat{\sigma}^2_{\mathrm{A,WLS}})^2 \times \\
&\quad (\hat{\Sigma}^{*-1}_{\mathrm{OLS}}\lambda_{\mathrm{g}})' \left(I - \hat{\Sigma}^{*-1}_{\mathrm{OLS}}\mathbf{1} \left((\hat{\Sigma}^{*-1}_{\mathrm{OLS}}\mathbf{1})'(\hat{\Sigma}^{*-1}_{\mathrm{OLS}}\mathbf{1}) \right)^{-1} \mathbf{1}'\hat{\Sigma}^{*-1}_{\mathrm{OLS}} \right) \hat{\Sigma}^{*-1}_{\mathrm{OLS}}\lambda_{\mathrm{g}}.
\end{aligned}
$$

where the second line shows the computation to be half the generalized explained sum of squares (Buse, 1973, 1979) of an auxiliary model, the weighted least squares regression of f^*_{OLS} on λ_{g}, with weights determined by $\hat{\Sigma}^*_{\mathrm{OLS}}$.

Score Test

The score test (Rao, 2008), also known as the Lagrange multiplier test, is a quadratic form of the score (the gradient of the log likelihood) and the expected Fisher's information, each evaluated under the null hypothesis. Among the tests that we consider, the score test is the least computationally demanding procedure, as it only requires estimation of the null model. For $H_0 : \sigma^2_{\mathrm{A}} = 0$, the score test with the transformed likelihood function is:

$$
\begin{aligned}
T_{\mathrm{S}} &= \frac{\lambda'_{\mathrm{g}}\Sigma^{*-2}_{\mathrm{OLS}}f^*_{\mathrm{OLS}} - \lambda'_{\mathrm{g}}\Sigma^{*-1}_{\mathrm{OLS}}\mathbf{1}}{CU'\Sigma^{*-2}_{\mathrm{OLS}}UC'} \\
&= \frac{1}{2}\left(\frac{\hat{\sigma}^2_{\mathrm{A,OLS}}}{\hat{\sigma}^2_{\mathrm{OLS}}} \right)^2 \lambda'_{\mathrm{g}} \left(I - \frac{\mathbf{1}'\mathbf{1}}{N} \right) \lambda_{\mathrm{g}},
\end{aligned}
$$

where $\hat{\sigma}^2_{\mathrm{OLS}} = (\hat{\epsilon}_{\mathrm{OLS}})'\hat{\epsilon}_{\mathrm{OLS}}/N$ is the OLS naive residual variance estimator. Similar to the Wald test, T_{S} can be obtained as half the regression sum of squares of an auxiliary model, the (unweighted) regression of $f^*/\hat{\sigma}^2_{\mathrm{A,OLS}}$ on λ_{g}. As it only involves the fitted null model, it isn't associated with a WLS or ML estimate.

We note that Wald and score tests for a null hypothesis value lying on the boundary of parameter space can take a special form (Freedman, 2007; Molenberghs and Verbeke, 2007; Morgan *et al.*, 2007; Verbeke and Molenberghs, 2007; Silvapulle, 1992; Silvapulle and Silvapulle, 1995; Verbeke and Molenberghs, 2003). However, for our model (3.1), the standard version is appropriate if the score function is positive at the boundary value and otherwise set to zero. As any negative score values are suppressed by our non-negative constrained estimates $\hat{\theta}_{\mathrm{OLS}}$ (3.11) and $\hat{\theta}_{\mathrm{WLS}}$ (3.13), our tests are implicitly zero when needed, and thus the appropriate Wald and score tests are as given above.

All three of the LRT, Wald, and score tests procedures are asymptotically equivalent but have different small-sample performance, which we evaluate below. These tests are considered to follow asymptotically a $50 : 50$ mixture of χ^2 distri-

butions with 1 and 0 DF, where 0 a DF χ^2 is a point mass at 0 (Chernoff, 1954; Self and Liang, 1987; Stram and Lee, 1994; Dominicus *et al.*, 2006; Verbeke and Molenberghs, 2003), although it has been shown that 0 values can occur with a higher frequency, and the standard 50:50 result will tend to produce conservative inferences (Blangero *et al.*, 2013; Crainiceanu and Ruppert, 2004c; Shephard, 1993; Shephard and Harvey, 1990).

Goldfeld and Quandt (GQ) Test

Instead of standard likelihood theory, an alternative approach to heritability hypothesis testing can be derived from tests of heteroscedasticity. This follows for the transformed model, since the null hypothesis of no heritability corresponds to homoscedasticity of the transformed phenotype variance ($\mathsf{Var}(\epsilon^*) = \sigma^2 I$). Thus, rejection of the hypothesis of homoscedasticity implies a rejection of the hypothesis of zero heritability. One class of such tests require an explicit, hypothesized form for the heterogeneous variance. Another, type called "nonconstrutive" does not require such explicit models; one example is the Goldfeld and Quandt (1965) (GQ) test, which we propose as a test for non-zero heritability.

The GQ test partitions observations into 2 groups, A & B, based on a variable that should explain any heterogeneous variance. The test statistic then compares the ratio of OLS residual mean squares:

$$T_{\mathrm{GQ}} = \frac{\hat{\epsilon}^{*'}_{\mathrm{A}} \hat{\epsilon}^{*}_{\mathrm{A}} / (n_{\mathrm{A}} - 1)}{\hat{\epsilon}^{*'}_{\mathrm{B}} \hat{\epsilon}^{*}_{\mathrm{B}} / (n_{\mathrm{B}} - 1)}, \tag{3.14}$$

where subscript A refers to the high variance group, subscript B to low variance group, $\hat{\epsilon}^*_{\mathrm{A}}$ are the residuals from regressing elements of Y^* in group A on corresponding rows of X^*, and likewise for $\hat{\epsilon}^*_{\mathrm{B}}$, finally, n_{A} and n_{B} are the number of observations in each respective group. With Gaussian errors and under a null hypothesis of homoscedasticity, T_{GQ} follows a F-distribution with degrees of freedom $\nu_1 = n_{\mathrm{B}} - p$ and $\nu_2 = n_{\mathrm{A}} - p$, where p is the number of columns in X^*.

For the transformed data Y^*, the kinship eigenvalues order the variance of the observations when $\sigma^2_{\mathrm{A}} > 0$. Thus we propose to define the two groups based on $\lambda_{gi} > 1$ and $\lambda_{gi} \leq 1$, where we make use of the fact $\sum_i \lambda_{gi}/N = \mathrm{trace}(2\Phi)/N = 1$.

This test is only able to detect non-zero heritability and cannot produce estimates of h^2. On the other hand, the parametric null distribution of (3.14) does not depend on the mixture approximation and large sample properties, and its implementation is straightforward. To our knowledge, this is the first proposed use of a heteroscedasticity test to create an exact (non-asymptotic), non-iterative test

of heritability.

3.1.4 Permutation Test for Heritability Inference

Permutation methods can be used to construct the null sampling distribution which can be used to produce P-values and thresholds. For the model with only additive genetic and environmental variance components, the null hypothesis of no heritability implies fully independent data. Thus, if there were no nuisance variables (X), a permutation test could be conducted by freely permuting the data (Y). With covariates, we must permute suitable residuals, as detailed below.

To conduct inference on σ^2_{A} in the presence of the nuisance parameters β and σ^2_{E}, we draw inspiration from various methods for permutation methods for the GLM (Winkler *et al.*, 2014). For example, there are several different permutation schemes when testing a strict subset of all GLM regression parameters. One approach is to permute only the column of interest in the design matrix. This approach, due to Draper and Stoneman (1966) could be restated as isolating the portion of the model affected by the null hypothesis, and then only permuting that portion. This is the motivation for our first permutation strategy (P1), where we repeatedly fit the model, but randomly permute kinship each time.

Another approach is to use the reduced, null hypothesis model to generate residuals, permute these residuals, and use them as surrogate null data to be re-analyzed (Freedman and Lane, 1983). For the GLM, this is the recommended approach (Winkler *et al.*, 2014), and corresponds to an ideal test where nuisance effects are removed from the data, leaving what should be only unstructured data (under the null) ready to be permuted. This is the motivation for permutation scheme (P2).

Finally, another approach to GLM permutation testing is to use the full, alternative hypothesis model to generate residuals, and then use these residuals as surrogate null data to be re-fit (ter Braak, 1992). This approach has the merit of removing all systematic variation from the data before permutation. This is the motivation for our third and fourth strategies (P3 & P4).

Partial Model Permutation (P1) We implement approach P1 by permuting just the aspect of the model tested by the H_0. For the untransformed model this corresponds to permuting the model's covariance term to be

$$2\sigma^2_{\mathrm{A}} P\Phi P' + \sigma^2_{\mathrm{E}} I,$$

where P is one of $N!$ possible $N \times N$ permutation matrices. For the transformed model, the permutated covariance takes the form

$$\sigma^2_{\mathrm{A}} P D_{\mathrm{g}} P' + \sigma^2_{\mathrm{E}} I.$$

Null model residual permutation (P2) For P2 we generate residuals under $H_0 : \sigma^2_{\mathrm{A}} = 0$, i.e. OLS residuals $\hat{\epsilon}_{\mathrm{OLS}}$ (3.12). Then we permute these residuals, and add-back nuisance (fixed) effects to generate new H_0 realizations $\tilde{Y}^*$:

$$\tilde{Y}^* \quad = \quad X^* \hat{\beta}_{\mathrm{OLS}} + P \hat{\epsilon}^*_{\mathrm{OLS}}, \tag{3.15}$$

where the tilde ($\tilde{}$) accent denotes one of many realizations, which in turn are fit with the model under consideration.

Full model residual permutation (P3) For P3, we start with full model residuals, i.e. either $\hat{\epsilon}_{\mathrm{ML}}$ or $\hat{\epsilon}_{\mathrm{WLS}}$, depending on the estimator used. Then we permute these residuals, and add-back nuisance to generate new null hypothesis realizations; e.g., for WLS:

$$\tilde{Y}^* \quad = \quad X^* \hat{\beta}_{\mathrm{WLS}} + P \hat{\epsilon}^*_{\mathrm{WLS}}. \tag{3.16}$$

and analogously for ML. Again, each realisation $\tilde{Y}$ is fit to the current model.

Full model whitened residual permutation (P4) P4 is like P3, but we go a step further and create residuals that are whitened before permutation. For example, for WLS:

$$\tilde{Y}^* \quad = \quad P(\hat{\Sigma}^{*-1/2} \hat{\epsilon}^*_{\mathrm{WLS}}), \tag{3.17}$$

and analogously for ML. Again, each realisation is fit to the current model.

In total we have introduced five different test procedures and four permutation strategies, summarized in Table 3.2.

Multiple Testing Correction

Whether inference is conducted voxel-wise or cluster-wise, the use of use of an uncorrected $\alpha = 5\%$ level leads to an excess of false positives. False Discovery Rate (FDR) correction, controlling the expected proportion of false positives among all detections, is easily applied based on uncorrected P-values alone (Genovese *et al.*,

2002). As uncorrected permutation cluster-wise P-values require an assumption of stationarity (though see Salimi-Khorshidi *et al.* (2010)), FDR is generally only applied with voxel-wise P-values. Familywise error rate (FWE) correction, controlling the chance of one or more false positives across the whole set (family) of tests (Nichols and Hayasaka, 2003) requires the distribution the maximum statistic, easily computed for either voxels or cluster size with permutation (Nichols and Holmes, 2002).

Table 3.2: Comparison of Tests for Heritability Inference.

Tests	h^2 Estimates	Distribution	Type	Optimization	Permutation
$T_{\text{L,ML}}$	✓	50:50 χ^2_1 and 0	Asymptotic	ML	P1, P2, P3, P4
$T_{\text{W,ML}}$	✓	50:50 χ^2_1 and 0	Asymptotic	ML	P1, P2, P3, P4
$T_{\text{W,WLS}}$	✓	50:50 χ^2_1 and 0	Asymptotic	WLS	P1, P2, P3, P4
T_{S}	✓	50:50 χ^2_1 and 0	Asymptotic	OLS	P1, P2, P3, P4
T_{GQ}	✗	F_{n_2-p,n_1-p}	Exact	OLS	P1, P2, P3, P4

Proposed test procedures: The score test (T_{S}), the Wald test and its variants in term of a WLS estimators ($T_{\text{W,WLS}}$) and ML estimators ($T_{\text{W,ML}}$), and the LRTs in terms of the transformed model ($T_{\text{L,ML}}$). ML optimization denotes iterative optimization until convergence; WLS a 1-step of Newton's method; and OLS an estimate based on (unweighted) least squares.

3.2 Evaluation

3.2.1 Simulation Studies

We conduct various simulation studies to evaluate proposed methods for heritability inference on the transformed model. The first study considers estimator bias and variance for the different methods. The second study measures the accuracy of parametric and permutation inference methods. Finally, the third study evaluates FWE control in an image-wise setting for voxel and cluster-wise inferences.

In all simulations, the response variable is assumed to be $Y = X\beta + \epsilon$ where ϵ follows $N(0, \Sigma)$, $\Sigma = h^2(2\Phi) + (1 - h^2)I$. The design matrix X consists of an intercept, a linear trend vector X_1 and a quadratic vector X_2 between 1 and -1, with $\beta = [0, 0, 10]$. Kinship structure Φ is based on real pedigrees (each described below), and the simulations considered a range of true heritabilities ($h^2 = 0, 0.2, 0.4, 0.6, 0.8$).

Simulation 1 This simulation evaluates the bias, standard deviation and mean squared error (MSE) of the heritability estimators (ML and WLS). The pedigrees and sample sizes used are shown in Table 3.3; we used pedigrees from the 10th

Genetics Analysis Workshop (GAW10) (MacCluer *et al.*, 1997) and from the GOBS dataset (described below). Univariate data Y was simulated as per the Gaussian model described above, and 10,000 realizations were used.

Table 3.3: Datasets used in Simulation 1.

Datasets	Number of Pedigrees	Sample Size
GAW10	2	138
GAW10	9	626
GOBS	73	858
GAW10	23	1497

Simulation 2 This simulation assesses the false positive rates for each method, on the basis of both parametric and permutation methods. For this analysis we used 2 pedigrees from the GAW10 dataset with 138 subjects; the small sample size was used to 'stress test' the methods. Univariate data Y was simulated as per the Gaussian model described above, 10,000 realizations were used, and 500 permutations for each nonparametric procedure. On the basis of Simulation 1 and 2, 'winner' tests and a permutation strategy were chosen and fed into the 3rd simulation study.

Simulation 3 Image simulations were conducted under the null hypothesis ($h^2 = 0$) on a $96 \times 96 \times 20$ image that the response variable for each voxel are simulated as described above, smoothed with a Gaussian filter with a Full Width at Half Maximum of 4mm. To avoid edge effects, larger images were simulated, smoothed and then truncated. For each realisation we collected empirical null distributions of maximum statistic and maximum cluster size to compute FWE P-values; we considered different cluster forming thresholds (parametric uncorrected P-value$= 0.05, 0.01, 0.005, 0.001$). We generated 5,000 realizations and used 500 permutations with each synthetic dataset. A range of true signals ($h^2 = 0.2, 0.4, 0.6, 0.8$) are induced to the center of simulated images to evaluate Monte Carlo power of voxel wise and spatially informed statistics via calculating family wise true positive rate (FWTP).

3.2.2 Application in Diffusion Tensor Imaging Data

We used data from the Genetics of Brain Structure and Function Study (GOBS) (Olvera *et al.*, 2011; McKay *et al.*, 2014) to perform voxel and cluster-wise FA heritability inference in healthy subjects. The sample comprised 859 Mexican-American

individuals from 73 extended pedigrees (average size 17.2 people, range = 1247). The sample was 59% female (351 men/508 women) and had a mean age of 43.2 (SD = 15.0; range = 1985). All participants provided written informed consent on forms approved by the Institutional Review Boards at the University of Texas Health Science Center San Antonio (UTHSCSA) and Yale University.

Diffusion imaging was performed at the Research Imaging Center, UTHSCSA, on a Siemens 3 T Trio scanner using a multi-channel phased array head coil. A single-shot single refocusing spin-echo, echo-planar imaging sequence was used to acquire diffusion-weighted data with a spatial resolution of $1.7 \times 1.7 \times 3.0mm$. The sequence parameters were: TE/TR = 87/8000 ms, FOV = 200 mm, 55 isotropically distributed diffusion weighted directions, two diffusion weighting values, b = 0 and 700 s/mm2 and three b = 0 (non-diffusion-weighted) images.

ENIGMA-DTI protocols for extraction of tract-wise average FA values were used. These protocols are detailed elsewhere (Jahanshad *et al.*, 2013) and are available online `http://enigma.ini.usc.edu/protocols/dti-protocols/`. Briefly, FA images from HCP subjects were non-linearly registered to the ENIGMA-DTI target brain using FSLs FNIRT (Jahanshad *et al.*, 2013). This target was created as a minimal deformation target based on images from the participating studies as previously described (Jahanshad et al., 2013b). The data were then processed using FSLs tract-based spatial statistics (TBSS; `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS`) analytic method (Smith *et al.*, 2006) modified to project individual FA values on the hand-segmented ENIGMA-DTI skeleton mask. The protocol, target brain, ENIGMA-DTI skeleton mask, source code and executables, are all publicly available (`http://enigma.ini.usc.edu/ongoing/dti-working-group/`). The FA values are normalized across individuals by inverse Gaussian transform (Servin and Stephens, 2007; Allison *et al.*, 1999) to ensure normality assumption. Finally, we analyzed the voxel and cluster-wise FA values with applying along the ENIGMA skeleton mask. To validate our proposed methods for heritability estimation and inference for imaging data, we applied them on GOBS dataset.

3.3 Results

3.3.1 Univariate Heritability Simulation Results

Simulation 1 Figure 3.1 compares WLS and ML heritability estimators for various designs and effect sizes, in terms of mean, standard deviation (SD) and mean squared error (MSE), for 10,000 Monte Carlo realizations. Large sample theory dictates that ML should provide best performance, and indeed it has least bias and

smallest standard deviation, but the (non-iterative) WLS has MSE's that are only slightly larger. As expected, when the sample size is increased WLS and ML heritability estimators reach almost same performance. While the WLS estimator bias is worse (more negative) than that of ML, the absolute magnitude of bias is small in large samples.

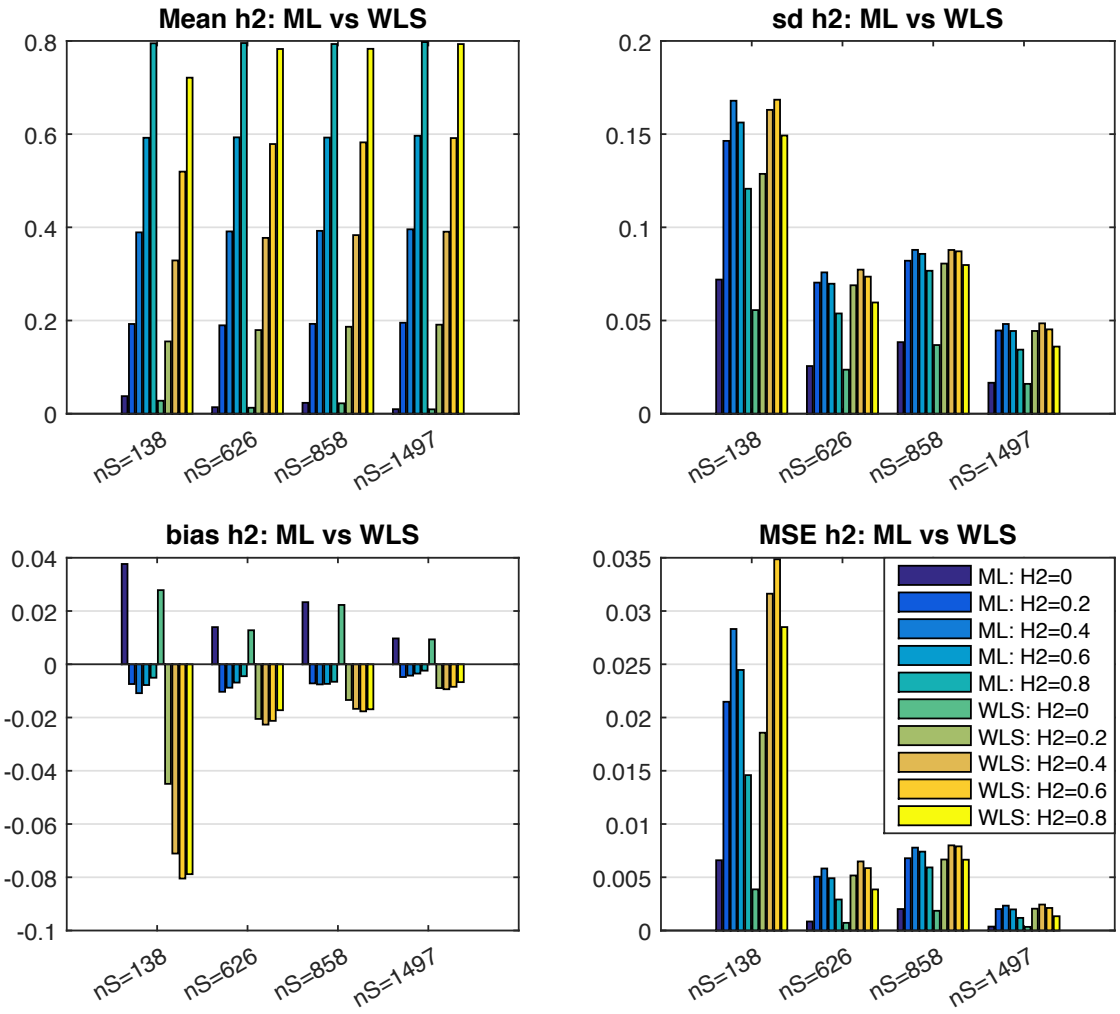


Figure 3.1: Simulation 1 results, comparing ML and WLS behaviour in terms of mean estimate (top left; true h^2 varies on abscissa within clusters), standard deviation (SD; top right), bias (lower left), and mean squared error (MSE; bottom right). See Table 3.3 for details of each pedigree; nS denotes number of subjects. WLS has worse bias than ML, but small in absolute magntidue, leading to quite similar MSE for large samples.

Simulation 2 This simulation assesses the accuracy of parametric null distributions, either a 50:50 χ^2 mixture or F distribution, and power. Under H_0,

all false positive rates (Table 3.4) are conservative except T_{GQ}. The LRT and score tests have Type I error rates that are closer to the nominal level than the Wald tests for the simulated null data ($h^2 = 0$) but none of them in the MC confidence interval (4.57% – 5.42%). Also, the WLS Wald tests had lower error rates than ML Wald tests. In terms of power, the same pattern exists between tests and the LRT and T_{GQ} are the most powerful ones.

The conservative false positive rates are attributable to asymptotic null distributions. In particular, the 50:50 mixture approximation has recently been shown to be conservative, which we confirm here. On the other hand, parametric null distribution of T_{GQ} does not depend on a mixture approximation and, under a normality assumption, it follows F-distribution exactly; this is likely why GQ had the most accurate false positive rate (4.36%).

Table 3.4: Simulation 2 result, comparing parametric rejection rates (percent), 5% nominal. For GAW10 data with 2 families, 138 subjects, 10,000 realizations. GQ test has most accurate false positive rate, LRT with ML ($T_{\mathrm{L,ML}}$) is most powerful; both GQ (T_{GQ}) and score (T_{S}) test have good power (95% MC CI for 0.05, i.e. for the null case is (4.57%, 5.42%)).

	True Effect (h^2)				
Test	0	0.2	0.4	0.6	0.8
T_{S}	3.76	40.66	76.76	94.32	98.94
$T_{\mathrm{W,WLS}}$	1.56	26.94	73.46	95.62	99.64
$T_{\mathrm{W,ML}}$	2.50	33.00	77.74	94.84	97.54
$T_{\mathrm{L,ML}}$	3.16	42.28	81.80	96.40	98.90
T_{GQ}	4.36	35.60	78.22	96.50	99.70

Figures 3.2 and 3.3 show the performance of permutation inference, with rejection rates and power for different effect sizes under the various permutation strategies. Figure 3.2 shows that, generally permutation strategy P1 is more conservative than P2, P3 and P4. Moreover the error rates in terms of P2 is close to the nominal level. Although the permutation strategy P4 has higher rejection rates, they still fall within the Monte Carlo confidence interval (4.57%–5.43%) except for $T_{\mathrm{W,ML}}$.

With respect to power, Figure 3.3 shows that again P2, P3 and P4 are generally superior to P1 for various effect sizes. In addition P2, P3 and P4 have almost same performance, all within the Monte Carlo confidence bounds.

Based on all of these results, we selected T_{S}, $T_{\mathrm{W,WLS}}$ and T_{GQ} and P2 as the computationally most efficient tests to be considered in the image-wise simulations.

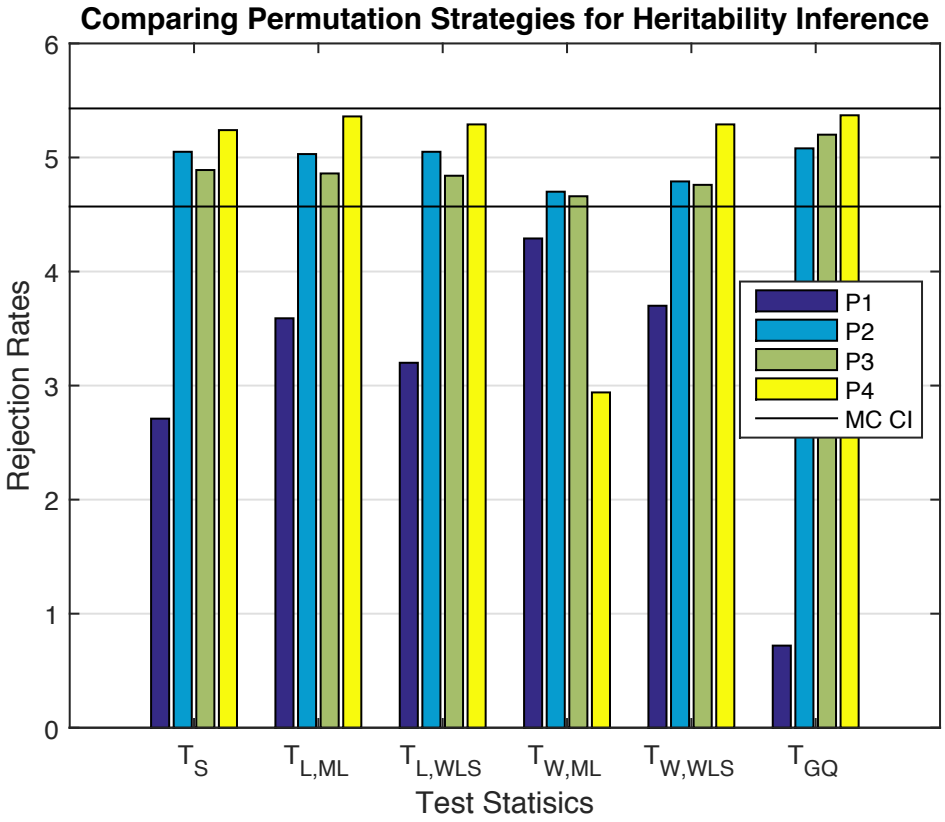


Figure 3.2: Simulation 2 results, false positive rates for heritability permutation inference, 5% nominal. Based on GAW10 data with 2 families, 138 subjects, 10,000 realizations, 500 permutations each realisation. Monte Carlo confidence interval (MC CI) is (4.57%,5.43%). Permutation schemes P2-P4 generally seem to work well, while $T_{\mathrm{W,ML}}$ tends to be conservative.

3.3.2 Image-wise Simulation Results

Simulation 3 This simulation evaluates false positive rate control in the more challenging image-wise setting, for both voxel and cluster-wise heritability inference. Figure 3.4 shows the P-P plot of uncorrected P-values, plotted as $-\log_{10} P$-values. Except for modest conservativeness ($P \approx 10^{-2.5}$), and of course the truncation due to limited permutations (500 permutations, minimal P-value of 0.002, maximum $-\log_{10} P$-value of 2.69), the accuracy is quite good over-all. Figure 3.5 show that FWE-corrected P-values are also accurate, with slight conservativeness with the GQ test. For the 5% level specifically, voxel-wise FWE for the score, the Wald and the GQ tests were 5.08%, 5.44% and 5.4% respectively, well within the Monte Carlo 95% CI, (4.40%–5.60%).

Figure 3.6 shows cluster-wise FWE rates for different cluster forming thresh-

olds. All rates are nominal except for the higher cluster forming thresholds of $T_{\mathrm{W,WLS}}$ (P=0.005 & P=0.001). The cluster-forming thresholds come from the parametric null distribution, and Figure 3.4 shows severe conservativeness for $T_{\mathrm{W,WLS}}$'s parametric P-values. For example, that Figure shows that when a P=0.001 uncorrected threshold is used for $T_{\mathrm{W,WLS}}$, the actual false positive rate is less than 0.0001. This effect, combined with variation of effective false positive rate of the cluster-forming threshold over permutations, could explain this slight anticonservativeness. Figure 3.7 compares the selected tests maximum cluster size P-values based on different cluster forming thresholds with their theoretical values; again $T_{\mathrm{W,WLS}}$ behaviour for large cluster forming thresholds shows slightly inflated rejection rates.

3.3.3 Real Data Analysis

Voxel-wise FA heritability estimation and inference for the GOBS study are shown with ML and WLS estimators, creating four test statistic images: $T_{\mathrm{L,ML}}$, T_{S}, $T_{\mathrm{W,WLS}}$, and T_{GQ}; permutation scheme P2 was used to compute uncorrected and FWE-corrected P-values. Figure 3.8 shows histograms of h^2_{ML} (top) and h^2_{ML} (bottom), showing generally the same distribution of heritability over the white matter skeleton. Figure 3.10 shows h^2 estimates on the TBSS skeleton. Figure 3.9 directly compares WLS and ML heritability estimates with a scatter plot, showing a slight but consistent trend towards underestimation of h^2_{WLS} relative to h^2_{ML}, consistent with simulation (Fig. 3.1).

Voxel-wise uncorrected $-\log_{10}$ P-values from T_{S}, $T_{\mathrm{W,WLS}}$, T_{GQ} and $T_{\mathrm{L,ML}}$ based on P2 are compared in Figure 3.11. Considering $T_{\mathrm{L,ML}}$ as a reference (on the abscissa), $T_{\mathrm{W,WLS}}$ and T_{GQ} are generally less sensitive than $T_{\mathrm{L,ML}}$ (Figure 3.11 middle and right panels), consistent with the simulations above. However, T_{S} was more comparable with $T_{\mathrm{L,ML}}$ (Figure 3.11 left panel). Level 5% FWE-corrected statistic thresholds for T_{S}, $T_{\mathrm{W,WLS}}$, $T_{\mathrm{L,ML}}$ and T_{GQ} are 39.92, 18.31, 24.27 and 1.72, respectively, producing significant voxel counts of 8521, 1043, 7418 and 2446, respectively, out of 117,139 voxels.

Cluster-wise inference results for cluster forming thresholds corresponded to uncorrected P-value=0.01% are shown in Table 3.5 the tests that we consider. Level 5% FWE-corrected cluster size thresholds for T_{S}, $T_{\mathrm{W,WLS}}$, $T_{\mathrm{L,ML}}$ and T_{GQ} are 265, 98, 142 and 135 voxels, respectively. For voxel-wise inference, Figure 3.12, the score test was most similar to ML's LRT, and likewise for cluster-wise inference, Figure 3.13.

Table 3.5: Real data results, cluster-wise inferences with different methods.

Method	Total # Clusters	#Significant Clusters	Largest Cluster Size	Smallest Corrected P-value
$T_{\mathrm{L,ML}}$	1770	22	24246	0.0005
$T_{\mathrm{W,WLS}}$	1725	19	3643	0.0003
T_{S}	1689	11	31250	0.0003
T_{GQ}	1751	20	4383	0.0003

Cluster-wise Inference for $T_{\mathrm{L,ML}}$, $T_{\mathrm{W,WLS}}$, T_S and T_{GQ}. Based 858 subjects from GOBS and 3,000 permutations.

Table 3.6: Computation times. Comparison of running times for a dataset with 138 subjects, 2 families, (GAW10 kinship) and 184,320 voxels. Run on Intel(R) core(TM) i7-2600 CPU @ 3.4 GH and 16 GB RAM.

Statistics	Univariate Trait	Image-wise Trait
$T_{\mathrm{L,ML}}$	1 Sec	8 Hr
$T_{\mathrm{W,WLS}}$	0.005 Sec	2 Sec
T_{S}	0.005 Sec	2 Sec
T_{GQ}	0.004 Sec	1.5 Sec

3.4 Discussion & Conclusions

We have proposed a number of computationally efficient tests for heritability with family data. To our knowledge this is the first work that enable practitioners to study brain phenotypes heritability in each voxel without confronting an intense computational burden. Our methods are based on the eigensimplified model of Blangero *et al.* (2013), most of which can be implemented with auxiliary models, corresponding to regressing squared OLS residuals on the kinship matrix eigenvalues.

For heritability estimation our WLS method, based on one step of Newton's method, was a fast and reasonable approximation to fully iterated ML, ideal for application to brain image data.

For heritability inference, we found that parametric P-values for LRT, Wald and score methods were all conservative, likely due to the untenable i.i.d. assumption underlying the 50:50 χ^2 mixture approximation. As an alternative, permutation test error rates were much closer than parametric one to the nominal level. Notably, all of our simulations included fixed effects covariates (X).

The GQ heteroscesdicity test, adapted here for heritability detection, had good performance in simulation, with the best false positive control and respectable power, but on the real data was dramatically different (see Fig 3.13(d)) and apparently less powerful.

Image wise simulation results showed FWE-corrected voxel- and cluster-wise

inference was valid at the 5% level for T_{S} and T_{GQ}, permutation scheme P2. In real data, the P-values for T_{GQ} were less similar to the LRT results than the score or Wald test, and was less sensitive over all. The GQ test's power depends on the cut point used to define the two groups, though we did not investigate further. On balance we suggest the use of T_{S} for standard neuroimaging inference tool including voxel and cluster-wise inference.

Running time for different test statistics that were presented in Table 3.6 based on a benchmark with Intel(R) core(TM) i7-2600 CPU @ 3.4 GH and 16 GB RAM feature confirms that the empirical null distribution of explained sum of squares of auxiliary model (T_{S}) under the permutation scheme P2 can be derived substantially faster than $T_{\mathrm{L,ML}}$, the classic test statistic for heritability inference. Although the sample size plays an important role in running time, we believe that T_{S} can be derived significantly faster than the other tests, since it does not depend on numerical optimization. Hence, the whole permutation distribution can be derived easily, either for a univariate trait or a multivariate spatially dependent neuroimaging data accounting explicitly for family wise error.

Finally, we note that yet-more computationally efficient estimates can be obtained by conditioning on the over-all variance estimate, $\hat{\sigma}^2$, which leads to a 1-parameter variance model. However, in initial simulations we found this lead to greater bias in h^2 and specifically h^2 estimates in excess of 1.0. Thus we retained the 2-parameter variance model.

In conclusion, our results present a novel inference techniques to be implemented in the genetic imaging analysis software like SOLAR-Eclipse (`http://www.nitrc.org/projects/se_linux`). These methods provide fast inference procedure on millions of phenotypes, filtering a small number of elements for further investigation with more computational intense tools. In future work we will extend these tools for inference on covariates, in particular permutation-based tests for voxel-wise GWAS analysis for family based data.

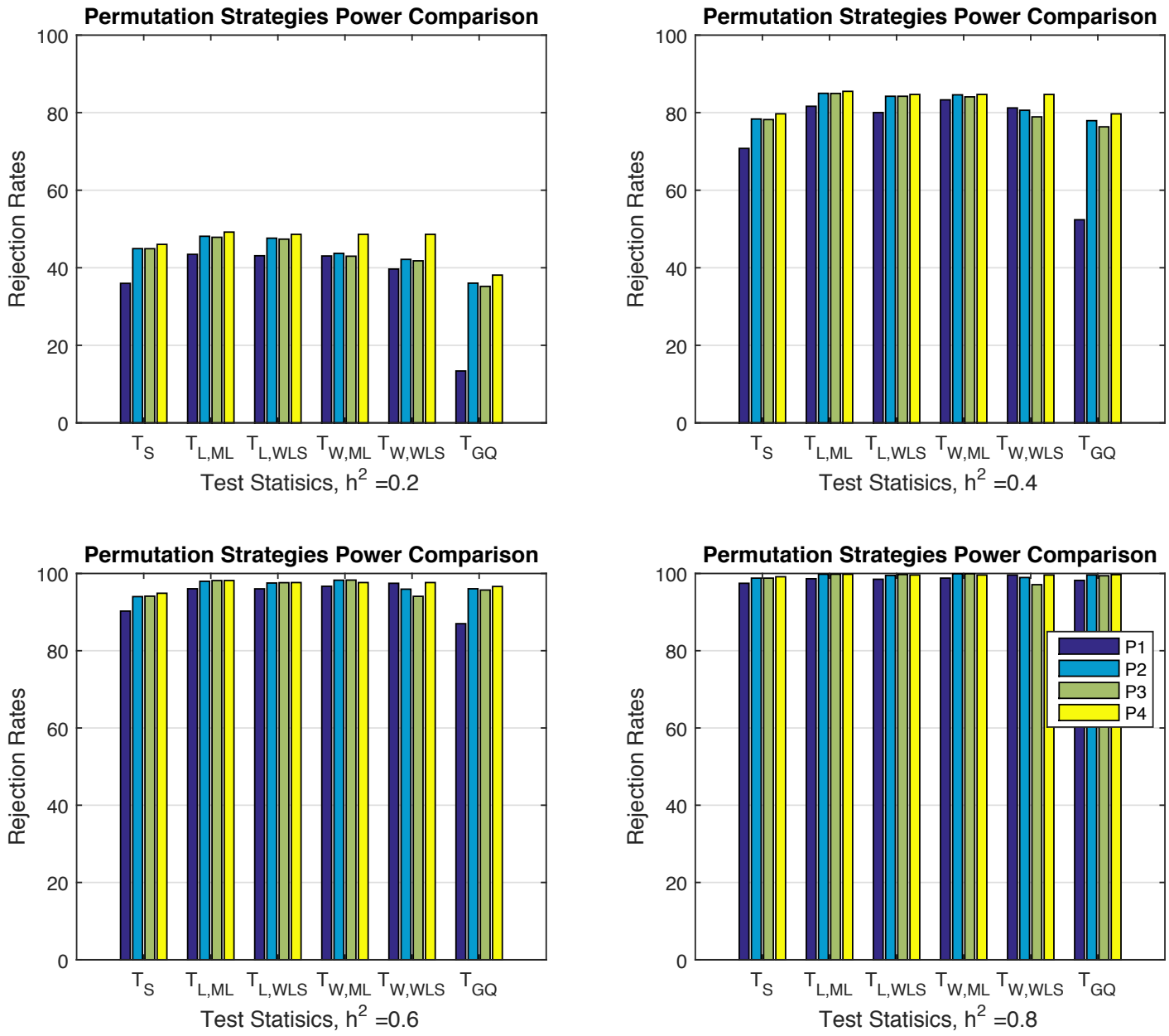


Figure 3.3: Simulation 2 results, power for heritability permutation inference. For GAW10 data with 2 families, 138 subjects, 10,000 realizations, 500 permutations each realisation. Monte Carlo confidence interval varies with true rejection rate; for 40% it is (39.0%,41.0%), for 80% it is (79.2%,80.8%)

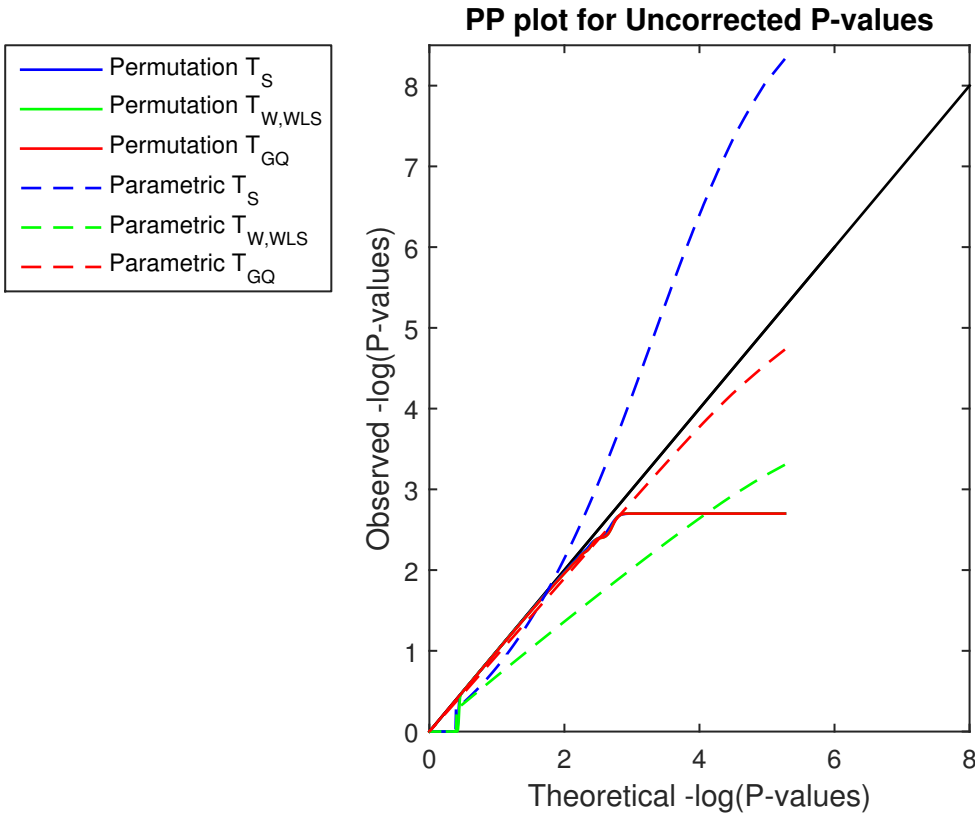


Figure 3.4: Simulation 3 results, $-\log_{10}$ PP Plot for uncorrected parametric and permutation P-values for our proposed test statistics. Permutation P-values are valid (solid lines), though are bounded below by 1/500 (above by 2.70 in $-\log_{10} P$), the smallest possible permutation P-value for the 500 permutations used. The permutation P-values are overplotted here, and only the permutation T_{GQ} is visible. Parametric P-values for the non-asymptotic GQ test (dashed red line) perform well, while the parametric score test's P-values (dashed blue line) are severely anticonservative (invalid) and Wald test P-values (dashed green line) are severely conservative. Different behavior is seen for P-values larger than 0.5 (smaller than 0.70 in $-\log_{10} P$) as tests giving $\approx 50\%$zero values produce $\approx 50\%$ P-values of 1 (0 in $-\log_{10} P$). Results based on GAW10 data with 2 families, 138 subjects, 5,000 realizations, 500 permutations each realisation, and $96 \times 96 \times 20$ images with 4mm FWHM smoothing.

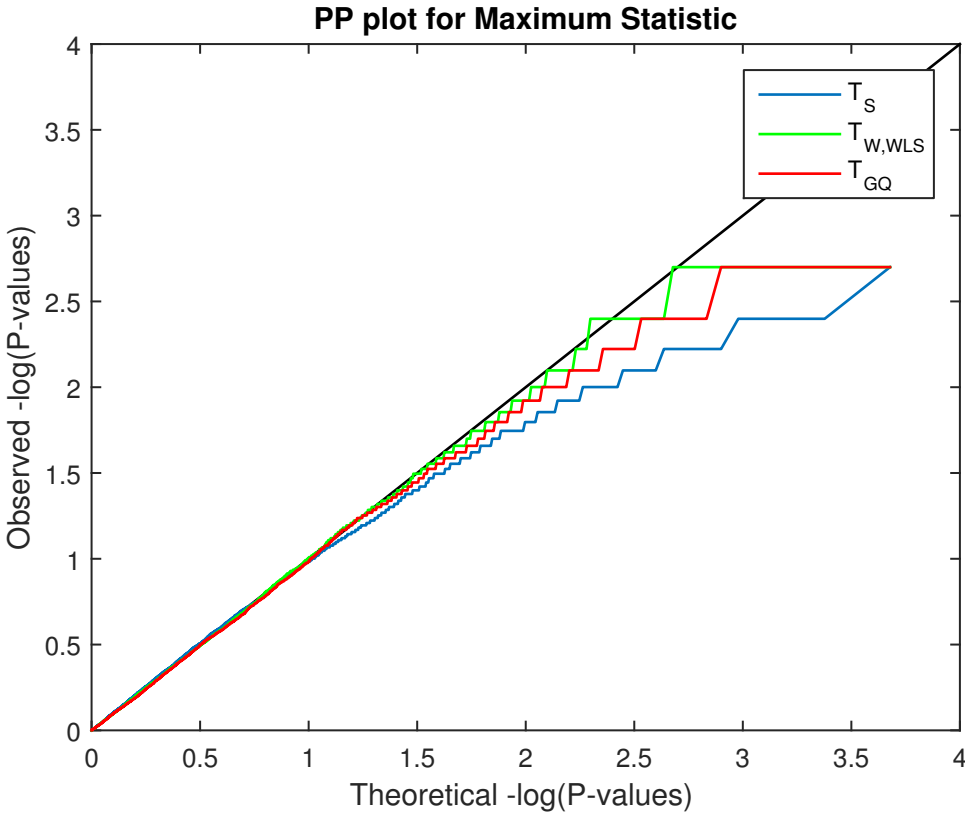


Figure 3.5: Simulation 3 results, $-\log_{10}$ PP plot for voxel-wise FWE permutation P-values under the null hypothsis, for three of our proposed test statistics. Each FWE P-value is for the maximum voxel-wise test statistic in each realised dataset. All three test statistics produce valid P-values, though are bounded below by 1/500 (above by 2.70 in $-\log_{10} P$). The Wald test's FWE it slightly conservative, and score a bit more so. Results based on GAW10 data with 2 families, 138 subjects, 5,000 realizations, 500 permutations each realisation, and $96 \times 96 \times 20$ images with 4mm FWHM smoothing.

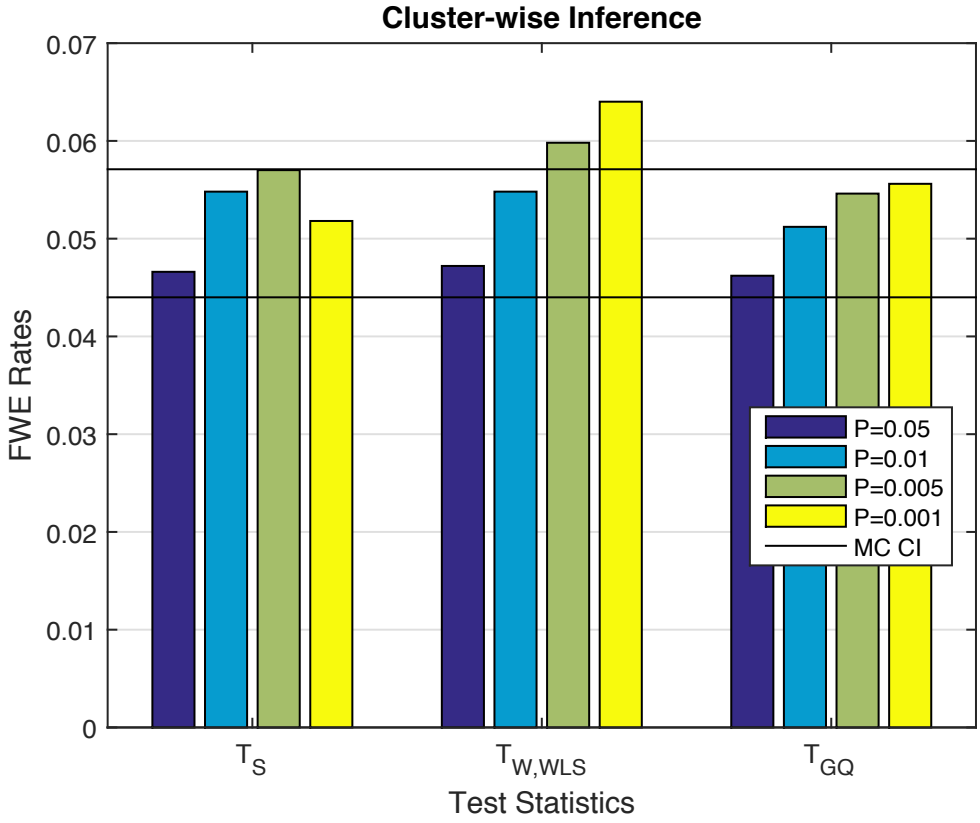


Figure 3.6: Simulation 3 results, FWE error rates for cluster-wise permutation heritability inference under the null hypothesis, for three of our proposed test statistics. Score and GC test have nominal false positive rates, while the Wald test is anticonservative for high (uncorrected P of 0.005 & 0.001) clustering forming thresholds. This is likely due to use of parametric cluster-forming threshold; see text for more discussion. Results based on GAW10 data with 2 families, 138 subjects, 5,000 realizations, 500 permutations each realisation. Monte Carlo 95% confidence interval (4.40%,5.60%).

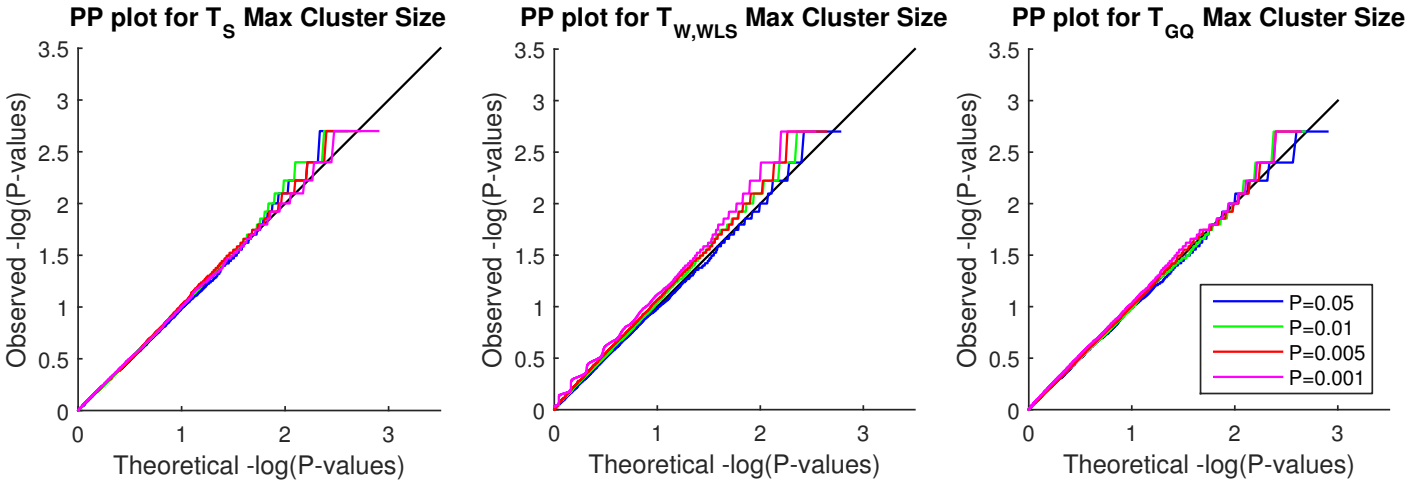


Figure 3.7: Simulation 3 results, $-\log_{10}$ PP plots for cluster-wise FWE permutation P-values under the null hypothesis, for three of our proposed test statistics. Each FWE P-value is for the maximum cluster size in each realised dataset. GQ has most accurate FWE P-values, followed by the score test; Wald is slightly anticonservative for high cluster forming thresholds; see text for discussion. For GAW10 data with 2 families, 138 subjects, 5,000 realizations, 500 permutations each realisation (MC CI=(4.40,5.60)).

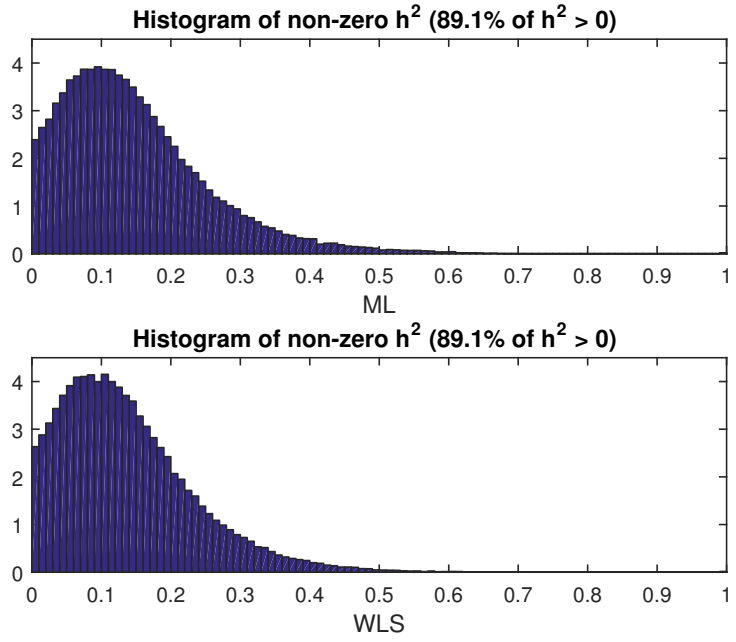


Figure 3.8: Real data results, comparison of voxel-wise heritability estimates from ML and WLS estimates. The histograms show that the estimates from the two methods are largely similar.

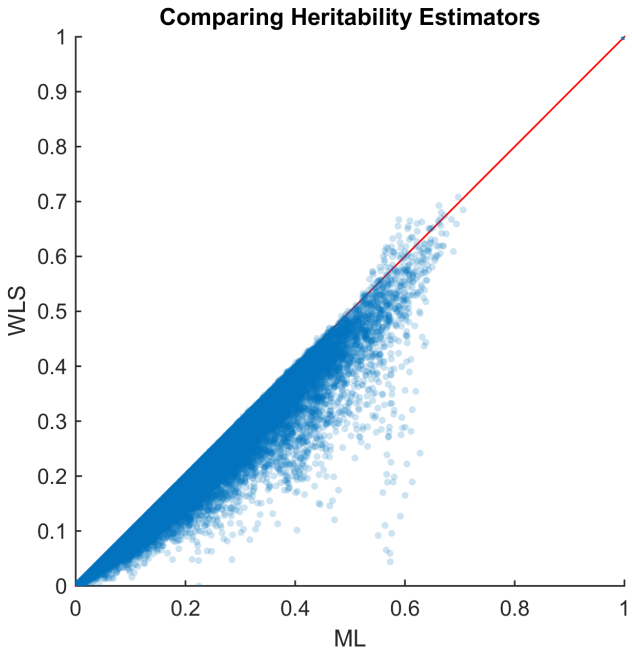
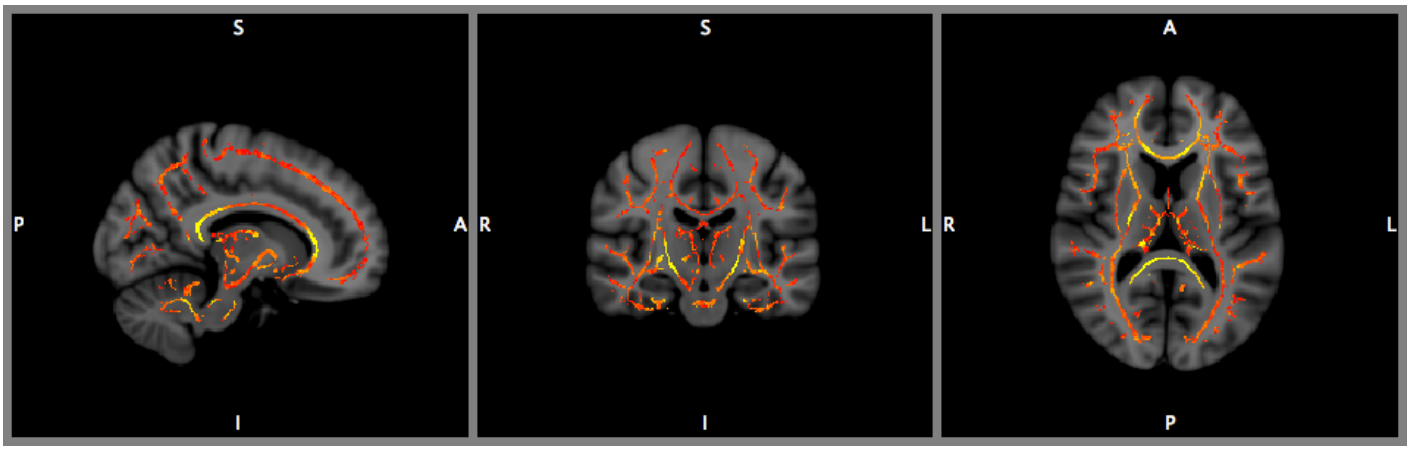
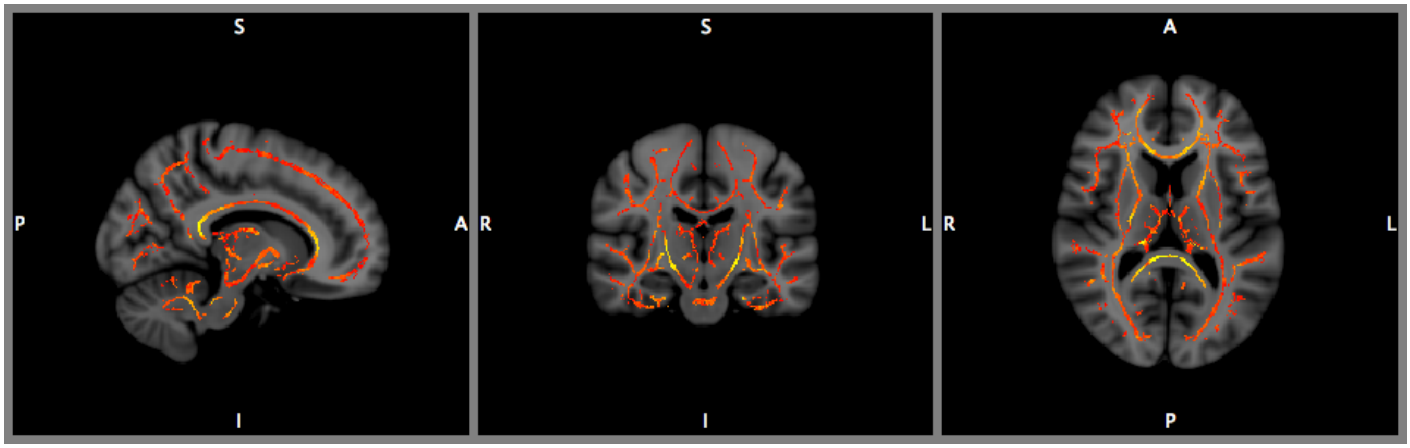


Figure 3.9: Real data results, scatterplot of voxel-wise heritability estimates from ML and WLS estimates. The two methods are largely similar, though ML is almost always larger than WLS estimates.



(a) Voxel wise ML heritability Estimation



(b) Voxel wise WLS heritability Estimation

Figure 3.10: Real data results, voxel-wise heritability estimates for ML (top) and WLS (bottom). Heritability shown in hot-metal color scale, intensity range [0,0.5] for both, overlaid on MNI reference brain. Differences only apparent in highest FA areas.

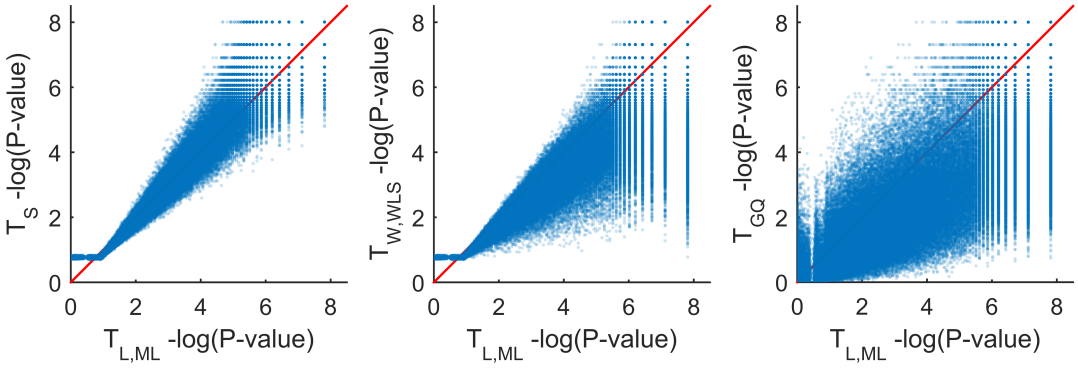
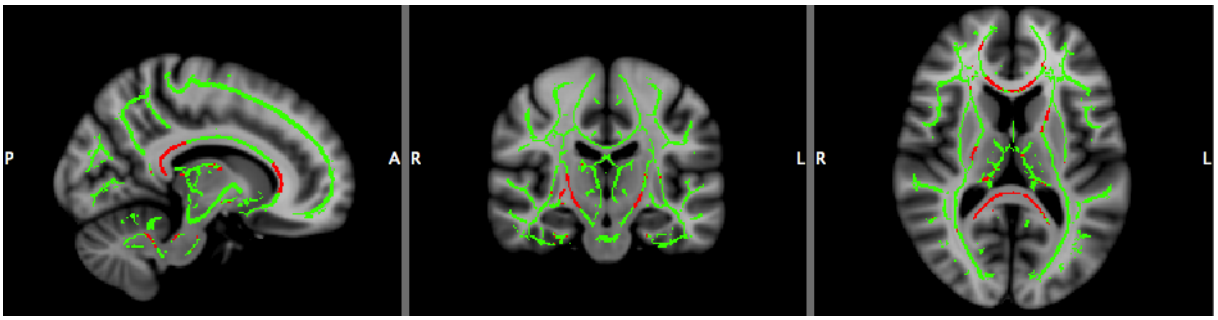
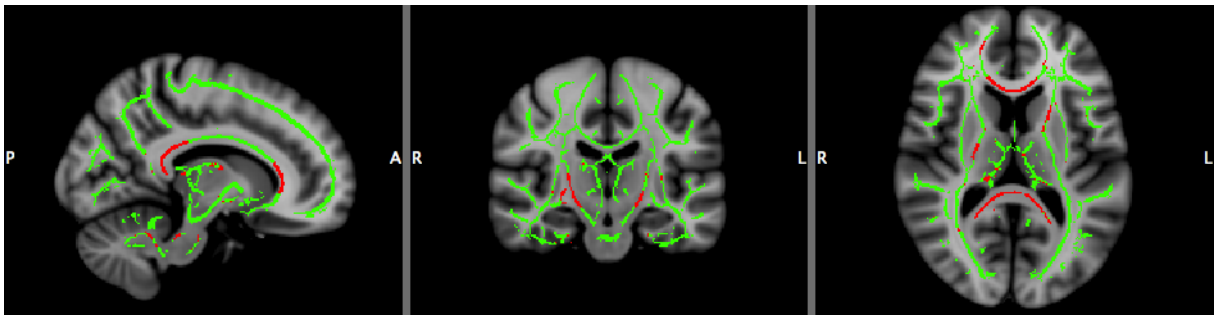


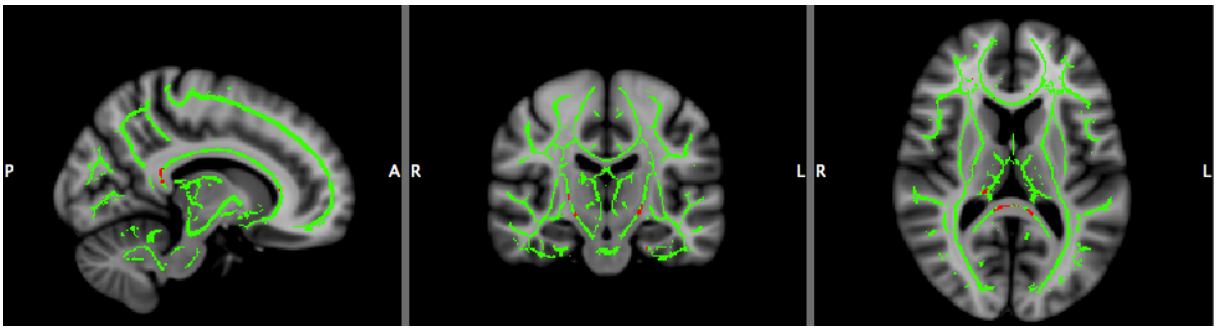
Figure 3.11: Real data results, scatter plots of voxel-wise uncorrected $-\log_{10} P$-values for score, WLS Wald and GQ tests vs. the ML LRT test. Score P-values are most faithful representation of the ML LRT P-values, while WLS Wald P-values tend to be more conservative; GQ P-values are much more different and generally more conservative.



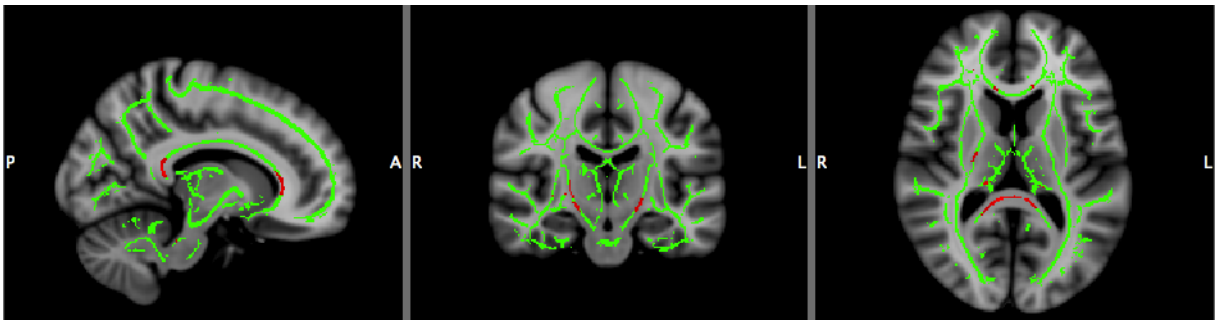
(a) LRT for ML estimator ($T_{\mathrm{L,ML}}$)



(b) Score Test (T_S)

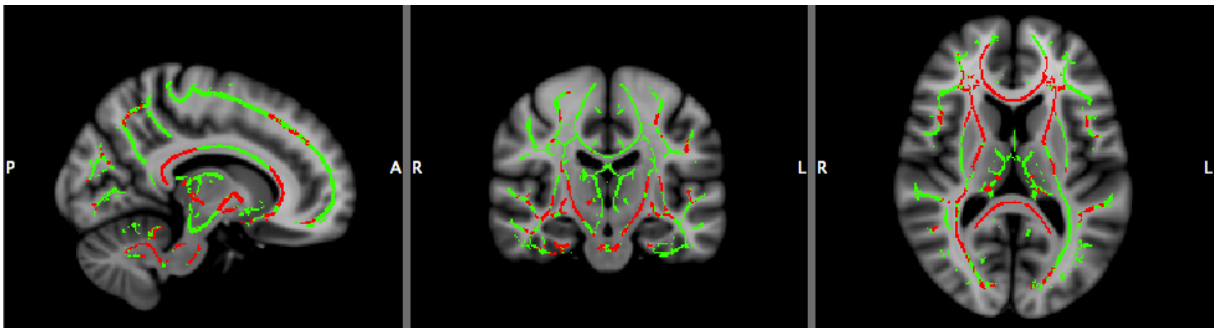


(c) WALD test for WLS estimator ($T_{\mathrm{W,WLS}}$)

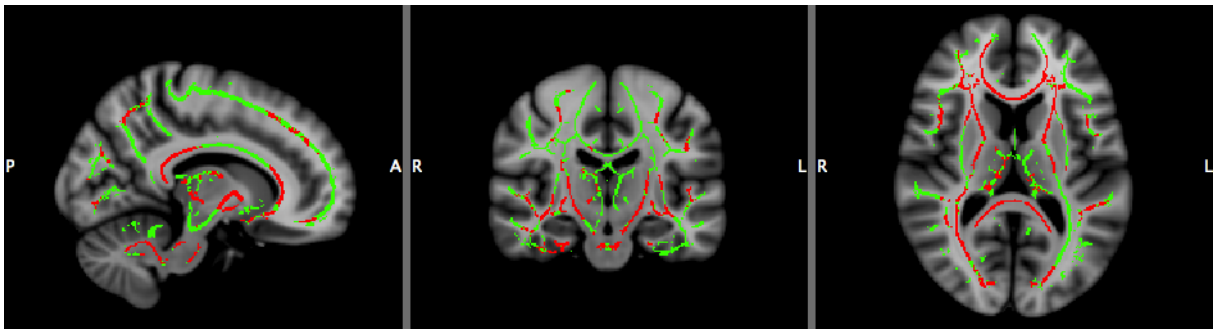


(d) GQ Test (T_{GQ})

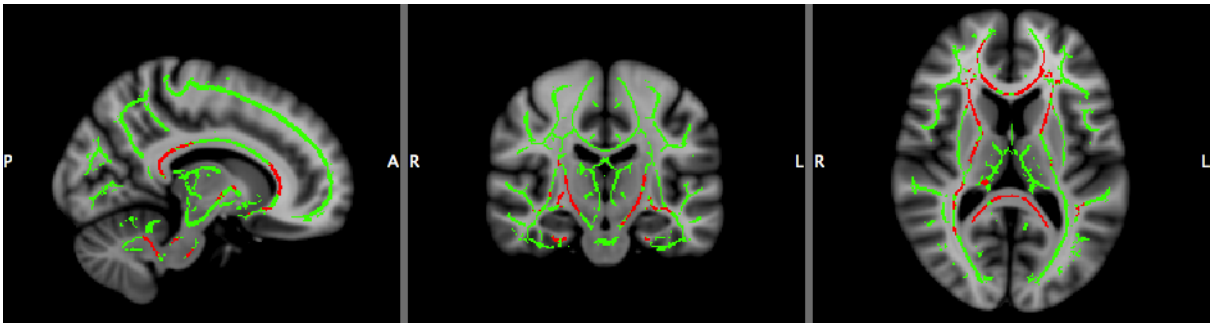
Figure 3.12: Real data results, voxel-wise 5% FWE significant heritability, for 4 different methods. Full skeleton and significant voxels are in green and red, respectively. The non-iterative score test gives very similar results to the ML (fully iterated) LRT, with the other 2 methods being less sensitive.



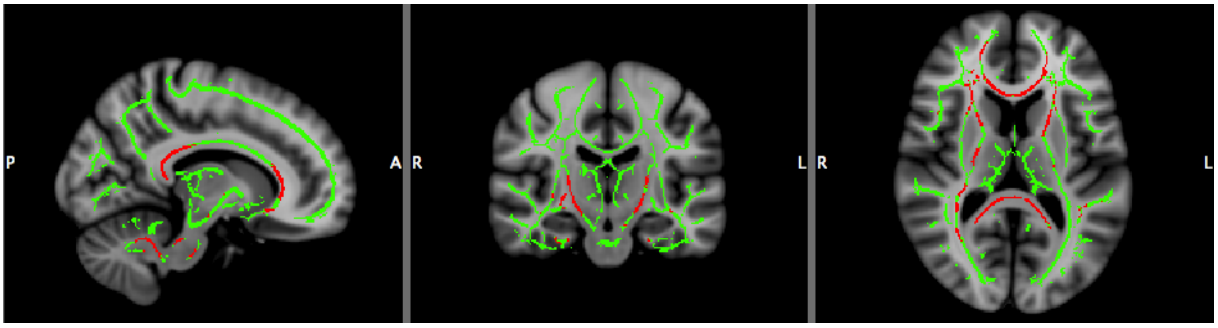
(a) LRT for ML estimator ($T_{\mathrm{L,ML}}$)



(b) Score Test (T_S)



(c) WALD test for WLS estimator ($T_{\mathrm{W,WLS}}$)



(d) GQ Test (T_{GQ})

Figure 3.13: Real data results, cluster-wise 5% FWE significant heritability, for 4 different methods, cluster-forming threshold parametric uncorrected P=0.01. Full skeleton and significant voxels are in green and red, respectively. Methods appear more similar, but again the non-iterative score test is most similar to the ML LRT result.

Chapter 4

Fast and Powerful Genome-wide Association Analysis

Genome wide association (GWA) analyses of brain imaging phenotypes are used to advance our understanding of the biological basis of psychiatric and neurological disorders. As discussed in the section 2.3, the current trend in GWA is the use of a linear mixed effect model (LMM) that can account for cryptic/family relatedness and population stratification. However, using mixed effect models for high-dimensional imaging phenotypes GWA presents enormous challenges in terms of computational intensity and the need to account multiple testing in both the imaging and genetic domain. Here we present method that makes mixed models practical with high-dimensional traits by a combination of a transformation applied to the data and model, and the use of a non-iterative variance component estimator. With such speed enhancements permutation tests are feasible, which allows family-wise error control for any type of inference, and specifically inference on powerful spatial tests like the cluster size statistic.

To our knowledge, the fastest implementation of LMM, FaST-LMM (Lippert *et al.*, 2011a), transforms the phenotype and LMM model with genetic similarity matrix (GSM) eigenvector matrix following covariance matrix re-parametrisation based on the ratio of residual to the genetic variance components (see section 2.3.1 for more details). The eigenvector matrix diagonalisation along with the covariance matrix re-parametrisation based on a single parameter speeds up the optimisation time substantially. In FaST-LMM the covariance matrix is estimated only under the null hypothesis of no marker effect, and then a generalized least squares (GLS) is applied to estimate the marker effect and the likelihood ratio test is used for hypothesis testing. While small sample size behaviour of this approach has not

been validated, using it for association analysis of imaging phenotypes with only, say 100, subjects might not be valid. In addition to concerns about the finite sample validity, FaST-LMM requires numerical optimisation for each phenotype that make it computationally intensive or essentially impractical for large-scale imaging phenotypes.

The key to our method is the projection of the phenotype and the model to a lower dimension space, along with a score test for inference. This projection is based on the eigenvectors of the adjusted GSM for the fixed effect nuisance terms. In this setting, the projected phenotype likelihood function is equivalent to that used with restricted maximum likelihood (REML) of the LMM. While both models have the same statistical properties, our particular projection provides several computational benefits that reduces LMM complexity dramatically as follows. As we described in chapter 3 , the diagonalised covariance allows a non-iterative 1-step variance component estimator, taking the form of a weighted regression of squared projected data on eigenvalues of the GSM adjusted for nuisance fixed effect terms; going forward we call this approach WLS REML. In our previous work, this non-iterative variance component estimation approach provided a 10^5 fold speed up over the fully converged iterative approach. Further speed ups are available in the GWA setting, as the regression of the 1-step estimation, applied over entire images and calculating the score test over many SNPs can be vectorised; in other words, variance component estimation for entire images and score test calculation for entire chromosomes are formulated as basic matrix operations that are rapidly executed by high-level languages like Matlab. Finally, the simplicity and fast evaluation of the score test makes permutation computationally feasible, allowing us to compute maximum statistic null distributions to control the FWE, accounting for the number of tests conducted over all image elements and markers. In addition, two permutation schemes can be defined, free and restricted, where in the latter case the permutation is restricted to exchangeability blocks defined based on the eigenvalues distribution.

The remainder of this chapter is organized as follows. In the next section we detail the statistical model used and describe each of our proposed methods. The simulation framework used to evaluate the methods, and the real data analysis used for illustration are described in evaluation section. We then present and interpret results, and offer concluding remarks.

4.1 Theory

At each voxel/element, a LMM for the genetic association for N individuals can be written as:

$$Y = X_1\beta_1 + X_2\beta_2 + g + \epsilon, \tag{4.1}$$

where Y is the N-vector of the measured phenotype; X_1 is a N-vector of a given marker's minor allele count, implementing an additive genetic model; X_2 is the $N \times (P-1)$ matrix containing an intercept and fixed effect nuisance variables like age and sex; β_1 is the scalar genetic effect; β_2 is the $(P-1)$-vector of nuisance regression coefficients and g is the N-vector of latent (unobserved) additive genetic effects; and ϵ is the N-vector of residual errors. The trait covariance, $\text{var}(Y) = \text{var}(g + \epsilon) = \Sigma$ can be written

$$\Sigma = \sigma_A^2(2\Phi) + \sigma_E^2 I, \tag{4.2}$$

where σ_A^2 and σ_E^2 are the additive genetic and the environmental variance components, respectively; I is the identity matrix; and 2Φ is the GSM matrix where element (i, j) is calculated as:

$$\phi_{i,j} = \frac{1}{M}\sum_{k=1}^{M} \frac{(x_{ik} - 2p_k)(x_{jk} - 2p_k)}{2p_k(1-p_k)},$$

where x_{ik} is the minor allele count of the i-th subject's k-th marker, coded as coded as 0, 1 or 2; p_k is frequency of the k-th marker; and M is the total number of markers.

Under the assumption that the the data follows a multivariate normal distribution, the model specified by Equations (4.2) and (4.1) have a log-likelihood of

$$\ell_{\text{ML}}(\beta_{\text{ML}}, \Sigma_{\text{ML}}; Y, X) = -\frac{1}{2}\left[\text{Const} + \log(|\Sigma|) + (Y - X\beta)'\Sigma^{-1}(Y - X\beta)\right], \tag{4.3}$$

and a Restricted Maximum likelihood (REML) function of

$$\ell_{\text{REML}}(\Sigma_{\text{REML}}; Y, X) = -\frac{1}{2}[\text{Const} - \log|X'X| + \log|\Sigma| + \log|X'\Sigma^{-1}X| + Y'PY], \tag{4.4}$$

where $X = [X_1 \quad X_2]$ and $\beta = [\beta_1 \quad \beta_2]$ are the full design matrix of fixed effects and

their parameter estimate vector, respectively, and $P = \Sigma^{-1}\left(I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right)$, the projection matrix. The fixed effect parameters are estimated using generalized least squares (GLS)

$$\hat{\beta}_{\text{REML}} = (X'\hat{\Sigma}^{-1}_{\text{REML}}X)^{-1}X'\hat{\Sigma}^{-1}_{\text{REML}}Y,$$

where $\hat{\Sigma}^{-1}_{\text{REML}}$ comes from optimised REML function (Eq. (4.5)).

Several algorithms have been proposed to accelerate ML or REML optimisation by transforming the model with the eigenvectors of the GSM and/or using a different covariance matrix parametrisation (Ganjgahi *et al.*, 2015; Blangero *et al.*, 2013; Lippert *et al.*, 2011a; Kang *et al.*, 2008). Here we consider standard additive model covariance matrix parametrisation (Eq. (4.2)) as we can efficiently estimate it with our 1-step, regression based approach (Ganjgahi *et al.*, 2015).

4.1.1 Simplified REML Function

The simplified ML function for LMM is discussed in chapter 3. For completeness, we review shortly the simplified ML function, to be next followed by development of the simplified REML function. The simplified ML function is obtained by transforming the data and model with an orthogonal transformation S, the matrix of eigenvectors of 2Φ that crucially coincide with the eigenvectors of Σ:

$$S'Y \quad = \quad S'X\beta + S'g + S'\epsilon$$

which we write as

$$Y^* \quad = \quad X^*\beta + g^* + \epsilon^*, \tag{4.5}$$

where Y^* is the transformed data, X^* is the transformed covariate matrix, g^* and ϵ^* are the transformed random components. The diagonalising property of the eigenvectors then gives a simplified form for the variance:

$$\text{var}(\epsilon^*) = \Sigma^* = \sigma^2_{\text{A}}D_{\text{g}} + \sigma^2_{\text{E}}I,$$

where Σ^* is the variance of the transformed data and $D_{\text{g}} = \text{diag}\{\lambda_{\text{g}i}\}$ is a diagonal matrix of the eigenvalues of 2Φ.

The log likelihood takes on the exact same form as Equation (4.3) for Y^*,

X^*, β and Σ^*, except is easier to work with since Σ^* is diagonal:

$$\ell_{\mathrm{ML}}(\beta, \sigma_{\mathrm{A}}, \sigma_{\mathrm{E}}; Y^*, X^*) = -\frac{1}{2}\left[N\log(2\pi) + \sum_{i=1}^{N}\log(\sigma_{\mathrm{A}}^2\lambda_{\mathrm{g}i} + \sigma_{\mathrm{E}}^2) + \sum_{i=1}^{N}\frac{(y_i^* - x_i^*\beta)^2}{\sigma_{\mathrm{A}}^2\lambda_{\mathrm{g}i} + \sigma_{\mathrm{E}}^2}\right],$$

where y_i^* is the i-th element of Y^*, and x_i^* is the i-th row of X^*.

It is clear from Equation (4.5) that computational burden of REML function is more substantial than ML function (Eq. (4.3)) even for small datasets. Here we introduce an orthogonal transformation that accelerates the REML function optimisation substantially. To obtain such transformation, we first show that the likelihood function of transformed data with any orthogonal residualising matrix, that is a matrix that maps Y to the null space of X, is exactly the same as the REML function. Then, a simplifed form of the REML function (Eq. (4.5)) can be obtained using a particular transformation that makes the covariance matrix of transformed data diagonal.

Let $M = I - X(X'X)^{-1}X'$ be the residual forming matrix based on the fixed effects regressors. Since M is idempotent, it can be diagonalised

$$M = AA', \tag{4.6}$$
$$A'A = I, \tag{4.7}$$

where A is the $N \times (N-P)$ matrix of the eigenvectors of M corresponding to the non-zero eigenvalues. Crucially, A also residualises the data, because it is orthogonal to the design matrix X:

$$\begin{aligned} A'X &= A'AA'X \\ &= A'MX = 0. \end{aligned}$$

Hence $A'Y \sim N(0, A'\Sigma A)$ and the log likelihood of the transformed data is

$$\ell(A'Y, \Sigma) = -\frac{1}{2}\left[\mathrm{Const} + \log|A'\Sigma A| + Y'A(A'\Sigma A)^{-1}A'Y\right]. \tag{4.8}$$

In Appendix A we show that this is equivalent to Equation (4.5), and thus we can use the eigenvectors of the residual-forming matrix to build the REML log likelihood.

As A is not unique, we seek to find one that diagonalises the covariance of the residualised data. The transformation matrix could be derived from eigendecomposition of GSM adjusted for the fixed effect covariates as follows:

$$M(2\Phi)M = S_{\mathrm{r}}D_{\mathrm{g_r}}S_{\mathrm{r}}',$$

where $D_{\mathrm{g_r}} = \mathrm{diag}\{\lambda_{g_r i}\}$ is the $(N-P)\times(N-P)$ diagonal matrix of non-zero eigenvalues; and S_{r} is the $N\times(N-P)$ matrix of eigenvectors that corresponds to non-zero eigenvalues. Firstly, S_r is a valid A, because its columns are orthogonal $S_r'S_r = I$ and $M = S_rS_r'$ (Kang *et al.*, 2008). Thus we define the projected polygenic model by pre-multiplying S_{r}' both sides of polygenic model (Eq. (4.1)):

$$\begin{aligned} S_{\mathrm{r}}'Y &= S_{\mathrm{r}}'X + S_{\mathrm{r}}'g + S_{\mathrm{r}}'\epsilon, \\ S_{\mathrm{r}}'MY &= S_{\mathrm{r}}'MX + S_r'Mg + S_{\mathrm{r}}'M\epsilon. \end{aligned}$$

which we write as:

$$Y_{\mathrm{r}}^* = g_{\mathrm{r}}^* + \epsilon_{\mathrm{r}}^*, \tag{4.9}$$

where Y_{r}^*, g_{r}^* and ϵ_{r}^* are $N-P$ projected phenotype, genetic and residual vectors, respectively. In this fashion, the projected phenotype covariance matrix becomes diagonal:

$$\begin{aligned} \mathrm{cov}(Y_{\mathrm{r}}^*) &= \Sigma_{\mathrm{r}}^* \\ &= \mathrm{cov}(S_{\mathrm{r}}'Y) \\ &= \mathrm{cov}(S_{\mathrm{r}}'MY) \\ &= S_{\mathrm{r}}'(M\Sigma M)S_{\mathrm{r}} \\ &= S_{\mathrm{r}}'(\sigma_A^2 M(2\Phi)M + \sigma_E^2 M)S_{\mathrm{r}} \\ &= \sigma_A^2 S_{\mathrm{r}}'(S_{\mathrm{r}}D_{\mathrm{g_r}}S_{\mathrm{r}}')S_{\mathrm{r}} + \sigma_E^2 S_{\mathrm{r}}'(S_{\mathrm{r}}S_{\mathrm{r}}')S_{\mathrm{r}} \\ &= \sigma_A^2 D_{\mathrm{g_r}} + \sigma_E^2 I, \end{aligned}$$

where we have used the identity $S_{\mathrm{r}}'M = S_{\mathrm{r}}'$. That is, therefore the projected data, Y_{r}^*, loglikelihood takes on a simpler form:

$$\ell_{\mathrm{REML}}(\sigma_{\mathrm{A}}^2, \sigma_{\mathrm{E}}^2; Y_{\mathrm{r}}^*) = -\frac{1}{2}\left[\mathrm{Const} + \sum_{i=1}^{N-P}\log(\lambda_{\mathrm{g_r}i}\sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{E}}^2) + \sum_{i=1}^{N-P}\frac{y_{\mathrm{r}i}^{*2}}{\lambda_{\mathrm{g_r}i}\sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{E}}^2}\right] \tag{4.10}$$

where $y_{\mathrm{r}i}^{*2}$ is the square of the i-th element of Y_{r}^*.

It is clear from the Equation (4.10) that working with the simplified version of REML is computationally easier than the original one (Eq. (4.5)). Beside accelerating the REML optimisation, this approach facilitates performing likelihood ratio test for fixed effects (βs) and leads to a computationally efficient estimator and test statistic, described below.

4.1.2 REML and ML Parameter Estimation

We choose Fisher's scoring method to optimize the simplified ML and REML functions because it leads to computationally efficient variance component estimators. The score and the expected Fisher information matrices for the simplified models can be expressed as:

$$\begin{aligned} S_{\mathrm{ML}}(\beta,\theta) &= \begin{bmatrix} X^{*\prime}\Sigma^{*-1}\epsilon^{*} \\ -\frac{1}{2}\left[U'\Sigma^{*-1}1 - U'\Sigma^{*-2}\epsilon^{*2}\right] \end{bmatrix}, \\ I_{\mathrm{ML}}(\beta,\theta) &= \begin{bmatrix} X^{*\prime}\Sigma^{*-1}X^{*} & 0 \\ 0 & \frac{1}{2}U'\Sigma^{*-2}U \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} S_{\mathrm{REML}}(\theta) &= -\frac{1}{2}\left[U_{\mathrm{r}}'\Sigma_{\mathrm{r}}^{*-1}1 - U_{\mathrm{r}}'\Sigma_{\mathrm{r}}^{*-2}Y_{\mathrm{r}}^{*2}\right], \\ I_{\mathrm{REML}}(\theta) &= \frac{1}{2}U_{\mathrm{r}}'\Sigma_{\mathrm{r}}^{*-2}U_{\mathrm{r}}, \end{aligned}$$

where $\theta = (\sigma^2_{\mathrm{E}}, \sigma^2_{\mathrm{A}})$; $U = [\mathbf{1}, \lambda_{\mathrm{g}}]$ and $U_{\mathrm{r}} = [\mathbf{1}_{\mathrm{r}}, \lambda_{\mathrm{g_r}}]$ are $N \times 2$ and $(N-P)\times 2$ matrices; and λ_{g} is the vector of eigenvalues of (2Φ); $\lambda_{\mathrm{g_r}}$ the vector of eigenvalues of $M(2\Phi)M$; $\mathbf{1}$ and $\mathbf{1}_{\mathrm{r}}$ are N and $(N-P)$-vectors of one, respectively; Y_{r}^{*2} is the element wise product of Y_{r}^{*}; and ϵ^{*2} is the element wise product of ϵ^{*}. Following Fisher's scoring method it can be shown that at each iteration, maximum likelihood estimation of β and θ are updated based on WLS regression of Y^{*} on X^{*} and ϵ^{*2} on U, respectively (see Chapter 3 for more details), as follows:

$$\begin{aligned} \hat{\beta}_{\mathrm{ML},j+1} &= \left(X^{*\prime}(\hat{\Sigma}_j^{*})^{-1}X^{*}\right)^{-1} X^{*\prime}(\hat{\Sigma}_j^{*})^{-1}Y^{*}, \\ \hat{\theta}_{\mathrm{ML},j+1} &= \max\left\{0,\ \left(U'(\hat{\Sigma}_j^{*2})^{-1}U\right)^{-1} U'(\hat{\Sigma}_j^{*2})^{-1}\hat{\epsilon}_j^{*2}\right\}, \end{aligned} \tag{4.11}$$

and restricted maximum likelihood estimation of θ is updated based on WLS regression of Y_{r}^{*2} on U_{r} as follows:

$$\hat{\theta}_{\mathrm{REML},j+1} = \max\left\{0,\ \left(U_{\mathrm{r}}'(\hat{\Sigma}_{\mathrm{r}j}^{*2})^{-1}U_{\mathrm{r}}\right)^{-1} U_{\mathrm{r}}'(\hat{\Sigma}_{\mathrm{r}j}^{*2})^{-1}Y_{\mathrm{r}}^{*2}\right\},$$

where j indexes iteration; Σ_j^{*2} and $\Sigma_{\mathrm{r}j}^{*2}$ are constructed with $\theta_{\mathrm{ML},j}$ and $\theta_{\mathrm{REML},j}$ respectively; ϵ_j^{*2} is the element-wise square of $\epsilon_j^{*} = Y^{*} - X^{*}\beta_{\mathrm{ML},j}$; Y_r^{*2} is the element-wise square of Y_r^{*}; and the variance parameters θ must be positive, hence the maximum operator. As usual, these updates are iterated until convergence criteria holds.

It has been shown that when the initial value is a consistent estimator, the estimator based on the first iteration is asymptotically normal and consistent (Amemiya, 1977). Such initial value for $\hat{\beta}_{\text{ML}}$ and $\hat{\theta}_{\text{ML}}$ could be derived from OLS regression coefficients of Y^* on X^* and squared residuals on U, respectively:

$$\begin{aligned}\hat{\beta}_{\text{ML,OLS}} &= \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y^*, \\ \hat{\theta}_{\text{ML,OLS}} &= \max\left\{0,\ \left(U'U\right)^{-1} U'\epsilon^{*2}\right\}.\end{aligned}$$

For REML, initial values for $\hat{\theta}_{\text{REML,OLS}}$ can be found as OLS regression coefficient of Y_{r}^{*2} on U_{r}:

$$\hat{\theta}_{\text{REML,OLS}} = \max\left\{0,\ \left(U_{\text{r}}'U_{\text{r}}\right)^{-1} U_{\text{r}}'Y_{\text{r}}^{*2}\right\}.$$

Hence our 1-step, non-iterative estimators are:

$$\hat{\beta}_{\text{ML,WLS}} = \left(X^{*\prime}(\hat{\Sigma}^*_{\text{OLS}})^{-1}X^*\right)^{-1} X^{*\prime}(\hat{\Sigma}^*_{\text{OLS}})^{-1}Y^*, \tag{4.12}$$

$$\hat{\theta}_{\text{ML,WLS}} = \max\left\{0,\ \left(U'(\hat{\Sigma}^{*2}_{\text{OLS}})^{-1}U\right)^{-1} U'(\hat{\Sigma}^{*2}_{\text{OLS}})^{-1}\epsilon^{*2}_{\text{OLS}}\right\}, \tag{4.13}$$

$$\hat{\theta}_{\text{REML,WLS}} = \max\left\{0,\ \left(U_{\text{r}}'(\hat{\Sigma}^{*2}_{\text{OLS,r}})^{-1}U_{\text{r}}\right)^{-1} U_{\text{r}}'(\hat{\Sigma}^{*2}_{\text{OLS,r}})^{-1}Y_{\text{r}}^{*2}\right\}, \tag{4.14}$$

where $\hat{\Sigma}^*_{\text{OLS}}$ and $\hat{\Sigma}^*_{\text{OLS,r}}$ are formed by $\hat{\theta}_{\text{ML,OLS}}$ and $\hat{\theta}_{\text{REML,OLS}}$ respectively, and $\hat{\epsilon}^{*2}_{\text{OLS}}$ is the element-wise square of $\hat{\epsilon}^*_{\text{OLS}} = Y^* - X^*\hat{\beta}_{\text{OLS}}$. Going forward, we will use "ML" or "REML" to refer to the iterated estimators and "WLS" to refer to these one-step estimators.

4.1.3 Association Testing

The Score, likelihood ratio (LRT) and the Wald tests can be used for the genetic association testing using either ML or REML functions of the model in Equation (4.1).

The score statistic (Rao, 2008) that requires the value of score and information matrices under the the null hypothesis constraint ($H_0 : \beta_1 = 0$) for the simplified ML model (Eq. 4.5) can be written

$$T_{\text{S,ML}} = \tilde{\epsilon}'_{\text{ML}}\tilde{\Sigma}^{*-1}_{\text{ML}}X_1^*\left[C'(X^{*\prime}\tilde{\Sigma}^{*-1}_{\text{ML}}X^*)^{-1}C\right]X_1^{*\prime}\tilde{\Sigma}^{*-1}_{\text{ML}}\tilde{\epsilon}_{\text{ML}},$$

where C is a $P\times 1$ contrast vector; $X^* = [X_1^* \quad X_2^*]$ encompasses the full transformed covariate matrix; $\tilde{\epsilon}'_{\text{ML}}$ and $\tilde{\Sigma}_{\text{ML}}$ are the ML residual and covariance matrix estimates

under the null hypothesis constraint. The score statistic for the projected model (Eq. (4.9)) can be derived like $T_{\text{S,ML}}$ following the projection with respect to the H_0 fixed effects, i.e. nuisance, terms X_2,

$$T_{\text{S,REML}} = Y_{\text{r}}^{*\prime}\tilde{\Sigma}_{\text{r}}^{*-1}X_{1\text{r}}^{*}(X_{1\text{r}}^{\prime *}\tilde{\Sigma}_{\text{r}}^{-1}X_{1\text{r}}^{*})^{-1}X_{1\text{r}}^{\prime *}\tilde{\Sigma}_{\text{r}}^{*-1}Y_{\text{r}}^{*},$$

where $Y_{\text{r}}^{*} = S_{2\text{r}}^{\prime}Y$ and $X_{1\text{r}}^{*} = S_{2\text{r}}^{\prime}X_1$ are $(N - P + 1)$-vectors of the projected phenotype and allele frequency count, respectively; and the projection matrix $S_{2\text{r}}$ is comprised of the eigenvectors of $M_2(2\Phi)M_2$ with non-zero eigenvalues, $M_2 = I - X_2(X_2X_2^{\prime})^{-1}X_2^{\prime}$; and $\tilde{\Sigma}_{\text{r}}^{*-1}$ is the projected model covariance matrix estimation under the null model constraint.

Likelihood Ratio Test The LRT (Neyman and Pearson, 1933) statistic is twice the difference of the optimised log-likelihoods, unrestricted minus H_0-restricted. For ML this requires optimizing the likelihood function twice, once under the null $H_0 : \beta_1 = 0$, once under the alternative. We denote the test statistic for this test $T_{\text{L,ML}}$. A well-known aspect of REML is that it cannot be used to tests of fixed effects, since the null hypothesis would represent a change of the projection that defines the REML model. However, we can consistently use the same projection $S_{2\text{r}}$, under the unrestricted and restricted models, to diagonalise our covariance and carry out such a hypothesis test. To be precise, the unrestricted model is

$$S_{2\text{r}}^{\prime}Y = S_{2\text{r}}^{\prime}X_1\beta_1 + S_{2\text{r}}^{\prime}g + S_{2\text{r}}^{\prime}\epsilon,$$

where $S_{2\text{r}}^{\prime}X_2\beta_2 = 0$ by the construction of $S_{2\text{r}}$, and the restricted model is

$$S_{2\text{r}}^{\prime}Y = S_{2\text{r}}^{\prime}g + S_{2\text{r}}^{\prime}\epsilon.$$

Following the same procedure as ML, the test statistic is denoted by $T_{\text{L,REML}}$.

The Wald Test For a scalar parameter, the Wald test (Rao, 2008) is the parameter estimate divided by the standard deviation of the estimate under an unrestricted model. For an vector parameter β and contrast C, it takes the form

$$T_W \quad = \quad C\hat{\beta}(C(X^{\prime}\hat{\Sigma}^{-1}X)C^{\prime})^{-1}\hat{\beta}^{\prime}C^{\prime}$$

where $\hat{\beta}$ and $\hat{\Sigma}^{-1}$ are the parameter estimations under the alternative hypothesis; this form holds for both ML and REML. A test for genetic association testing can be calculated either using fully converged or 1-step variance component estimators.

In the parametric framework, all of the aforementioned tests null distribution follow chi square distribution with one degree of freedom asymptotically.

4.1.4 Inference Using Permutation Test

In neuroimaging the permutation test is a standard tool to conduct inference while controlling the family wise error rate (FWE) (Nichols and Holmes, 2002). It only requires an assumption of exchangeability, that the joint distribution of the data is invariant to permutation, and can be provide exact inference in the absences of nuisance variables, or approximately exact control with nuisance variables (Winkler *et al.*, 2014). Control of the FWE of a voxel-wise or cluster-wise statistic is obtained from a maximum distribution of the corresponding statistic (see Section (2.7.1) for details). It has been well known that naive use of permutation test for the genetic association testing, ignoring dependence structure between individuals, leads to invalid inferences (Abney, 2015). Here we propose two permutation schemes that account for dependence explained by our model, one free and one restricted permutation approach.

Free Permutation

The genetic association testing in the context of LMM using a permutation test requires proper handling of fixed effect and random effect nuisance variables to respect the exchangeablity assumption. There has been several permutation schemes for regression coefficient hypothesis testing in presence of fixed effect nuisance covariates when the errors are independent (Winkler *et al.*, 2014). However little work has been done for fixed effect inference using permutation test in linear mixed models where the error term is correlated.

Free Permutation for the simplified ML Model: For the simplified model (Eq. 4.5) we create permuted data $\tilde{Y}^*$ using the reduced, null model residuals and use them as surrogate null data to be re-analysed as follows:

$$\tilde{Y}^* = X_2^* \hat{\beta}_2 + P\hat{\epsilon},$$

where P is one of $N!$ possible $N \times N$ permutation matrices; $\hat{\beta}_2$ is estimated using the Equations (4.11) or (4.12); $\hat{\epsilon}$ denotes the residuals under the null hypothesis $H_0 : \beta_1 = 0$, calculated using either fully converged or 1-step estimators and the tilde accent on the data ($\tilde{Y}^*$) denotes one of many null hypothesis realisations. The reduced null model is not exchangeable due heteroscedasticity of Σ^*, and the non-

constant diagonal of D_{g} specifically. To account for this, in each permutation step, we fit the simplified model (4.5) with the permuted covariance matrix

$$\mathrm{cov}(\tilde{Y}^*) = P\Sigma^* P' = \sigma^2_{\mathrm{A}} P D_{\mathrm{g}} P' + \sigma^2_{\mathrm{E}} I,$$

and with sufficient numbers of permuted realisations the empirical null distribution of maximum score, LRT and the Wald tests (or cluster-size, after thresholding one of these test statistics) can be derived.

Free Permutation for Simplified REML model: An alternate permutation scheme could be built based on projecting the LMM model (Eq. (4.1)) to the lower dimension space with respect to the null hypothesis reduced model, i.e. only the nuisance fixed effect terms. Similar to the description of the LRT above (Section 4.1.3), we define a reduced transformation with $S_{2\mathrm{r}}$ based on the non-trivial eigenvectors of $M_2(2\Phi)M_2$:

$$\tilde{Y}^*_{\mathrm{r}} = X^*_{1\mathrm{r}}\beta_1 + g^*_{\mathrm{r}} + \epsilon^*_{\mathrm{r}},$$

where $\tilde{Y}^*_{\mathrm{r}} = PY^*_{\mathrm{r}}$ is the permuted data under this reduced model; $Y^*_{\mathrm{r}} = S'_{2\mathrm{r}}Y$ is the reduced transformed data; $X^*_{1\mathrm{r}} = S'_{2\mathrm{r}}X_1$ is as above, the reduced transformed additive genetic effect; $g^*_{\mathrm{r}} = S'_{2\mathrm{r}}g$ and $\epsilon^*_{\mathrm{r}} = S'_{2\mathrm{r}}\epsilon$ are the latent genetic effect and random error terms, respectively, after the reduced transformation. In this setting, like the free permutation for simplified ML model, permuted covariance matrix

$$\mathrm{cov}(PY^*_{\mathrm{r}}) = \sigma^2_{\mathrm{A}} P D_{\mathrm{g_r}} P' + \sigma^2_{\mathrm{E}} I$$

is fitted in each permutation step. However, under the null hypothesis of no genetic effect, estimated random effects for permuted phenotype are exactly as same as the un-permuted one hence variance components only require to be estimated once.

Restricted Permutation: exchangeability blocks For Family Data

In the free permutation approaches we permute despite the lack of changeability, but then correspondingly permute the covariance structure to account for this. An alternate approach is to define exchangeability blocks where observations within each block are regarded as exchangeable. Precisely, with exchangeability blocks, the assumption is that permutations altering the order of observations only within each block preserve the null hypothesis distribution of the full data.

While not feasible for the original model (4.1), in the simplified ML (4.5) or

simplified REML (4.9) model we can define approximate exchangeability blocks. In simplified models the sorted eigenvalues arrange the observations in order of variance (increasing or decreasing, depending on software conventions). Hence blocks of contiguous observations Y^* or Y_{r}^* will have variance that is as similar as possible and will be, under the null hypothesis, approximately exchangeable.

Specifically, we propose to define exchangeability blocks such that the range of D_{g} or $D_{\mathrm{g_r}}$ values within a block is no greater than 0.01. Permutation for simplified models then takes place as exactly as defined in free permutation following calculating the test statistic using estimated covariance matrix from unpermuted data.

4.1.5 Efficient score statistic implementation for vectorized images

To fully exploit the computational advantage of our non-iterative, reduced-dimension projected model estimation method we require a vectorised algorithm. That is, even without iteration, the method will be relatively slow if the evaluation of the estimates is so complex that each phenotype must be looped over one-by-one. For fast evaluation with a high-level language like Matlab, the estimation process for a set of phenotypes must be cast as a series of matrix algebra manipulations.

In this section we develop the vectorised algorithm for association one chromosome's worth of SNPs and all image voxels/elements (subject to memory constraints). To avoid proximal contamination (Lippert *et al.*, 2011a) (see section 2.3.1 for formal definition) and efficient implementation of LMM, we follow leave one chromosome out approach where all markers on a chromosome being tested are excluded from the GSM (Widmer *et al.*, 2014; Yang *et al.*, 2014).

Let $\mathbf{Y}_{\mathrm{r}}$ and $\mathbf{X}_{\mathrm{r}}$ be a $(N-P)\times V$ and $(N-P)\times G$ matrices of projected traits and allele frequencies respectively, where V and G are number of elements in image and number of SNPs the tested chromosome, respectively. The score test requires parameter estimation under the null hypothesis constraints, and since X_2 is the same for all SNPs, the estimated covariance matrix will be the same all markers the chromosome. Thus the covariance matrix only need to be estimated once as follows:

$$\begin{aligned}
\mathbf{F} &= \mathbf{Y}_{\mathrm{r}} \odot \mathbf{Y}_{\mathrm{r}}, \\
\theta &= \max((U_{\mathrm{r}}' U_{\mathrm{r}})^{-1} U_{\mathrm{r}}' \mathbf{F}, \mathbf{0}), \\
W &= \mathbf{1}_{NV} \oslash ((U_{\mathrm{r}}\theta) \odot (U_{\mathrm{r}}\theta)),
\end{aligned}$$

where $\mathbf{F}$ and $\mathbf{Y}_{\mathrm{r}}$ are $(N-P)\times V$ matrices, where each column of $\mathbf{Y}_{\mathrm{r}}$ corresponds

to Y_{r}^* (Eq. (4.9)) for one image element and $\mathbf{F}$ is the element-wise squaring of $\mathbf{Y}_{\mathrm{r}}$; $\odot$ denotes Hadamard product; $\oslash$ denotes element-wise division; θ is the $2 \times V$ matrix of OLS solutions which is matrix counterpart of $\hat{\theta}_{\mathrm{REML,OLS}}$; $\mathbf{0}$ is the $2 \times V$ matrix of zeros; and here $\max(\cdot, \cdot)$ is an element-wise maximum between the two operands, evaluating to a $2 \times V$ matrix; W and $\mathbf{1}_{NV}$ are the $(N-P) \times V$ matrices, where each column of W is $\mathrm{diag}(\hat{\Sigma}^{*-2}_{\mathrm{OLS,r}})$ for the corresponding image element and $\mathbf{1}_{NV}$ is a matrix of ones. We define the following notations to estiamte the variance components for the vectorized image:

$$\begin{aligned} A &= \mathbf{1}_V W, \\ B &= D'_{\mathrm{g}} W, \\ C &= (D'_{\mathrm{g}} \odot \lambda'_{\mathrm{g_r}}) W, \\ D &= \mathbf{1}_V (W \odot F), \\ E &= D'_{\mathrm{g}} (W \odot F), \end{aligned}$$

where $\mathbf{1}_V$ is the length-V column vector of ones. In this setting

$$\begin{aligned} \boldsymbol{\sigma}^2_{\mathrm{A}} &= \max((-B \odot D + A \odot E) \oslash (A \odot C - B \odot B), \mathbf{0}), \\ \boldsymbol{\sigma}^2_{\mathrm{E}} &= \max\left((C \odot D - B \odot E) \oslash (A \odot C - B \odot B), \mathbf{0}\right), \\ \boldsymbol{\Sigma}^{-1} &= U_{\mathrm{r}} \begin{bmatrix} \mathbf{1}_V \oslash \boldsymbol{\sigma}^2_{\mathrm{E}} \\ \mathbf{1}_V \oslash \boldsymbol{\sigma}^2_{\mathrm{A}} \end{bmatrix}, \end{aligned}$$

where $\boldsymbol{\sigma}^2_{\mathrm{A}}$ and $\boldsymbol{\sigma}^2_{\mathrm{E}}$ are the length-V column vectors of genetic and environmental variance components, respectively; and $\boldsymbol{\Sigma}^{-1}$ is a $(N-P) \times V$ matrix which here each column of $\boldsymbol{\Sigma}^{-1}$ is the element-wise reciprocal of the diagonal of the variance matrix of the corresponding image element's data $\mathbf{Y}_{\mathrm{r}}$ for each element of image. In this fashion, the score statistic matrix for all markers being tested and the vectorized image can be expressed as:

$$\mathbf{T}_{\mathrm{S,R}} = \left[(\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \odot \mathbf{Y}_{\mathrm{r}})) \odot (\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \odot \mathbf{Y}_{\mathrm{r}}))\right] \oslash \left[(\mathbf{X} \odot \mathbf{X})\boldsymbol{\Sigma}^{-1}\right], \tag{4.15}$$

where $\mathbf{T}_{\mathrm{S,R}}$ is a $G \times V$ matrix of score statistics for all SNPs and traits.

4.2 Evaluation

4.2.1 Simulations

Intensive simulation studies are conducted to evaluate proposed methods for association estimation and testing. The aim of the first study is to compare fully converged and 1-step random effect estimators based on the simplified ML and REML functions. In the second study, the performance of various test statistics for the association testing are compared using a fully converged or 1-step random effect estimators for ML and REML functions. Finally, we compare FaST-LMM (Lippert *et al.*, 2011a) to our preferred test, the score test based on the simplified REML function, $\mathbf{T}_{\mathrm{S,REML}}$, using both false positive error rates and empirical power using simulated genetic markers.

In all simulations the response variable is assumed to follow $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \Sigma)$ and $\Sigma = \sigma^2_{\mathrm{A}}(2\Phi) + (1 - \sigma^2_{\mathrm{A}})I$, giving a unit variance phenotype. As above, the design matrix is partitioned $X = [X_1 X_2]$, where X_1 is the allele count per subject for a given marker, and X_2 are all other non-genetic fixed effects. In our simulations, X_1 is based on simulated marker, where each marker has a reference allele frequency sampled from a uniform distribution on $[0.1, 0.9]$. The X_2 matrix has 3 columns, an intercept, a linear trend from -1 to 1, and the element-wise square of the linear trend. Kinship matrices from a family study, genetic analysis workshop 10 (GAW10), and genetic similarity matrix from simulated genetic markers for a sample of unrelated individuals with different sizes were chosen to set the covariance, for a range of genetic variances, $\sigma^2_{\mathrm{A}} = 0, 0.2, 0.4, 0.6\&0.8$. Specifically, the Cholesky decomposition of Σ was used to premultiply i.i.d normal random variables with 5000 realisations.

4.2.2 Real Data

To validate our proposed methods for association estimation and inference for imaging data, we applied them on a dataset from healthy and schizophrenic individuals to perform ROI and voxel-wise genome wide association analysis using cluster wise inference. The sample was 54% healthy individual (175 control/155 schizophrenic) and had a mean age of 39.12 (SD= 14.9) where 50% of the sample is male.

Diffusion Tensor Imaging

Imaging data was collected using a Siemens 3T Allegra MRI (Erlangen, Germany) using a spin-echo, EPI sequence with a spatial resolution of $1.7 \times 1.7 \times 4.0$ mm. The

sequence parameters were: TE/TR=87/5000ms, FOV=200mm, axial slice orientation with 35 slices and no gaps, twelve isotropically distributed diffusion weighted directions, two diffusion weighting values (b=0 and 1000 s/mm2), the entire protocol repeated three times.

ENIGMA-DTI protocols for extraction of tract-wise average FA values (2.6.2) were used. These protocols are detailed elsewhere (Jahanshad *et al.*, 2013) and are available online `http://enigma.ini.usc.edu/protocols/dti-protocols/`. Briefly, FA images from subjects were non-linearly registered to the ENIGMADTI target brain using FSL's FNIRT (Jahanshad *et al.*, 2013). This target was created as a minimal de-formation target based on images from the participating studies as previously described (Jahanshad et al., 2013b). The data were then processed using FSL's tract-based spatial statistics (TBSS; `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS`) analytic method (Smith *et al.*, 2006) modified to project individual FA values on the hand-segmented ENIGMADTI skeleton mask. The protocol, target brain, ENIGMADTI skeleton mask, source code and executables, are all publicly available (`http://enigma.ini.usc.edu/ongoing/dti-working-group/`). The FA values are normalized across individuals by inverse Gaussian transform (Servin and Stephens, 2007; Allison *et al.*, 1999) to ensure normality assumption. Finally, we analyzed the voxel and cluster-wise FA values with applying along the ENIGMA skeleton mask.

Genetic Quality Control

In this study only genotyped Single Nucleotide Polymorphisms (SNPs) from genome-wide information were included in the analysis. Visual inspection of multi-dimensional scaling analysis was used to extract individuals with European ancestry. SNPs from individuals with European ancestry that do not meet any of the following quality criteria were excluded: genotype call rate 95%, significant deviation from Hardy-Weinberg equilibrium $p < 10^{-6}$ and minor allele frequency 0.1 was used to ensure that sufficient numbers of subjects would be found in our sample in each genotypic group (homozygous major allele, heterozygous, homozygous minor allele) using an additive genetic model. After all quality control steps, 962,885 out of 1000,000 SNPs remain for genome-wide association analysis.

4.3 Result

Simulation Results

The first simulation study on the accuracy of σ^2_A estimation with the 1-step approach is similar to their fully converged counterparts (Figure 4.1). When the data are independent ($\sigma^2_A = 0$), the methods are indistinguishable in terms of bias and mean squared error (MSE). When $\sigma^2_A > 0$, the fully converged methods have less bias, but the difference is modest in absolute value; in terms of MSE, the 1-step methods have just slightly worse performance. The first simulation also shows good performance of fixed effect (β_1) estimation (Figure 4.2). Both the 1-step and fully converged similar bias and MSE, with WLS REML again closely follows fully converged REML.

Simulation results on false positive rates for the hypothesis test $H_0 : \beta_1 = 0$ are shown in Figure 4.3(a). These show that parametric results using either simplified ML or REML functions, lay within the Monte Carlo confidence interval (MCCI) regardless of simulation settings and random effect estimator. The similar findings are obtained for power comparisons (Figure 4.3(b)) where all parametric statistics have almost the same power. Like the parametric approach, we found that both permutation schemes, free or restricted permutations to exchangeability blocks, control the false positives at the nominal level (Figure 4.4(a)), 5%, and could provide almost the same power (Figure 4.4(b)) for all statistics either based on the simplified ML or REML functions. However, for all test statistics and σ^2_A, the free permutation scheme is slightly more powerful than the restricted permutation test consistently.

The score test null distributions for $H_0 : \beta_1 = 0$ based on the simplified models using the fully converged and non-iterative variance component estimators are valid and indistinguishable (Figure 4.5(a) and 4.5(b)). However, the latter is much faster to calculate. Based on all of these results, we selected the score test based on the simplified REML function as the computationally most efficient test to be considered for genome-wide simulations and real data analysis.

Genome wide simulations were conducted to compare the parametric P-values from FaST-LMM and the score test based on the simplified REML using non-iterative variance component estimator in terms of false positives and power. The simulation results reveal that both approaches provide overall valid error rates (FaST-LMM=4.94% and The score test = 4.89%, Figure 4.6(a)). Power simulation shows that FaST-LMM and the score test have largely similar power, however, FaST-LMM is slightly more powerful (Figure 4.6(b)) and slower (Running time quantification is detailed later).

Association Analysis of FA data

We performed GWA of whole brain FA data, either using a whole brain parcellation of 22 regions of interest (ROIs) or voxel by voxel, to compare performance of the score test based on the simplified REML function using 1-step and fully converged random effect estimators; we also evaluate the use of ordinary least square (OLS) with MDS as nuisance fixed effects regressors to controlling for population structure in GWAS with unrelated individuals.

The random effect estimators, 1-step and fully converged based on simplified REML function were compared directly in Figure (4.7) with a scatter plot, showing an apparent trade-off between accuracy and running time where the non-iterative method has lower estimates of σ^2_{A} for some regions. Even with likely random effect underestimation using the non-iterative method, association statistic comparisons reveal strong concordance, with both approaches having almost the same performance (Figure 4.8).

Figure 4.9 shows different approaches to genomic control, showing that, regardless of random effect estimation method, the score test based on the simplified REML has smaller genomic control values than OLS with MDS nuisance regressors for all ROIs consistently. The genomic control of OLS with MDS nuisance regressors is poor, while both fully converged and 1-step estimators have similar values close to unity.

Figure 4.10 compares QQ-plot of association statistics between our model and OLS with MDS. These plots show either an identical distribution or slightly larger values for the OLS approach; however, as showing Figure 4.9, the OLS approach has poor genomic control and after adjustment (Figure 4.11) we get essentially identical results.

A permutation test was used to derive FWE corrected P-values for 22 ROIs and 962,885 SNPs to assess association significance. Although none of the $22 \times 962,885 \approx 20$ million statistics passed the permutation based FWE threshold ($\chi^2_1 = 32.08$), simply applying Bonferroni correction to the usual GWA P-value (5×10^{-8}) yields to stranger threshold ($\chi^2_1 = 35.72$), indicating the potential improved power from a permutation-based inference that accounts for dependency among the tests.

Finally we performed voxel-wise genome-wide association analysis of 60,000 voxel with almost 1 million genetic markers, using our proposed method, the score test based on the simplified REML function with the WLS REML random effect estimator for association testing. Cluster-wise inference was performed where the permutation test has been used to derive FWE maximum cluster size null distribu-

tion, however, neither voxel nor cluster wise inference pass the FWE threshold.

Benchmarking and running times

We compared running time between FaST-LMM which to our knowledge is the fastest implementation of LMM and the score test using WLS REML random effect estimator. The comparison was done using simulated and read data on Intel(R) core(TM) i7-2600 CPU @3.4 GH and 16GB RAM. Parametric association testing of 5000 phenotypes with 6000 simulated markers using sample of 300 individuals took 1 hour with FaST-LMM, however, efficient implementation of score test (Eq. (4.12)) only took 3 seconds. Also, the whole genome association of all 22 ROIs using Equation (4.12) requires 2 minuets. On real data, whole genome association on 22 ROIs, required 440 minutes using FaST-LMM while our approach took only 2 minutes.

4.4 Discussion

Imaging genetics is an active research area that concerns studying the impact of genetic factor on neuroimaging phenotypes to gain better understanding of brain in healthy and illness. This field has moved from establishing a heritable phenotype to find genetic markers that are associated with imaging phenotypes. Despite emerging world-wide consortia to boost GWA studies power using the largest possible sample sizes, there is a compelling need for computationally efficient analytic techniques that controls the population structure at the site level and provides powerful analysis.

It is the linear mixed model's effectiveness in controlling population structure inspired us to use it for association of high dimensional imaging phenotypes. However it presents challenges in terms of computational intensity and elevated false positive risk. Accelerated LMMs procedures are demanded by the growing sample sizes with whole genome resequencing data. However, association of imaging phenotypes require computationally efficient LMM that handles large number of models fit for each element of image.

To tackle this problem, we introduced an orthogonal transformation that substantially reduced LMM complexity for GWA specifically with high dimensional imaging phenotypes.

The equivalence between projected model and REML function helped us to reduce complexity of association testing. Specifically, the projection reduces the information matrix to a scaler that enables efficient implemention of score test for

association of vectorized image. Furthure improvements in speed can be achieved by using the WLS REML random effect estimator that we found to be more accurate than the WLS ML estimator.

We conducted intensive simulation studies to compare performance of the proposed methods. We evaluated a broad set of test statistics for association testing using the simplified ML and REML functions accompanied by 1-step and fully converged random effect estimators. The 1-step random effect estimator using simplified REML function provides more accurate approximation of the fully converged one in comparison to the WLS ML variance component estimator. The simulation and real data analysis shows that only minor differences in marker effect estimation and association test statistics between 1-step and fully converged random effect estimator. However, the former requires less computational resources. Also, we could not observe any appreciable differences between statistics performances in terms of the error rate and power using the GSM from unrelated individuals or kinship matrix from a family study.

The WLS REML random effect estimator is fast enough to be used to estimate voxel-wise heritability. Although the proposed 1-step random effect estimator is not as accurate as fully converged one, it can be used for filtering a small number of elements for further investigation with more computational intense tools.

Furthermore, higher genomic control for association testing restricted to individuals with European ancestry using OLS regression model perhaps indicating inefficiency of this approach population structure control in comparison to LMM.

The score test based on the simplified REML function is chosen for further investigations because it only requires variance component estimation under the null hypothesis constraint which is same for all markers. Furthermore, efficient implementation of score test for vectorized image accelerates association testing part of LMM.

Null distribution of score test with WLS REML is almost as same as fully converged one that enables performing parametric test hence mitigate permutation test. The score test only require covariance matrix under the null hypothesis hence it is same for all markers for each phenotype so it should be estimated only once for all phenotypes

Our contribution in accelerating LMM can be seen in two steps. First, covariance matrix estimation using WLS REML random effect estimator speeds up the procedure up to 10^9-fold (Ganjgahi *et al.*, 2015). Further improvement in speed can be obtained by using the score test based on simplified REML function. Our proposed method allows efficient implementation that reduces running time up to

200-fold. In addition, the efficient implementation of score test is fast enough that allows using permutation test to control family-wise error rate for number of elements in image and number of markers.

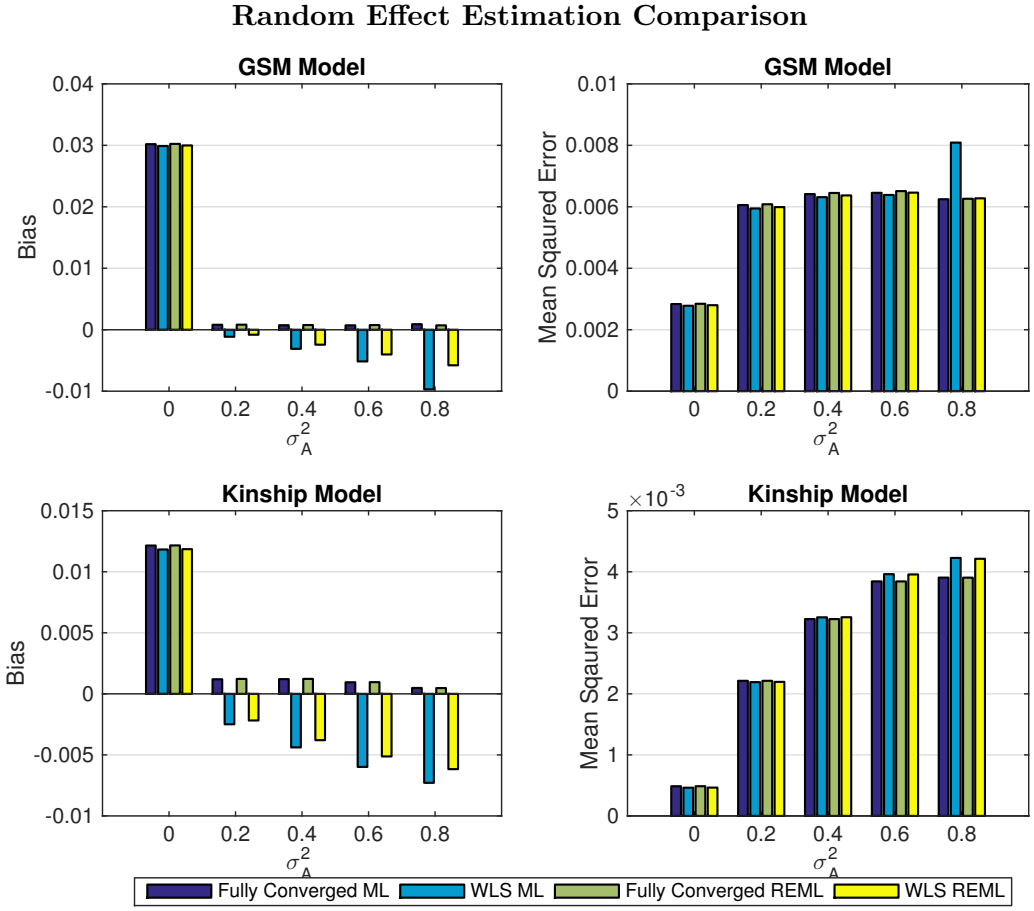


Figure 4.1: Simulation 1 results, comparing non-iterative and fully converged random effect estimators bias (left column) and mean squared error (right column) using the simplified ML or REML in terms of 5000 realisations, for different level of genetic random effect σ^2_A. The results are based on a GSM constructed from 1200 unrelated individuals (top row) and kinship matrix from GAW 10 with 23 families and 1497 individuals (bottom row). While the 1-step estimators generally (ML or REML) have more bias than fully converged ones, WLS REML has less bias than WLS ML, and in terms of MSE there is a relatively small difference in performance among all the methods.

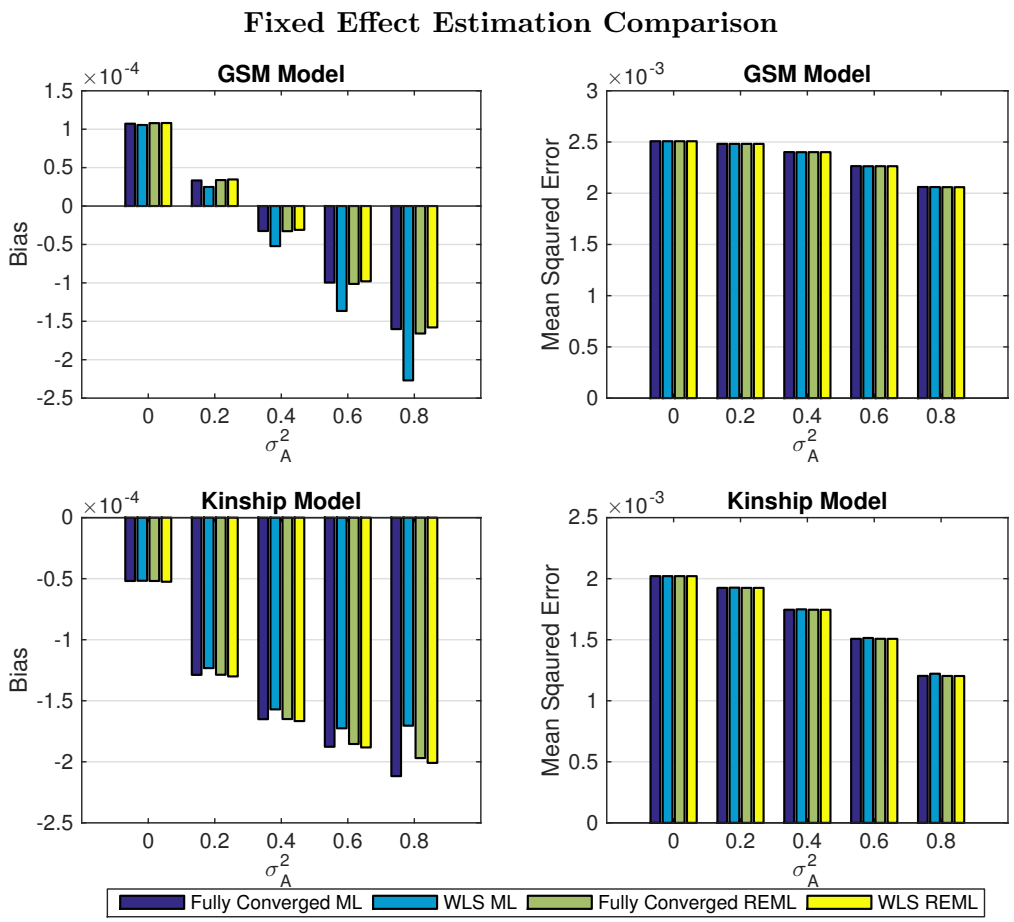
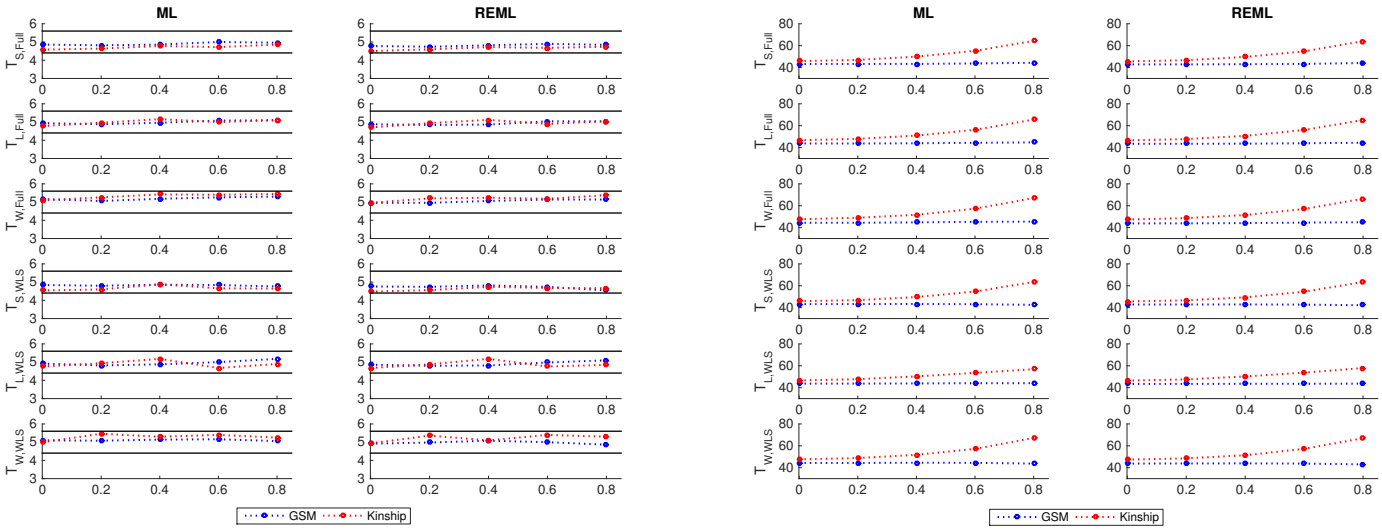


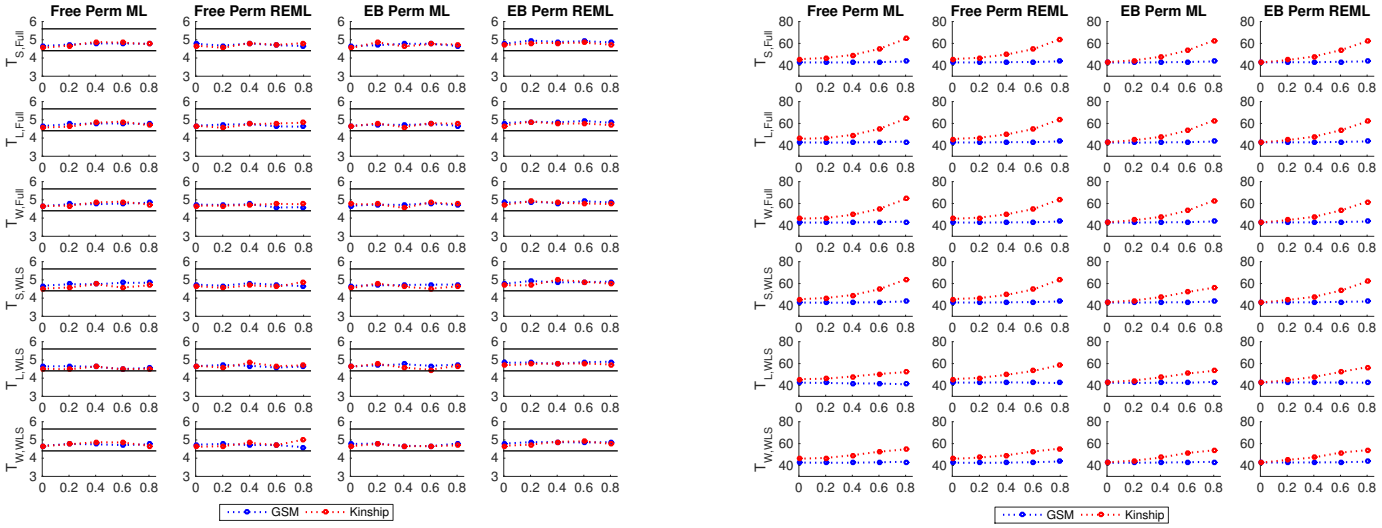
Figure 4.2: Simulation 1, comparing marker effect (fixed effect, β_1) estimators bias (left column) and mean squared error (right column) using the simplified ML or REML in terms of 5000 realisations, for different level of genetic random effect σ^2_A where $\beta_1 = 0$. The results are based on a GSM constructed from 1200 unrelated individuals (top row) and kinship matrix from GAW 10 with 23 families and 1497 individuals (bottom row). The accuracy of marker effect estimation comparison using non-iterative and fully converged variance component estimators based on the ML and REML reveals that fixed effect estimation using WLS REML variance component estimator has almost identical performance as the fully converged one over different levels of genetic variance.



(a) Parametric eror rate comparisons

(b) Parametric power comparisons

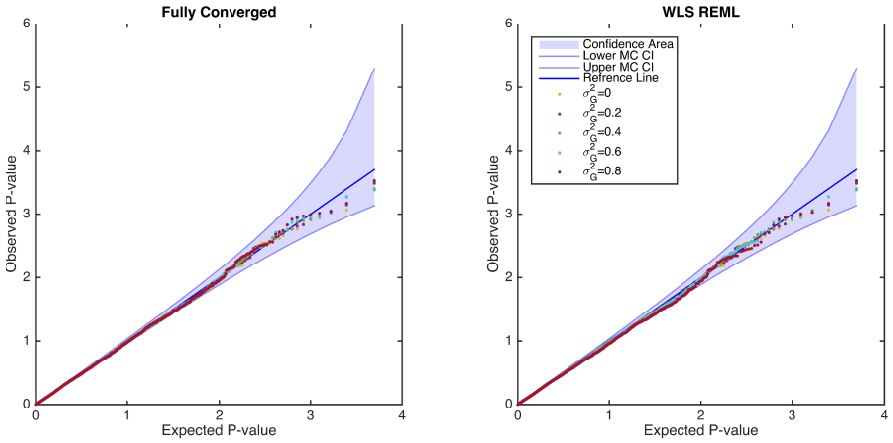
Figure 4.3: Simulation 2, comparing proposed statistics parametric error rates, 5% nominal (left panel) and power (right panel) based on simulation using either a GSM from 300 unrelated individuals or a kinship from GOBS study with 171 individuals and 10 families and 5000 realisations. The first three rows in each panel correspond to association statistics using the fully converged random effect estimator and the rest show the result using the non-iterative random effect estimator. Monte Carlo confidence interval is (4.40%, 5.60%). Regardless of kinship matrix in the simulation and variance component estimator, non-iterative or fully converged, all statistic have the same performances.



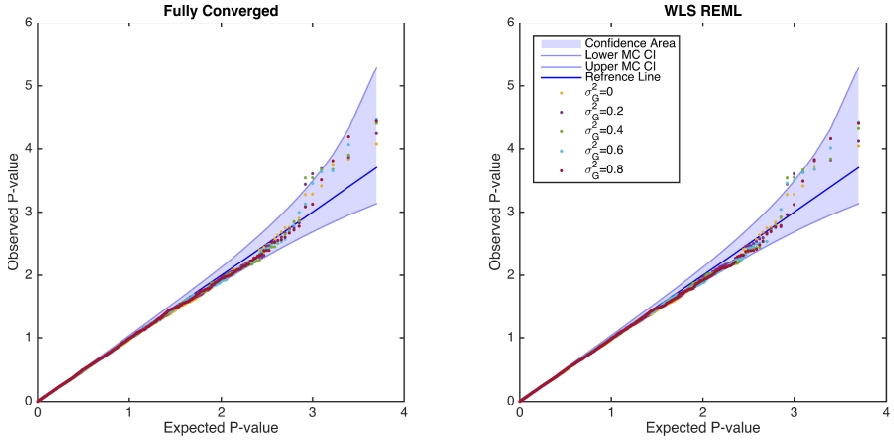
(a) Permutation error rate comparisons

(b) Permutation power comparisons

Figure 4.4: Simulation 2, comparing proposed statistics permutation based error rates, 5% nominal (left panel) and power (right panel) based on simulation using either a GSM from 300 unrelated individuals or a kinship from GOBS study with 171 individuals and 10 families and 5000 realisations and 500 permutations each realisations. The first three rows in each panel correspond to association statistics using the fully converged random effect estimator and the rest show the result using the non-iterative random effect estimator. Monte Carlo confidence interval is (4.40%, 5.60%). Despite the kinship matrix in the simulation and variance component estimator, non-iterative or fully converged, all statistic have the same performances. Both permutation schemes could control the error rate at the nominal level, however free permutation is slightly more powerful than the restricted permutation.

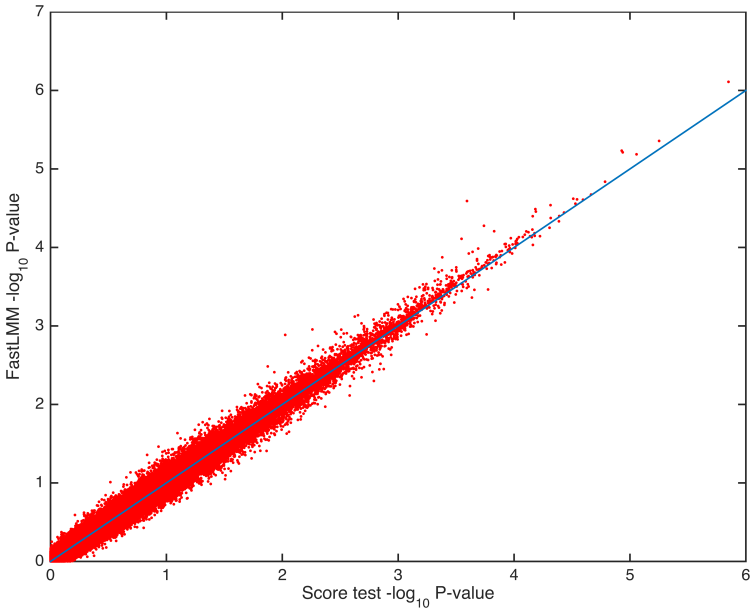


(a) GSM from 300 unrelated individuals

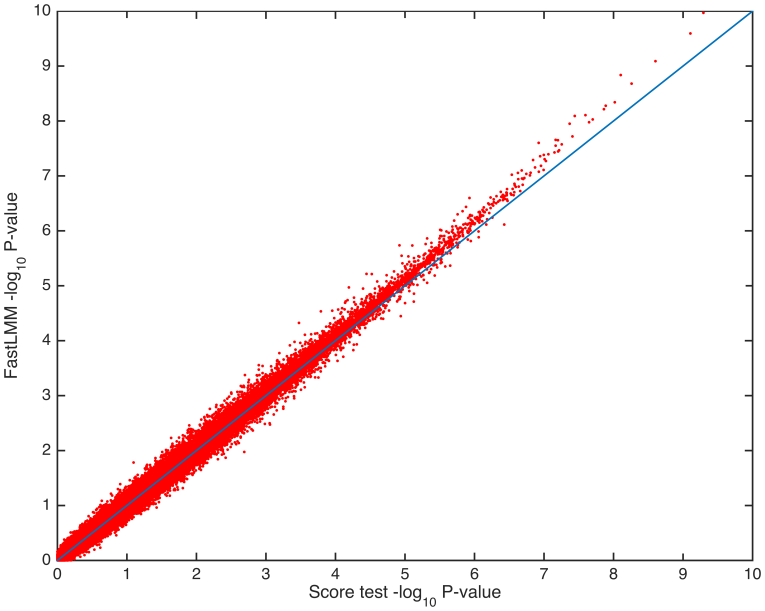


(b) GOBS kinship matrix, 171 individuals and 10 Families

Figure 4.5: Simulation 3, comparing score statistic parametric null distribution for H_0 : $\beta_1 = 0$ derived from the simplified REML function using non-iterative and fully converged random effect estimator, for the GSM (top row) and a kinship from GOBS study (bottom row). There is no apparent difference between the two random effect estimators, and both are consistent with a valid (uniform) P-value distribution.

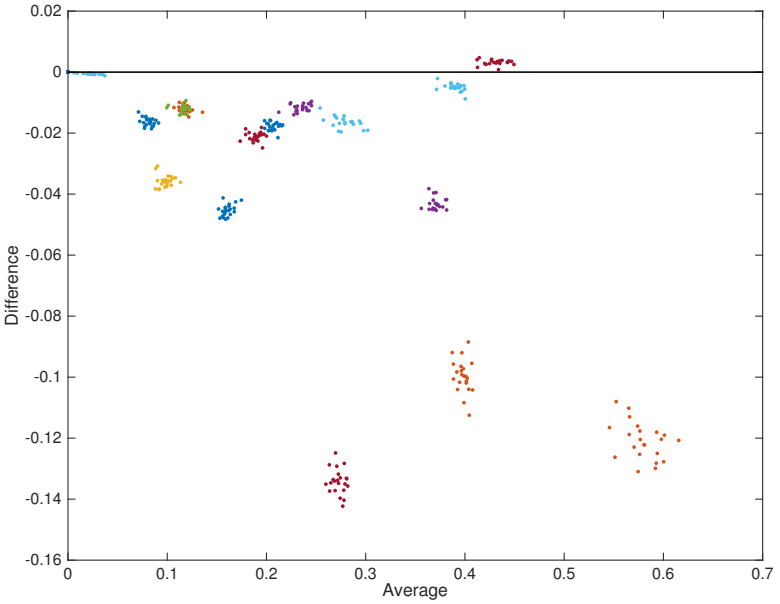


(a) Null comparison

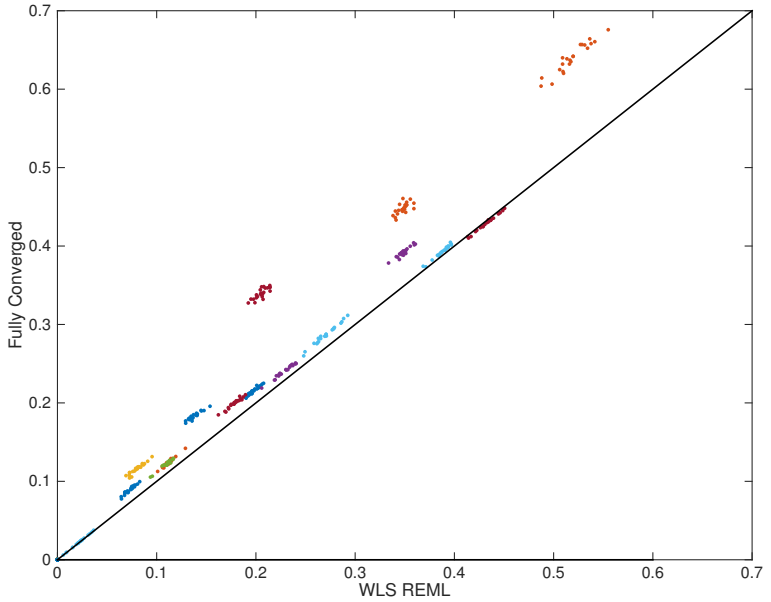


(b) Power comparison

Figure 4.6: Simulation 4, comparing parametric false positive rates for $H0 : \beta_1 = 0$ with FaST-LMM's LRT and the score test based on 1-step optimisation of the simplified REML function, using 100 random markers and 5000 realisations. The overall error rates for FaST-LMM and the score test are 4.94% and 4.89%, respectively, for nominal 5% where MCCI is (4.40%,5.60%). While overall power is largely similar for both approaches (FaST-LMM = 15.25% and The score 15.22%), FaST-LMM is slightly more powerful but 200-fold slower.



(a) Random effect estimation comparisons using Bland-Altman plot.



(b) Random effect estimation comparisons using scatter plot.

Figure 4.7: Real data analysis, comparing 1-step and fully converged random effect estimators of σ^2_{A} based on the simplified REML function. Colors represent random effect estimation at different regions for all 22 chromosomes. The Bland-Altman (a) and the scatter plot (b) show consistent trend towards underestimation of random effect using non-iterative method, though this apparent increased accuracy comes with a 10^9-fold greater computation time.

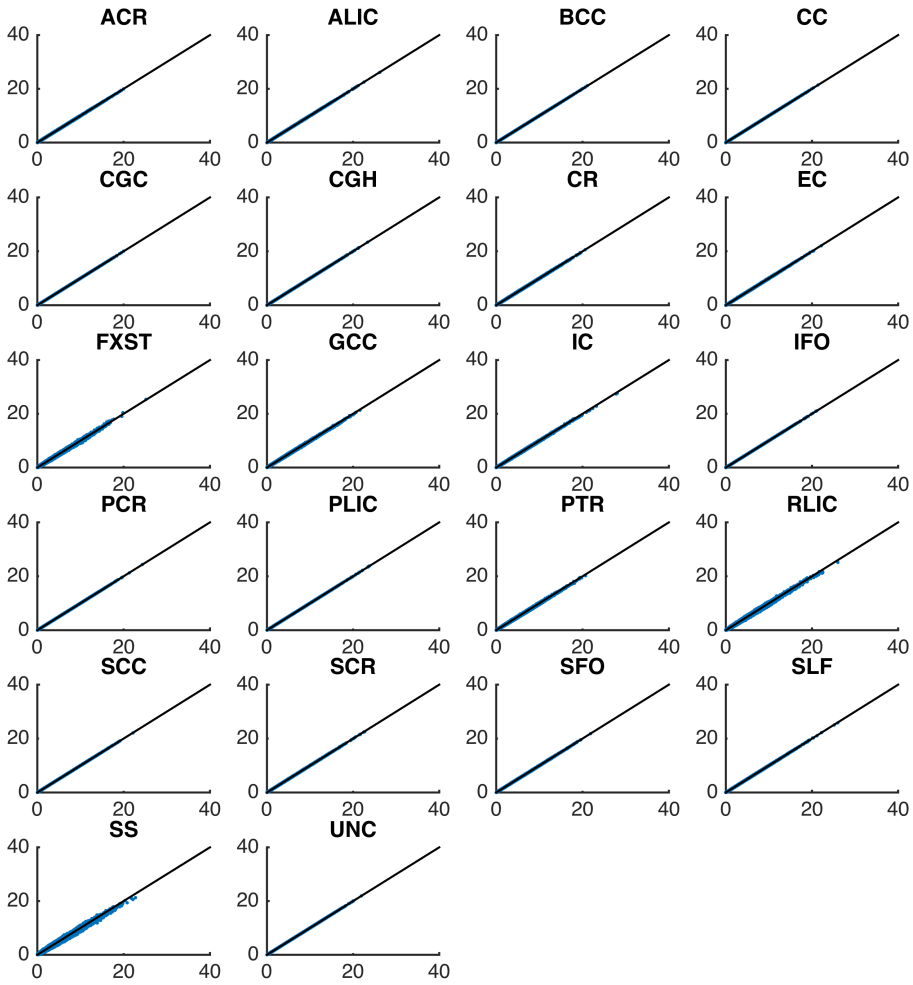


Figure 4.8: Real data analysis, comparing values of the score test for association testing ($H_0 : \beta_1 = 0$) using non-iterative and fully converged random effect estimators. Each plot represents a ROI where x-axis shows score test using WLS REML estimator and y-axis represents score test using the fully converged random effect estimator. The two approaches are almost identical.

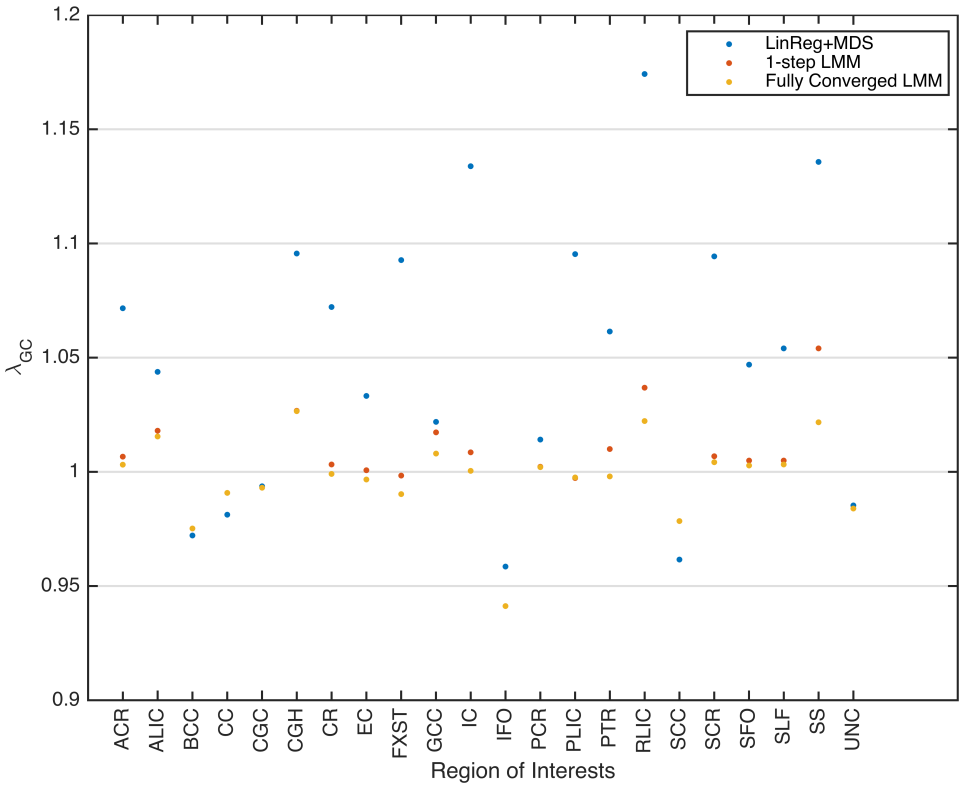


Figure 4.9: Real data analysis, comparing the genomic control values of the score test based on the simplified REML function using either fully converged random effect estimator (Fully converged LMM, yellow dots) or the WLS REML random effect estimator (1-step LMM, red dots) with the linear regression with MDS as nuisance fixed effects (LinReg+MDS, blue dots) for 22 ROIs in the CEU sample. Our proposed method consistently gives smaller genomic factor regardless of random effect estimation method.

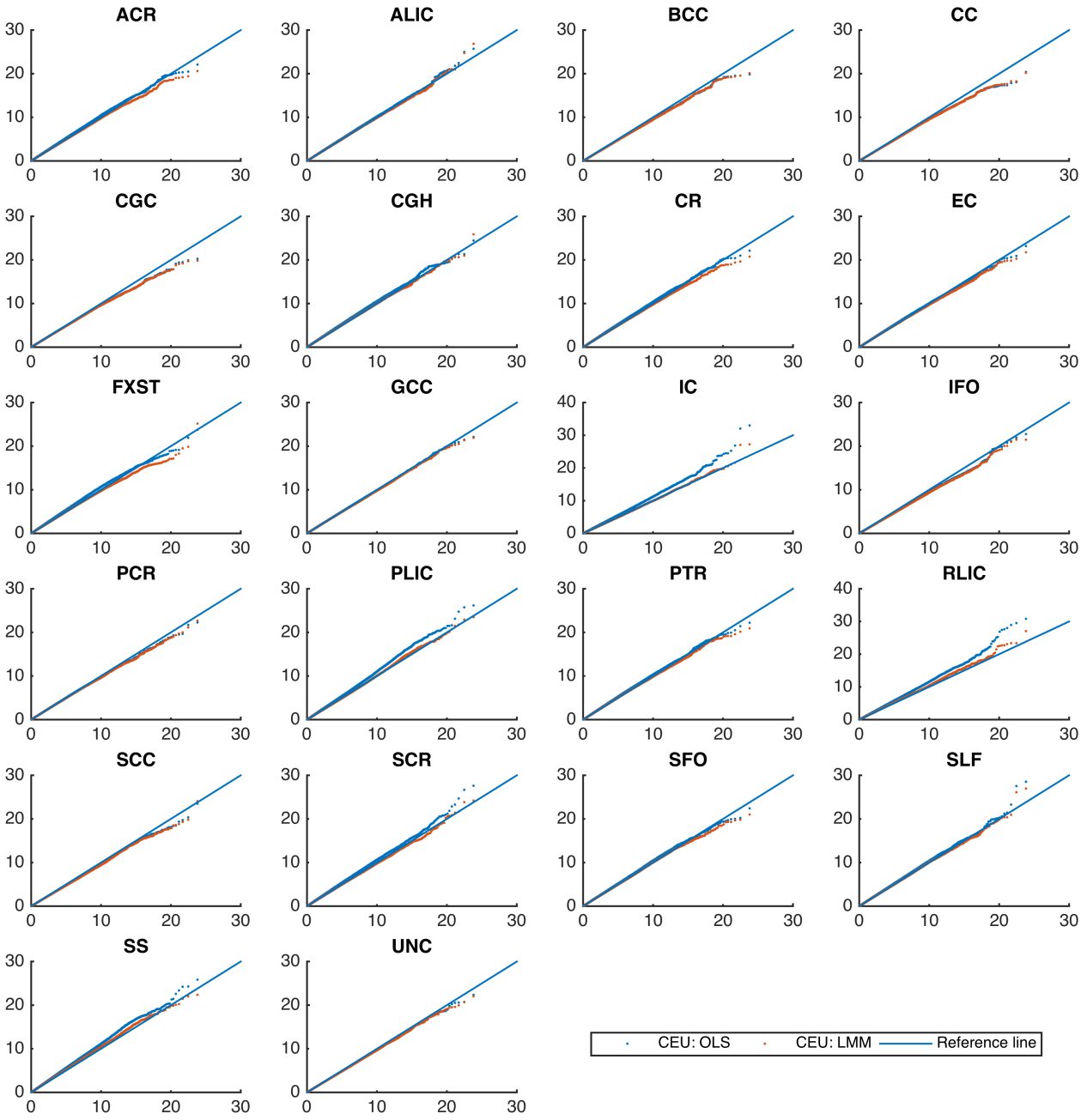


Figure 4.10: Real data analysis, QQ plot for comparing the score test based on the simplified REML function using the WLS REML random effect estimator with the linear regression with MDS as nuisance fixed effects. Each plot corresponds to different ROIs. These plots show either an identical distribution or slightly larger values for the OLS approach. However the OLS approach has poor genomic control (Figure 4.9).

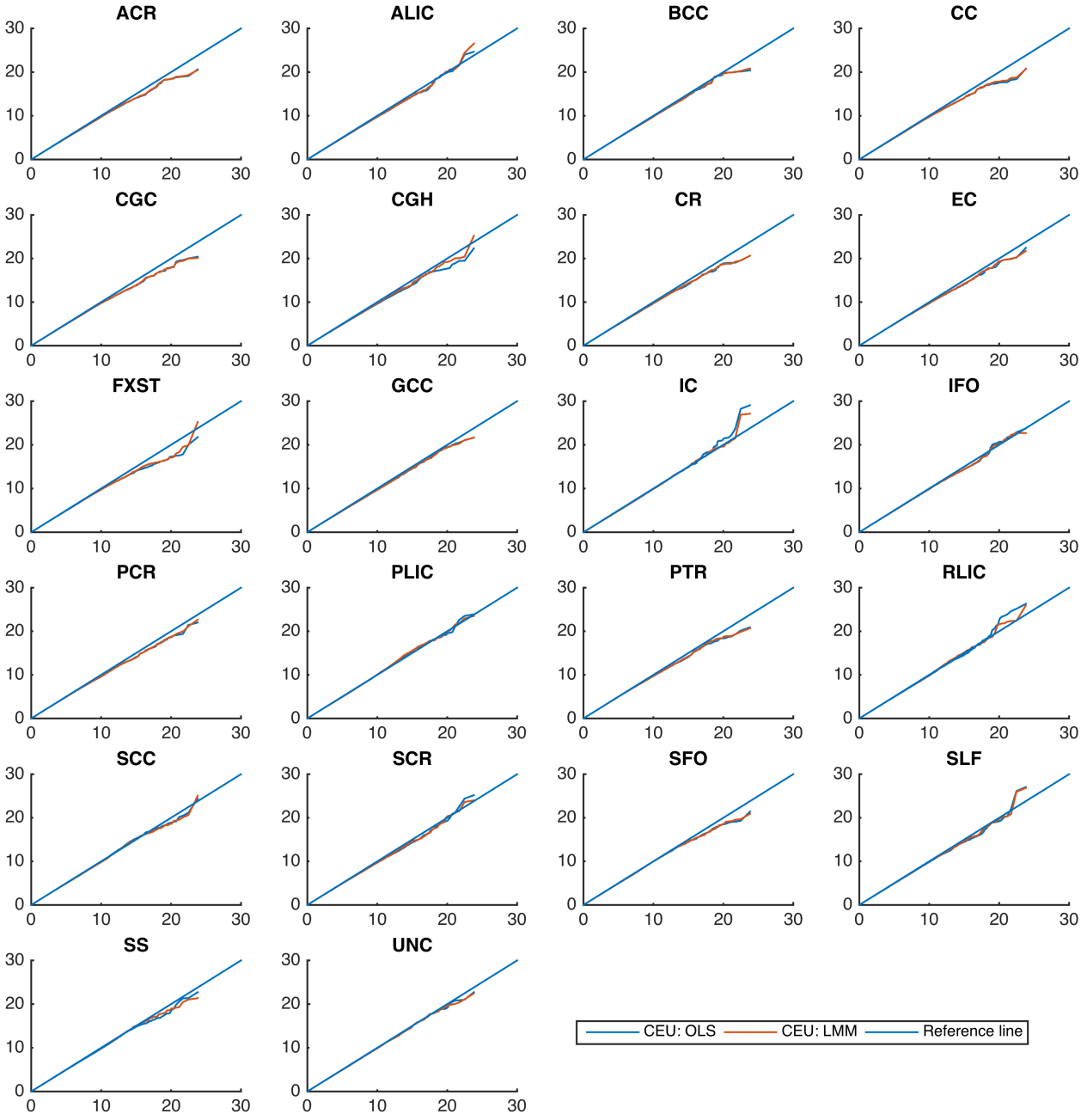


Figure 4.11: Real data analysis, QQ plot for comparing the adjusted association statistics for genomic control values. Each plot corresponds to different ROIs. These plots show after adjustment we get essentially identical results for the score test based on the simplified REML function using the WELS REML random effect estimator and the OLS approach.

Chapter 5

Bivariate Genetic Analysis

Multivariate genetic analysis have been used widely in the quantitative genetic field, for instance to assess peliotropy or genetic correlation. Genetic analysis of correlated traits could increase power of GWA study and boost the accuracy of heritability estimates because of peliotropy. The Multivariate Linear mixed effect model (mvLMM) is the statistical method that can be used for this purpose, however fitting the mvLMM involves multidimensional optimisation where, in the simplest case for the bivariate model, it requires optimising a potentially non-convex likelihood function with at least 6 unknown covariance parameters.

Several algorithms have been proposed to optimise the maximum or restricted likelihood functions including Fisher's scoring, Newton or EM algorithms. Despite these advances in optimisation, they could be either very slow or suffer from convergence failure. For instance, methods based on information matrix like Newton or Fisher's scoring methods could suffer from convergence problems, or EM type algorithm might be slow. Furthermore these methods do not take into account positive-definite condition of covariance components, hence parameter estimates could be invalid. To overcome aforementioned issues in mvLMM optimisation here we propose a two stage optimisation procedures where Fisher's scoring method is accompanied by a non-linear constrained optimisation where the covariance components positive-definiteness is imposed as constraint.

The reminder of this chapter is as follows. We first outline the theory, then the proposed method is evaluated using a simulation study, following discussion.

5.1 Theory

We consider simple bivariate LMM model where there is no missing data, though with out loss of generality this development could be extended to k-variate traits. For a fixed effects design matrix that is the same for both traits we have the model

$$Y = X\beta + g + \epsilon, \tag{5.1}$$

where $Y = [Y_1 \quad Y_2]'$ is the $2N$-vector of stacked traits; Y_i, $i = 1, 2$, is the N-vector of phenotype vector for trait i; X is the $2N \times p$ stacked covariate matrix; $g \sim N(0, \Sigma_g \otimes (2\Phi))$ where (2Φ) is the kinship matrix; $\otimes$ denotes kronecker product; and $\epsilon \sim N(0, \Sigma_e \otimes I_N)$ where I_N is the identity matrix. In this setting, the traits covariance matrix ($\Sigma = \text{cov}(Y_1, Y_2)$) is modelled as the sum of the genetic and environmental covariance components

$$\Sigma = \Sigma_g \otimes (2\Phi) + \Sigma_e \otimes I,$$

where $\text{cov}(g, \epsilon) = 0$ and Σ_g and Σ_e are 2×2 matrices with the following form:

$$\Sigma_g = \begin{bmatrix} \sigma^2_{g,11} & \sigma^2_{g,12} \\ \sigma^2_{g,12} & \sigma^2_{g,22} \end{bmatrix}, \quad \Sigma_e = \begin{bmatrix} \sigma^2_{e,11} & \sigma^2_{e,12} \\ \sigma^2_{e,12} & \sigma^2_{e,22} \end{bmatrix}, \tag{5.2}$$

where $\sigma^2_{g,ii}, \sigma^2_{e,ii}$, $i = 1, 2$, are the genetic and environmental variance components, respectively; $\sigma^2_{g,12}$ and $\sigma^2_{e,12}$ are the genetic and environmental covariance components, respectively. Under an assumption that data follows multivariate normal distribution, the parameters can be estimated by maximizing the loglikelihood function:

$$\ell\left(\Sigma_g, \Sigma_e, \beta \middle| Y, X\right) = -\frac{1}{2}\left(\text{Const} + \ln|\Sigma| + (Y - X\beta)'\,\Sigma^{-1}\,(Y - X\beta)\right). \tag{5.3}$$

Simplified Bivariate LMM function

Computational burden of likelihood function in Eq (5.3) can be substantial even for small datasets, in particular the determinant of Σ must be computed, along with a quadratic form of Σ with the residuals. As discussed in chapter 3, the phenotype and model (Eq. (5.1)) can be transformed using with eigenvector matrix of kinship matrix to simplify the univariate LMM likelihood function. Here the orthogonal transformation matrix S, which satisfies $(2\Phi) = SD_gS'$, is used to transform each

trait and univariate model separately

$$(S' \otimes I_2)Y = (S' \otimes I_2)X\beta + (S' \otimes I_2)g + (S' \otimes I_2)\epsilon$$

which we write as

$$Y^* \quad = \quad X^*\beta + g^* + \epsilon^*, \tag{5.4}$$

where Y^* is the transformed phenotype vector; X^* are the transformed fixed effect covariates; g^* is the transformed genetic random component; and ϵ^* is the transformed residual. In this setting, $\text{cov}(g^*) = \Sigma_g \otimes D_g$ where $D_g = \text{diag}\{\lambda_{gi}\}$ is a diagonal matrix of the kinship matrix eigenvalues, $\text{cov}(\epsilon^*) = \Sigma_e \otimes I_N$ and the covariance matrix of simplified model can be written as

$$\Sigma^* = \Sigma_g \otimes D_g + \Sigma_e \otimes I. \tag{5.5}$$

Finally, the likelihood function becomes sum of N bivariate normal distributions as follows.

$$\ell(\beta, \Sigma_g, \Sigma_e | Y^*, X^*) = -\frac{1}{2}\left[\text{Const} + \sum_{i=1}^{N} \ln|\Sigma_i| + \sum_{i=1}^{N}(Y_i^* - X_i^*\beta)'\Sigma_i^{-1}(Y_i^* - X_i^*\beta)\right] \tag{5.6}$$

where $\Sigma_i = \lambda_{gi}\Sigma_g + \Sigma_e$; Y_i^* is the 2-vector of stacked transformed phenotypes for subject i; and $X_i^* = [X_{i1}^* \quad X_{i2}^*]'$ is the $2 \times p$ matrix of transformed covariates for subject i.

5.1.1 Likelihood Optimisation

Several algorithms have been proposed to maximize the mvLMM likelihood function (Eqns, (5.3) or (5.6)), including methods based on extension of Newton-Rophson method (Jennrich and Sampson (1976)), Fisher's scoring method (Lange *et al.* (1976)) or EM type algorithms. These methods do not take into account the constraint that covariance components must be semi-positive definite. Constrained non-linear optimization methods can alternatively be used to optimize the likelihood function, however they can be quite slow. Another approach is Sequential Quadratic Programming (SQP), a variant of Newton's method for optimisation with non-linear constraints. SQP uses line search approach for optimisation where the direction of search involves optimising set of sub-problems. At each iteration, an approximation of Lagrangian function Hessian is used to generate a set of quadratic functions, subject to linear constraints, whose solution define the search direction in the parameter

space. However, we found that SQP performance might be influenced by the choice of initial value. Hence we propose to perform two step optimization where in the first step Fisher's scoring method is used to perform the optimisation and then the solution is used as the initial value for Sequential Quadratic Programming (SQP) to ensure a semi-positive definite answer.

In the first step of our proposed method we use Fisher's scoring to estimate the parameters. Each iteration of scoring method corresponds to solving the following linear systems:

$$\mathbf{B}\theta = \mathbf{d}, \tag{5.7}$$

where $\theta = \left(\sigma^2_{g,11}, \sigma^2_{e,11}, \sigma^2_{g,22}, \sigma^2_{e,22}, \sigma^2_{g,12}, \sigma^2_{e,12}\right)'$ is the vector of parameter estimates; $\mathbf{d}$ denotes a vector of quadratic forms; and $\mathbf{B}$ is the information matrix. That is, the ith element of $\mathbf{d}$ is

$$d_i = (Y - X\beta)'\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}(Y - X\beta)$$

and the ijth element of $\mathbf{B}$ is

$$b_{ij} = \mathrm{tr}(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}).$$

Perhaps the most computationally expensive part of scoring method (Eq. (5.7)) is the evaluation of Σ^{-1} in each iteration. Here we adopt the approach of Meyer (1985) who proposed to use a transformation that accelerates mvLMM optimisation significantly. In each iteration, a canonical transformation is used to decorrelate the covariance components and convert the mvLMM into a set of univariate LMMs. The canonical parameters are updated and back-transformed to the original scale. In our first step of our proposed procedure, we build upon that work to speed up simplified model (Eq. (5.4)) optimisation using scoring method.

Despite decorrelating individuals by premultiplying the kinship matrix eigenvector S (Eq. (5.4)) that makes each block in covariance matrix diagonal, in each iteration, further simplification in Σ^{-1} evaluation can be achieved by decorrelating the traits using current covariance components estimates. When Σ_g and Σ_e are semi-positive definite matrices there exist a transformation, so called canonical transformation (Q), that diagonalize the both covariance matrices simultaneously. To derive the canonical transformation, we start with eigendecomposition of Σ_e and

define P as follows:

$$\begin{aligned} P &= B^{-1/2}U', \\ \Sigma_e &= UBU', \end{aligned}$$

where U is the 2×2 eigenvector matrix of Σ_e; and B is the 2×2 matrix of eigenvalues. Then the transformation can be derived using eigenvectors of $P\Sigma_g P'$ and $B^{-1/2}$,

$$\begin{aligned} P\Sigma_g P' &= M\Lambda M', \\ Q &= M'B^{-1/2}M, \end{aligned}$$

where M is the 2×2 eigenvector matrix; $\Lambda = \text{diag}\{\lambda_i\}$ is the 2×2 diagonal matrix of eigenvalues; and Q is the canonical transformation that decorrelates the traits,

$$\begin{aligned} Q\Sigma_g Q' &= \Lambda, \\ Q\Sigma_e Q' &= I. \end{aligned}$$

Where as the simplified mvLMM model (Eq. (5.4)) decorrelates the individuals, this partially canonical transformation decorrelates the traits:

$$(Q \otimes I_N)Y^* = X^*(Q \otimes I_p)\beta + (Q \otimes I_N)g^* + (Q \otimes I_N)\epsilon$$

which we write as

$$Y^c = X^*\beta^c + g^c + \epsilon^c, \tag{5.8}$$

where Y^c is the $2N$-vector of phenotype in canonical scale where Y_i^c, $i = 1, 2$, are now uncorrelated; β^c is the p-vector of regression coefficients; $g^c \sim N(0, \Lambda \otimes D_g)$ is the genetic effect; and $\epsilon^c \sim N(0, I)$ is the residual effect, all in the canonical scale. In this fashion, the covariance matrix of transformed phenotype in the canonical scale can be expressed as

$$\Sigma^c = \Lambda \otimes D_g + I.$$

The correlation among the canonical parameters remains only between corresponding elements of the genetic and environmental variance (Eq. (5.2)). Specifically the

information matrix has the form

$$\mathbf{B} = \begin{bmatrix} E\begin{bmatrix} \frac{\partial^2 \ell}{\partial^2 \sigma^{2c}_{g,11}} & \frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,11} \partial \sigma^{2c}_{e,11}} \\ \frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,11} \partial \sigma^{2c}_{e,11}} & \frac{\partial^2 \ell}{\partial^2 \sigma^{2c}_{e,11}} \end{bmatrix} & 0 & 0 \\ 0 & E\begin{bmatrix} \frac{\partial^2 \ell}{\partial^2 \sigma^{2c}_{g,22}} & \frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,22} \partial \sigma^{2c}_{e,22}} \\ \frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,22} \partial \sigma^{2c}_{e,22}} & \frac{\partial^2 \ell}{\partial^2 \sigma^{2c}_{e,22}} \end{bmatrix} & 0 \\ 0 & 0 & E\begin{bmatrix} \frac{\partial^2 \ell}{\partial^2 \sigma^{2c}_{g,12}} & \frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,12} \partial \sigma^{2c}_{e,12}} \\ \frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,12} \partial \sigma^{2c}_{e,12}} & \frac{\partial^2 \ell}{\partial^2 \sigma^{2c}_{e,12}} \end{bmatrix} \end{bmatrix}$$

where $\sigma^{2c}_{g,ij}$ and $\sigma^{2c}_{e,ij}$ $(i,j = 1,2)$ are the genetic and residual covariance components in the canonical scale respectively; $E[.]$ denotes expectation with respect to multivariate normal density. In the canonical scale, the quadratic form for genetic components $(\sigma^{2c}_{g,ij}, \quad i,j = 1,2)$ become

$$d_{ij} = \begin{bmatrix} \epsilon_1^{'c}\Sigma_1^{-1c} & 0 \\ 0 & \epsilon_2^{'c}\Sigma_2^{-1c} \end{bmatrix} D_g \otimes I_{ij} \begin{bmatrix} \Sigma_1^{-1c}\epsilon_1^c & 0 \\ 0 & \Sigma_2^{-1c}\epsilon_2^c \end{bmatrix},$$

and quadratic terms for environmental components $(\sigma^{2c}_{e,ij}, \quad i,j = 1,2)$ in the canonical scale can be written as

$$d_{ij} = \begin{bmatrix} \epsilon_1^{'c}\Sigma_1^{-1c} & 0 \\ 0 & \epsilon_2^{'c}\Sigma_2^{-1c} \end{bmatrix} I \otimes I_{ij} \begin{bmatrix} \Sigma_1^{-1c}\epsilon_1^c & 0 \\ 0 & \Sigma_2^{-1c}\epsilon_2^c \end{bmatrix},$$

where ϵ_i^c is the N-vector of residual for trait i in the canonical scale, Σ_i^{-1c} is the $N \times N$ inverse covariance matrix for trait i in the canonical scale where $\Sigma_i^c = \lambda_i D_g + I$ and I_{ij} is the 2×2 matrix of zeros that ij and ji elements are 1. In this fashion, the information matrix $(\mathbf{B})$ elements in the canonical scale can be expressed as:

$$\begin{aligned} E\left[\frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,ij} \partial \sigma^{2c}_{g,ij}}\right] &= \mathrm{tr}\left(\Sigma^{-1c}\frac{\partial \Sigma^c}{\partial \sigma^{2c}_{g,ij}}\Sigma^{-1c}\frac{\partial \Sigma^c}{\partial \sigma^{2c}_{g,ij}}\right), \\ &= \mathrm{tr}\left(\mathrm{diag}(\Sigma_i^{-1c})(I_{i,j} \otimes D_g)\mathrm{diag}\left(\Sigma_i^{-1c}\right)(I_{i,j} \otimes D_g)\right), \\ E\left[\frac{\partial^2 \ell}{\partial \sigma^{2c}_{g,ij} \partial \sigma^{2c}_{e,ij}}\right] &= \mathrm{tr}\left(\Sigma^{-1c}\frac{\partial \Sigma^c}{\partial \sigma^{2c}_{g,ij}}\Sigma^{-1c}\frac{\partial \Sigma^c}{\partial \sigma^{2c}_{e,ij}}\right), \\ &= \mathrm{tr}\left(\mathrm{diag}(\Sigma_i^{-1c})(I_{i,j} \otimes D_g)\mathrm{diag}(\Sigma_i^{-1c})(I_{i,j} \otimes I)\right), \\ E\left[\frac{\partial^2 \ell}{\partial \sigma^{2c}_{e,ij} \partial \sigma^{2c}_{e,ij}}\right] &= \mathrm{tr}\left(\Sigma^{-1c}\frac{\partial \Sigma^c}{\partial \sigma^{2c}_{e,ij}}\Sigma^{-1c}\frac{\partial \Sigma^c}{\partial \sigma^{2c}_{e,ij}}\right), \end{aligned}$$

$$= \quad \mathrm{tr}\left(\mathrm{diag}(\Sigma_i^{-1c})(I_{i,j} \otimes I)\mathrm{diag}(\Sigma_i^{-1c})(I_{i,j} \otimes I)\right),$$

where $\mathrm{diag}(\Sigma_i^{-1c})$ represents block diagonal matrix. With some use of matrix algebra, variance, covariance parameters in the canonical scale can be estimated as follows:

$$\begin{aligned}
\begin{bmatrix} \sigma^{2c}_{g,11} \\ \sigma^{2c}_{e,11} \end{bmatrix} &= (U'\Sigma_1^{-2c}U)^{-1}U'\Sigma_1^{-2c}\epsilon_1^{2c}, \\
\begin{bmatrix} \sigma^{2c}_{g,22} \\ \sigma^{2c}_{e,22} \end{bmatrix} &= (U'\Sigma_2^{-2c}U)^{-1}U'\Sigma_2^{-2c}\epsilon_2^{2c}, \\
\begin{bmatrix} \sigma^{c}_{g,12} \\ \sigma^{c}_{e,12} \end{bmatrix} &= (U'\Sigma_1^{-1c}\Sigma_2^{-1c}U)^{-1}U'\Sigma_1^{-1c}\Sigma_2^{-1c}\epsilon_1^{c}\epsilon_2^{c},
\end{aligned}$$

where $U = [1 \quad D_g]$, 1 is the N-vector of ones, $\Sigma_i^c = 1 + \lambda_i D_g$ is the covariance matrix for each trait in the canonical space. In the canonical scale, variance and covariance parameters are updated by regressing squared and cross product of decorrelated residuals on eigenvalues of kinship matrix respectively. Once parameters are updated in canonical scale, the parameter estimates in the original scale can be obtained by applying the inverse transformation to the updated parameters in the canonical scale.

$$\begin{aligned}
\Sigma_g &= Q^{-1}\Sigma_g^c Q{-1}', \\
\Sigma_e &= Q^{-1}\Sigma_e^c Q{-1}'.
\end{aligned}$$

The aforementioned procedure is iterated until the following criteria meet:

$$\begin{aligned}
\frac{\ell^{j+1} - \ell^j}{\ell^j} &\leq 10^{-6}, \\
\sqrt{\frac{\sum_{i=1}^6(\theta_i^{j+1} - \theta_i^j)^2}{\sum_{i=1}^6(\theta_i^{j+1})^2}} &\leq 10^{-6},
\end{aligned}$$

where ℓ^{j+1} is the likelihood value at $(j+1)$th iteration; θ_i^{j+1}, $i = 1, \cdots, 6$, is the ith element of parameter estimate vector θ at the $(j+1)$th iteration. Then the parameter estimates are fed into optimising the simplified likelihood with constraint, where covariance components are imposed to be semi-positive definite:

$$\begin{aligned}
|\mathrm{cov}_g| &\leq \sqrt{\sigma^2_{g1}\sigma^2_{g2}}, \\
|\mathrm{cov}_e| &\leq \sqrt{\sigma^2_{e1}\sigma^2_{e2}},
\end{aligned}$$

$$0 \leq h_i^2 \quad \leq \quad 1.$$

The two steps optimisation can be summarised as follows

1. Obtain initial value for Σ_g and Σ_e
2. Find the canonical transformation Q.
3. Update the parameter estimates in the canonical space.
4. Back-transform the parameter estimates to the original scale.
5. Iterate step 2-4 until relative change of loglikelihood is less than 10^{-8}.
6. Apply SQP algorithm using the initial value from step 5.

5.1.2 Genetic Correlation Inference

The genetic correlation can be estimated following phenotypic covariance decomposition in the Equation (5.2) as

$$\rho_g = \frac{\widehat{\text{cov}}_g}{\sqrt{\hat{\sigma}_{g1}^2 \hat{\sigma}_{g2}^2}}$$

where $\widehat{\sigma}_{gi}^2 \quad i = 1, 2$ is the estimated genetic variance for each trait, $\hat{\text{cov}}_g$ is the estimated covariance component. The Likelihood ratio test (LRT) can be used for genetic correlation testing where $\text{H}_0 : \rho_g = 0$ hypothesis testing is equivalent to $\text{H}_0 : \text{cov}_g = 0$. The LRT follows χ^2 with 1 degree of freedom asymptotically.

Permutation Test

The empirical null distribution of LRT for ρ_g testing are built by fitting the simplified bivariate model (Eq. (5.4)) with covariance matrix that off-diagonal part is permuted in each permutation step.

$$\begin{aligned} Y^* &= (I \otimes X^*)\beta + g^* + \epsilon^* \\ \text{cov}(Y^*) &= \begin{bmatrix} \sigma_{g1}^2 D_g + \sigma_{e1}^2 I & \text{cov}_g P D_g P' + \text{cov}_e I \\ \text{cov}_g P D_g P' + \text{cov}_e I & \sigma_{g2}^2 D_g + \text{cov}_e I \end{bmatrix} \end{aligned}$$

5.2 Evaluation

We conducted a simulation study to evaluate and compare the error rate and convergence performance of two step optimisation procedure with scoring method and

constraint optimisation. The response variable is assumed to be $Y = X\beta + \epsilon$ where ϵ follows $N(0, \Sigma)$, $\Sigma = \Sigma_g \otimes (2\Phi) + \Sigma_e \otimes I$. The design matrix X is same for both traits and consists of an intercept, a linear trend vector X_1 and a quadratic vector X_2 between 1 and -1, with $\beta = [0, 0, 10]$. Kinship structure Φ is based on real pedigrees from GOBS study with 18 families and 413 individuals and the simulations considered a range of true covariance components.

Table 5.1: Variance and Covariance terms used in Simulation.

Setting	σ^2_{g1}	σ^2_{g2}	cov_g	σ^2_{e1}	σ^2_{e2}	cov_e
S1	0.80	0.80	0	0.2	0.2	0.1
S2	0.65	0.65	0	0.35	0.35	0.40
S3	0.80	0.30	0	0.20	0.70	0.40

5.3 Results

Convergence failure comparison (Table 5.2) between scoring method, constraint optimisation and two step optimisation revealed that two steps optimisation potentially could be used as a stable numerical optimisation procedure that gives valid error rate in the Monte Carlo confidence interval. Bias and mean squared error of genetic covariance estimation is provided in Table 5.3. The parametric and permutation based error rate for simulation study S1 is 4.8% and 5.02% which are in the Monte Carlo confidence interval (4.22%, 5.78%).

Table 5.2: Comparing different optimisation methods error rate and convergence failure based on 3000 realisations, Monte Carlo confidence interval is (4.22%, 5.78%).

Optimisation Method	Parametric Rejection Rate	Convergence Failure Rate
Scoring	12 %	15%
Constraint	25%	20%
Two step	4.5%	0

Table 5.3: Genetic covariance bias and mean squared error from two step optimisation.

Setting	Bias	Mean Squared Error
S1	-0.0059	0.0070
S2	0.0540	0.0093
S3	0.0050	0.0055

5.4 Discussion

In this chapter, optimisation methods for mvLMM model were evaluated. Fitting a mvLMM is computationally intensive and involves a multidimensional optimization of a potentially non-convex function that may have multiple local optima. Also current optimisation methods do not take into account positive-definite condition of covariance components, hence parameter estimates could be invalid. Sequential Quadratic Optimisation (SQP), a variant of Newton's method for optimisation with non-linear constraints can be used to force parameter estimate lies in the valid parameter space.

The simplified mvLMM were obtained to reduce the mvLMM likelihood function complexity by decorrelating the individuals within each trait using the kinship matrix eigenvector. Fisher's scoring method and SQP performances in optimising the simplified mvLMM have been evaluated. The scoring method for optimising the simplified mvLMM can be speed up in each iteration using canonical transformation in each iteration that further decorrelates the traits using current values of Σ_g and Σ_e. Based on the simulation study, we found scoring method or constrained optimisation alone could not provide stable solution for simple bivariate case. Also we found that the SQP performance is influenced by the choice of initial value and it is slower than the scoring method.

A two-step procedure was proposed for the simplified mvLMM optimisation, scoring method following SQP, the variant of non-linear constraint optimisation summarised as follows. First, the scoring method was used to obtain a parameter estimates that the simplified mvLMM likelihood relative change is less than 10^{-8}. In the second step, the covariance components semi-positive definite condition were imposed assure obtaining estimates in the parameter space using the initial value from the previous step. Based on the simulation study, we found that two steps optimisation could provide valid error rate and prevents convergence failures for simple bivariate case. However, more evaluation on other simulation settings and real data analysis is required to evaluate the two step optimisation stability.

Chapter 6

Conclusion and Future Work

The last decades of neuroimaging research have brought immense insight into our understating of the human brain function and structure. Imaging Genetics is a relatively new and growing discipline that concerns explaining individual differences in neuroimaging phenotypes using genetic and environmental factors. The linear mixed effect model (LMM) is a widely used and flexible statistical tool that accommodates longitudinal & repeated measures data and data from dependent (e.g. related) individuals. We have focused on LMMs because they are the key ingredient for the genetic analysis of quantitative traits, including evaluating overall genetic effect, i.e. heritability, and finding genetic markers that are associated with phenotypes. However, parameter estimates and hypothesis testing rely on numerical optimisation which is prone to convergence failure and computationally intensive.

Despite advances in the field of quantitative genetic to accelerate LMM likelihood optimisation driven by growing sample sizes and deep sequencing, genetic analysis of high dimensional imaging phenotypes at each voxel/element of image presents a challenge both in terms of computational intensity and the need to account for elevated false positive risk because of massive multiple testing issues. Furthermore, for neuroimaging spatial statistics, like familywise error (FWE) corrected inference with either voxel- or cluster-wise inference, the relevant parametric null distributions are intractable. While Random field theory (Worsley *et al.*, 1992; Friston *et al.*, 1994b; Nichols and Hayasaka, 2003) results exist for χ^2 images (Cao, 1999), they are not applicable here as the test statistic image can not be expressed as a linear combination of component error fields. Hence, there is a compelling need for alternative inference procedures that make fewer assumptions. Permutation tests are a type of non-parametric procedure that can be used to derive empirical null distribution of maximum voxel or cluster extend. However, deploying permutation

test for fixed or random effects hypothesis testing in LMM is impractical because it depends on numerical optimisation.

The overall goal of this thesis was to present computationally efficient linear mixed effect model for genetic analyses of high dimensional imaging phenotype that is consider multiple testing problem. The key to this work is to use orthogonal transformations that simplify the likelihood and restricted likelihood functions following the approximate, non-iterative variance component estimator that substantially reduces LMM complexity.

In chapter 3, heritability estimation computational intensity and inference inaccuracy were addressed. Fast, non-iterative heritability estimator was introduced based on simplified likelihood function to reduce computational intensity of heritability estimation for high dimensional imaging data. Simulation study and real data analysis revealed that WLS and fully converged heritability estimators have comparable performances in terms of bias and mean squared error. A suit of test statistics including the Wald, LRT and the score test performances for the sake of inference were evaluated using intensive simulation studies in terms of rejection rates. We found that parametric inference could be invalid because of untenable $50:50$ mixture of chi squared assumptions. However, permutation test could provide valid and more powerful results. The score test based on the simplified likelihood function which is the explained sum of square of regressing squared residuals on kinship matrix eigenvalues was introduced as an alternative to the LRT for heritability hypothesis testing. While it does not depend on numerical optimisation, it could be used to perform standard spatially oriented inferences accompanying with the permutation test to derive maximum statistic empirical null distribution.

LMM efficiency in controlling population structure and avoiding spurious associations in admixed samples motived us to use it for association analysis of high-dimensional imaging phenotypes. Advances in reducing LMM complexity in GWA concern growing sample sizes and number of genetic markers being tested. However, increased computational intensity that arises from fitting LMM at each voxel of whole image and FWE corrected p-values for number of correlated tested has not been addressed yet.

In chapter 4, LMM complexity in variance component estimation and performing association testing at millions of markers were reduced using the orthogonal transformation that simplifies the restricted likelihood function. The non-iterative, 1-step random effect estimator was introduced using projected polygenic model using eigendecomposition of kinship matrix adjusted for nuisance fixed effect terms. We evaluated the Wald, LRT and the score tests performances for association testing

based on the simplified ML or REML functions that we could not observe appreciable differences among them. Moreover, the simplified REML function enabled us to implement efficient score test that is fast enough to perform voxel-wise GWA accompanied by cluster-wise inference. Real data analysis revealed that including MDS components as a nuisance fixed effects in multiple regression model could not control the population structure even in a sample of individual with European ancestry. However, the mixed effect approach provided genomic control close to one that indicates good control of population structure in GWA studies.

In chapter 5 genetic analyses of correlated traits were discussed to estimate genetic correlation or evaluating peliotropy effect in GWA studies. The mvLMM likelihood function could be potentially non-convex which makes the multidimensional optimisation challenging. Two steps optimisation, Fisher's scoring method following nonlinear constrained optimisation, were introduced to overcome this issue. The simulation study results showed that two steps procedure could avoid convergence failure and provide valid parametric error rate. Furthermore, a permutation test for genetic correlation hypothesis testing was introduced to derive maximum statistic null distribution for FWE corrected error rates.

Our proposed non-iterative random effect estimator that accelerates heritability and GWA analyses works on a LMM model with only one random effect. However, for some designs like twin studies, there is a need to fit LMM with more than one random effects, i.e a heritability study with ACE model. Hence developing an orthogonal transformation that diagonalises the LMM covariance matrix with more than one random effects could be the subject of further investigations.

Peilotropy plays an important role in discovering genetic markers that are associated with multiple phenotypes. However the mvLMM parameter estimation is the subject of optimizing a potentially non-convex multidimensional likelihood function. Developing a fast and stable optimisation algorithm for the mvLMM, makes multivariate genome wide association analysis with imaging phenotype feasible. Moreover, genetic correlation estimation and hypothesis testing with imaging phenotype require substantial amount of computation. Developing computationally efficient, preferably non iterative, estimator and corresponding test statistic that its empirical null distribution could be derived using the permutation test could be the subject of future research.

Appendix A

Assuming that X is a full column rank, then $[A \quad X]$ is the $N \times N$ square non-singular matrix. To show that the project data loglikelihood is the same as the REML function (Eq. (4.5)) we use the following identity:

$$\begin{bmatrix} A' \\ X' \end{bmatrix} \Sigma \begin{bmatrix} A & X \end{bmatrix} = \begin{bmatrix} A'\Sigma A & A'\Sigma X \\ X'\Sigma A & X'\Sigma X \end{bmatrix}. \tag{A.1}$$

Taking the determinant of LHS of the Equation (A.1) gives us:

$$\begin{aligned} \left| \begin{bmatrix} A' \\ X' \end{bmatrix} \Sigma \begin{bmatrix} A & X \end{bmatrix} \right| &= \left| \begin{bmatrix} A' \\ X' \end{bmatrix} \right| |\Sigma| \left| \begin{bmatrix} A & X \end{bmatrix} \right| \\ &= |\Sigma| \left| \begin{bmatrix} A' \\ X' \end{bmatrix} \begin{bmatrix} A & X \end{bmatrix} \right| \end{aligned} \tag{A.2}$$

The RHS determinant of the Equation (A.1) using the block matrix determinant rule can be written as

$$\left| \begin{bmatrix} A'\Sigma A & A'\Sigma X \\ X'\Sigma A & X'\Sigma X \end{bmatrix} \right| = \left| A'\Sigma A \right| \left| X'\Sigma X - X'\Sigma A (A'\Sigma A)^{-1} A'\Sigma X \right|. \tag{A.3}$$

Hence taking the determinant from the identity in the Equation (A.1) gives us

$$|\Sigma| \begin{vmatrix} A'A & A'X \\ X'A & X'X \end{vmatrix} = \left| A'\Sigma A \right| \left| X'\Sigma X - X'\Sigma A (A'\Sigma A)^{-1} A'\Sigma X \right|,$$

where using $A'A = I$, $A'X = 0$, it can be shown that

$$\log |A'\Sigma A| = \log |\Sigma| + \log |X'\Sigma^{-1}X| - \log |X'X|.$$

Finally, using $A(A'\Sigma A)^{-1}A = P$ (Searle *et al.*, 2009, M4.f p. 451), it is clear that Equations (4.5) and (4.8) are equivalent. Hence, transformed data likelihood function is exactly as same as the REML function.

Bibliography

Abney, M. (2015). Permutation testing in the presence of polygenic variation. *Genetic Epidemiology*, **39**(4), 249–258.

Allison, D. B., Neale, M. C., Zannolli, R., Schork, N. J., Amos, C. I., and Blangero, J. (1999). Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *American journal of human genetics*, **65**(2), 531–44.

Almasy, L. and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics*, **62**(5), 1198–211.

Almasy, L., Dyer, T. D., and Blangero, J. (1997). Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genetic epidemiology*, **14**(6), 953–8.

Amemiya, T. (1977). A note on a heteroscedastic model. *Journal of Econometrics*, **6**(3), 365 – 370.

Amos, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *American journal of human genetics*, **54**(3), 535–43.

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometry?the methods. *NeuroImage*, **11**(6), 805 – 821.

Balding, D. (2006). A tutorial on statistical methods for population association studies. *Nat Rev Genet*, **7**(10), 781–791.

Basser, P., Mattiello, J., and LeBihan, D. (1994). {MR} diffusion tensor spectroscopy and imaging. *Biophysical Journal*, **66**(1), 259 – 267.

Blangero, J. and Almasy, L. (1997). Multipoint oligogenic linkage analysis of quantitative traits. *Genetic Epidemiology*, **14**(6), 959–964.

Blangero, J., Williams, J. T., and Almasy, L. (2000). Quantitative trait locus mapping using human pedigrees. *Human Biology*, **72**(1), 35–62.

Blangero, J., Diego, V. P., Dyer, T. D., Almeida, M., Peralta, J., Kent, J. W., Williams, J. T., Almasy, L., and Göring, H. H. H. (2013). *A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees.*, volume 81. Academic Press.

Blokland, G. A., McMahon, K. L., Hoffman, J., Zhu, G., Meredith, M., Martin, N. G., Thompson, P. M., de Zubicaray, G. I., and Wright, M. J. (2008). Quantifying the heritability of task-related brain activation and performance during the n-back working memory task: A twin fmri study. *Biological Psychology*, **79**(1), 70 – 79.

Brouwer, R. M., Mandl, R. C., Peper, J. S., van Baal, G. C. M., Kahn, R. S., Boomsma, D. I., and Pol, H. E. H. (2010). Heritability of {DTI} and {MTR} in nine-year-old children. *NeuroImage*, **53**(3), 1085 – 1092.

Burggren, A., Zeineh, M., Ekstrom, A., Braskie, M., Thompson, P., Small, G., and Bookheimer, S. (2008). Reduced cortical thickness in hippocampal subregions among cognitively normal apolipoprotein e e4 carriers. *NeuroImage*, **41**(4), 1177 – 1183.

Buse, A. (1973). Goodness of fit in generalized least squares estimation. *The American Statistician*, **27**(3), 106–108.

Buse, A. (1979). Goodness-of-fit in the seemingly unrelated regressions model: a generalization. *Journal of Econometrics*, **10**.

Buse, A. (1984). Tests for additive heteroskedasticity: Goldfeld and quandt revisited. *Empirical Economics*, **9**(4), 199–216.

Cao, J. (1999). The size of the connected components of excursion sets of χ^2, t and F fields. *Advances in Applied Probability*, **31**(3), 579–595.

Cardon, L. R. and Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, **361**(9357), 598 – 604.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, **25**(3), pp. 573–578.

Chiang, M.-C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., Klunder, A. D., Toga, A. W., McMahon, K. L., de Zubicaray, G. I.,

Wright, M. J., Srivastava, A., Balov, N., and Thompson, P. M. (2009). Genetics of brain fiber architecture and intellectual performance. *The Journal of Neuroscience*, **29**(7), 2212–2224.

Chiang, M.-C., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Hickie, I., Toga, A. W., Wright, M. J., and Thompson, P. M. (2011). Genetics of white matter development: A {DTI} study of 705 twins and their siblings aged 12 to 29. *NeuroImage*, **54**(3), 2308 – 2317.

Cho, Y. S., Go, M. J., Kim, Y. J., Heo, J. Y., Oh, J. H., Ban, H.-J., Yoon, D., Lee, M. H., Kim, D.-J., Park, M., *et al.* (2009). A large-scale genome-wide association study of asian populations uncovers genetic factors influencing eight quantitative traits. *Nature genetics*, **41**(5), 527–534.

Crainiceanu, C. (2008). Likelihood ratio testing for zero variance components in linear mixed models. In D. Dunson, editor, *Random Effect and Latent Variable Model Selection*, volume 192 of *Lecture Notes in Statistics*, pages 3–17. Springer New York.

Crainiceanu, C. M. and Ruppert, D. (2004a). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*, **91**(1), 35 – 52.

Crainiceanu, C. M. and Ruppert, D. (2004b). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(1), 165–185.

Crainiceanu, C. M. and Ruppert, D. (2004c). Restricted likelihood ratio tests in nonparametric longitudinal models. *Statistica Sinica*, **14**(3), 713–730.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38.

den Braber, A., Bohlken, M. M., Brouwer, R. M., van 't Ent, D., Kanai, R., Kahn, R. S., de Geus, E. J. C., Hulshoff Pol, H. E., and Boomsma, D. I. (2013). Heritability of subcortical brain measures: A perspective for future genome-wide association studies. *NeuroImage*, **83C**, 98–102.

Devlin, B. and Roeder, K. (1999). Genomic control for association studies. *Biometrics*, **55**(4), 997–1004.

Dominicus, A., Skrondal, A., Gjessing, H., Pedersen, N., and Palmgren, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics*, **36**(2), 331–340.

Draper, N. and Stoneman, D. (1966). Testing for the Inclusion of Variables in Linear Regression by a Randomisation Technique. *Technometrics*, **8**(4), 695–699.

Drikvandi, R., Verbeke, G., Khodadadi, A., and Partovi Nia, V. (2013). Testing multiple variance components in linear mixed-effects models. *Biostatistics (Oxford, England)*, **14**(1), 144–59.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. *Statistical Science*, **19**(4), 676–685.

Falconer, D. and Mackay, T. (1996). *Introduction to Quantitative Genetics*. Longman.

Fischl, B. and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, **97**(20), 11050–11055.

Fisher, R. (1935). *The Design of Experiments*. Oliver and Boyd.

Fitzmaurice, G. M., Lipsitz, S. R., and Ibrahim, J. G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics*, **63**(3), 942–6.

Freedman, D. and Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics*, **1**(4), 292–298.

Freedman, D. a. (2007). How Can the Score Test Be Inconsistent? *The American Statistician*, **61**(4), 291–295.

Friston, K., Holmes, A., Worsley, J.-P. Poline, C.D. Frith, , and Frackowiak, R. (1994a). Statistical parametric maps in functional imaging: a general linear approach. *. . . brain mapping*.

Friston, K., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2002). Classical and bayesian inference in neuroimaging: Theory. *NeuroImage*, **16**(2), 465 – 483.

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. (1994b). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, **1**(3), 210–220.

Ganjgahi, H., Winkler, A. M., Glahn, D. C., Blangero, J., Kochunov, P., and Nichols, T. E. (2015). Fast and powerful heritability inference for family-based neuroimaging studies. *NeuroImage*, **115**, 256 – 268.

Genovese, C. R., Lazar, N. A., and Nichols, T. E. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, **15**(4), 870–878.

Glahn, D. C., Thompson, P. M., and Blangero, J. (2007). Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. *Human Brain Mapping*, **28**(6), 488–501.

Goldfeld, S. and Quandt, R. (1965). Some Tests for Homoscedasticity. *Journal of the American Statistical ...*, **60**(310), 539–547.

Good, P. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. Springer.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**(2), 383–385.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**(358), 320–338.

Helgason, A., Yngvadóttir, B., Hrafnkelsson, B., Gulcher, J., and Stefánsson, K. (2005). An icelandic example of the impact of population structure on association studies. *Nature genetics*, **37**(1), 90–95.

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivieres, S., Jahanshad, N., Toro, R., and et al. (2015). Common genetic variants influence human subcortical brain structures. *Nature*, **520**(7546), 224–229.

Holmes, A. P., Blair, R. C., Watson, J. D., and Ford, I. (1996). Nonparametric analysis of statistic images from functional mapping experiments. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, **16**(1), 7–22.

Hopper, J. L. and Mathews, J. D. (1982). Extensions to multi- variate normal models for pedigree analysis. *Annals of Human Genetics*, **46**, 373–383.

Hua, X., Leow, A. D., Parikshak, N., Lee, S., Chiang, M.-C., Toga, A. W., Jr, C. R. J., Weiner, M. W., and Thompson, P. M. (2008). Tensor-based morphometry

as a neuroimaging biomarker for alzheimer's disease: An {MRI} study of 676 ad, mci, and normal subjects. *NeuroImage*, **43**(3), 458 – 469.

Jahanshad, N., Kochunov, P. V., Sprooten, E., Mandl, R. C., Nichols, T. E., Almasy, L., Blangero, J., Brouwer, R. M., Curran, J. E., de Zubicaray, G. I., Duggirala, R., Fox, P. T., Hong, L. E., Landman, B. A., Martin, N. G., McMahon, K. L., Medland, S. E., Mitchell, B. D., Olvera, R. L., Peterson, C. P., Starr, J. M., Sussmann, J. E., Toga, A. W., Wardlaw, J. M., Wright, M. J., Pol, H. E. H., Bastin, M. E., McIntosh, A. M., Deary, I. J., Thompson, P. M., and Glahn, D. C. (2013). Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the enigma-dti working group. *NeuroImage*, **81**(0), 455 – 469.

Jennrich, R. and Sampson, P. (1976). Newton-raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, **18**(1), 11–17.

Kadri, N. K., Guldbrandtsen, B., Srensen, P., and Sahana, G. (2014). Comparison of genome-wide association methods in analyses of admixed populations with complex familial relationships. *PLoS ONE*, **9**(3), e88926.

Kang, H. M., Zaitlen, N. a., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. **178**(3), 1709–23.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. a., Kong, S.-Y., Freimer, N. B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, **42**(4), 348–54.

Kochunov, P., Glahn, D., Lancaster, J., Winkler, A., Smith, S., Thompson, P., Almasy, L., Duggirala, R., Fox, P., and Blangero, J. (2010). Genetics of microstructure of cerebral white matter using diffusion tensor imaging. *NeuroImage*, **53**(3), 1109 – 1116.

Kochunov, P., Glahn, D., Nichols, T., Winkler, A., Hong, E., Holcomb, H., Stein, J., Thompson, P., Curran, J., Carless, M., Olvera, R., Johnson, M., Cole, S., Kochunov, V., Kent, J., and Blangero, J. (2011a). Genetic analysis of cortical thickness and fractional anisotropy of water diffusion in the brain. *Frontiers in Neuroscience*, **5**(120).

Kochunov, P., Glahn, D., Lancaster, J., Thompson, P., Kochunov, V., Rogers, B., Fox, P., Blangero, J., and Williamson, D. (2011b). Fractional anisotropy of cerebral white matter and thickness of cortical gray matter across the lifespan. *NeuroImage*, **58**(1), 41 – 49.

Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sproote, E., Nichols, T., Hong, L., Behrens, T., Andersson, E. J.and Yacoub, Ugurbil, K., Brouwer, C., Landman, B., Braber, A., Almassy, L., Fox, P., Olvera, R., Blangero, J., DC., G., and Van Essen, D. (2014a). Heritability of fractional anisotropy in human white matter: A comparison of human connectome project and enigma-dti data. *NeuroImage*. In Review.

Kochunov, P., Jahanshad, N., Sprooten, E., Nichols, T. E., Mandl, R. C., Almasy, L., Booth, T., Brouwer, R. M., Curran, J. E., de Zubicaray, G. I., Dimitrova, R., Duggirala, R., Fox, P. T., Hong, L. E., Landman, B. A., Lemaitre, H., Lopez, L. M., Martin, N. G., McMahon, K. L., Mitchell, B. D., Olvera, R. L., Peterson, C. P., Starr, J. M., Sussmann, J. E., Toga, A. W., Wardlaw, J. M., Wright, M. J., Wright, S. N., Bastin, M. E., McIntosh, A. M., Boomsma, D. I., Kahn, R. S., den Braber, A., de Geus, E. J., Deary, I. J., Pol, H. E. H., Williamson, D. E., Blangero, J., van 't Ent, D., Thompson, P. M., and Glahn, D. C. (2014b). Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: Comparing meta and megaanalytical approaches for data pooling. *NeuroImage*, **95**(0), 136 – 150.

Koten, J. W., Wood, G., Hagoort, P., Goebel, R., Propping, P., Willmes, K., and Boomsma, D. I. (2009). Genetic contribution to variation in cognitive function: An fmri study in twins. *Science*, **323**(5922), 1737–1740.

Kremen, W. S., Prom-Wormley, E., Panizzon, M. S., Eyler, L. T., Fischl, B., Neale, M. C., Franz, C. E., Lyons, M. J., Pacheco, J., Perry, M. E., Stevens, A., Schmitt, J. E., Grant, M. D., Seidman, L. J., Thermenos, H. W., Tsuang, M. T., Eisen, S. a., Dale, A. M., and Fennema-Notestine, C. (2010). Genetic and environmental influences on the size of specific brain regions in midlife: the vetsa mri study. *NeuroImage*, **49**(2), 1213–23.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**(4), 963–974.

Lange, K. (2003). *Mathematical and statistical methods for genetic analysis*. Springer, 2nd edition.

Lange, K., Westlake, J., and Spence, M. A. (1976). Extensions to pedigree analysis. III. variance components by the scoring method. *Annals of Human Genetics*, **39**(4), 485–91.

Lee, O. E. and Braun, T. M. (2012). Permutation tests for random effects in linear mixed models. *Biometrics*, **68**(2), 486–493.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., and Heckerman, D. (2011a). FaST linear mixed models for genome-wide association studies. *Nature Methods*, **8**(10), 833–837.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., Heckerman, D., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., and Heckerman, D. (2011b). Improved linear mixed models for genome-wide association studies. *Nature methods*, **8**(6), 833–5.

Lippert, C., Quon, G., Kang, E. Y., Kadie, C. M., Listgarten, J., and Heckerman, D. (2013). The benefits of selecting phenotype-specific variants for applications of mixed models in genomics. *Scientific reports*, **3**, 1815.

Listgarten, J., Lippert, C., and Heckerman, D. (2013). FaST-LMM-Select for addressing confounding from spatial structure and rare variants. *Nature Genetics*, **45**(5), 470–471.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, **74**(4), 817–827.

MacCluer, J. W., Blangero, J., Dyer, T. D., and Speer, M. C. (1997). GAW10: simulated family data for a common oligogenic disease with quantitative risk factors. *Genetic epidemiology*, **14**(6), 737–42.

Matthews, S. C., Simmons, A. N., Strigo, I., Jang, K., Stein, M. B., and Paulus, M. P. (2007). Heritability of anterior cingulate response to conflict: An fmri study in female twins. *NeuroImage*, **38**(1), 223 – 227.

McKay, D., Knowles, E., Winkler, A., Sprooten, E., Kochunov, P., Olvera, R., Curran, J., Kent, J. JackW., Carless, M., Gring, H., Dyer, T., Duggirala, R., Almasy, L., Fox, P., Blangero, J., and Glahn, D. (2014). Influence of age, sex and genetic factors on the human brain. *Brain Imaging and Behavior*, **8**(2), 143–152.

Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, pages 153–165.

Molenberghs, G. and Verbeke, G. (2007). Likelihood Ratio, Score, and Wald Tests in a Constrained Parameter Space. *The American Statistician*, **61**(1), 22–27.

Morgan, B. J. T., Palmer, K. J., and Ridout, M. S. (2007). Negative Score Test Statistic. *The American Statistician*, **61**(4), 285–288.

Mumford, J. and Nichols, T. (2006). Modeling and inference of multisubject fmri data. *Engineering in Medicine and Biology Magazine, IEEE*, **25**(2), 42–51.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **231**, pp. 289–337.

Nichols, T. E. and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, **12**(5), 419–446.

Nichols, T. E. and Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, **15**(1), 1–25.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, **15**(1), 1–25.

Ober, C., Abney, M., and McPeek, M. S. (2001). The genetic dissection of complex traits in a founder population. *The American Journal of Human Genetics*, **69**(5), 1068–1079.

Olvera, R., Bearden, C., Velligan, D., Almasy, L., Carless, M., Curran, J., Williamson, D., Duggirala, R., Blangero, J., and Glahn, D. (2011). Common genetic influences on depression, alcohol, and substance use disorders in mexican-american families. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, **156**(5), 561–568.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545–554.

Pievani, M., Rasser, P., Galluzzi, S., Benussi, L., Ghidoni, R., Sabattoli, F., Bonetti, M., Binetti, G., Thompson, P., and Frisoni, G. (2009). Mapping the effect of {APOE} ?4 on gray matter loss in alzheimer's disease in vivo. *NeuroImage*, **45**(4), 1090 – 1098.

Pirinen, M., Donnelly, P., and Spencer, C. C. A. (2013). Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *Ann. Appl. Stat.*, **7**(1), 369–390.

Polk, T. A., Park, J., Smith, M. R., and Park, D. C. (2007). Nature versus nurture in ventral visual cortex: A functional magnetic resonance imaging study of twins. *The Journal of Neuroscience*, **27**(51), 13921–13925.

Potkin, S. G., Turner, J. A., Guffanti, G., Lakatos, A., Fallon, J. H., Nguyen, D. D., Mathalon, D., Ford, J., Lauriello, J., Macciardi, F., and FBIRN (2009a). A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophrenia Bulletin*, **35**(1), 96–108.

Potkin, S. G., Guffanti, G., Lakatos, A., Turner, J. A., Kruggel, F., Fallon, J. H., Saykin, A. J., Orro, A., Lupoli, S., Salvi, E., Weiner, M., Macciardi, F., and for the Alzheimer's Disease Neuroimaging Initiative (2009b). Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for alzheimer's disease. *PLoS ONE*, **4**(8), 1–15.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, **38**(8), 904–909.

Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews*, **11**(June), 459–463.

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, **67**(1), 170 – 181.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., *et al.* (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**(3), 559–575.

Rao, C. R. (2008). *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., and Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.

Rimol, L. M., Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Fischl, B., Franz, C. E., Hagler, D. J., Lyons, M. J., Neale, M. C., Pacheco, J., Perry, M. E., Schmitt, J. E., Grant, M. D., Seidman, L. J., Thermenos, H. W., Tsuang, M. T., Eisen, S. a., Kremen, W. S., and Dale, A. M. (2010). Cortical thickness is influenced by regionally specific genetic factors. *Biological psychiatry*, **67**(5), 493–9.

Sabatti, C., Service, S. K., Hartikainen, A.-L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C. G., Zaitlen, N. A., Varilo, T., Kaakinen, M., *et al.* (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature genetics*, **41**(1), 35–46.

Salimi-Khorshidi, G., Smith, S. M., and Nichols, T. E. (2010). Adjusting the effect of nonstationarity in cluster-based and TFCE inference. *NeuroImage*, **54**(3), 2006–2019.

Samuh, M. H., Grilli, L., Rampichini, C., Salmaso, L., and Lunardon, N. (2012). The use of permutation tests for variance components in linear mixed models. *Communications in Statistics - Theory and Methods*, **41**(16-17), 3020–3029.

Searle, S. R., Casella, G., and McCulloch, C. E. (2009). *Variance components*, volume 391. John Wiley & Sons.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**(398), pp. 605–610.

Servin, B. and Stephens, M. (2007). Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLoS Genet*, **3**(7).

Shephard, N. (1993). Maximum likelihood estimation of regression models with stochastic trend components. *Journal of the American Statistical Association*, **88**(422), pp. 590–595.

Shephard, N. G. and Harvey, A. C. (1990). On the probability of estimating a deterministic component in the local level model. *Journal of Time Series Analysis*, **11**(4), 339–347.

Silvapulle, M. J. (1992). Robust wald-type tests of one-sided hypotheses in the linear model. *Journal of the American Statistical Association*, **87**(417), 156–161.

Silvapulle, M. J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association*, **90**(429), 342–349.

Smith, S. and Nichols, T. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, **44**(1), 83–98.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., and Behrens, T. E. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, **31**(4), 1487 – 1505.

Stein, J. L., Hua, X., Morra, J. H., Lee, S., Hibar, D. P., Ho, A. J., Leow, A. D., Toga, A. W., Sul, J. H., Kang, H. M., Eskin, E., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M. J., Craig, D. W., Gerber, J. D., Allen, A. N., Corneveaux, J. J., Stephan, D. A., Webster, J., DeChairo, B. M., Potkin, S. G., Jr., C. R. J., Weiner, M. W., and Thompson, P. M. (2010a). Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in alzheimer's disease. *NeuroImage*, **51**(2), 542 – 554.

Stein, J. L., Hua, X., Lee, S., Ho, A. J., Leow, A. D., Toga, A. W., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M. J., Craig, D. W., Gerber, J. D., Allen, A. N., Corneveaux, J. J., DeChairo, B. M., Potkin, S. G., Weiner, M. W., and Thompson, P. M. (2010b). Voxelwise genome-wide association study (vgwas). *NeuroImage*, **53**(3), 1160 – 1174. Imaging Genetics.

Stein, J. L., Medland, S. E., Vasquez, A. A., Hibar, D. P., Senstad, R. E., and Winkler, e. a. Anderson M (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet*, **44**(5), 552–561.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, **50**(4), pp. 1171–1177.

Svishcheva, G. R., Axenovich, T. I., Belonogova, N. M., van Duijn, C. M., and Aulchenko, Y. S. (2012). Rapid variance componentsbased method for whole-genome association analysis. *Nature Genetics*, **44**(10), 1166–1170.

ter Braak, C. J. (1992). *Permutation Versus Bootstrap Significance Tests in Multiple Regression and Anova*, volume 376 of *Lecture Notes in Economics and Mathematical Systems*. Springer Berlin Heidelberg.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**(2), pp. 254–262.

Verbeke, G. and Molenberghs, G. (2007). What Can Go Wrong With the Score Test? *The American Statistician*, **61**(4), 289–290.

Voight, B. F. and Pritchard, J. K. (2005). Confounding from cryptic relatedness in case-control association studies. *PLoS Genet*, **1**(3).

Weir, B. S., Anderson, A. D., and Hepler, A. B. (2006). Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics*, **7**(10), 771–780.

Widmer, C., Lippert, C., Weissbrod, O., Fusi, N., Kadie, C., Davidson, R., Listgarten, J., and Heckerman, D. (2014). Further improvements to linear mixed models for genome-wide association studies. *Scientific reports*, **4**, 6874.

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., and Glahn, D. C. (2010). Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, **53**(3), 1135–1146.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, **92C**, 381–397.

Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of {FMRI} data. *NeuroImage*, **14**(6), 1370 – 1386.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**(6), 900–918.

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics*, **46**(2), 100–6.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., and Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, **38**(2), 203–8.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. a., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., and Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature genetics*, **42**(4), 355–360.

Zhou, X. and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature genetics*, **44**(7), 821–4.

Zhou, X. and Stephens, M. (2014). Efficient algorithms for multivariate linear mixed models in genome-wide association studies. *Nature methods*, **11**(4), 407.