The Annals of Statistics
2016, Vol. 44, No. 5, 2127–2160
DOI: 10.1214/15-AOS1421

ON HIGH-DIMENSIONAL MISSPECIFIED MIXED MODEL ANALYSIS IN GENOME-WIDE ASSOCIATION STUDY

BY JIMING JIANG[1,∗], CONG LI[2,†], DEBASHIS PAUL[3,∗],
CAN YANG[4,‡] AND HONGYU ZHAO[5,§]

*University of California, Davis**, *Takeda Pharmaceuticals International Co.*†,
Hong Kong Baptist University‡ *and Yale University*§

We study behavior of the restricted maximum likelihood (REML) estimator under a misspecified linear mixed model (LMM) that has received much attention in recent genome-wide association studies. The asymptotic analysis establishes consistency of the REML estimator of the variance of the errors in the LMM, and convergence in probability of the REML estimator of the variance of the random effects in the LMM to a certain limit, which is equal to the true variance of the random effects multiplied by the limiting proportion of the nonzero random effects present in the LMM. The asymptotic results also establish convergence rate (in probability) of the REML estimators as well as a result regarding convergence of the asymptotic conditional variance of the REML estimator. The asymptotic results are fully supported by the results of empirical studies, which include extensive simulation studies that compare the performance of the REML estimator (under the misspeci-

---

Tribute: On January 9, 2016, Professor Peter Gavin Hall, one of the greatest minds in Mathematical Statistics, former Co-Editor of the Annals of Statistics, left us at the age of 65. While we are still grieving, and trying to recover from this irreplaceable loss, we feel extremely lucky to have been born around the time that we have had the opportunity to learn from Professor Hall, not only through his many papers and books but also from interactions with him on various occasions, and even to know him personally. All of us recall the most remarkable characters of Professor Hall being his extreme kindness, and interests in listening to research ideas and problems, and sharing his insights and thoughts. As a former Co-Editor of the Annals, Professor Hall had given his heart to this journal. Some of us recall seeing him reading new AOS submissions while lying in bed in the intensive care unit of UC Davis Hospital. Professor Hall will be greatly missed, but he shall forever live in our hearts.

Received June 2014; revised November 2015.

[1] Supported in part by the NSF Grants DMS-08-09127, SES-1121794, and the NIH Grant R01-GM085205A1.

[2] Supported in part by the NIH Grant R01-GM59507.

[3] Supported in part by the NSF Grants DMS-11-06690 and DMS-1407530.

[4] Supported in part by Grant #61501389 from National Science Funding of China, Grant #22302815 from the Hong Kong Research Grant Council, and Grant FRG2/14-15/069 from Hong Kong Baptist University (the Hong Kong RGC grant HKBU).

[5] Supported in part by the NIH Grant R01-GM59507, the CTSA Grant UL1-RR024139, and the Department of Veterans Affairs (VA Cooperative Studies Program).

MSC2010 subject classifications. Primary 62J99; secondary 60B20.

Key words and phrases. Asymptotic property, heritability, misspecified LMM, MMMA, random matrix theory, REML, variance components.

fied LMM) with other existing methods, and real data applications (only one example is presented) that have important genetic implications.

1. Introduction. Genome-wide association study (GWAS), which typically refers to examination of associations between up to millions of genetic variants in the genome and certain traits of interest among unrelated individuals, has been very successful for detecting genetic variants that affect complex human traits/diseases in the past eight years. According to the web resource of GWAS catalog [Hindorff et al. (2009); http://www.genome.gov/gwastudies], as of October 2013, more than 11,000 single-nucleotide polymorphisms (SNPs) have been reported to be associated with at least one trait/disease at the genome-wide significance level (p-value $\leq 5 \times 10^{-8}$), many of which have been validated/replicated in further studies. However, these significantly associated SNPs only account for a small portion of the genetic factors underlying complex human traits/diseases [Manolio et al. (2009)]. For example, human height is a highly heritable trait with an estimated heritability of around 80%, that is, 80% of the height variation in the population can be attributed to genetic factors [Visscher, Hill and Wray (2008)]. Based on large-scale GWAS, about 180 genetic loci have been reported to be significantly associated with human height [Allen et al. (2010)]. However, these loci together can explain only about 5–10% of variation of human height [Allen et al. (2010), Manolio et al. (2009), Visscher (2008)]. This "gap" between the total genetic variation and the variation that can be explained by the identified genetic loci is universal among many complex human traits/diseases and is referred to as the "missing heritability" [Maher (2008), Manolio (2010), Manolio et al. (2009)].

One possible explanation for the missing heritability is that many SNPs jointly affect the phenotype, while the effect of each SNP is too weak to be detected at the genome-wide significance level. To address this issue, Yang et al. (2010) used a linear mixed model (LMM)-based approach to estimate the total amount of human height variance that can be explained by all common SNPs assayed in GWAS. They showed that 45% of the human height variance can be explained by those SNPs, providing compelling evidence for this explanation: A large proportion of the heritability is not "missing," but rather hidden among many weak-effect SNPs. These SNPs may require a much larger sample size to be detected. The LMM-based approach was also applied to analyze many other complex human traits/diseases [e.g., metabolic syndrome traits, Vattikuti, Guo and Chow (2012); and psychiatric disorders, Lee et al. (2012), Cross-Disorder Group of Psychiatric Genomics Consortium (2013)] and similar results have been observed.

Statistically, the heritability estimation based on the GWAS data can be casted as the problem of variance component estimation in high-dimensional regression, where the response vector is the phenotypic values and the design matrix is the standardized genotype matrix (to be detailed below). One needs to estimate the residual variance and the variance that can be attributed to all of the variables in the design matrix. In a typical GWAS data set, although there may be many weak-effect SNPs [e.g., $\sim 10^3$, Stahl et al. (2012)] that are associated with the phenotype, they are still only a small portion of the total number SNPs (e.g., $10^5 \sim 10^6$). In other words, using a statistical term, the true underlying model is sparse. However, the LMM-based approach used by Yang et al. assumes that the effects of all the SNPs are nonzero. It follows that the assumed LMM is misspecified. In spite of the huge impact of its results in the genetics community, the misspecified LMM-based approach has not yet been rigorously justified. In this paper, we provide theoretical justification of the misspecified LMM in high-dimensional variance component estimation by investigating the asymptotics of the restricted maximum likelihood [REML; e.g., Jiang (2007)] estimator as both the sample size and the dimension of the vector of random effects tend to infinity. The results of our theoretical study imply consistency of the REML estimators of some of the important genetic quantities, such as the heritability, in spite of the model misspecification. We also study convergence rate and asymptotic variance property of the REML estimator. The theoretical results are fully supported by the results of our empirical studies. Our study not only provides theoretical support for the recent discoveries in human genetics made by the LMM but also, for the first time, introduces the notion of misspecified mixed model analysis (MMMA) and its asymptotic properties.

In addition to the significant impact of variance estimation in the genetic community, the problem of estimating the residual variance in the high-dimensional setting has drawn much attention recently. First, the problem is interesting in its own right, as addressed in some recent papers [Fan, Guo and Hao (2012), Reid, Tibshirani and Friedman (2013)]. Second, the significance tests for the estimated coefficients in sparse regression [Javanmard and Montanari (2013), Lockhart et al. (2013)] require an estimator of the residual variance. Our results open another door for the variance estimation in high-dimensional regression.

1.1. *Misspecified LMM and REML estimation.* Consider a LMM that can be expressed as

$$y = X\beta + \tilde{Z}\alpha + \varepsilon, \tag{1}$$

where y is an $n \times 1$ vector of observations; X is a $n \times q$ matrix of known covariates; β is a $q \times 1$ vector of unknown regression coefficients (the fixed effects); $\tilde{Z} = p^{-1/2}Z$, where Z is an $n \times p$ matrix whose entries are random variables. Furthermore, α is a $p \times 1$ vector of random effects that is distributed as $N(0, \sigma_\alpha^2 I_p)$, I_p being the p-dimensional identity matrix, and ε is an $n \times 1$ vector of errors that is distributed as $N(0, \sigma_\varepsilon^2 I_n)$, and α, ε and Z are independent. See Section 6 for discussion regarding the normality assumption about the random effects and errors. The estimation of σ_ε^2 is among the main interests. Without loss of generality, assume that X is full rank.

The LMM (1) is what we call assumed model. In reality, however, only a subset of the random effects are nonzero. More specifically, we have $\alpha = \{\alpha'_{(1)}, 0'\}'$, where $\alpha_{(1)}$ is the vector of the first m components of α ($1 \le m \le p$), and 0 is the $(p-m) \times 1$ vector of zeros. Correspondingly, we have $\tilde{Z} = [\tilde{Z}_{(1)} \tilde{Z}_{(2)}]$, where $\tilde{Z}_{(j)} = p^{-1/2} Z_{(j)}$, $j = 1, 2$, $Z_{(1)}$ is $n \times m$, and $Z_{(2)}$ is $n \times (p-m)$. Therefore, the true LMM can be expressed as

$$y = X\beta + \tilde{Z}_{(1)}\alpha_{(1)} + \varepsilon. \tag{2}$$

With respect to the true model (2), the assumed model (1) is misspecified. We shall call the latter a misspecified LMM or mis-LMM. However, this may not be known to the investigator, who would proceed with the standard mixed model analysis [e.g., Jiang (2007), Chapter 1] to obtain estimates of the model parameters, based on (1). This is what we referred to as MMMA. In this paper, we will be focusing on REML method [e.g., Jiang (2007), Section 1.3.2]. Furthermore, following Jiang (1996), we consider estimation of σ_ε^2 and the ratio $\gamma = \sigma_\alpha^2/\sigma_\varepsilon^2$. According to Jiang [(2007), Section 1.3.2], the REML estimator of γ, denoted by $\hat{\gamma}$, is the solution to the equation

$$\frac{y' P_\gamma \tilde{Z}\tilde{Z}' P_\gamma y}{\operatorname{tr}(P_\gamma \tilde{Z}\tilde{Z}')} = \frac{y' P_\gamma^2 y}{\operatorname{tr}(P_\gamma)}, \tag{3}$$

where $P_\gamma = V_\gamma^{-1} - V_\gamma^{-1} X (X' V_\gamma^{-1} X)^{-1} X' V_\gamma^{-1}$ with $V_\gamma = I_n + \gamma \tilde{Z}\tilde{Z}'$. Equation (3) is combined with another REML equation, which can be expressed as

$$\sigma_\varepsilon^2 = \frac{y' P_\gamma^2 y}{\operatorname{tr}(P_\gamma)}, \tag{4}$$

to obtain the REML estimator of σ_ε^2, namely, $\hat{\sigma}_\varepsilon^2 = y' P_{\hat{\gamma}}^2 y / \operatorname{tr}(P_{\hat{\gamma}})$.

In the context of mixed effects models, asymptotic behavior of the REML estimators is well established [Cressie and Lahiri (1993), Das (1979), Jiang (1996), Richardson and Welsh (1994)]. Note that the standard LMM is a conditional model, on the X and Z; hence, in particular, the matrix Z is nonrandom. However, this difference is relatively trivial. A more important difference is, as noted, that the LMM (1) is misspecified. Nevertheless, what appears to be striking is that the estimator $\hat{\sigma}_\varepsilon^2$ is, still, consistent. On the other hand, the estimator $\hat{\gamma}$ converges in probability to a constant limit, although the limit may not be the true γ. In spite of the inconsistency of $\hat{\gamma}$, when it comes to estimating some important quantities of genetic interest, such as the heritability (see below), REML still provides the right answer. Before presenting any theoretical results, we first illustrate with a numerical example that also highlights the practical relevance of our theoretical study.

1.2. *A numerical illustration.*

In GWAS, SNPs are high-density bi-allelic genetic markers. Loosely speaking, each SNP can be considered as a binomial random variable with two trials and the probability of "success" is defined as "allele frequency" in genetics. Accordingly, the genotype for each SNP can be coded as either 0, 1 or 2. In our simulation, we first simulate the allele frequencies for p SNPs, $\{f_1, f_2, \dots, f_p\}$, from the Uniform$[0.05, 0.5]$ distribution, where f_j is the allele frequency of the jth SNP. We then simulate the genotype matrix $U \in \{0, 1, 2\}^{n\times p}$, with rows corresponding to the sample/individual and columns corresponding to the SNP. Specifically, for the jth SNP, the genotype value of each individual is sampled from $\{0, 1, 2\}$ according to probabilities $(1 - f_j)^2$, $2f_j(1 - f_j)$, and f_j^2, respectively. After that, each column of U is standardized to have zero mean and unit variance, and the standardized genotype matrix is denoted as Z. Let $\tilde{Z} = p^{-1/2}Z$. In Yang et al. (2010), an LMM was used to describe the relationship between a phenotypic vector y and the standardized genotype matrix $\tilde{Z}$:

$$(5)\qquad y = 1_n\mu + \tilde{Z}\alpha + \varepsilon, \qquad \alpha \sim N(0, \sigma_\alpha^2 I_p), \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n),$$

where 1_n is the $n \times 1$ vector of 1's, μ is an intercept, α is the vector of random effects, I_n is the $n \times n$ identity matrix, and ε is the vector of errors. An important quantity in genetics is "heritability," defined as the proportion of phenotypic variance explained by all genetic factors. For convenience, we assume that all of the genetic factors have been captured by the SNPs in GWAS. Under this assumption, the heritability can be characterized via the variance components in model (5):

$$(6)\qquad h^2 = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}.$$

Note that the definition of heritability by (6) assumes that $\alpha_j \sim N(0, \sigma_\alpha^2)$ for all $j \in \{1, 2, \dots, p\}$. However, in reality, only a subset of the SNPs is associated with the phenotype. A correct model therefore is

$$(7)\qquad y = 1_n\mu + \tilde{Z}_{(1)}\alpha_{(1)} + \varepsilon, \qquad \alpha_{(1)} \sim N(0, \sigma_\alpha^2 I_m), \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n),$$

where m is the total number of SNPs that are associated with the phenotype, $\alpha_{(1)}$ is the subvector of α corresponding to the nonzero components that are associated with the SNPs, and $\tilde{Z}_{(1)} = p^{-1/2}Z_{(1)}$, $Z_{(1)}$ being the submatrix of Z corresponding to the associated SNPs. In this case, the heritability should instead be given by

$$(8)\qquad h_{\text{true}}^2 = \frac{(m/p)\sigma_\alpha^2}{(m/p)\sigma_\alpha^2 + \sigma_\varepsilon^2}.$$

In practice, it is impossible to identify all of the m SNPs due to the limited sample size. Therefore, we follow model (7) while simulating the phenotypic values, but pretend that we do not know which SNPs are associated with the phenotype. This means that we simply use all the SNPs in Z to estimate the variance components, σ_α^2 and σ_ε^2 in model (5). The estimated heritability is then obtained as

$$(9)\qquad \hat{h}^2 = \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2}.$$

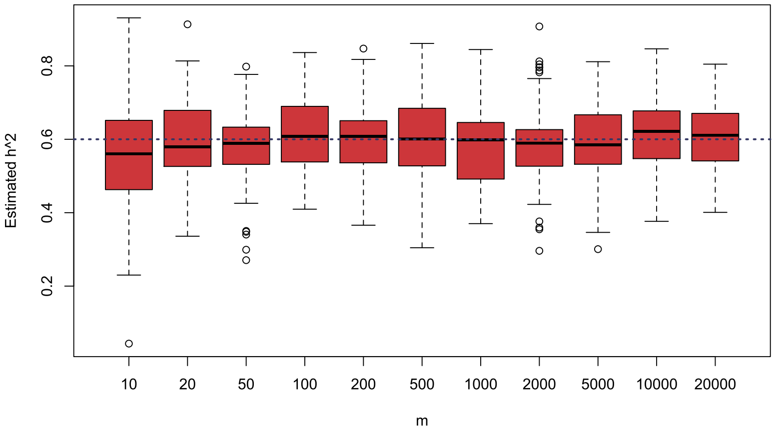

FIG. 1. *Heritability–REML provide right answer despite model misspecification.*

In this illustrative simulation, we fixed $n = 2000$, $p = 20{,}000$, $\sigma_\varepsilon^2 = 0.4$ and varied m from 10 to 20,000. We also set the variance component $\sigma_\alpha^2 = 0.6p/m$ so that the proportion of phenotypic variance explained by genetic factors $h_{\text{true}}^2 = 0.6$, based on (8). We repeated the simulation 100 times. As shown in Figure 1, there is almost no bias in the estimated h^2 regardless of the underlying true model, whether it is sparse (i.e., m/p is close to zero) or dense (i.e., m/p is close to one). This suggests that the REML works well in providing unbiased estimator of the heritability despite the model misspecification.

1.3. *Outline of theoretical results.* Throughout this paper, we assume that q, the dimension of β, is fixed, while n, p and m increase. For the simplicity of illustration, let us first assume that $n, p, m \to \infty$ such that

$$\frac{n}{p} \longrightarrow \tau, \qquad \frac{m}{p} \longrightarrow \omega, \tag{10}$$

where $0 < \tau, \omega \le 1$ are constants. Note that τ is the limiting ratio of the sample size and the number of random effects, while ω is the limiting proportion of the nonzero random effects. First consider the case where the entries of Z are i.i.d. The point is that the more realistic case where the entries of Z are standardized (see below) can be handled by utilizing the results for the i.i.d. case, and some inequalities on the difference, or perturbation (see below), between the two cases.

Suppose that the true variance components, $\sigma_\alpha^2, \sigma_\varepsilon^2$ are positive, and (10) holds. Then, (i) with probability tending to one, there is a REML estimator, $\hat{\gamma}$, such that $\hat{\gamma} \xrightarrow{\mathrm{P}} \omega\gamma_0$, where γ_0 is the true γ; (ii) $\hat{\sigma}_\varepsilon^2 \xrightarrow{\mathrm{P}} \sigma_{\varepsilon 0}^2$, where $\hat{\sigma}_\varepsilon^2$ is the REML estimator given by (4) with $\gamma = \hat{\gamma}$, as in (i), and $\sigma_{\varepsilon 0}^2$ is the true σ_ε^2.

As far as the consistency is concerned, condition (10) can be relaxed to

$$\liminf\left(\frac{m \wedge n}{p}\right) > 0, \qquad \limsup\left(\frac{m \vee n}{p}\right) \le 1, \tag{11}$$

so that, with probability tending to one, that there exist REML estimators, $\hat{\gamma}, \hat{\sigma}_{\varepsilon}^2$, such that (i) $\hat{\sigma}_{\varepsilon}^2 \xrightarrow{\mathrm{P}} \sigma_{\varepsilon 0}^2$, in other words, the REML estimator of σ_{ε}^2 is consistent; and (ii) the adjusted REML estimator of γ is consistent, that is, $(p/m)\hat{\gamma} \xrightarrow{\mathrm{P}} \gamma_0$.

Note. The latest asymptotic result may explain what has been observed in Figure 1. Note that the estimated heritability, (9), can be written as

$$\hat{h}^2 = \frac{(m/p)(p/m)\hat{\gamma}}{1 + (m/p)(p/m)\hat{\gamma}}. \tag{12}$$

On the other hand, the true heritability, (8), can be written as

$$h_{\text{true}}^2 = \frac{(m/p)\gamma_0}{1 + (m/p)\gamma_0}. \tag{13}$$

Because $(p/m)\hat{\gamma}$ converges in probability to γ_0, when we replace the $(p/m)\hat{\gamma}$ in (12) by γ_0, the resulting first-order approximation of (12) is exactly (13). It should also be noted that condition (11) requires that the limiting lower bound be positive. This may explain why the bias for $m = 10$ in Figure 1 is much more significant compared to other cases, because the ratio m/p in this case, $10/20{,}000 = 0.0005$, is fairly close to zero.

As mentioned, the asymptotic results can be extended to the case where the design matrix, Z, for the random effects is standardized. Let $U = (u_{ik})_{1 \le i \le n, 1 \le k \le p}$ whose entries are i.i.d. Define $Z = (z_{ik})_{1 \le i \le n, 1 \le k \le p}$, where $z_{ik} = (u_{ik} - \bar{u}_k)/s_k$ with $\bar{u}_k = n^{-1} \sum_{i=1}^n u_{ik}$ and $s_k^2 = (n-1) \sum_{i=1}^n (u_{ik} - \bar{u}_k)^2$. In other words, the new Z matrix has the sample mean equal to 0 and sample variance equal to 1 for each column. We then define $\tilde{Z} = p^{-1} Z$, and proceed as in (1). Also, as noted, in GWAS, the entries of U are generated from a discrete distribution which assigns the probabilities $\theta^2, 2\theta(1-\theta), (1-\theta)^2$ to the values $0, 1, 2$, where θ is prespecified so that $\theta \in (0.05, 0.5)$; however, there is also interest in the case where the entries of U are normal. Under the discrete distribution, it makes no difference if we standardize the discrete distribution so that is has mean 0 and variance 1, so, without loss of generality, the entries of U are $u_{ik} = (d_{ik} - \mu)/\sigma$, where d_{ij} has the above discrete distribution, $\mu = \mathbb{E}(d_{ik}) = 2(1-\theta)$, and $\sigma^2 = \operatorname{var}(u_{ik}) = 2\theta(1-\theta)$.

Both the Gaussian and discrete cases can be treated under the framework of the following broader class of distributions [e.g., Hsu, Kakade and Zhang (2012)]. Let $\xi_1, \dots, \xi_n$ be random variables. We say $\xi = (\xi_i)_{1 \le i \le n}$ is sub-Gaussian if there exists $\sigma > 0$ such that for all $\lambda \in R^n$ we have $\mathbb{E}(e^{\lambda' \xi}) \le e^{|\lambda|^2 \sigma^2 / 2}$. The asymptotic results regarding the MMMA are extended to the sub-Gaussian class.

In addition to the consistency results, we also study convergence rate and asymptotic variance property of the REML estimator under the mis-LMM. The results provide further insights into the asymptotic behavior of these estimators.

2. Preliminaries. A key component for our proofs is the following celebrated result in random matrix theory [e.g., Paul and Aue (2014)]. Let Z be an $n \times p$ matrix whose entries are i.i.d., complex-valued random variables with mean 0 and variance 1, where $n \to \infty$ as $p \to \infty$ such that $n/p \to \tau$, as in (10). We are interested in the asymptotic behavior of the empirical spectral distribution (ESD) of $S = p^{-1}ZZ'$, defined as

$$F^S(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{(\lambda_k \le x)}, \qquad x \in \mathbb{R},$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of S.

LEMMA 2.1 (Marčenko–Pastur law). *Suppose* (10) *holds. Then, as* $p \to \infty$, *the ESD of* S *converges almost surely* (*a.s.*) *in distribution to the Marčenko–Pastur* (*M–P*) *law,* F_τ, *whose p.d.f. is given by*

$$\varphi_\tau(x) = \frac{1}{2\pi\tau x}\sqrt{\{b_+(\tau) - x\}\{x - b_-(\tau)\}},$$

if $b_-(\tau) \le x \le b_+(\tau)$, *and* $\varphi_\tau(x) = 0$ *elsewhere, where* $b_\pm(\tau) = (1 \pm \sqrt{\tau})^2$.

A result that is frequently referred to is the following corollary of Lemma 2.1, which is a consequence of convergence in distribution [e.g., Jiang (2010), page 45].

COROLLARY 2.1. *Under the assumptions of Lemma* 2.1, *we have, for any positive integer* l, $n^{-1}\operatorname{tr}(S^l) \xrightarrow{\text{a.s.}} \int_{b_-(\tau)}^{b_+(\tau)} x^l \varphi_\tau(x)\,dx$ *as* $p \to \infty$.

The next result is regarding the extreme eigenvalues of S [e.g., Bai (1999), Theorem 2.16]. Let $\lambda_{\min}(S)$ [resp., $\lambda_{\max}(S)$] denote the smallest (largest) eigenvalues of S.

LEMMA 2.2. *Suppose that, in addition to the assumptions of Lemma* 2.1, *the fourth moment of the entries of* Z *are finite. Then we have, as* $p \to \infty$, $\lambda_{\min}(S) \xrightarrow{\text{a.s.}} b_-(\tau)$ *and* $\lambda_{\max}(S) \xrightarrow{\text{a.s.}} b_+(\tau)$.

Let $\xi_1, \dots, \xi_n$ be random variables. We say $\xi = (\xi_i)_{1 \le i \le n}$ is sub-Gaussian if there exists $\sigma > 0$ such that for all $\lambda \in \mathbb{R}^n$ we have $\mathbb{E}(e^{\lambda'\xi}) \le e^{|\lambda|^2\sigma^2/2}$. The Gaussian distribution, of course, is a member of the sub-Gaussian class. The following is a restatement of Lemma 5.5 of Vershynin (2012).

LEMMA 2.3. *A random variable* ξ *is sub-Gaussian if any of the following equivalent conditions hold*:

(I) $\mathbb{E}(e^{\xi^2/K_1^2}) < \infty$ *for some* $0 < K_1 < \infty$;

(II) $\{\mathbb{E}(|\xi|^q)\}^{1/q} \le K_2\sqrt{q}$ *for all* $q \ge 1$, *for some* $0 < K_2 < \infty$.
If, moreover, $\mathbb{E}(\xi) = 0$, *then the following is equivalent to* (I) *and* (II):
(III) $\mathbb{E}(e^{t\xi}) \le e^{t^2 K_3^2}$ *for all* $t \in R$, *for some* $0 < K_3 < \infty$.

Define the *sub-Gaussian norm* of a random variable ξ as

$$\|\xi\|_{\psi_2} \equiv \sup_{q \ge 1}\{q^{-1/2}(\mathbb{E}|\xi|^q)^{1/q}\}. \tag{14}$$

Clearly, by (II) of Lemma 2.3, ξ is a sub-Gaussian random variable if and only if $|\xi|_{\psi_2} < \infty$. One of the useful characteristics of sub-Gaussianity is that it is preserved under linear combinations. Specifically, we have the following result.

LEMMA 2.4 [Vershynin (2012), Lemma 5.9]. *Suppose that* $X_1, \dots, X_n$ *are independent sub-Gaussian random variables, and* $b_1, \dots, b_n \in \mathbb{R}$ *are nonrandom. Then* $\sum_{i=1}^n b_i X_i$ *is sub-Gaussian and, for some* $C > 0$, *we have*

$$\left\|\sum_{i=1}^n b_i X_i\right\|_{\psi_2}^2 \le C \sum_{i=1}^n b_i^2 \|X_i\|_{\psi_2}^2.$$

Lemma 2.4 follows easily from the equivalent characterizations in Lemma 2.3, specifically, by using the moment generating function. The following simple corollary is very useful for our applications.

COROLLARY 2.2. *Let* $X_1, \dots, X_n$ *be independent with* $\max_{1 \le i \le n} \|X_i\|_{\psi_2} \le K < \infty$. *Then* $\sum_{i=1}^n b_i X_i$ *is sub-Gaussian and, for some* $C > 0$, *we have*

$$\left\|\sum_{i=1}^n b_i X_i\right\|_{\psi_2}^2 \le C K^2 \left(\sum_{i=1}^n b_i^2\right).$$

The following result, due to Rudelson and Vershynin (2013), is a concentration inequality for quadratic forms involving a random vector with independent sub-Gaussian components. It is referred to as *Hanson–Wright inequality*. For any matrix Q of real entries, the spectral norm of Q is defined as $\|Q\| = \lambda_{\max}^{1/2}(Q'Q)$ and the Euclidean norm is defined as $\|Q\|_2 = \sqrt{\mathrm{tr}(Q'Q)}$.

PROPOSITION 2.1. *Let* $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)'$, *where the* ξ_i*'s are independent random variables satisfying* $\mathbb{E}(\xi_i) = 0$ *and* $\max_{1 \le i \le n} \|\xi_i\|_{\psi_2} \le K < \infty$. *Let* Q *be an* $n \times n$ *matrix. Then, for some constant* $c > 0$, *we have, for any* $t > 0$,

$$\mathbb{P}\{|\boldsymbol{\xi}'Q\boldsymbol{\xi} - \mathbb{E}(\boldsymbol{\xi}'Q\boldsymbol{\xi})| > t\} \le 2\exp\left\{-c\min\left(\frac{t^2}{K^4\|Q\|_2^2}, \frac{t}{K^2\|Q\|}\right)\right\}.$$

In the settings that we are interested in, we have $\mathbb{E}(\xi_i^2) = 1$ for all i and so $\mathbb{E}(\boldsymbol{\xi}' Q \boldsymbol{\xi})$ reduces to $\mathrm{tr}(Q)$.

The next result, well known in random matrix theory [e.g., Bai and Silverstein (2010); Sections A.5, A.6], is regarding perturbation of the ESD.

LEMMA 2.5. *For any $n \times p$ matrices A, B we have*:

(i) $\|F^{AA'} - F^{BB'}\| \le n^{-1}\,\mathrm{rank}(A - B)$, *where for a real-valued function g on $\mathbb{R}$*, $\|g\| = \sup_{x\in\mathbb{R}} |g(x)|$;

(ii) $L^4(F^{AA'}, F^{BB'}) \le 2n^{-2}(\|A\|_2^2 + \|B\|_2^2)\|A - B\|_2^2$, *where the Levy distance between two distributions*, F *and* G *on* $\mathbb{R}$, *is defined as* $L(F, G) = \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \le G(x) \le F(x + \varepsilon) + \varepsilon\}$.

The following result is implied by Lemma 2 of Bai and Yin (1993).

LEMMA 2.6. *Suppose that $X_{ij}, i, j = 1, 2, \ldots$ are i.i.d. with $\mathbb{E}(X_{11}^2) < \infty$. Then, we have* $\max_{1\le j\le n} |\bar{X}_j - \mathbb{E}(X_{11})| \xrightarrow{\text{a.s.}} 0$, *where* $\bar{X}_j = n^{-1}\sum_{i=1}^n X_{ij}$.

Lemmas 2.5 and 2.6 are used to study the asymptotic ESD of symmetric random matrices involving the standardized design matrix. Note that the standardized design matrix can be expressed as $Z = (U - \bar{u} \otimes 1_n)D_s^{-1}$, where $\bar{u} = (\bar{u}_1, \ldots, \bar{u}_p)$, and $D_s = \mathrm{diag}(s_1, \ldots, s_p)$ (where $\otimes$ denotes the Kronecker product). Let A be the matrix associated with the REML estimation (see the beginning of the proof of Theorem 3.1 below). Consider $\Psi = p^{-1}\zeta\zeta'$, where $\zeta = A'Z$ and A is $n \times (n - q)$ satisfying $A'X = 0$ and $A'A = I_{n-q}$. The following corollary is proved in Section 7.

COROLLARY 2.3. *Under the assumptions of Lemma 2.1, the ESD of Ψ converges a.s. in distribution to the M–P law. Furthermore, under the assumptions of Lemma 2.2, $\lambda_{\min}(\Psi)$ and $\lambda_{\max}(\Psi)$ converge a.s. $b_-(\tau)$ and $b_+(\tau)$, respectively.*

3. Main theoretical results.

First, we state a result regarding the consistency of the misspecified REML estimator of σ_ε^2, $\hat{\sigma}_\varepsilon^2$, and convergence in probability of the misspecified REML estimator of γ, $\hat{\gamma}$. Throughout this section, the design matrix, Z, is assumed to be standardized, as described near the end of Section 1, where the entries of U are i.i.d. sub-Gaussian.

THEOREM 3.1. *Suppose that the true $\sigma_\alpha^2, \sigma_\varepsilon^2$ are positive, and (10) holds. Then*:

(i) *With probability tending to one, there is a REML estimator, $\hat{\gamma}$, such that $\hat{\gamma} \xrightarrow{\mathrm{P}} \omega\gamma_0$, where γ_0 is the true γ.*

(ii) $\hat{\sigma}_\varepsilon^2 \xrightarrow{\mathrm{P}} \sigma_{\varepsilon 0}^2$, *where $\hat{\sigma}_\varepsilon^2$ is (4) with $\gamma = \hat{\gamma}$, as in* (i), *and $\sigma_{\varepsilon 0}^2$ is the true σ_ε^2.*

REMARK 3.1. It is interesting to note that the limit of $\hat{\gamma}$ in (i) depends on ω, but not τ. More specifically, the limit is equal to the true γ multiplied by ω, the limiting proportion of the nonzero random effects [see the remark below (10)]. The result seems totally intuitive.

REMARK 3.2. On the other hand, part (ii) of Theorem 3.1 states that the REML estimator of σ_ε^2 is consistent in spite of the model misspecification.

As far as the consistency of $\hat{\sigma}_\varepsilon^2$ is concerned, condition (10) can be relaxed. We state this as a corollary of Theorem 3.1.

COROLLARY 3.1. *Suppose that, in Theorem* 3.1, *condition* (10) *is weakened to* (11). *Then, with probability tending to one, there are REML estimators*, $\hat{\gamma}, \hat{\sigma}_\varepsilon^2$, *such that*:

(i) $\hat{\sigma}_\varepsilon^2 \xrightarrow{\mathrm{P}} \sigma_{\varepsilon 0}^2$, *in other words, the REML estimator of* σ_ε^2 *is consistent*;

(ii) *The adjusted REML estimator of* γ *is consistent, that is*, $(p/m)\hat{\gamma} \xrightarrow{\mathrm{P}} \gamma_0$.

Another consequence of Theorem 3.1 may be regarded as an extension of the well-known result on consistency of the REML estimator [e.g., Jiang (1996)], which is based on conditioning on Z.

COROLLARY 3.2. *Suppose that* $m = p$, *that is, the LMM is correctly specified. Then, as* $n, p \to \infty$ *such that* (11) *holds with* $m \wedge n$ *and* $m \vee n$ *replaced by* n, *there are REML estimators* $\hat{\gamma}$ *and* $\hat{\sigma}_\varepsilon^2$ *such that* $\hat{\gamma} \xrightarrow{\mathrm{P}} \gamma_0$ *and* $\hat{\sigma}_\varepsilon^2 \xrightarrow{\mathrm{P}} \sigma_{\varepsilon 0}^2$; *in other words, the REML estimators are consistent without conditioning on* Z.

Given the consistency of $\hat{\sigma}_\varepsilon$, more precise asymptotic behavior of the latter is of interest. As noted, the estimation of σ_ε^2 is also of main practical interest. The following result establishes convergence rate of the REML estimator of σ_ε^2 as well as that of the adjusted REML estimator of γ.

In order to describe this result, we need to introduce certain functions. Define

$$h_{k,l}(\gamma) = \int \frac{x^l}{(1+\gamma x)^k} \varphi_\tau(x)\, dx, \tag{15}$$

where φ_τ denotes the p.d.f. of the Marčenko–Pastur law with the parameter $\tau \in (0, 1]$. For ease of reference, we also define,

$$\begin{aligned} f_1(\gamma) &= h_{2,1}(\gamma), & f_2(\gamma) &= h_{2,0}(\gamma), \\ g_1(\gamma) &= h_{1,1}(\gamma), & g_2(\gamma) &= h_{1,0}(\gamma). \end{aligned} \tag{16}$$

THEOREM 3.2. *Suppose that in the assumptions of Theorem 3.1, condition (10) is strengthened to*

$$\sqrt{n}\left|\frac{n}{p}-\tau\right| \to 0, \qquad \sqrt{n}\left|\frac{m}{p}-\omega\right| \to 0, \tag{17}$$

and U has independent sub-Gaussian entries with zero mean, unit variance and bounded sub-Gaussian norm, $\alpha_i \overset{i.i.d.}{\sim} N(0,\sigma_{\alpha 0}^2)$, while $\varepsilon_i \overset{i.i.d.}{\sim} N(0,\sigma_{\varepsilon 0}^2)$. Then, with $\gamma_* := \omega\gamma_0$, we have

$$\sqrt{n}(\hat{\gamma}-\gamma_*) \Longrightarrow N\big(0, 2\Xi_1(\gamma_*,\tau,\omega)\big), \tag{18}$$

$$\sqrt{n}(\hat{\sigma}_\varepsilon^2-\sigma_{\varepsilon 0}^2) \Longrightarrow N\big(0, 2\sigma_{\varepsilon 0}^4\Xi_2(\gamma_*,\tau,\omega)\big), \tag{19}$$

where $\Longrightarrow$ denotes convergence in distribution, $\Xi_1(\gamma_*,\tau,\omega)$ equals

$$\gamma_*^2\left(\frac{f_2(\gamma_*)}{g_2(\gamma_*)}-\frac{f_2(\gamma_*)}{g_2(\gamma_*)}\right)^{-2} \times\left[\frac{H_{2,2;1,1}(\gamma_*,\tau,\omega)}{(h_{1,1}(\gamma_*))^2}-2\frac{H_{2,2;1,0}(\gamma_*,\tau,\omega)}{h_{1,1}(\gamma_*)h_{1,0}(\gamma_*)}+\frac{H_{2,2;0,0}(\gamma_*,\tau,\omega)}{(h_{1,0}(\gamma_*))^2}\right],$$

and $\Xi_2(\gamma_*,\tau,\omega)$ equals

$$\frac{H_{2,2;1,1}(\gamma_*,\tau,\omega)}{(h_{1,0}(\gamma_*))^2}+2\rho_*\left(\frac{H_{3,2;2,1}(\gamma_*,\tau,\omega)}{h_{1,0}(\gamma_*)h_{1,1}(\gamma_*)}-\frac{H_{2,2;1,1}(\gamma_*,\tau,\omega)}{(h_{1,0}(\gamma_*))^2}\right) +\rho_*^2\left(\frac{H_{3,3;2,2}(\gamma_*,\tau,\omega)}{(h_{1,1}(\gamma_*))^2}+\frac{H_{2,2;1,1}(\gamma_*,\tau,\omega)}{(h_{1,0}(\gamma_*))^2}-2\frac{H_{3,2;2,1}(\gamma_*,\tau,\omega)}{h_{1,0}(\gamma_*)h_{1,1}(\gamma_*)}\right),$$

with $f_j(\gamma)$ and $g_j(\gamma)$ as in (16), $H_{k,l;s,t}(\gamma,\tau,\omega)$ is given in Proposition 3.1, and

$$\rho_* = \frac{\gamma_*(h_{2,1}(\gamma_*)/h_{1,0}(\gamma_*)-2h_{3,2}(\gamma_*)/h_{1,0}(\gamma_*))}{(h_{2,0}(\gamma_*)/h_{1,0}(\gamma_*)-h_{2,1}(\gamma_*)/h_{1,1}(\gamma_*))}.$$

REMARK 3.3. We can weaken the conditions of Gaussianity of ε_i's and α_i's in Theorem 3.2 to that they have finite fourth moment, and that $\mathbb{E}(\varepsilon_i^4)=3\sigma_{\varepsilon 0}^4$ while $\mathbb{E}(\alpha_i^4)=3\sigma_{\alpha 0}^4$.

The following result gives the form of the functions $H_{k,l;s,t}(\gamma,\tau,\omega)$. The detailed descriptions are cumbersome, and hence parts of the constituent functions are defined in the supplementary material [Jiang et al. (2016)].

PROPOSITION 3.1. *Under the conditions of Theorem 3.2, for integers k,l,s,t such that $1 \le k,l \le 3$ and $0 \le s < k$, $0 \le t < l$, and for any $\gamma > 0$, we have*

$$\frac{1}{n}\operatorname{tr}\big(\Sigma_\gamma^{-k}\bar{U}^s\Sigma_{1,0}\Sigma_\gamma^{-l}\bar{U}^t\Sigma_{1,0}\big) \xrightarrow{\mathrm{P}} H_{k,l;s,t}(\gamma,\tau,\omega), \tag{20}$$

where $\xrightarrow{\mathrm{P}}$ *denotes convergence in probability, and*

$$\begin{aligned}H_{k,l;s,t}(\gamma,\tau,\omega) &= H^{(1)}_{k,l;s,t}(\gamma,\tau)+2(\omega\gamma_0)H^{(2)}_{k,l;s,t}(\gamma,\tau)\\ &\quad+(\omega\gamma_0)^2H^{(3)}_{k,l;s,t}(\gamma,\tau)+\frac{\tau}{\omega}(\omega\gamma_0)^2H^{(4)}_{k,l;s,t}(\gamma,\tau),\end{aligned}\tag{21}$$

where

$$H^{(1)}_{k,l;s,t}(\gamma,\tau)=h_{k+l,s+t}(\gamma),$$

$$H^{(2)}_{k,l;s,t}(\gamma,\tau)=\frac{(-1)^{s+t}}{\gamma^{s+t}}\sum_{j=0}^{s+t}(-1)^j\binom{s+t}{j}\eta_{k+l-j}(\gamma,\tau),$$

$$H^{(3)}_{k,l;s,t}(\gamma,\tau)=\frac{(-1)^{s+t}}{\gamma^{s+t}}\sum_{j_1=0}^{s}\sum_{j_2=0}^{t}(-1)^{j_1+j_2}\binom{s}{j_1}\binom{t}{j_2}\kappa_{k-j_1,l-j_2}(\gamma,\tau),$$

$$\begin{aligned}H^{(4)}_{k,l;s,t}(\gamma,\tau)&=\frac{(-1)^{s+t}}{\gamma^{s+t}}\sum_{j_1=0}^{s}\sum_{j_2=0}^{t}(-1)^{j_1+j_2}\binom{s}{j_1}\binom{t}{j_2}\eta_{k-j_1}(\gamma,\tau)\eta_{l-j_2}(\gamma,\tau)\\ &=\frac{(-1)^{s+t}}{\gamma^{s+t}}\left(\sum_{j_1=0}^{s}(-1)^{j_1}\binom{s}{j_1}\eta_{k-j_1}(\gamma,\tau)\right)\\ &\quad\times\left(\sum_{j_2=0}^{t}(-1)^{j_2}\binom{t}{j_2}\eta_{l-j_2}(\gamma,\tau)\right),\end{aligned}$$

where the functions $\kappa_{k,l}(\gamma,\tau)$ *and* $\eta_k(\gamma,\tau)$ *are as defined in* (S.3) *and* (S.4) *in the supplementary material* [*Jiang et al.* (2016)].

The proofs of the theorems are given in Section 7.

REMARK 3.4. Note that in the description of the function $H_{k,l;s,t}(\gamma^*,\tau,\omega)$, only the last term involves ω and this term is proportional to τ/ω. All of the other parameter, or function of parameters, that are involved in the asymptotic variances, namely, $\sigma^2_{\varepsilon 0}$ and $\gamma_*=\omega\gamma_0$, can be consistently estimated. At least for one particular case of important practical interest, the dependence on ω in the asymptotic variances is negligible; in other words, the term that depends on ω is of lower order than the leading term, which only involves $\sigma^2_{\varepsilon 0}$ and γ_*. This important case is when $n/p\to 0$, or $\tau=0$; in other words, the total number of SNPs, p, is much larger than the sample size, n. This is, indeed, a very practical situation in GWAS. In such a case, inference can be made by estimating the asymptotic variance. Another result of practical interest is that, if there is information about a lower bound for ω, then an upper bound of the asymptotic variance can be easily obtained, and the result holds regardless of the ratio n/p ($\to 0$ or otherwise). Such an upper bound is useful in obtaining, for example, conservative confidence intervals for the parameters of interest.

REMARK 3.5. Although, throughout this paper, we have assumed that the dimension of β, q, is fixed (see the beginning of Section 1.3), the proofs show that the results of Theorems 3.1 and 3.2 remain valid as long as $q = o(\sqrt{n})$. Another consequence of the latter condition is following. Throughout this paper, the matrix of covariates, X in (1), is considered fixed. This is equivalent to the assumption that X and Z, ε are independent. However, as long as $q = o(\sqrt{n})$, the independence of X and Z is asymptotically ignorable in that the results of Theorems 3.1 and 3.2 continue to hold even if X is not independent with Z. This is because the REML procedure depends on X only through the matrix A, which has the property that $A'X = 0$ and $A'A = I_{n-q}$. Furthermore, as argued near the end of the proof of Theorem 3.1 (see Section 7.3), what is actually at play is the matrix $AA' = I_n - P_X$, and P_X has rank $q = o(\sqrt{n})$. It turns out that, under the latter condition, P_X is ignorable in all of our asymptotic arguments; in other words, one can replace AA' by I_n and the results do not change.

4. More simulation studies.

To demonstrate our theoretical results numerically, we carry out more comprehensive simulation study following the same procedures as described in Section 1.2. The h^2 was also set at 0.6 ($\sigma_e^2 = 0.4$ and $\gamma = 1.5$). We fix the ratio $\tau = n/p = 0.1$ and varied $\omega = m/p$ from 0.001 to 1. We examine the performance of the REML, under the mis-LMM, in estimating γ and σ_e^2 as n varies from 1000 to 5000. The performance of the adjusted REML estimator of γ for $\omega = 0.01$ is shown in Figure 2. It appears that the adjusted

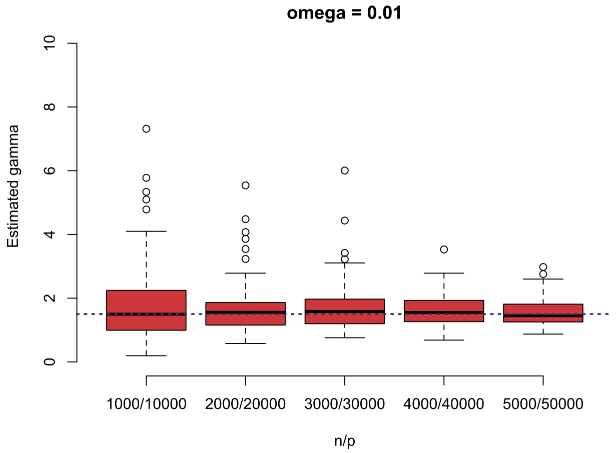

FIG. 2. *Adjusted REML estimation of γ for different n and p ($\omega = 0.01$).*

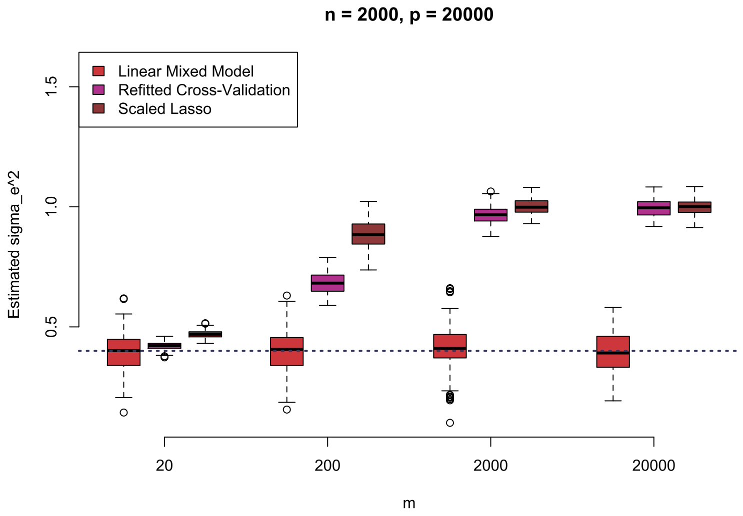

FIG. 3. *Comparison of estimators of* σ_ε^2 *with refitted C.V. and scaled lasso for different* m ($\omega = 0.01$).

REML always gives nearly unbiased estimate of γ, confirming our observations in Section 1.2 and theoretical results, namely, part (ii) of Corollary 3.1. More importantly, as both n and p increase (with n/p fixed at 0.1), the standard deviation of the estimate decreases.

As noted, several other methods for high dimensional variance estimation have been proposed recently. As a comparison, we examine the performances of two of these methods, refitted cross validation (c.v.) [Fan, Guo and Hao (2012)] and scaled lasso [Sun and Zhang (2012)], in estimating σ_e^2 under the misspecified LMM. The results for $n = 2000$, $p = 20{,}000$ are shown in Figure 3. Again, the REML estimator appears to be unbiased regardless of the value of m. On the other hand, the competing methods tend to have much larger bias, especially when m is large. This is not surprising because the competing methods are largely based on the sparsity assumption that m is relatively small compared to p. Indeed, when $m = 20$, the biases and standard deviations of the competing methods are quite small. In the latter case, the competing method may outperform the REML in terms of mean squared error (MSE). However, the REML performs well consistently across a much broader range of m, as demonstrated by Figure 3.

5. Real data example. LMM is nowadays commonly used in the genetics community for heritability estimation of complex traits [Visscher et al. (2012)], including anthropometric traits [Yang et al. (2010)], metabolic syndrome traits [Vattikuti, Guo and Chow (2012)], and psychiatric disorders [Cross-Disorder Group of the Psychiatric Genomics Consortium (2013), Lee et al. (2012), Yang et al. (2016)]. Here, we provide a real data example by using LMM to estimate heritability of body mass index (BMI). We downloaded COGA and SAGE datasets from dbGaP [accession number: phs000125.v1.p1 (COGA) and phs000092.v1.p1 (SAGE)].

First, we remove the duplicated samples in COGA and SAGE. Second, we remove samples without height and weight information because BMI is of our interest here. Third, we exclude relatives because these samples can inflate the heritability estimation [Yang et al. (2010)]. As a result, a total of $n = 2294$ individuals from European ancestry remain after these steps. To avoid artifacts from genotyping in our estimation, we apply stringent quality control for the genotype data from these individuals. Specifically, we remove SNPs with a missing rate > 0.01. We test for Hardy–Weinberg equilibrium and exclude SNPs with p-value < 0.001. SNPs with minor allele frequency (MAF) $< 5\%$ are also removed to focus on the analysis of common variants. After these quality control steps, $p = 728{,}000$ SNPs remain for analysis.

After sample and genotype cleaning, we apply the LMM approach to estimate the heritability of BMI. Specifically, we normalize the genotype matrix such that it has zero (sample) mean and unit (sample) variance, for each column, denoted as Z. We then use $\tilde{Z} = p^{-1/2}Z$ as the design matrix for the random effects. As for the matrix X for the fixed effects, we include, in addition to the intercept, the first ten principal component scores computed from $\tilde{Z}\tilde{Z}'$, known as the genetic similarity matrix, to account for the influence of population stratification. It should be noted that, strictly speaking, the X matrix here is not independent with Z. However, by Remark 3.5, this dependence does not affect our asymptotic results, as long as $q = o(\sqrt{n})$. For the current data, $\sqrt{n}$ is about 48 and q is 11, so the condition may be considered satisfied. Another note, from a practical point of view, is regarding the normalization of the genotype matrix so that each SNP had zero mean and unit variance. This is according to the common practice of LMM application to GWAS [Yang et al. (2010); Yang et al. (2011)]. Although heritability is not originally defined on the normalized genotypes, heritability estimation based on normalized genotype data explicitly assumes that the genetic variants with lower allele frequencies tend to have larger effect sizes. Speed et al. (2012) carefully examined heritability estimation under this assumption and their results suggested that this assumption could give the most stable heritability estimation in the presence of a misspecified distribution of effect sizes. As part of the results, we obtain the REML estimates as $\hat{\sigma}_\alpha^2 = 6.119$ with a standard error (s.e.) of 4.292, and $\hat{\sigma}_\varepsilon^2 = 25.149$ with a s.e. of 4.287, which results in the estimated heritability of $\hat{h}^2 = 19.6\%$ [see (9)] with a s.e. of 13.6%.

As a comparison, we use the refitted c.v. method to estimate the residual variance (see Section 4). Specifically, we randomly partition the data into two-groups (with equal sample sizes). We use the first half of the data to select the top $K = \lceil n/\log(n) \rceil = 296$ SNPs, and then estimate the residual variance associated with the second half of the data based on those selected SNPs. We repeat the random partitioning 50 times and the estimated residual variance almost equals to the sample phenotype variance. The result given by the refitted cross-validation method may suggest that genetics has little contribution to the phenotype, which could further lead to the phenomenon of "missing heritability." However, the results given by LMM suggest that genetic factors can explain a substantial proportion of phenotypic variance. More importantly, the heritability of BMI estimated by LMM (about 16.5%–22.9%) has been replicated based on several independent datasets [Yang et al. (2011), Zaitlen et al. (2013)].

6. Discussion. The LMM that is used for GWAS in this paper is relatively simple. Of course, there are aspects of the model that can be improved from practical points of view. For example, the i.i.d. assumption about the random effects, α, may not be met in cases where the SNPs are dense, in which case it may not be unrealistic to expect that the markers are in linkage equilibrium. If there is only one marker in each region having impacts on the phenotype of interest, one may be able to "thin" the SNPs to a reduced set so that there is less dependency among the markers, which does not practically affect the estimate of the overall genetic contribution to trait variations. In this case, the theoretical framework established in this paper would still stand valid, and the estimates would be reasonable.

A real challenging situation is when there are multiple markers in a region having impacts on the phenotypes and these markers are in linkage disequilibrium. From a regression perspective with the objective of estimating regression coefficients, the presence of colinearity of these markers will make the effect size estimates questionable. From a variance component point of view, the presence of linkage disequilibrium among these "true" markers as well as makers in the region with "null" effects will make it very challenging to estimate the variance contributions from the region. So, even if the random effects are i.i.d., the colinearity in the design matrix with a mix of "true" and "null" variables may raise issues to statistical inference, even in standard regression analysis. Some methods have been proposed to address potential issues due to linkage disequilibrium [e.g., Speed et al. (2012)], and these methods and their refinement may lead to more unbiased estimates. Therefore, regardless of the i.i.d. assumption, linkage disequilibrium will pose potential challenges in estimating the variance explained by genetic factors, and the "optimal" solution will depend on an appropriate model that can truly capture how different markers contribute to the phenotypes. While this should become more clear in the future, we do not yet have the knowledge to do so.

Similarly, homoscedasticity implied by the i.i.d. assumption may be too strong in some cases. Again, the real challenge is that the "true" underlying genetic model that affects our traits is largely unknown. In assessing associations between traits of multiple rare variants from the same region, many methods have been proposed to weigh in the allele frequencies differently [e.g., Madsen and Browning (2009)]. However, the relative performance of different weighting schemes still needs to be comprehensively evaluated by empirical data; it is likely that there will not be an universal rule on this issue. We take the approach analyzed in this paper as a first step to address this very complex problem. In this regard, George Box is often quoted as once famously saying that "essentially, all models are wrong, but some are useful" [Box (1979)]. It is true that the model we have in hand is not perfect; yet, it has been found useful in many GWAS studies, as noted in Section 1. A main purpose of this paper is to give answer to why the seemingly simple LMM works, in spite of some misspecification. We are confident that, as more knowledge becomes available, a more appropriate, and possibly complex model will emerge, and prove to be practically more efficient than the current model. When the time is right, a related asymptotic theory will be established.

In some cases, the number of nonzero random effects, m, may be of interest, so a related question is whether one can "consistently" estimate m. As explained in the Introduction, it is reasonable to assume that many SNPs jointly affect the phenotype. In other words, the number m is large. More precisely, we assume that the number of SNPs that have effects is a fraction of p, the total number of SNPs. Therefore, it would make more sense to consider estimation of the ratio $\omega = m/p$. However, as it turns out, this estimation problem is ill-posed under the true LMM considered in this paper. There have been studies on estimating the proportion of nonzero means in high-dimensional problems. See, for example, Cai, Jin and Low (2007), Jin and Cai (2007), Jin (2008). However, none of these existing results applies to our case. Furthermore, our own investigation shows that the ratio ω is, in fact, asymptotically nonidentifiable, under the true LMM.

To explain why this is necessarily the case, first note that the result of our paper showed that, at least to the first order approximation, the behavior of the REML estimators, depends on two quadratic forms of y, which are the differences between the two sides of equations (3) and (4), and their partial derivatives (with respect to the parameters) evaluated at the true parameters. Note that the matrices involved in the quadratic forms depending on Z, which asymptotically follows the random matrix theory. In other words, the asymptotic behavior of REML depends on the local behavior of these two quadratic forms. In our paper, we showed that, at least to the first-order approximation (which is all that matters so far as consistency is concerned), the local behaviors of those two quadratic forms depend only on σ_ε^2 and $\omega\sigma_\alpha^2$. Note that the second quantity depends on ω and σ_α^2 through their, product, but not separately. In fact, this is how we proved (see the next section), and why we can prove, consistency of REML estimators for exactly these two functions of the parameters and not something else. This means that, in particular, asymptotically, one cannot identify the parameters ω and σ_α^2 separately using the REML method (under the true LMM).

One may argue that REML is just one method of estimation; perhaps, somewhere, there is different method that can consistently estimate ω? The fact is, although there are alternative methods, the REML is widely considered the best method of variance components estimation in LMM [e.g., Jiang (1996, 1997, 2007), Sections 1.3.2, 1.8.3 and the reference therein]. Typically, consistency property of one good estimator reveals (asymptotic) identifiability of functions of parameters that is universal, regardless of the method of estimation.

To further support our argument, let us consider an alternative method of estimation. One of the earlier methods of estimation in LMM is the method of moments (or ANOVA estimation; for example, Jiang [(2007), Section 1.5.1]). Jiang (2003) proposed an extension called empirical method of moments (EMM) to estimate higher moments of the random effects in LMM. For the most part, EMM is the same as the method of moments (MM) except that some of the parameters in the MM equation are replaced by consistent estimators obtained using other method(s). Consider a special case where the entries of Z are independent $N(0, 1)$. We then let $\tilde{Z} = p^{-1/2}Z$, etc., as in Section 1.1. Let A be a symmetric $n \times n$ matrix satisfying $AX = 0$. It can be shown that

$$\mathbb{E}\left\{\frac{y'Ay}{\mathrm{tr}(A)}\right\} = \sigma_\varepsilon^2 + \omega\sigma_\alpha^2, \tag{22}$$

where $\omega = m/p$ is the proportion that we wish to estimate. Note that, in (22), the expectation is unconditional, that is, with respect to α, ε and Z. Next, let a be an $n \times 1$ vector such that $a'X = 0$. By Lemma 2.1 of Jiang (2007), it can be shown that

$$\mathbb{E}\left(\frac{a'y}{|a|}\right)^4 = 3\left\{\sigma_\varepsilon^4 + 2\omega\sigma_\alpha^2\sigma_\varepsilon^2 + \omega\sigma_\alpha^4\left(\omega + \frac{2}{p}\right)\right\}. \tag{23}$$

Again, the expectation in (23) is unconditional. We now specify the choices of A and a in (22), (23). Let $A = P_{X^\perp} = I_n - P_X$, where $P_X = X(X'X)^{-1}X'$. Note that $\mathrm{tr}(P_{X^\perp}) = \mathrm{rank}(P_{X^\perp}) = n - q$, where $q = \mathrm{rank}(X)$. Let $a_1, \dots, a_{n-q}$ be an orthonormal base of $\mathcal{L}(X)^\perp$, the linear space that is orthogonal to $\mathcal{L}(X)$, the linear space spanned by the columns of X. From (22), we have

$$\mathbb{E}\left(\frac{y'P_{X^\perp}y}{n-q}\right) = \sigma_\varepsilon^2 + \omega\sigma_\alpha^2; \tag{24}$$

and, from (23), we have

$$\mathbb{E}(a_k'y)^4 = 3\left\{\sigma_\varepsilon^4 + 2\omega\sigma_\alpha^2\sigma_\varepsilon^2 + \omega\sigma_\alpha^4\left(\omega + \frac{2}{p}\right)\right\}, \qquad 1 \le k \le n - q. \tag{25}$$

It follows, by taking average over $k = 1, \dots, n - q$ in (25), that

$$\mathbb{E}\left\{\frac{1}{n-q}\sum_{k=1}^{n-q}(a_k'y)^4\right\} = 3\left\{\sigma_\varepsilon^4 + 2\omega\sigma_\alpha^2\sigma_\varepsilon^2 + \omega\sigma_\alpha^4\left(\omega + \frac{2}{p}\right)\right\}. \tag{26}$$

If σ_ε^2 were known, one would use the MM by removing the expectation signs on the left sides of (24) and (26), leading to the MM equations. Namely, let $M_2 = y'P_{X^\perp}y/(n-q)$ and $M_4 = (n-q)^{-1}\sum_{k=1}^{n-q}(a_k'y)^4$. Then the MM equations are

$$\begin{cases} M_2 = \sigma_\varepsilon^2 + \omega\sigma_\alpha^2; \\ M_4 = 3[\sigma_\varepsilon^4 + 2\omega\sigma_\alpha^2\sigma_\varepsilon^2 + \omega\sigma_\alpha^4\{\omega + (2/p)\}]. \end{cases} \tag{27}$$

The equation system leads to the solution

$$\omega = \frac{2(M_2 - \sigma_\varepsilon^2)^2}{p\{(M_4/3) - M_2^2\}}. \tag{28}$$

Again, if σ_ε^2 were known, (28) would be the MM estimator of ω. Now, because σ_ε^2 is not known, we replace it in (28) by its REML estimator, $\hat{\sigma}_\varepsilon^2$, which is proved to be consistent (Theorem 3.1). The result is the EMM estimator of ω, given by

$$\hat{\omega} = \frac{2(M_2 - \hat{\sigma}_\varepsilon^2)^2}{p\{(M_4/3) - M_2^2\}}. \tag{29}$$

At first, the estimator $\hat{\omega}$ seems to have a neat closed-form expression, but it is not well-behaved (we have empirical results to support this claim). Why? The reason is simple: The moment equations, (24) and (26), asymptotically depend only on σ_ε^2 and $\omega\sigma_\alpha^2$. This is obviously true for (24). As for (26), note that the right-hand side can be written as

$$3\{\sigma_\varepsilon^4 + 2(\omega\sigma_\alpha^2)\sigma_\varepsilon^2 + (\omega\sigma_\alpha^2)^2\} + \frac{6}{p}(\omega\sigma_\alpha^2)\omega^{-1}. \tag{30}$$

In a way, the expression does have a little bit dependency on ω through ω^{-1}, in addition to σ_ε^2 and $\omega\sigma_\alpha^2$, but the dependency on ω^{-1} is to the order of $O(p^{-1})$; in contrast, the rest of the terms in (30) are all $O(1)$. Thus, as $p \to \infty$, the dependency on ω will disappear, and these equations depend only on σ_ε^2 and $\omega\sigma_\alpha^2$. In fact, this is, again, the very reason why we can consistently estimate these two functions of parameters, but not something else.

To play the devil's advocate, let us take another look at (29), note that the numerator converges in probability to $2(\omega\sigma_\alpha^2)^2$. Thus, one would love to see that the denominator converges, in probability, to $2\omega\sigma_\alpha^4$, because that is the mean of the denominator, but the convergence does not take place. In fact, it can be shown that

$$\frac{M_4}{3} - M_2^2 = \left(\frac{2}{p}\right)\omega\sigma_\alpha^4 + O(p^{-1/2}).$$

So, when multiplied by p, the denominator of (29) is equal to

$$2\omega\sigma_\alpha^4 + O(p^{1/2}). \tag{31}$$

Because the second term in (31) is of higher order than the first term, the convergence in probability is not going to happen.

The definition of LMM in Section 1 assumes that the random effects, α, and errors, ε, are normal. Such an assumption may not hold in some applications. Note that the normality assumption for the random effects and errors is needed for the REML estimators to be the "real" REML estimators because, without the normality assumption, one does not have an analytic expression for the likelihood function. All of the standard packages for LMM are built on this analytic expression. Thus, without the normality assumption, all of the standard packages are not computing the real REML estimators; instead, they compute something else. Then some related question are: (i) What is this something else? and (ii) whether it is something useful. Of course, one could still try to work on the non-Gaussian likelihood, even if it does not have an analytic form, and likely to be messy, but such an approach is rarely of interest in practice. Thus, we may focus on questions (i) and (ii).

The answer to (i) is that, without the normality, the REML estimators obtained under the normality are called Gaussian REML estimators. They are a special case of quasi-likelihood estimators [e.g., Heyde (1997)]. As for (ii), in fact, Gaussian REML estimators are used almost all the time, in practice, even although one can rarely be sure about the normality of the random effects and errors. Such a practice is justified, in particular, by a series of papers published in the 1990s [Jiang (1996, 1997)], in which the author showed that, even without the normality assumption, the Gaussian REML estimators of the variance components are still consistent and asymptotically normal, provided that the random effects and errors are i.i.d., with zero means and finite fourth moments. This is, of course, part of the standard LMM theory, which assumes that the Z matrix is nonrandom.

In a way, our case goes beyond the standard LMM theory by assuming that Z is not only random but also high-dimensional. Thus, we have to check our proofs to make sure that our main results indeed go through without the normality assumption, with the understanding that we are considering the Gaussian REML estimators in such a case. After checking the proofs, we are assured of such a conclusion. Note that the conditional distribution of y given Z, which is where the normality of α and ε plays a role, is only used in a relatively minor early part of the proof for both Theorems 3.1 and 3.2, that is, to show that the difference between a quadratic form, say, $y'By$, and its conditional expectation given Z goes to zero in probability. The majority part of each proof then deals with $\mathbb{E}(y'By|Z)$ using the random matrix theory. To show that $y'By - \mathbb{E}(y'By|Z) \xrightarrow{\mathrm{P}} 0$, by Chebyshev's inequality, one needs to deal with $\mathrm{var}(y'By)$ when y is non-Gaussian. We then refer to the following result which is derived from Lemma 2.1 of Jiang (1997).

LEMMA 6.1. *Let* $\xi_i, 1 \le i \le n$ *be i.i.d. with* $\mathbb{E}(\xi_i) = 0$, $\mathbb{E}(\xi_i^2) = 1$, *and* $\mathbb{E}(\xi_i^4) < \infty$. *Then, with* $\xi = (\xi_i)_{1 \le i \le n}$ *and A being any constant symmetric matrix, we have*

$$\mathrm{var}(\xi' A \xi) \le \Big[2 \vee \max_{1 \le i \le n} \{\mathrm{var}(\xi_i^2)\}\Big] \mathrm{tr}(A^2). \tag{32}$$

Note that, if the ξ_i's are normal, the left-hand side of (32) is equal to $2\,\mathrm{tr}(A^2)$. Furthermore, any case of $\mathrm{var}(y'By|Z)$ can be reduced to that Lemma 6.1 is dealing with, and only upper bounds are needed. The implication is that, to come up with the same results needed for the proof without the normality assumption, all one needs is the finiteness of the fourth moments of the random effects and errors.

Regarding the numerical stability of the solution to equation (3), Jiang (2000) proved that, under situations much more general than the LMM considered in this paper, the solution to the REML equation is guaranteed to have kth moment for any positive integer k. Thus, from a theoretical point of view, the solution to equation (3) is guaranteed stable. Furthermore, according to our empirical experience based on a huge number of simulation studies and real-data analyses, the numerical procedure to solve the REML equation almost always converges.

7. Proofs.

7.1. *Proof of Corollary* 2.3. Note that $\zeta = (M - L)D_s^{-1}$, where $M = A'U$ and $L = A'\bar{u} \otimes 1_n$, and that M is $(n-q) \times p$ whose entries are independent sub-Gaussian, with mean 0, variance 1, and $A'A = I_{n-q}$. Furthermore, write $\tilde{M} = M/\sqrt{p}$ and $\tilde{L} = L/\sqrt{p}$. By Lemma 2.1, the ESD of $\tilde{M}\tilde{M}'$ converges a.s. in distribution to the M–P law. On the other hand, write $\tilde{B} = \tilde{M} - \tilde{L}$ and note that $\operatorname{rank}(\tilde{L}) \le \operatorname{rank}(\bar{u} \otimes 1_n) = 1$. Thus, by (i) of Lemma 2.5, we have $\|F^{\tilde{B}\tilde{B}'} - F^{\tilde{M}\tilde{M}'}\| \le (n-q)^{-1}$; hence, the ESD of $\tilde{B}\tilde{B}'$ converges a.s. in distribution to the M–P law, and $\lambda_{\min}(\tilde{B}\tilde{B}')$ and $\lambda_{\max}(\tilde{B}\tilde{B}')$ converge a.s. to $b_-(\tau)$ and $b_+(\tau)$, respectively.

Next, write $\tilde{A} = (\tilde{M} - \tilde{L})D_s^{-1}$. By (ii) of Lemma 2.5, we have $L^4(F^{\tilde{A}\tilde{A}'}, F^{\tilde{B}\tilde{B}'}) \le 2(n-q)^{-2}(\|\tilde{A}\|_2^2 + \|\tilde{B}\|_2^2)\|\tilde{A} - \tilde{B}\|_2^2$. Note that $\|\tilde{B}\|_2^2 = \operatorname{tr}(\tilde{B}\tilde{B}') = \operatorname{tr}(\tilde{M}\tilde{M}') - 2\operatorname{tr}(\tilde{L}\tilde{M}') + \operatorname{tr}(\tilde{L}\tilde{L}')$. By Lemma 2.1, we have $\operatorname{tr}(\tilde{M}\tilde{M}') = \operatorname{tr}(p^{-1}MM') = (n-q)O_{\text{a.s.}}(1)$, where $O_{\text{a.s.}}(1)$ denotes a term that is bounded almost surely. We have

$$\begin{aligned}
|\operatorname{tr}(\tilde{L}\tilde{M}')| &= \frac{1}{n}\left|\operatorname{tr}\{(1_n' p^{-1}UU' \otimes 1_n)(AA')\}\right| \\
&= \frac{1}{n}\left|\operatorname{tr}(1_n' p^{-1}UU'AA' \otimes 1_n)\right| \\
&= \frac{1}{n}\left|1_n' p^{-1}UU'AA'1_n\right| \\
&\le \frac{1}{n}\sqrt{1_n'(p^{-1}UU')^2 1_n}\sqrt{1_n'(AA')^2 1_n} \\
&\le \lambda_{\max}(p^{-1}UU')\lambda_{\max}(AA') \\
&= \lambda_{\max}(p^{-1}UU'),
\end{aligned}$$

which is $O_{\text{a.s.}}(1)$ by Lemma 2.2, and $\text{tr}(\tilde{L}\tilde{L}') \le \lambda_{\max}(p^{-1}UU') = O_{\text{a.s.}}(1)$. It follows that $\|\tilde{B}\|_2^2 = (n-q)O_{\text{a.s.}}(1)$. Also, we have $\|\tilde{A}\|_2^2 = \text{tr}(\tilde{B}D_s^{-2}\tilde{B}') \le \lambda_{\max}(D_s^{-2})\|\tilde{B}\|_2^2 = \|\tilde{B}\|_2^2/\min_{1\le j\le p} s_j^2$. By Lemma 2.6, we have $\max_{1\le j\le p}|s_j^2 - 1| \xrightarrow{\text{a.s.}} 0$; hence, we have $(\min_{1\le j\le p} s_j^2)^{-1} = O_{\text{a.s.}}(1)$. It follows that $\|\tilde{A}\|_2^2 = (n-q)O_{\text{a.s.}}(1)$. Finally, we have $\|\tilde{A} - \tilde{B}\|_2^2 = \text{tr}\{\tilde{B}(I_p - D_s^{-1})^2\tilde{B}'\} \le \lambda_{\max}\{(I_p - D_s^{-1})^2\}\|\tilde{B}\|_2^2$, and

$$\begin{aligned}\lambda_{\max}\{(I_p - D_s^{-1})^2\} &\le \frac{(\max_{1\le j\le p}|s_j^2 - 1|)^2}{(\min_{1\le j\le p} s_j^2 + \min_{1\le j\le p} s_j)^2} \\ &= o_{\text{a.s.}}(1).\end{aligned} \tag{33}$$

It follows that $\|\tilde{A} - \tilde{B}\|_2^2 = (n-q)o_{\text{a.s.}}(1)$. Thus, we have $L^4(F^{\tilde{A}\tilde{A}'}, F^{\tilde{B}\tilde{B}'}) = o_{\text{a.s.}}(1)$, hence the ESD of $\tilde{A}\tilde{A}'$ converges a.s. in distribution to the M–P law.

Note that $\tilde{A}\tilde{A}' = \tilde{B}\tilde{B}' + \Delta$ with $\Delta = \tilde{B}(D_s^{-2} - I_p)\tilde{B}$; hence, $\lambda_{\max}(\tilde{A}\tilde{A}') \ge \lambda_{\max}(\tilde{B}\tilde{B}') - \|\Delta\|$ and $\lambda_{\max}(\tilde{A}\tilde{A}') \le \lambda_{\max}(\tilde{B}\tilde{B}') + \|\Delta\|$ [e.g., Jiang (2010), page 167; also using the fact that $\lambda_{\max}(M) \le \|M\|$ and $\lambda_{\min}(M) \ge -\|M\|$ for any symmetric matrix M]. Similarly, we have $\lambda_{\min}(\tilde{A}\tilde{A}') \ge \lambda_{\min}(\tilde{B}\tilde{B}') - \|\Delta\|$ and $\lambda_{\min}(\tilde{A}\tilde{A}') \le \lambda_{\min}(\tilde{B}\tilde{B}') + \|\Delta\|$. It remains to show that $\|\Delta\| \xrightarrow{\text{a.s.}} 0$, but this follows from

$$\|\Delta\| \le \|\tilde{B}\|^2\|D_s^{-2} - I_p\| \le \frac{\max_{1\le j\le p}|s_j^2 - 1|}{\min_{1\le j\le p} s_j^2}\lambda_{\max}(\tilde{B}\tilde{B}') \xrightarrow{\text{a.s.}} 0.$$

7.2. *Notation.* Some notation will be used throughout the next two subsections. Most of these have been introduced before; we summarize below for convenience. Recall that A is an $n \times (n-q)$ matrix with $A'X = 0$ and $A'A = I_{n-q}$. We write $Z = [Z_{(1)}\ Z_{(2)}]$, where $Z_{(1)}$ is $n \times m$ and $Z_{(2)}$ is $n \times (p-m)$, $\tilde{Z} = p^{-1/2}Z$, and $\tilde{Z}_{(j)} = p^{-1/2}Z_{(j)}$, $j = 1, 2$. Also, we have $\tilde{y} = y - X\beta = \tilde{Z}_{(1)}\alpha_{(1)} + \varepsilon$ so that $\tilde{y}|Z \sim N(0, \sigma_{\varepsilon 0}^2 V_{1,0})$, where $V_{1,0} = I_n + \gamma_0\tilde{Z}_{(1)}\tilde{Z}_{(1)}'$; similarly, $\Sigma_{1,0} = A'V_{1,0}A = I_{n-q} + \gamma_0 A'\tilde{Z}_{(1)}\tilde{Z}_{(1)}'A$. Moreover, let $\zeta = A'Z$; $\bar{U} = p^{-1}\zeta\zeta' = \tilde{\zeta}\tilde{\zeta}'$ with $\tilde{\zeta} = p^{-1/2}\zeta$; $V_\gamma = I_n + \gamma\tilde{Z}\tilde{Z}' = I_n + (\gamma/p)ZZ'$; $P_\gamma = A\Sigma_\gamma^{-1}A' = V_\gamma^{-1} - V_\gamma^{-1}X(X'V_\gamma^{-1}X)^{-1}X'V_\gamma^{-1}$ with $\Sigma = \Sigma_\gamma = I_{n-q} + \gamma\bar{U}$ [e.g., Jiang (2007), page 13]; $G = G_\gamma = -(\partial/\partial\gamma)\Sigma_\gamma^{-1} = \Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}$. Define $b_1(\gamma) = \text{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0})$, $b_2(\gamma) = \text{tr}(\Sigma_\gamma^{-2}\Sigma_{1,0})$, $c_1(\gamma) = \text{tr}(\Sigma_\gamma^{-1}\bar{U})$, $c_2(\gamma) = \text{tr}(\Sigma_\gamma^{-1})$, $s(\gamma) = y'P_\gamma^2 y/\text{tr}(P_\gamma) = \tilde{y}P_\gamma^2\tilde{y}/\text{tr}(P_\gamma)$, and $\Delta(\gamma) = y'B_\gamma y = \tilde{y}'B_\gamma\tilde{y}$ with

$$B = B_\gamma = \frac{P_\gamma\tilde{Z}\tilde{Z}'P_\gamma}{\text{tr}(P_\gamma\tilde{Z}\tilde{Z}')} - \frac{P_\gamma^2}{\text{tr}(P_\gamma)}.$$

We shall also write $\gamma_* = \omega\gamma_0$.

7.3. *Proof of Theorem* 3.1. Our approach is to first consider a simplified version of Theorem 3.1, in which the entries of Z are i.i.d. $N(0,1)$, and then extend the proof by explaining how to relax the restriction.

Part (i). First consider the asymptotic behavior of $\hat{\gamma}$. For any fixed $\gamma > 0$, write $\Delta = \Delta(\gamma)$ and $B = B_\gamma$ for notational simplicity. Note that ζ is $(n-q) \times p$, whose entries are independent $N(0,1)$. Straight calculation, and Corollary 2.1, show that $\mathrm{tr}(P_\gamma) = \mathrm{tr}(\Sigma^{-1}) = O_{\mathrm{P}}(n)$, and $\mathrm{tr}(P_\gamma \tilde{Z}\tilde{Z}') = \mathrm{tr}(\Sigma^{-1}\bar{U}) = O_{\mathrm{P}}(n)$.

Next, write $\Delta = \mathbb{E}(\Delta|Z) + \Delta - \mathbb{E}(\Delta|Z) = \Delta_1 + \Delta_2$. By the normal theory [e.g., Jiang (2007), page 238], it can be shown that $\mathrm{var}(\Delta|Z) = 2\sigma_{\varepsilon 0}^2 \mathrm{tr}(DV_{1,0}DV_{1,0})$, where $D = A(C_1 - C_2)A'$ with $C_1 = \Sigma^{-1}\bar{U}\Sigma^{-1}/c_1$, $c_1 = \mathrm{tr}(\Sigma^{-1}\bar{U})$, $C_2 = \Sigma^{-2}/c_2$, and $c_2 = \mathrm{tr}(\Sigma^{-1})$. By Corollary 2.1, we have $c_j = O_{\mathrm{P}}(n)$, $j = 1, 2$. On the other hand, we have $\mathrm{tr}(DV_{1,0}DV_{1,0}) = \mathrm{tr}[\{(C_1 - C_2)\Sigma_{1,0}\}^2] = \mathrm{tr}\{(C_1\Sigma_{1,0})^2\} - 2\,\mathrm{tr}(C_1\Sigma_{1,0}C_2\Sigma_{1,0}) + \mathrm{tr}\{(C_2\Sigma_{1,0})^2\}$;

$$\mathrm{tr}\{(C_1\Sigma_{1,0})^2\} \le c_1^{-2}\,\mathrm{tr}\{(\Sigma^{-1}\bar{U}\Sigma^{-1}\Sigma_0)^2\} = O_{\mathrm{P}}(n^{-1}),$$

by Corollary 2.1, where Σ_0 is Σ with γ replaced by γ_0;

$$\mathrm{tr}\{(C_2\Sigma_{1,0})^2\} \le c_2^{-2}\,\mathrm{tr}\{(\Sigma^{-1}\Sigma_0\Sigma^{-1})^2\} = O_{\mathrm{P}}(n^{-1}) \quad \text{and}$$

$$\mathrm{tr}(C_1\Sigma_{1,0}C_2\Sigma_{1,0}) = (c_1c_2)^{-1}\,\mathrm{tr}(\Sigma^{-1}\bar{U}\Sigma^{-1}\Sigma_{1,0}\Sigma^{-2}\Sigma_{1,0}) = O_{\mathrm{P}}(n^{-1}).$$

It follows that $\mathrm{var}(\Delta|Z) = O_{\mathrm{P}}(n^{-1})$, hence, for any $\delta > 0$, we have $\mathbb{P}\{|\Delta - \mathbb{E}(\Delta|Z)| > \delta|Z\} \le \delta^{-2}\mathrm{var}(\Delta|Z) \xrightarrow{\mathrm{P}} 0$, as $n \to \infty$. Thus, by the dominated convergence theorem, we have $\mathbb{P}\{|\Delta - \mathbb{E}(\Delta|Z)| > \delta\} \to 0$, $\forall \delta > 0$, implying $\Delta_2 = o_{\mathrm{P}}(1)$.

Next, we have $\Delta_1 = \mathbb{E}(\Delta|Z) = \sigma_{\varepsilon 0}^2(b_1/c_1 - b_2/c_2)$, $b_1 = \mathrm{tr}(\Sigma^{-1}\bar{U}\Sigma^{-1}\Sigma_{1,0})$, $b_2 = \mathrm{tr}(\Sigma^{-2}\Sigma_{1,0})$, and c_1, c_2 are defined earlier. By Lemma 2.1, we have

$$\frac{c_1}{n-q} = \frac{1}{n-q}\sum_{k=1}^{n-q}\frac{\lambda_k}{1+\gamma\lambda_k} \xrightarrow{\text{a.s.}} \int_{b_-(\tau)}^{b_+(\tau)} \frac{x f_\tau(x)}{1+\gamma x}\,dx, \tag{34}$$

where λ_k, $1 \le k \le n-q$ are the eigenvalues of $\bar{U}$. Similarly, we have

$$\frac{c_2}{n-q} = \frac{1}{n-q}\sum_{k=1}^{n-q}\frac{1}{1+\gamma\lambda_k} \xrightarrow{\text{a.s.}} \int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\,dx. \tag{35}$$

Also, we have $b_1 = \mathrm{tr}(\Sigma^{-1}\bar{U}\Sigma^{-1}) + \gamma_0\,\mathrm{tr}\{\Sigma^{-1}\bar{U}\Sigma^{-1}\bar{U}_{(1)}\}$, and

$$\frac{\mathrm{tr}(\Sigma^{-1}\bar{U}\Sigma^{-1})}{n-q} = \frac{1}{n-q}\sum_{k=1}^{n-q}\frac{\lambda_k}{(1+\gamma\lambda_k)^2} \xrightarrow{\text{a.s.}} \int_{b_-(\tau)}^{b_+(\tau)} \frac{x f_\tau(x)}{(1+\gamma x)^2}\,dx. \tag{36}$$

On the other hand, note that $\mathrm{tr}\{\Sigma^{-1}\bar{U}\Sigma^{-1}\bar{U}_{(1)}\} = p^{-1}\sum_{k=1}^m \zeta_k' G\zeta_k$, where ζ_k is the kth column of ζ, and $G = \Sigma^{-1}\bar{U}\Sigma^{-1}$. Write $\Sigma = \Sigma_{-k} + (\gamma/p)\zeta_k\zeta_k'$, where $\Sigma_{-k} = I_{n-q} + (\gamma/p)\sum_{l \ne k}\zeta_l\zeta_l'$. Using a matrix identity [e.g., Sen and Srivastava

(1990), page 275], we have $\Sigma^{-1} = \Sigma_{-k}^{-1} - (\gamma/p)\{1 + (\gamma/p)u_k\}^{-1}\Sigma_{-k}^{-1}\zeta_k\zeta_k'\Sigma_{-k}^{-1}$, where $u_k = \zeta_k'\Sigma_{-k}^{-1}\zeta_k$. Thus, after some tedious derivation, we have the expression

$$\zeta_k' G\zeta_k = \frac{u_k^2 + v_k}{p\{1 + (\gamma/p)u_k\}^2}, \tag{37}$$

where $v_k = \zeta_k'\Sigma_{-k}^{-1}U_{-k}\Sigma_{-k}^{-1}\zeta_k$ and $U_{-k} = \sum_{l\neq k}\zeta_l\zeta_l'$. Note that ζ_k is independent with Σ_{-k}. Thus, by Proposition 2.1, we have, for any $1 \le k \le m$ and $t > 0$,

$$\begin{aligned} &\mathbb{P}\{|u_k - \mathrm{tr}(\Sigma_{-k}^{-1})| > t|\Sigma_{-k}\} \\ &\quad \le 2\exp\left\{-c\min\left(\frac{t^2}{K^4\|\Sigma_{-k}^{-1}\|_2^2}, \frac{t}{K^2\|\Sigma_{-k}^{-1}\|}\right)\right\}, \end{aligned} \tag{38}$$

where c and K are some positive constants. If we let

$$t = t_{m,k} = K^2\max\left(\sqrt{\frac{2\log(m)}{c}}\|\Sigma_{-k}^{-1}\|_2, \frac{2\log(m)}{c}\|\Sigma_{-k}^{-1}\|\right),$$

then, it is seen that the min in (38) is $\ge 2\log(m)/c$. It follows that $\mathbb{P}\{t_{m,k}^{-1}|u_k - \mathrm{tr}(\Sigma_{-k}^{-1})| > 1|\Sigma_{-k}\} \le 2/m^2$, $1 \le k \le m$; hence,

$$\mathbb{P}\left\{\max_{1\le k\le m} t_{m,k}^{-1}|u_k - \mathrm{tr}(\Sigma_{-k}^{-1})| > 1\right\} \le \frac{2}{m}. \tag{39}$$

On the other hand, we have $\|\Sigma_{-k}^{-1}\| \le 1$, and $\|\Sigma_{-k}^{-1}\|_2 \le \sqrt{\mathrm{tr}(\Sigma^{-2}) + 8} = O_{\mathrm{P}}(\sqrt{n})$, by Corollary 2.1. It follows by (10) that

$$\max_{1\le k\le m}|u_k - \mathrm{tr}(\Sigma_{-k}^{-1})| = O_{\mathrm{P}}(\sqrt{n\log n}). \tag{40}$$

Similarly, write $V_k = \Sigma_{-k}^{-1}U_{-k}\Sigma_{-k}^{-1}$. By a similar argument, it can be shown that

$$\max_{1\le k\le m}|v_k - \mathrm{tr}(V_k)| = O_{\mathrm{P}}(n\sqrt{n\log n}). \tag{41}$$

Also, by an earlier expansion, it can be shown that

$$|\mathrm{tr}(\Sigma_{-k}^{-1}) - \mathrm{tr}(\Sigma^{-1})| = \frac{(\gamma/p)\zeta_k'\Sigma_{-k}^{-2}\zeta_k}{1 + (\gamma/p)u_k} \le 1. \tag{42}$$

It follows, by (40) and (42), that

$$\max_{1\le k\le m}|u_k - \mathrm{tr}(\Sigma^{-1})| = O_{\mathrm{P}}(\sqrt{n\log n}). \tag{43}$$

Furthermore, by the same expansion, and (42), it can be shown that

$$|\mathrm{tr}(V_k) - \mathrm{tr}(\Sigma^{-1}U\Sigma^{-1})| \le 8p\lambda_{\max}(\bar{U}) + (1 + 2\sqrt{2})u_k \le O_{\mathrm{P}}(n), \tag{44}$$

where the O_{P} does not depend on k. It follows, by (41) and (44), that

$$\max_{1\le k\le m}\left|v_k - p\,\mathrm{tr}(\Sigma^{-1}\bar{U}\Sigma^{-1})\right| = O_{\mathrm{P}}(n\sqrt{n\log n}). \tag{45}$$

By (37), (43) and (45), it can be shown that $a_1 - O_{\mathrm{P}}(\sqrt{\log n/n}) < \zeta_k' G\zeta_k/(n-q) < a_1 + O_{\mathrm{P}}(\sqrt{\log n/n})$, where the O_{P}'s do not depend on k, and

$$a_1 = \left\{1+\gamma\left(\frac{n-q}{p}\right)\frac{\mathrm{tr}(\Sigma^{-1})}{n-q}\right\}^{-2} \times\left[\left(\frac{n-q}{p}\right)\left\{\frac{\mathrm{tr}(\Sigma^{-1})}{n-q}\right\}^2 + \frac{\mathrm{tr}(\Sigma^{-1}\bar{U}\Sigma^{-1})}{n-q}\right].$$

It then follows, by Lemma 2.1, that $\mathrm{tr}\{\Sigma^{-1}\bar{U}\Sigma^{-1}\bar{U}_{(1)}\}/(n-q) \xrightarrow{\mathrm{P}} \omega d_1$, where

$$d_1 = \left\{1+\gamma\tau\int_{b_-(\tau)}^{b_+(\tau)}\frac{\varphi_\tau(x)}{1+\gamma x}\,dx\right\}^{-2} \times\left[\tau\left\{\int_{b_-(\tau)}^{b_+(\tau)}\frac{\varphi_\tau(x)}{1+\gamma x}\,dx\right\}^2 + \int_{b_-(\tau)}^{b_+(\tau)}\frac{x\varphi_\tau(x)}{(1+\gamma x)^2}\,dx\right].$$

Therefore, we have $b_1/c_1 \xrightarrow{\mathrm{P}} (f_1+\gamma_0\omega d_1)/g_1$.

By a similar argument, we have $b_2/c_2 \xrightarrow{\mathrm{P}} (f_2+\gamma_0\omega d_2)/g_2$, where

$$d_2 = \left\{1+\gamma\tau\int_{b_-(\tau)}^{b_+(\tau)}\frac{\varphi_\tau(x)}{1+\gamma x}\,dx\right\}^{-2}\int_{b_-(\tau)}^{b_+(\tau)}\frac{\varphi_\tau(x)}{(1+\gamma x)^2}\,dx.$$

We have proved that Δ_1 converges in probability to a constant limit. The next thing we do is to determine the limit, in a different way. This is because the expression of the limit given above involving the d's is a bit complicated, from which it is not easy to make a conclusion. To this end, it is easy to show that $0 \le b_j/c_j \le (\gamma_0/\gamma)\vee 1$, $j=1,2$. Thus, by the dominated convergence theorem, $\mathbb{E}(b_j/c_j)$ converges to the same limit as b_j/c_j, $j=1,2$. On the other hand, it can be shown that

$$\mathbb{E}\left(\frac{b_1}{c_1}\right) = \mathbb{E}\left\{\frac{\mathrm{tr}(G)}{c_1}\right\} + \gamma_0\left(\frac{m}{p}\right)\mathbb{E}\left\{\frac{\mathrm{tr}(G\bar{U})}{c_1}\right\}, \tag{46}$$

$$\mathbb{E}\left(\frac{b_2}{c_2}\right) = \mathbb{E}\left\{\frac{\mathrm{tr}(\Sigma^{-2})}{c_2}\right\} + \gamma_0\left(\frac{m}{p}\right)\mathbb{E}\left\{\frac{\mathrm{tr}(\Sigma^{-2}\bar{U})}{c_2}\right\}. \tag{47}$$

Furthermore, it is easy to show that $0 \le \mathrm{tr}(G)/c_1 \le 1$, $0 \le \mathrm{tr}(G\bar{U})/c_1 \le \gamma^{-1}$, $0 \le \mathrm{tr}(\Sigma^{-2})/c_2 \le 1$, and $0 \le \mathrm{tr}(\Sigma^{-2}\bar{U})/c_2 \le \gamma^{-1}$. Thus, by Lemma 2.1 and, again, the dominated convergence theorem, the right-hand sides of (46) and (47) converge to the limit l_1, l_2, respectively, where $l_j = u_j + \gamma_0\omega w_j$, $u_j = f_j/g_j$, $j=1,2$, $w_1 = \int_{b_-(\tau)}^{b_+(\tau)}\{x^2\varphi_\tau(x)/(1+\gamma x)^2\}\,dx/g_1$, and $w_2 = f_1/g_2$. Thus, with a little bit of algebra, it follows that the limit of Δ_1 is $\sigma_{\varepsilon 0}^2\{(\gamma_*/\gamma)-1\}(u_2-u_1)$, and $u_2-u_1>0$ by a well-known inequality [e.g., Jiang (2010), pages 147–148].

Finally, recall that $\Delta = \Delta(\gamma)$. Thus, in conclusion, we have shown that $\Delta(\gamma)$ converges in probability to a constant limit, which is >0, $=0$, or <0 depending on whether γ is $<\gamma_*$, $=\gamma_*$, or $>\gamma_*$. This proves (i).

Part (ii). Write $\xi = A'\tilde{y}$. We have

$$s'(\gamma) = \frac{\mathrm{tr}(G)\xi'\Sigma^{-2}\xi}{\{\mathrm{tr}(\Sigma^{-1})\}^2} - 2\frac{\xi' G\Sigma^{-1}\xi}{\mathrm{tr}(\Sigma^{-1})}.$$

It is easy to show that $\mathbb{E}(\xi'\xi) \le \sigma_{\varepsilon 0}^2(1+\gamma_0)(n-q)$. Thus, we have $0 \le \xi'\Sigma^{-2}\xi \le \xi'\xi = O_{\mathrm{P}}(n-q)$, $0 \le \xi' G\Sigma^{-1}\xi \le \lambda_{\max}(\bar{U})\xi'\xi = O_{\mathrm{P}}(n-q)$, by Lemma 2.2, and $\mathrm{tr}(G) \le \lambda_{\max}(\bar{U})(n-q)O_{\mathrm{P}}(n-q)$. Furthermore, for any $0<\gamma\le 2\gamma_0$, we have $(n-q)^{-1}\mathrm{tr}(\Sigma^{-1}) \ge (n-q)^{-1}\mathrm{tr}\{(I_{n-q}+2\gamma_0\bar{U})^{-1}\} \xrightarrow{\text{a.s.}} \int_{b_-(\tau)}^{b_+(\tau)}\{1+2\gamma_0 x\}^{-1} f_\tau(x)\,dx > 0$, by Lemma 2.1. Note that the O_{P}'s here do not depend on γ. It follows that $\sup_{0<\gamma\le 2\gamma_0}|s'(\gamma)| = O_{\mathrm{P}}(1)$. Therefore, by the Taylor expansion, we have $\hat{\sigma}_\varepsilon^2 = s(\gamma_*) + s'(\tilde{\gamma})(\hat{\gamma}-\gamma_*) = s(\gamma_*) + o_{\mathrm{P}}(1)$, by part (i) of Theorem 3.1, where $\tilde{\gamma}$ lies between γ_* and $\hat{\gamma}$.

Next, by the proof of part (i), it is easy to show that, with $\gamma=\gamma_*$, we have $s(\gamma) = \sigma_{\varepsilon 0}^2(b_2/c_2) + O_{\mathrm{P}}(n^{-1/2})$, and $b_2/c_2 \xrightarrow{\mathrm{P}} l_2$, where l_2 is defined in the proof of part (i) with $\gamma=\gamma_*$. It follows that $l_2 = u_2+\gamma w_2 = (f_2+\gamma f_1)/g_2 = 1$. This proves part (ii).

We have proved the theorem under the assumption that the entries of Z are independent $N(0,1)$. We now explain how the result can be extended under more general conditions. The first extension is to the case where the entries of Z are i.i.d. sub-Gaussian. The only place in the proof where the normality was used was in the early going of part (i), where the normality of Z implied that the entries of $\zeta = A'Z$ are also independent $N(0,1)$. However, the way A is involved is always through $AA' = P_{X^\perp} = I - P_X$, where $P_X = X(X'X)^{-1}X'$, and P_X has rank q, which is fixed (see the beginning of Section 1.3). It turns out that P_X is negligible in the sense that, after replacing AA' by I, the $(n\times n)$ identity matrix, the difference does not affect the order of the approximation in every single place throughout the proof. Furthermore, when A is replaced by I, the entries of ζ are clearly i.i.d., and the rest of the proof applies without any change to the case where the entries of Z are independent sub-Gaussian. This extends the result to the latter case.

The next extension is to the case of standardized design matrix. Using the preliminary results, namely, Lemmas 2.5, 2.6 and Corollary 2.3, it can be shown that, the difference induced by the standardization is negligible in the same sense.

All the extensions have been verified, step-by-step, throughout the proof to make sure that the results of Theorem 3.1 remain valid for the case where Z is the standardized design matrix as described in Section 1.3 (also above Corollary 2.3), where the entries of U are i.i.d. sub-Gaussian. The detailed verifications, which are tedious, are omitted.

7.4. *Proof of Theorem* 3.2. Recall that $\hat{\gamma}$ solves equation (3), and $\hat{\sigma}_\varepsilon^2$ is given by the right-hand side of (4) with $\gamma = \hat{\gamma}$. It follows that $\Delta(\hat{\gamma}) = 0$ and $\hat{\sigma}_\varepsilon^2 = s(\hat{\gamma})$. Theorem 3.1 has established that $\hat{\gamma} \xrightarrow{\mathrm{P}} \gamma_*$. Because $\Delta(\hat{\gamma}) = 0$, by the Taylor series expansion, and some algebra, we have

$$\hat{\gamma} - \gamma_* = -\frac{\Delta(\gamma_*)}{\Delta'(\gamma_*)} + O_{\mathrm{P}}(|\Delta(\gamma_*)|^2). \tag{48}$$

Here, we also use the fact that $\Delta'(\gamma_*)$ converges in probability to a nonzero quantity. Indeed, from the proof of Theorem 3.1, it can be checked that $\Delta'(\gamma)$ converges in probability, for every fixed γ, to $\Delta'_\infty(\gamma)$, where

$$\Delta_\infty(\gamma) = \sigma_{\varepsilon 0}^2\left(\frac{\gamma_*}{\gamma} - 1\right)\left\{\frac{f_2(\gamma)}{g_2(\gamma)} - \frac{f_1(\gamma)}{g_1(\gamma)}\right\},$$

and the difference within the $\{\cdots\}$ is positive. It follows that

$$\Delta'_\infty(\gamma_*) = -\frac{\sigma_{\varepsilon 0}^2}{\gamma_*}\left\{\frac{f_2(\gamma_*)}{g_2(\gamma_*)} - \frac{f_1(\gamma_*)}{g_1(\gamma_*)}\right\} < 0. \tag{49}$$

Next, a Taylor series expansion of $s(\gamma)$ yields $\hat{\sigma}_\varepsilon^2 = s(\gamma_*) + s'(\gamma_*)(\hat{\gamma} - \gamma_*) + O(|\hat{\gamma} - \gamma_*|^2)$, which, combined with (48), leads to the expansion

$$\hat{\sigma}_\varepsilon^2 = s(\gamma_*) - \frac{s'(\gamma_*)}{\Delta'(\gamma_*)}\Delta(\gamma_*) + O_{\mathrm{P}}(|\Delta(\gamma_*)|^2). \tag{50}$$

Write $s(\gamma) = s_1(\gamma) + s_2(\gamma)$, where $s_1(\gamma) = \mathbb{E}\{s(\gamma)|Z\}$ and $s_2(\gamma) = s(\gamma) - s_1(\gamma)$. It was shown in the proof of Theorem 3.1 that $s_1(\gamma) = \sigma_{\varepsilon 0}^2\{b_2(\gamma)/c_2(\gamma)\}$. Also, we have the expression $s_2(\gamma) = \tilde{w}' D_\gamma \tilde{w} - \operatorname{tr}(D_\gamma)$, where

$$D_\gamma = \sigma_{\varepsilon 0}^2 \frac{\Sigma_{1,0}^{1/2}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}^{1/2}}{\operatorname{tr}(\Sigma_\gamma^{-1})} \quad \text{and} \quad \tilde{w} = \frac{\Sigma_{1,0}^{-1/2}A'\tilde{y}}{\sigma_{\varepsilon 0}}.$$

Note that $\tilde{w}|Z \sim N(0, I_{n-q})$. Also recall [from the proof of Theorem 3.1, part (i)] that $\Delta(\gamma) = \Delta_1(\gamma) + \Delta_2(\gamma)$ with $\Delta_1(\gamma) = \sigma_{\varepsilon 0}^2 \sum_{j=1}^2 (-1)^{j-1} b_j(\gamma)/c_j(\gamma)$, and, similarly, $\Delta_2(\gamma) = \tilde{w}' F_\gamma \tilde{w} - \operatorname{tr}(F_\gamma)$, where $F_\gamma = \sigma_{\varepsilon 0}^2 \Sigma_{1,0}^{1/2} H_\gamma \Sigma_{1,0}^{1/2}$ with

$$H_\gamma = \frac{\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}}{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U})} - \frac{\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}}{\operatorname{tr}(\Sigma_\gamma^{-1})}.$$

As in the proof of Theorem 3.1, part (ii), write $\xi = A'\tilde{y}$, and observe that

$$s'(\gamma) = \frac{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1})\xi'\Sigma_\gamma^{-2}\xi}{\{\operatorname{tr}(\Sigma_\gamma^{-1})\}^2} - \frac{2\xi'\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-2}\bar{U}\Sigma_\gamma^{-1}\xi}{\operatorname{tr}(\Sigma_\gamma^{-1})}.$$

We have $\mathbb{E}(\xi'\Sigma_\gamma^{-2}\xi|Z) = \sigma_{\varepsilon 0}^2 \operatorname{tr}(\Sigma_\gamma^{-2}\Sigma_{1,0})$, and

$$\mathbb{E}(\xi'\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-2}\bar{U}\Sigma_\gamma^{-1}\xi|Z) = \sigma_{\varepsilon 0}^2 \operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-2}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}).$$

With these, using similar derivations to the proof of Theorem 3.1, we conclude that

$$\begin{aligned} s'(\gamma) &\xrightarrow{\mathrm{P}} s'_\infty(\gamma) \\ &= \sigma_{\varepsilon 0}^2\biggl[\frac{h_{2,1}(\gamma)\{h_{2,0}(\gamma)+\gamma_* h_{2,1}(\gamma)\}}{\{h_{1,0}(\gamma)\}^2} \\ &\qquad - \frac{2\{h_{4,2}(\gamma)+\gamma_* h_{4,3}(\gamma)\}}{h_{1,0}(\gamma)}\biggr] \end{aligned} \tag{51}$$

[see (15) for notation], so that, using the fact that $h_{k,l}(\gamma) + \gamma h_{k,l+1}(\gamma) = h_{k-1,l}(\gamma)$, we have

$$s'(\gamma_*) = \sigma_{\varepsilon 0}^2\biggl(\frac{h_{2,1}(\gamma_*)}{\{h_{1,0}(\gamma_*)\}^2} - \frac{2h_{3,2}(\gamma_*)}{h_{1,0}(\gamma_*)}\biggr).$$

Thus, going back to (50), we can write

$$\begin{aligned} &\hat{\sigma}_\varepsilon^2 - \sigma_{\varepsilon 0}^2 \\ &\quad = s_2(\gamma_*) - \frac{s'_\infty(\gamma_*)}{\Delta'_\infty(\gamma_*)}\Delta_2(\gamma_*) \\ &\qquad + \sigma_{\varepsilon 0}^2\biggl\{\frac{b_2(\gamma_*)}{c_2(\gamma_*)} - 1\biggr\} - \sigma_{\varepsilon 0}^2\frac{s'_\infty(\gamma_*)}{\Delta'_\infty(\gamma_*)}\biggl\{\frac{b_1(\gamma_*)}{c_1(\gamma_*)} - \frac{b_2(\gamma_*)}{c_2(\gamma_*)}\biggr\} \\ &\qquad - \frac{s'(\gamma_*) - s'_\infty(\gamma_*)}{\Delta'_\infty(\gamma_*)}\Delta(\gamma_*) + \frac{\{\Delta'(\gamma_*) - \Delta'_\infty(\gamma_*)\}s'(\gamma_*)}{\Delta'(\gamma_*)\Delta'_\infty(\gamma_*)}\Delta(\gamma_*) \\ &\qquad + O_{\mathrm{P}}(|\Delta(\gamma_*)|^2). \end{aligned} \tag{52}$$

We shall argue that all of the terms on the right-hand side of (52) except those in the first line are $o_{\mathrm{P}}(n^{-1/2})$, while $\sqrt{n}$ times the term in the first line on the right-hand side of (52) converges weakly to a normal distribution with zero mean.

Once we establish that

$$\Delta(\gamma_*) = O_{\mathrm{P}}(n^{-1/2}), \tag{53}$$

then since $s'(\gamma_*) - s'_\infty(\gamma_*) = o_{\mathrm{P}}(1)$ and $\Delta'(\gamma_*) - \Delta'_\infty(\gamma_*) = o_{\mathrm{P}}(1)$, it immediately follows that the terms on the last two lines of (52) are $o_{\mathrm{P}}(n^{-1/2})$.

In order to establish (53), we need the following lemma.

LEMMA 7.1. *Suppose that* (10) *holds and let* $\Psi = p^{-1}ZZ'$ *and* $\tilde{\Sigma}_\gamma = I_n + \gamma\Psi$. *Then we have*

$$\operatorname{tr}(\Sigma_\gamma^{-k}\bar{U}^l) = \operatorname{tr}(\tilde{\Sigma}_\gamma^{-k}\Psi^l) + O_{\mathrm{P}}(1) \qquad \forall k \ge 1, l \ge 0. \tag{54}$$

The proof of Lemma 7.1, which is omitted, follows closely the note regarding AA' near the end of the proof of Theorem 3.1. The advantage of this lemma is that, because the entries of Z are independent sub-Gaussian with mean 0, unit variance, and bounded fourth moments, the behavior of the trace on the right-hand side of (54) is well studied. Indeed, we can use Theorem 9.10 of Bai and Silverstein (2010) on the asymptotic behavior of linear spectral statistics to claim that, for all $k \geq 1, l \geq 0$, we have

$$\left|\frac{1}{n}\operatorname{tr}(\tilde{\Sigma}_\gamma^{-k}\Psi^l) - \int \frac{x^l}{(1+\gamma x)^k}\varphi_{n/p}(x)\,dx\right| = O_{\mathrm{P}}(n^{-1}). \tag{55}$$

Equation (55), combined with (17), (15) and (54), imply that for all $k \geq 1, l \geq 0$, we have

$$\left|\frac{1}{n}\operatorname{tr}(\Sigma_\gamma^{-k}\bar{U}^l) - h_{k,l}(\gamma)\right| = O_{\mathrm{P}}(n^{-1}). \tag{56}$$

We have the following result describing the asymptotic behavior of the ratios $b_j(\gamma)/c_j(\gamma)$, $j = 0, 1$.

PROPOSITION 7.1. *Under the assumptions of Theorem 3.2, for any $\gamma > 0$, for $j = 1, 2$,*

$$\frac{b_j(\gamma)}{c_j(\gamma)} - \frac{b_{j,\infty}(\gamma)}{c_{j,\infty}(\gamma)} = o_{\mathrm{P}}(n^{-1/2}).$$

Proof of Proposition 7.1 is given in the supplementary material [Jiang et al. (2016)].

Observe that

$$\begin{aligned}
b_{1,\infty}(\gamma_*) &= h_{2,1}(\gamma_*) + \gamma_*\{1+\tau\gamma_* h_{1,0}(\gamma_*)\}^{-2}\{\tau h_{1,0}^2(\gamma_*) + h_{2,1}(\gamma_*)\},\\
b_{2,\infty}(\gamma_*) &= h_{2,0}(\gamma_*) + \gamma_*\{1+\tau\gamma_* h_{1,0}(\gamma_*)\}^{-2} h_{2,0}(\gamma_*),\\
c_{1,\infty}(\gamma_*) &= h_{1,1}(\gamma_*) \quad \text{and} \quad c_{2,\infty}(\gamma_*) = h_{1,0}(\gamma_*).
\end{aligned}$$

Also, by the proof of Theorem 3.1, part (ii), we have $b_{2,\infty}(\gamma_*)/c_{2,\infty}(\gamma_*) = 1$. Moreover, $\Delta_1(\gamma_*) \xrightarrow{\mathrm{P}} 0$, implying $b_{1,\infty}(\gamma_*)/c_{1,\infty}(\gamma_*) - b_{2,\infty}(\gamma_*)/c_{2,\infty}(\gamma_*) = 0$. In conjunction with Proposition 7.1, this shows that the terms in the second line of the right-hand side of (52) are $o_{\mathrm{P}}(n^{-1/2})$. Thus, $\Delta_1(\gamma_*) = o_{\mathrm{P}}(n^{-1/2})$. On the other hand, it is seen from the proof of Theorem 3.1 that $\Delta_2(\gamma_*) = O_{\mathrm{P}}(n^{-1/2})$. Therefore, (52) holds.

Next, by the expressions of $s_2(\gamma)$, $\Delta_2(\gamma)$, we can write the first line of the right-hand side of (52) as $Q = \tilde{w}'M\tilde{w} - \operatorname{tr}(M)$, where $M = D_{\gamma_*} - \{s'_\infty(\gamma_*)/\Delta'_\infty(\gamma_*)\} \times F_{\gamma_*}$. We have [e.g., Jiang (2007), page 238] $\mathbb{E}(Q|Z) = 0$ and $\operatorname{var}(Q|Z) = 2\operatorname{tr}(M^2)$. From the definitions of D_γ and F_γ, the following expressions can be derived:

$$\frac{\operatorname{tr}(D_\gamma^2)}{\sigma_{\varepsilon 0}^4} = \frac{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0})}{\{\operatorname{tr}(\Sigma_\gamma^{-1})\}^2}, \tag{57}$$

$$
\begin{aligned}
\frac{\operatorname{tr}(F_\gamma^2)}{\sigma_{\varepsilon 0}^4} = {} & \frac{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0})}{\{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U})\}^2} \\
& + \frac{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0})}{\{\operatorname{tr}(\Sigma_\gamma^{-1})\}^2} \\
& - 2\frac{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0})}{\operatorname{tr}(\Sigma_\gamma^{-1})\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U})},
\end{aligned} \tag{58}
$$

$$
\begin{aligned}
\frac{\operatorname{tr}(D_\gamma F_\gamma)}{\sigma_{\varepsilon 0}^4} = {} & \frac{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0})}{\operatorname{tr}(\Sigma_\gamma^{-1})\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U})} \\
& - \frac{\operatorname{tr}(\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0}\Sigma_\gamma^{-1}\bar{U}\Sigma_\gamma^{-1}\Sigma_{1,0})}{\{\operatorname{tr}(\Sigma_\gamma^{-1})\}^2}.
\end{aligned} \tag{59}
$$

When multiplied by n, each of the term above is bounded, and by Proposition 3.1, converges to a limit. Notice that $\rho_* = -s'_\infty(\gamma_*)/\Delta'_\infty(\gamma_*)$. Consequently, by making use of the symmetry of the matrix M and the fact that $\tilde{w}|M \sim N(0, I_{n-q})$, it is easily seen that conditionally on Z, $Q = s_2(\gamma_*) + \rho_*\Delta_2(\gamma_*)$ is a weighted sum of $n-q$ independent $\chi_1^2 - 1$ random variables, where the weights are the eigenvalues of M. Thus, by a standard approach, such as using characteristic functions, or a martingale central limit theorem, it is easily seen that $\sqrt{n}Q$ converges in distribution to the normal distribution given in (19).

We would like to point out that if α_i's and ε_i's are non-Gaussian, but satisfy the conditions of Remark 3.3, then this is the only step where the proof differs. Specifically, in this case the asymptotic normality of $\sqrt{n}Q$ may be obtained by making use of a central limit theorem for random quadratic forms [see, e.g., Theorem 7.2 of Bai and Yao (2008)].

Note that (53) also ensures $\hat{\gamma} - \gamma_* = O_{\mathrm{P}}(n^{-1/2})$ by virtue of (48), the convergence of $\Delta'(\gamma_*)$ to $\Delta'_\infty(\gamma_*)$, and (49). Moreover, from this representation of $\hat{\gamma} - \gamma_*$, and the decomposition of $\Delta(\gamma)$ in the beginning of the proof of Theorem 3.1, as well as the expression for $\operatorname{var}(\Delta(\gamma)|Z)$, using an argument similar to that used in proving asymptotic normality of $\hat{\sigma}_\varepsilon^2$, we obtain (18). This completes the proof of Theorem 3.2.

Acknowledgments. The authors wish to thank Professor Iain Johnstone for helpful discussion. The authors are grateful to a Co-Editor, an Associate Editor and three referees for their constructive comments and suggestions that have led to significant improvements of the results and presentation.

SUPPLEMENTARY MATERIAL

Supplementary material for "On high-dimensional misspecified mixed model analysis in genome-wide association studies" (DOI: 10.1214/15-AOS1421SUPP; .pdf). This document contains additional proof details.

REFERENCES

ALLEN, H. L., ESTRADA, K., LETTRE, G., BERNDT, S. I., WEEDON, M. N., RIVADENEIRA, F. et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467** 832–838.

BAI, Z. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. MR1711663

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

BAI, Z. and YAO, J.-F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053

BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

BOX, G. E. P. (1979). Some problems of statistics of everyday life. *J. Amer. Statist. Assoc.* **74** 1–4.

CAI, T. T., JIN, J. and LOW, M. G. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449. MR2382653

CRESSIE, N. and LAHIRI, S. N. (1993). The asymptotic distribution of REML estimators. *J. Multivariate Anal.* **45** 217–233. MR1221918

CROSS-DISORDER GROUP OF THE PSYCHIATRIC GENOMICS CONSORTIUM (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45** 984–994.

DAS, K. (1979). Asymptotic optimality of restricted maximum likelihood estimates for the mixed model. *Calcutta Statist. Assoc. Bull.* **28** 125–142. MR0586086

FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. MR2885839

HEYDE, C. C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Springer, New York. MR1461808

HINDORFF, L. A., SETHUPATHY, P., JUNKINS, H. A., RAMOS, E. M., MEHTA, J. P., COLLINS, F. S. and MANOLIO, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106** 9362–9367.

HSU, D., KAKADE, S. M. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17** 1–6. MR2994877

JAVANMARD, A. and MONTANARI, A. (2013). Confidence intervals and hypothesis testing for high-dimensional regression. Available at arXiv:1306.3171.

JIANG, J. (1996). REML estimation: Asymptotic behavior and related topics. *Ann. Statist.* **24** 255–286. MR1389890

JIANG, J. (1997). Wald consistency and the method of sieves in REML estimation. *Ann. Statist.* **25** 1781–1803. MR1463575

JIANG, J. (2000). A matrix inequality and its statistical application. *Linear Algebra Appl.* **307** 131–144. MR1741921

JIANG, J. (2003). Empirical method of moments and its applications. *J. Statist. Plann. Inference* **115** 69–84. MR1972940

JIANG, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer, New York. MR2308058

JIANG, J. (2010). *Large Sample Techniques for Statistics*. Springer, New York. MR2675055

JIANG, J., LI, C., PAUL, D., YANG, C. and ZHAO, H. (2016). Supplement to “On high-dimensional misspecified mixed model analysis in genome-wide association study.” DOI:10.1214/15-AOS1421SUPP.

JIN, J. (2008). Proportion of non-zero normal means: Universal oracle equivalences and uniformly consistent estimators. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 461–493. MR2420411

JIN, J. and CAI, T. T. (2007). Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102** 495–506. MR2325113

LEE, S. H., DECANDIA, T. R., RIPKE, S., YANG, J., SULLIVAN, P. F., GODDARD, M. E. et al. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44** 247–250.

LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. and TIBSHIRANI, R. (2013). A signicance test for the lasso. Available at arXiv:1301.7161.

MADSEN, B. E. and BROWNING, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5** e1000384.

MAHER, B. (2008). Personal genomes: The case of the missing heritability. *Nature* **456** 18–21.

MANOLIO, T. A. (2010). Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363** 166–176.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J. et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461** 747–753.

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718

REID, S., TIBSHIRANI, R. and FRIEDMAN, J. (2013). A study of error variance estimation in Lasso regression. Available at arXiv:1311.5274.

RICHARDSON, A. M. and WELSH, A. H. (1994). Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Austral. J. Statist.* **36** 31–43. MR1309503

RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** 9. MR3125258

SEN, A. and SRIVASTAVA, M. (1990). *Regression Analysis: Theory, Methods, and Applications*. Springer, , New York. MR1063855

SPEED, D., HEMANI, G., JOHNSON, M. R. and BALDING, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *Amer. J. Human Genetics* **91** 1011–1021.

STAHL, E. A., WEGMANN, D., TRYNKA, G., GUTIERREZ-ACHURY, J., DO, R., VOIGHT, B. F. et al. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44** 483–489.

SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166

VATTIKUTI, S., GUO, J. and CHOW, C. C. (2012). Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet.* **8** e1002637.

VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

VISSCHER, P. M. (2008). Sizing up human height variation. *Nat. Genet.* **40** 489–490.

VISSCHER, P. M., HILL, W. G. and WRAY, N. R. (2008). Heritability in the genomics era–concepts and misconceptions. *Nat. Rev. Genet.* **9** 255–266.

VISSCHER, P. M., BROWN, M. A., MCCARTHY, M. I. and YANG, J. (2012). Five years of GWAS discovery. *Amer. J. Human Genetics* **90** 7–24.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. and VISSCHER, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42** 565–569.

YANG, J., LEE, S. H., GODDARD, M. E. and VISSCHER, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Amer. J. Human Genetics* **88** 76–82.

YANG, C., LI, C., KRANZLER, H. R., FARRER, L. A., ZHAO, H. and GELERNTER, J. (2016). Exploring the genetic architecture of alcohol dependence in African-Americans via analysis of a genomewide set of common variants. *Hum. Genet.* **133** 617–624.

ZAITLEN, N., KRAFT, P., PATTERSON, N., PASANIUC, B., BHATIA, G., POLLACK, S. and PRICE, A. L. (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet*. **9** e1003520.

J. JIANG
D. PAUL
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
1 SHIELDS AVENUE
DAVIS, CALIFORNIA 95656
USA
E-MAIL: jimjiang@ucdavis.edu
debpaul@ucdavis.edu

C. LI
TAKEDA PHARMACEUTICALS INTERNATIONAL CO.
35 LANDSDOWNE ST, ROOM 7042
CAMBRIDGE, MASSACHUSETTS 02141
USA
E-MAIL: licong.jason@gmail.com

C. YANG
DEPARTMENT OF MATHEMATICS
HONG KONG BAPTIST UNIVERSITY
FONG SHU CHUEN BUILDING
HO SIN HANG CAMPUS
KOWLOON TONG KOWLOON
HONG KONG
E-MAIL: eeyang@hkbu.edu.hk

H. ZHAO
YALE SCHOOL OF PUBLIC HEALTH
YALE UNIVERSITY
60 COLLEGE STREET
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: hongyu.zhao@yale.edu