Tests for High Dimensional Regression Coefficients with Factorial Designs

Ping-Shou Zhong and Song Xi Chen *

Iowa State University, Iowa State University and Peking University

Abstract

We propose simultaneous tests for coefficients in high dimensional linear regression models with factorial designs. The proposed tests are designed for the "large p, small n" situations where the conventional F-test is no longer applicable. We derive the asymptotic distribution of the proposed test statistic under the high dimensional null hypothesis and various scenarios of the alternatives, which allow power evaluations. We also evaluate the power of the F-test under very mild dimensionality. The proposed tests are employed to analyze a micro-array data on Yorkshire Gilts to find significant gene ontology terms which are significantly associated with the thyroid hormone after accounting for the designs of the experiment.

Key Words and Phrases: Factorial Design; Gene-set test; High dimensional regression; Large p, small n; U-statistics.

*Ping-Shou Zhong is a Ph.D candidate, Department of Statistics, Iowa State University, Ames, IA 50011-1210 (E-mail: pszhong@iastate.edu). Song Xi Chen is Professor, Department of Statistics, Iowa State University, Ames, IA 50011-1210 (E-mail: songchen@iastate.edu) and Chair Professor, Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing 100651, China (E-mail: csx@gsm.pku.edu.cn).
The authors acknowledge support from NSF grants SES-0518904, DMS-0604563 and DMS-0714978. We also thank two anonymous referees and the Associate Editor for constructive comments and suggestions which have improved the presentation of the paper.

1. INTRODUCTION

The emergence of high-dimensional data, such as the gene expression values in microarray and the single nucleotide polymorphism (SNP) data, brings challenges to many traditional statistical methods and theory. One important aspect of the high-dimensional data under the regression setting is that the number of covariates greatly exceeds the sample size. For example, in microarray data, the number of genes (p) is in the order of thousands whereas the sample size (n) is much less, usually less than fifty due to limitation to replicate. This is the so-called "large-p, small-n" paradigm, which translates to a regime of asymptotics where $p \to \infty$ much faster than n. See Kosorok and Ma (2007), Fan, Hall and Yao (2007), Huang, Wang and Zhang (2007), Chen and Qin (2010) among others. Kosorok and Ma (2007) considered uniform convergence for a large number of marginal discrepancy measures targeted on univariate distributions, means and medians. Chen and Qin (2010) proposed a two sample test on high dimensional means. Both of these two aforementioned papers considered testing under "large-p, small-n" without a regression structure, which is the focus of the present paper. Much earlier, for more moderate dimensions, Portnoy (1984, 1985) had considered consistency and asymptotic normality for the M-estimators of linear regression coefficients when the dimension p of the covariates grows to infinity faster than the square root of the sample size n. The rates for p that Portnoy considered were $p = o(n/\log(p))$ for consistency and $p = o(n^{2/3}/\log(p))$ for asymptotic normality of the M-estimators.

Covariate selection for high dimensional linear regression has attracted much attention and has been intensively considered in recent years. Penalizing methods are alternatives to the traditional least square estimator for simultaneous variable selection and shrinkage estimation. These include the LASSO (Tibshirani, 1996) with a L_1-penalty, the bridge regression with a L_2-penalty (Frank and Friedman, 1996), the SCAD penalty proposed by Fan and Li (2001) and Candes and Tao (2007)'s Dantzig selector; see also Fan and Lv (2008) and Wang (2009) for other methods of variable selection. There is also a line of works on ANOVA with diverging number of treatments while the number of replications (cell sample sizes) is small and can be regarded as fixed. This includes the rank based nonparametric tests proposed by Brownie and Boos (1994), Boos and Brownie (1995), Akritas and Arnold (2000), Bathke and Lankowski (2005), Bathke and Harrar (2008), Harrar and Bathke (2008), and Wang and Akritas (2009). The problem can be viewed as "large p, fixed n" in contrast to the conventional "fixed p, large n" setting and the "large p, small n" paradigm we are considering.

This paper is aimed at developing simultaneous tests on linear regression coefficients that can accommodate high dimensionality and factorial designs. The latter is often encountered in statistical experiments especially those in biology, and there is no exception for high dimensional data. Testing hypotheses on the regression coefficients is a necessity in determining the effects of covariates on certain outcome variable. Our interest here is on testing the significance of a large number of covariates simultaneously. This is motivated by the latest need in biology to identify significant sets of genes (Subramanian *et.al*, 2005; Efron and Tibshirani, 2007; Newton *et.al*, 2007), which are associated with certain clinical outcome, rather than identifying individual gene. As the dimension of a gene-set ranges from a few to thousands, and the gene-sets can overlap as they share common genes, there are both high dimensionality and multiplicity in gene-set testing. In order to test for the significance of a gene-set, the P-value associated with a hypothesis regarding the regression coefficient corresponding to the gene-set is needed. This calls for multivariate tests for regression coefficients that can accommodate both high dimensionality and dependence among the covariates.

We propose tests for high dimensional regression coefficients for both simple random or factorial designs. A feature of the tests is that they do not require explicit relationships between the growth rates of p and n, which makes the tests adaptable to a wide range of high dimensionality. The tests also account for a variety of dependence among the high dimensional covariate. These together with their accommodation to factorial designs makes the tests more applicable in applications. The F-test is the conventional test for regression coefficients simultaneously under the normality and $p < n - 1$. We take the opportunity to study the F-test and find that it is adversely affected by an increasing dimension.

The paper is organized as follows. We first study the F-test and propose a new test statistic in Section 2 for simple random designs. Section 3 discusses some general properties of U-statistics under high dimensionality. Section 4 establishes the main properties of the proposed test. Extensions to factorial designs are made in Section 5. Section 6 reports results from simulation studies. Empirical analyses on a microarray dataset on Yorkshire Gilts with factorial designs are reported in Section 7. All technical details are relegated to the Appendix.

2. MODELS AND TEST STATISTICS

Consider a linear regression model

$$E(Y_i|X_i) = \alpha + X_i'\beta \quad \text{and} \quad Var(Y_i|X_i) = \sigma^2 \tag{2.1}$$

for $i = 1, \ldots, n$ where $X_1, \cdots, X_n$ are independent and identically distributed p-dimensional covariates and $Y_1, \cdots, Y_n$ are independent responses, β is the vector of regression coefficients, and α is a nuisance intercept. We do not impose any specific distribution on Y_i given X_i except when studying the F-test in the next subsection.

The true parameter (α, β) in the linear regression model is defined as

$$(\alpha, \beta) = \arg \min_{\tilde{\alpha} \in R^1, \tilde{\beta} \in R^p} E(Y_i - \tilde{\alpha} - X_i'\tilde{\beta})^2.$$

To make β identifiable, we assume that $\Sigma = \mathrm{Var(X_i)} > 0$. This is weaker than the sparse Riesz condition in Zhang and Huang (2008), which requires the eigenvalues of Σ are all bounded from below and above. The sparse Riesz condition is for the purpose of parameter estimation and variable selection, which are different from the agenda of this paper.

Our interest is in testing a high dimensional hypothesis

$$H_0 : \beta = \beta_0 \quad \text{vs} \quad H_1 : \beta \neq \beta_0 \tag{2.2}$$

for a specific $\beta_0 \in R^p$. For instance $\beta_0 = 0$ which arises in the context of gene-set testing with H_0 indicating a particular set of genes to be insignificant.

2.1 F-test and its performances under high dimensionality

When the conditional distribution of Y_i given X_i is normally distributed, the conventional test for (2.2) is the F-test when $p < n-1$. The F-statistic is a monotone function of the likelihood ratio statistic and is distributed as a non-central F distribution under the alternative (Anderson, 2003). It is interesting to know the power implication on the F-test when $p/n \to \rho \in (0,1)$ when both p and n diverge to infinity.

Let $U = (\mathbf{1}, X)$ which is assumed to be of full rank and $A = (\mathbf{0}, I_p)$, where $\mathbf{1}$ denotes the n-dimensional vector of $1's$. Let $\gamma^T = (\alpha, \beta^T)$ and $\gamma_0^T = (\alpha, \beta_0^T)$, then the null hypothesis in (2.2) becomes $H_0 : A\gamma = A\gamma_0$. The F statistic for testing H_0 (Rao *et al.*, 2008, p51) is

$$
\begin{aligned}
G_{n,p} &= \frac{(\hat{\gamma}-\gamma_0)'A'(A(U'U)^{-1}A')^{-1}A(\hat{\gamma}-\gamma_0)/p}{Y'(I_n - P_U)Y/(n-p-1)} \\
&= \frac{(\hat{\beta}-\beta_0)'(A(U'U)^{-1}A')^{-1}(\hat{\beta}-\beta_0)/p}{Y'(I_n - P_U)Y/(n-p-1)} \qquad (2.3)
\end{aligned}
$$

where $\hat{\gamma} = (\hat{\alpha}, \hat{\beta}')' = (U'U)^{-1}U'Y$ is the least square estimator of γ and $Y = (Y_1, \cdots, Y_n)'$. Under H_0, $G_{n,p} \sim F_{p,\, n-p-1}$. Hence, an α-level F-test rejects H_0 if $G_{n,p} > F_{p,\, n-p-1;\alpha}$, the upper α quantile of the $F_{p,\, n-p-1}$ distribution.

In this paper, we use I_m to denote the $m \times m$ identity matrix and $\Phi(\cdot)$ as the distribution function of N(0,1). To facilitate our analysis, like Bai and Saranadasa (1996), we assume that

> There exists a m-variate random vector $Z_i = (Z_{i1}, \cdots, Z_{im})'$ for some $m \geq p$ so that $X_i = \Gamma Z_i + \mu$, where Γ is a $p \times m$ matrix such that $\Gamma\Gamma' = \Sigma$, and $E(Z_i) = 0$, $\mathrm{Var}(Z_i) = I_m$; each Z_{il} has finite 8-th moment, $E(Z_{il}^4) = 3 + \Delta$ for some constant Δ; for any $\sum_{\nu=1}^{d} \ell_\nu \leq 8$ and $i_1 \neq \cdots \neq i_d$, $E(Z_{1i_1}^{\ell_1} Z_{1i_2}^{\ell_2} \cdots Z_{1i_d}^{\ell_d}) = E(Z_{1i_1}^{\ell_1})E(Z_{1i_2}^{\ell_2}) \cdots E(Z_{1i_d}^{\ell_d})$. (2.4)

Model (2.4) resembles a factor model where the p-variate X is linearly generated by a m-variate factor Z. However, unlike the factor model which assumes far less number of factors than p so as to achieve a dimension reduction, we assume here the number of factors m is at least as larger as p. Model (2.4) slightly differs from the one assumed in Bai and Saranadasa (1996) in relaxing their assumption of Z_i having independent components. We also requires the existence of the 8-th moments for Z_i.

The power property of the F-test when $p/n \to \rho \in (0,1)$ is depicted in the following theorem.

Theorem 1 *Assume* $Y_i|X_i \sim N(X_i'\beta, \sigma^2)$*, Model (2.4),* $(\beta-\beta_0)'\Sigma(\beta-\beta_0) = o(1)$ *and* $\rho_n = p/n \to \rho \in (0,1)$ *as* $n \to \infty$ *then* $\Omega_F(\|\beta-\beta_0\|)$*, the power of the F-test, satisfies*

$$\Omega_F(\|\beta-\beta_0\|) - \Phi\left(-z_\alpha + \sqrt{\frac{(1-\rho)n}{2\rho}}(\beta-\beta_0)'\Sigma(\beta-\beta_0)\right) \to 0. \tag{2.5}$$

We notice that the denominator of the F statistic (2.3) estimates σ^2. When p is closer to n, there are fewer degrees of freedom left to estimate σ^2. The impact of the dimensionality on the F-test is revealed in Theorem 1 by $\sqrt{(1-\rho)/\rho}$ being a decreasing function of ρ. Hence, the power is adversely impacted by an increased dimension even $p < n-1$, reflecting a reduced degree of freedom in estimating σ^2 when the dimensionality is close to the sample size.

2.2 A new test statistic

We have seen two limitations with the F-test under mild dimensionality above. One is that p can not be larger than $n-1$; and the other is the conditional normality assumption. To test for regression coefficients in the "large p, small n" paradigm without the normality assumption, we modify the F-statistic in two aspects. One is to remove the denominator as it is a major contributor to F-test's fragile power performance under even mild dimensionality as shown in Theorem 1. Another is to renovate the numerator to make it more effective in measuring the discrepancy between β and β_0. We note that when $\alpha = 0$, $\|Y - X\beta_0\|^2$ is a measure between β and β_0, whose expectation is $(\beta-\beta_0)'E(X'X)(\beta-\beta_0)+n\sigma^2$. To avoid the $n\sigma^2$ term, we consider $(Y_i - X_i'\beta_0)(Y_j - X_j'\beta_0)$ for $i \neq j$ and a U-statistic with $X_i'X_j(Y_i - X_i'\beta_0)(Y_j - X_j'\beta_0)$ as the kernel. Our proposal here is similar to the effort made in improving the Wald type F-statistics as demonstrated in Brunner, Dette and Munk (1997) and Ahmad, Brunner and Werner (2008).

When the nuisance parameter $\alpha \neq 0$, to remove α, we consider a U-statistic

$$T_{n,p} = \frac{1}{P_n^4} \sum^{*} \phi(i_1, i_2, i_3, i_4), \tag{2.6}$$

$$\text{where} \quad \phi(i_1, i_2, i_3, i_4) = \frac{1}{4}(X_{i_1} - X_{i_2})'(X_{i_3} - X_{i_4})\Delta_{i_1,i_2}\Delta_{i_3,i_4} \tag{2.7}$$

and $\Delta_{i,j} = Y_i - Y_j - (X_i - X_j)'\beta_0$. Through this paper, we use $\sum^{*}$ to denote summations over distinct indeces. For example, in (2.6), the summation is over the set $\{i_1 \neq i_2 \neq i_3 \neq i_4$, for $i_1, i_2, i_3, i_4 \in \{1, \cdots, n\}\}$ and $P_n^m = n!/(n-m)!$. As $T_{n,p}$ is invariant to location shifts in both X_i and Y_i. We assume, without loss of generality, that $\alpha = \mu = 0$ in the rest of the paper.

The set of conditions we use to regulate for the "large p, small n" is

$$p(n) \to \infty \ \textit{as} \ n \to \infty, \ \Sigma > 0 \ \textit{and} \ tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}. \tag{2.8}$$

These conditions do not impose any explicit relative growth rates between p and n, and they are quite mild. Assuming Σ being positive definite assures the identification of the regression coefficient. We allow some eigenvalues of Σ diverge to infinity as $p \to \infty$. If all the eigen-values are bounded, the last part of (2.8) is trivially true for any p.

3. U-STATISTICS UNDER HIGH DIMENSIONALITY

As $T_{n,p}$ is a U-statistic, we devote this section to discuss U-statistics for high dimensional data. The theory of U-statistics for fixed dimensional data, as pioneered by Hoeffding (1948), has been well documented; see Serfling (1980) and Lee (1990) for summaries. We will demonstrate below that, while some results in the classical U-statistic remain valid, others may not be directly applicable if p diverges.

Suppose $W_1, W_2, \cdots, W_n$ are independent and identically distributed observations from a distribution F on R^q, where q may diverge. Consider a U-statistic of s-th order for a fixed $s < n$

$$U_{n,q} = \frac{1}{\binom{n}{s}} \sum_{C_{n,s}} h(W_{i_1}, \cdots, W_{i_s}),$$

where $C_{n,s} = \{$all distinct combinations of $\{i_1, i_2, \cdots, i_s\}$ from $\{1, \cdots, n\}\}$. The kernel h is symmetric so that its value is invariant to the permutations of its arguments. Let $E\{h(W_1, \cdots, W_s)\} = \theta(F)$, say. In our current testing problem, $q = p+1$, $s = 4$ and $\theta(F) = \|\Sigma(\beta - \beta_0)\|^2$.

Let $h_c(w_1, \cdots, w_c) = E\{h(w_1, \cdots, w_c, W_{c+1}, \cdots, W_s)\}$ be projections of h to lower dimensional sample spaces, $\tilde{h} = h - \theta(F)$ and $\tilde{h}_c = h_c - \theta(F)$ for $c = 1, \cdots, s$. Let $g_c(w_1, \cdots, w_c) = \tilde{h}_c - \sum_{j=1}^{c-1} \sum_{1 \le i_1 < \cdots < i_j \le c} g_j(w_{i_1}, \cdots, w_{i_j})$ where $g_1(w_1) = \tilde{h}_1(w_1)$, and

$$M_{nc} = \sum_{1 \le i_1 < \cdots < i_c \le n} g_c(w_{i_1}, \cdots, w_{i_c}).$$

The following theorem provides the Hoeffding decompositions (Hoeffding, 1948) for $U_{n,q}$ and its variance respectively, which are valid regardless of q being fixed or diverging.

Proposition 1 *Assume* $E\{h^2(W_1, \cdots, W_s)\}$ *exist and let* $\zeta_c = \mathrm{Var(h_c)}$ *for* $c = 1, 2, \cdots, s$. *Then (i)* $\zeta_{c+1} \ge \zeta_c$; *(ii)*

$$U_{n,q} - \theta(F) = \sum_{c=1}^{s} \binom{s}{c} \binom{n}{c}^{-1} M_{nc} \tag{3.1}$$

and (iii)

$$\mathrm{Var(U_{n,q})} = \binom{\mathrm{n}}{\mathrm{s}}^{-1} \sum_{\mathrm{c=1}}^{\mathrm{s}} \binom{\mathrm{s}}{\mathrm{c}} \binom{\mathrm{n-s}}{\mathrm{s-c}} \zeta_{\mathrm{c}}. \tag{3.2}$$

The proof in Hoeffding (1948)(see also Serfling, 1980) is applicable even when q is increasing to infinity. Specifically, the result in (i) is implied by $E\{h_{c+1}(w_1, \cdots, w_c, W_{c+1})\} = h_c(w_1, \cdots, w_c)$ and

$$\zeta_{c+1} = E\{\mathrm{Var(h_{c+1}(W_1, \cdots, W_{c+1})|W_1, \cdots, W_c)}\} + \zeta_c.$$

The variance decomposition for the variance in (3.2) reflects the decomposition of the U-statistic in (3.1) as $\{M_{nc}, \mathscr{F}_c\}_{c \ge 1}$ forms a forward martingale where $\mathscr{F}_c$ denotes the σ-field generated by $\{W_1, \ldots, W_c\}$ and $\mathrm{Var(M_{nc}) = O(\zeta_c)}$.

When $q \to \infty$, unlike the fixed dimension cases, ζ_c may no longer be bounded and can diverge. This brings ambiguity in assessing the relative orders of terms in the decomposition (3.1). To appreciate this point, we note that if q is fixed, all ζ_c are bounded provided $\zeta_s < \infty$, hence the $(c+1)$-th term in the variance decomposition (3.2) is a smaller order of the c-th term. This means that the asymptotic behavior of the U-statistic is determined by the c-th term where c is the smallest integer such that $\zeta_c \neq 0$. However, if q diverges, ζ_c may diverge and a higher order projection $M_{n(c+1)}$ may be at the same order or higher than M_{nc}. Hence, for high dimensional data, the leading order terms of the U-statistics may consist of multiple terms.

As ζ_c is monotone non-decreasing, the following strategy may be applied to determine the dominant terms of $U_{n,q}$. We can start evaluating ζ_cs from the two ends, namely ζ_1 and ζ_s. If ζ_1 and ζ_s are of the same order, then $U_{n,q}$ will be dominated by the first term so that

$$U_{n,q} - \theta(F) = \binom{s}{1}\binom{n}{1}^{-1} M_{n1}\{1 + o_p(1)\}.$$

If ζ_s and ζ_1 are not the same order, but ζ_2 and ζ_s are, then $U_{n,q}$ will be dominated by the first two terms so that

$$U_{n,q} - \theta(F) = \sum_{c=1}^{2} \binom{s}{c}\binom{n}{c}^{-1} M_{nc}\{1 + o_p(1)\}.$$

This process can be continued until the dominating terms are found. We will employ this strategy on the proposed test statistic $T_{n,p}$ in the next section.

4. MAIN RESULTS

We first symmetrize ϕ defined in (2.7) by

$$h(W_i, W_j, W_k, W_l) = \tfrac{1}{3}\{\phi(i,j,k,l) + \phi(i,k,j,l) + \phi(i,l,j,k)\}$$

where $W_i = (X_i^{\tau}, \varepsilon_i)^{\tau}$ and $\varepsilon_i = Y_i - X_i'\beta_0$. Then,

$$T_{n,p} = \frac{1}{\binom{n}{4}} \sum_{C_{n,4}} h(W_i, W_j, W_k, W_l). \tag{4.1}$$

It can be shown that the projections of h are, respectively,

$$\begin{aligned} h_1(w_1) &= \tfrac{1}{2}(\beta-\beta_0)'(x_1x_1' + \Sigma)\Sigma(\beta-\beta_0) + \tfrac{1}{2}\varepsilon_1 x_1'\Sigma(\beta-\beta_0), \\ h_2(w_1, w_2) &= \tfrac{1}{6}\Big\{(\beta-\beta_0)'(x_1-x_2)(x_1-x_2)'\Sigma(\beta-\beta_0) \\ &+ (\varepsilon_1-\varepsilon_2)(x_1-x_2)'\Sigma(\beta-\beta_0) \\ &+ ((\beta-\beta_0)'(x_1x_1'+\Sigma) + \varepsilon_1 x_1')(\varepsilon_2 x_2 + (x_2x_2'+\Sigma)(\beta-\beta_0))\Big\} \end{aligned}$$

and

$$
\begin{aligned}
& h_3(w_1, w_2, w_3) \\
= & \tfrac{1}{12}\{(x_1 - x_2)'(\beta - \beta_0) + (\varepsilon_1 - \varepsilon_2)\}(x_1 - x_2)'\{(x_3 x_3' + \Sigma)(\beta - \beta_0) + x_3\varepsilon_3\} \\
& + \tfrac{1}{12}\{(x_1 - x_3)'(\beta - \beta_0) + (\varepsilon_1 - \varepsilon_3)\}(x_1 - x_3)'\{(x_2 x_2' + \Sigma)(\beta - \beta_0) + x_2\varepsilon_2\} \\
& + \tfrac{1}{12}\{(x_2 - x_3)'(\beta - \beta_0) + (\varepsilon_2 - \varepsilon_3)\}(x_2 - x_3)'\{(x_1 x_1' + \Sigma)(\beta - \beta_0) + x_1\varepsilon_1\}.
\end{aligned}
$$

Let $B_i = (\beta - \beta_0)'\Sigma^i(\beta - \beta_0)$ for $i = 1, 2, 3$, $A_0 = \Gamma'\Gamma$, $A_1 = \Gamma'(\beta - \beta_0)(\beta - \beta_0)'\Gamma$, $A_2 = \Gamma'\Sigma(\beta - \beta_0)(\beta - \beta_0)'\Sigma\Gamma$ and $A_3 = \Gamma'\Sigma\Gamma$. Derivations given in the Appendix show that $\zeta_1 = \frac{1}{4}\zeta_1^*$ and $\zeta_2 = \frac{1}{36}\zeta_2^*$ where

$$
\begin{aligned}
\zeta_1^* &= (B_1 + \sigma^2)B_3 + B_2^2 + \Delta tr(A_1 \circ A_2) \quad \text{and} \\
\zeta_2^* &= \sigma^4 tr(\Sigma^2) + 21B_2^2 + 22B_1B_3 + 22\sigma^2 B_3 + B_1^2 tr(\Sigma^2) + 2\sigma^2 tr(\Sigma^2)B_1 \\
&+ 2\Delta(B_1 + \sigma^2)tr(A_1 \circ A_3) + 20\Delta tr(A_1 \circ A_2) + \Delta^2 tr\{(A_0 \text{diag}(A_1))^2\},
\end{aligned}
$$

where $C \circ B = (c_{ij}b_{ij})$ for matrices $C = (c_{ij})$ and $B = (b_{ij})$, and $\text{diag}(A) = \text{diag}\{a_{11}, \cdots, a_{mm}\}$ for $A = (a_{ij})_{m\times m}$. The proof of the following theorem in the Appendix shows that $\{\zeta_c\}_{c=2}^4$ are of the same order. This means that the test statistic is dominated by the first two terms corresponding M_{n1} and M_{n2}.

Theorem 2 *Under Model (2.4) and as $n \to \infty$,*
(i) $E(T_{n,p}) = \|\Sigma(\beta - \beta_0)\|^2$ and $\mathrm{Var}(T_{n,p}) = \{\frac{4}{n}\zeta_1^* + \frac{2}{n(n-1)}\zeta_2^*\}\{1 + o(1)\}$;
(ii) $T_{n,p} - \|\Sigma(\beta - \beta_0)\|^2 = \{\frac{4^2}{n}M_{n1} + \frac{2\times 6^2}{n(n-1)}M_{n2}\}\{1 + o_p(1)\}$, where $E(M_{n1}^2) = \zeta_1$ and $E(M_{n2}^2) = \zeta_2 - 2\zeta_1$.

Under $H_0 : \beta = \beta_0$, $A_1 = A_2 = B_i = 0$ for $i = 1, 2, 3$. Thus, $\zeta_1 = 0$ and $T_{n,p}$ is a degenerate U-statistic dominated by M_{n2}. In this case,

$$
\mathrm{Var}(T_{n,p}) = \frac{2}{n(n-1)}\sigma^4 tr(\Sigma^2)\{1 + o(1)\}.
$$

This form of the variance for $T_{n,p}$ is also valid under a subclass of H_1 specified by

$$
\begin{aligned}
&(\beta - \beta_0)'\Sigma(\beta - \beta_0) = o(1) \quad \text{and} &\quad (4.2) \\
&(\beta - \beta_0)'\Sigma(\beta - \beta_0)(\beta - \beta_0)'\Sigma^3(\beta - \beta_0) = o\{n^{-1}tr(\Sigma^2)\}.
\end{aligned}
$$

As this subclass prescribes a smaller difference between β and β_0, we call it the local alternatives. Under the local alternatives, $\zeta_1 = o(n^{-1}\zeta_2)$ which means like the case under H_0, M_{n2} is also the dominating term while M_{n1} is of smaller order.

Theorem 3 *Assume Model (2.4) and Condition (2.8), then under either H_0 or the local alternatives (4.2), as $n \to \infty$,*

$$\frac{n}{\sigma^2\sqrt{2tr(\Sigma^2)}}(T_{n,p} - \|\Sigma(\beta - \beta_0)\|^2) \xrightarrow{d} N(0,1). \tag{4.3}$$

To formulate a test procedure based on $T_{n,p}$, we need to estimate $tr(\Sigma^2)$ and σ^2 appeared in the asymptotic variance. We will use the estimator of $tr(\Sigma^2)$ proposed in Chen, Zhang and Zhong (2010). Specifically, let $Y_{1n} = \frac{1}{P_n^2}\sum^{*}(X_{i_1}'X_{i_2})^2$, $Y_{2n} = \frac{1}{P_n^3}\sum^{*}X_{i_1}'X_{i_2}X_{i_2}'X_{i_3}$ and $Y_{3n} = \frac{1}{P_n^4}\sum^{*}X_{i_1}'X_{i_2}X_{i_3}'X_{i_4}$. Then an unbiased and ratio consistent estimator of $tr(\Sigma^2)$ is

$$\widehat{tr(\Sigma^2)} = Y_{1n} - 2Y_{2n} + Y_{3n}.$$

We note here that a closely related estimator, that only employs $Y_{1,n}$, has been proposed in Ahmad, Werner and Brunner (2008) for normally distributed X_i with zero mean. The estimator of σ^2 under H_0 is

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - X_i\beta_0 - \bar{Y} + \bar{X}\beta_0)^2. \tag{4.4}$$

Applying Theorem 3 and the Slutsky Theorem, the proposed test rejects H_0 at a significant level α if

$$nT_{n,p} \geq \sqrt{2\widehat{tr(\Sigma^2)}}\hat{\sigma}^2 z_\alpha, \tag{4.5}$$

where z_α is the upper-α quantile of $N(0,1)$.

Theorem 3 also implies that $\Omega_L(\|\beta - \beta_0\|)$, the asymptotic power of the proposed test under the local alternatives is

$$\Omega_L(\|\beta - \beta_0\|) \doteq \Phi\left(-z_\alpha + \frac{n\|\Sigma(\beta - \beta_0)\|^2}{\sqrt{2tr(\Sigma^2)}\sigma^2}\right). \tag{4.6}$$

The power is largely impacted by $\eta_n(\beta-\beta_0, \Sigma, \sigma^2) = n\|\Sigma(\beta-\beta_0)\|^2/\{\sqrt{2tr(\Sigma^2)}\sigma^2\}$, which may be viewed as a signal to noise ratio (SNR). In particular, the power converges to α if $\eta_n(\beta - \beta_0, \Sigma, \sigma^2) = o(1)$ which means that the test can not distinguish H_0 from the local alternative in this case. If it is of a larger order of 1, the power converges to 1, indicating consistency of the test.

Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ be the eigenvalues of Σ. Then, a sufficient condition for the test to have a non-trivial power is $\|\beta - \beta_0\| = O(n^{-1/2} S_\lambda^{1/4} \lambda_1^{-1})$ where $S_\lambda = \sum_{i=1}^p \lambda_i^2$. Suppose all the eigenvalues are bounded from zero and infinity, let $\delta_\beta = \|\beta - \beta_0\|/\sqrt{p}$ define "signal strength", then the test has non-trivial power if δ_β is of order $n^{-1/2}p^{-1/4}$. This is a smaller order than $n^{-1/2}$, the corresponding "signal" strength for the fixed dimensional case.

We can also evaluate power of the proposed test under other scenarios of H_1 such that

$$(\beta - \beta_0)'\Sigma(\beta - \beta_0) \text{ is not o}(1) \tag{4.7}$$

violating the first part of (4.2) in the specification of the local alternatives. We will demonstrate in the Appendix that under two situations of (4.7), the proposed test can achieve at least 50% power.

5. GENERALIZATION TO FACTORIAL DESIGNS

So far we have assumed that $\{(X_i, Y_i)\}_{i=1}^n$ is a simple random sample. However, in many scientific studies, observations are obtained via certain designs of experiments. For example, a randomized factorial design was used in a microarray study that we will analyze in the next section. In this section, we provide an extension of the proposed high dimensional regression test to accommodate factorial designs.

For ease of expedition, we will concentrate on two way factorial designs with two factors A and B, where A has I levels and B has J levels. Let c indicate a cell for $c = 1, \cdots, IJ$, which has n_c observations in the cell. The observations (X'_{ijk}, Y_{ijk}) in the $i-$th level of A and $j-$th level of B satisfy a linear model

$$E(Y_{ijk}|X_{ijk}) = \alpha_0 + \gamma_i + \theta_j + \gamma\theta_{ij} + X'_{ijk}\beta, \quad k = 1, \cdots, n_c, \tag{5.1}$$

where γ_i represent for the effect of A, θ_j for that of B, and $\gamma\theta_{ij}$ for their interactions. These effects could be either random effects or fixed effects. Our purpose in this section is to generalize the test given in Section 4 for

$$H_0 : \beta = \beta_0 \quad \text{vs} \quad H_1 : \beta \neq \beta_0 \tag{5.2}$$

for Model (5.1) while treating $(\alpha_0, \gamma_i, \theta_j, \gamma\theta_{ij})$ as nuisance parameters.

Let $\mu_{ij} = \alpha_0 + \gamma_i + \theta_j + \gamma\theta_{ij}$. Model (5.1) can be written as

$$E(Y_{ijk}|X_{ijk}) = \mu_{ij} + X_{ijk}'\beta, \quad k = 1, \cdots, n_c. \tag{5.3}$$

Define $Y = (Y^{1'}, \cdots, Y^{IJ'})'$, $X = (X^{1'}, X^{2'}, \cdots, X^{IJ'})'$ where

$$X^c = (X_{ij1}, \cdots, X_{ijn_c})' := (X_{c1}, \cdots, X_{cn_c})'$$

and $Y^c = (Y_{ij1}, \cdots, Y_{ijn_c})' := (Y_{c1}, \cdots, Y_{cn_c})'$ for $c = (i-1)J + j$. Then,

$$E(Y|X) = D\alpha + X\beta, \tag{5.4}$$

where $D = \mathrm{I}_{IJ} \otimes \mathbf{1}_{n_c}$ is the design matrix, α corresponding to the cell means parameters μ_{ij}. Multiply $I - P_D$ on both sides of (5.4) where $P_D = D(D'D)^{-}D' = \mathrm{I}_{IJ} \otimes n_c^{-1}\mathbf{1}_{n_c}\mathbf{1}_{n_c}'$ is the projection matrix of D, we have

$$E\{(I - P_D)Y|X\} = (I - P_D)X\beta,$$

where we eliminate the nuisance parameters α in (5.4). So a natural generalization of $T_{n,p}$ to the factorial design is

$$T_{n,p} = \frac{1}{IJ}\sum_{c=1}^{IJ}(P_{n_c}^4)^{-1}\sum^{*}\phi(i,j,k,l), \tag{5.5}$$

where $\phi(i,j,k,l) = \frac{1}{4}(X_{ci} - X_{cj})'(X_{ck} - X_{cl})\Delta(i,j)\Delta(k,l)$, $\Delta(i,j) = \{Y_{ci} - Y_{cj} - (X_{ci} - X_{cj})'\beta_0\}$, and the second summation is over distinct observations in the c-th cell.

As an extension to Model (2.4), we assume in each cell

$$X_{ci} = \Gamma_c Z_{ci} + \mu_c, \tag{5.6}$$

where Γ_c is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_c\Gamma_c' = \Sigma_c = \mathrm{Var}(X_{ijk})$ for $c = (i-1)J + j$, and Z_{ci} are independent and identically distributed random vectors having the same qualifications as in Model (2.4). An extension of Condition (2.8) is

$$p(n_c) \to \infty \textit{ as } \min_c n_c \to \infty,\ \Sigma_c > 0 \textit{ and } tr(\Sigma_c^4) = o\{tr^2(\Sigma_c^2)\}. \tag{5.7}$$

For $c = 1, \cdots, IJ$, the factorial design version of the local alternative hypothesis (4.2) is

$$(\beta - \beta_0)'\Sigma_c(\beta - \beta_0) = o(1) \text{ and} \tag{5.8}$$
$$(\beta - \beta_0)'\Sigma_c(\beta - \beta_0)(\beta - \beta_0)'\Sigma_c^3(\beta - \beta_0) = o\{n_c^{-1} tr(\Sigma_c^2)\}.$$

The following corollary can be readily established by modifying the proof of Theorem 3.

Corollary 1 *Assume Model (5.6) and assumption (5.7), then under either H_0 or (5.8),*

$$\sigma_{fac,0}^{-1}\Big(T_{n,p} - \tfrac{1}{IJ}\textstyle\sum_{c=1}^{IJ} \|\Sigma_c(\beta - \beta_0)\|^2\Big) \xrightarrow{d} N(0,1), \tag{5.9}$$

where $\sigma_{fac,0}^2 = \frac{2\sigma^4}{(IJ)^2}\sum_{c=1}^{IJ} tr(\Sigma_c^2)/\{n_c(n_c-1)\}$.

Let $\widehat{tr(\Sigma_c^2)}$ be the analog of the $tr(\Sigma^2)$ estimator given in (4.4) and $\hat\sigma^2 = \frac{1}{IJ}\sum_{i,j}\frac{1}{n_c-1}\sum_{k=1}^{n_c}(Y_{ijk} - X'_{ijk}\beta_0 - \bar Y_{ij\cdot} + \bar X'_{ij\cdot}\beta_0)^2$, where $\bar Y_{ij\cdot} = \frac{1}{n_c}\sum_{k=1}^{n_c} Y_{ijk}$ and $\bar X_{ij\cdot} = \frac{1}{n_c}\sum_{k=1}^{n_c} X_{ijk}$. Then, an α-level test for the factorial design rejects H_0 if

$$T_{n,p} \geq \frac{\hat\sigma^2 z_\alpha}{(IJ)}\{2\sum_{c=1}^{IJ}\widehat{tr(\Sigma_c^2)}/\{n_c(n_c-1)\}\}^{1/2}.$$

Similar to our analysis in the Appendix for the simple random design, we can also evaluate the power of the test for two fixed alternatives under

$$(\beta - \beta_0)'\Sigma_c(\beta - \beta_0) \text{ is not } o(1) \text{ for any } c. \tag{5.10}$$

This evaluation is given in a longer version of this paper.

6. SIMULATION STUDY

We conducted numerical simulations to evaluate the finite sample performance of the proposed tests under both simple random and factorial designs. For comparison purposes, we also carried out simulation for the F-test and an Empirical Bayes (EB) test proposed by Goeman *et al.* (2009). The empirical Bayes test is formulated via a score test on the hyper-parameter of a prior distribution assumed on the regression coefficients. As it allows $p > n$, it is applicable for high dimensional data.

The first set of simulations were designed to evaluate the performance of the test for the linear regression model with the simple random designs:

$$Y_i = \alpha + X_i^{'}\beta + \varepsilon_i, \tag{6.1}$$

where $\mathrm{Var}(\varepsilon_i) = \sigma^2 = 4$. Two distributions were experimented for ε_i. One was $N(0,4)$; the other was a centralized gamma distribution with the shape parameter 1 and the scale parameter 0.5. The hypotheses to be tested were

$$H_0 : \beta = \mathbf{0}_{p\times 1} \quad vs \quad H_1 : \beta \neq \mathbf{0}_{p\times 1}.$$

Independent and identically distributed covariates $X_1, \cdots, X_n$ with $X_i = (X_{i1}, \cdots, X_{ip})'$ were generated according to a moving average model

$$X_{ij} = \rho_1 Z_{ij} + \rho_2 Z_{i(j+1)} + \cdots + \rho_T Z_{i(j+T-1)} + \mu_j, \quad j = 1, \cdots, p; \tag{6.2}$$

for some $T < p$. Here $Z_i = (Z_{i1}, \ldots, Z_{i(p+T-1)})'$ is a $(p+T-1)$-dimensional $N(0, I_{p+T-1})$ random vector, $\{\mu_j\}_{j=1}^{p}$ were fixed constants generated from the Uniform (2,3) distribution. The coefficients $\{\rho_l\}_{l=1}^{T}$ were generated independently from the Uniform $(0,1)$ distribution and were kept fixed once generated. Model (6.2) implied that $\Sigma = \left(\sum_{k=1}^{T-|j-l|} \rho_k \rho_{k+|j-l|} I\{|j-l| < T\}\right)$. Hence the correlation among X_{ij} and X_{il} were determined by $|j-l|$ and T. We chose two values of T, 10 and 20, to generate different levels of dependence. The auto-correlation functions for model (6.2) are displayed in Figure 1.

Two configurations of the alternative hypothesis H_1 were experimented. One allocated half of the β-components of equal magnitude to be non-zeros, the so-called the "non-sparse case". The other has only five non-zero components of equal magnitude, the so-called "sparse case". In both cases, we fixed $\|\beta\|^2$ at three levels: $0.02, 0.04$ and 0.06. To gain information on the performance of the proposed test, we consider two settings regarding p and n. One is $p < n$, which allowed F-test; and the other one is $p >> n$. In the first setting, we set $\rho_n = p/n = (0.85, 0.90, 0.95)$, where $p = 34, 54, 76$ and $n = 40, 60, 80$ respectively. For the setting of $p >> n$, we chose $p = 310, 400$ and 550, which was increased exponentially, according to $p = \exp(n^{0.4}) + 230$ for $n = 40, 60, 80$ respectively.

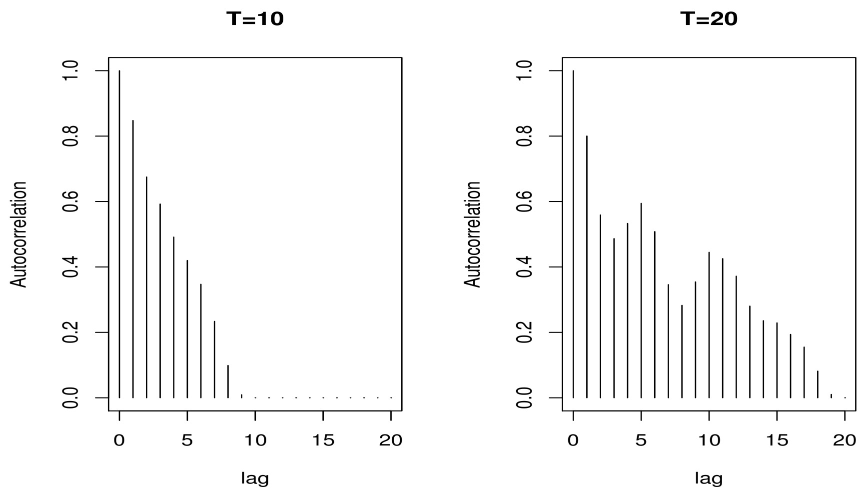

Figure 1: The auto-correlation functions for series $\{X_{ij}\}_{j=1}^{p}$.

Tables 1 and 2 summarize the empirical sizes and powers of the proposed tests as well as those for the F-tests and EB tests with the normally and the centralized gamma distributed residuals for $p < n$. The empirical sizes of the proposed tests, EB tests and the F-tests were quite reasonably around 0.05. We find that the proposed tests consistently outperformed the EB and the F-tests for both normally and gamma distributed residuals, for different levels of dependence (T=10 or 20), and for both the sparse and the non-sparse settings. In particular, in the sparse setting, although there were some reduction of power for all three tests, the power reduction in the F-test was the most significant. The empirical power of the proposed test was quite responsive to the signal to the noise ratio (SNR), which is $n\|\Sigma(\beta-\beta_0)\|^2/\{\sqrt{2tr(\Sigma^2)}\sigma^2\}$, in all the settings. We also computed the theoretical power (reported in a longer version of the paper) given in (4.6) derived from Theorem 3 under the so-called local alternatives. It was found that there was a good agreement between the empirical power and the theoretical power when the SNR was relatively small. This makes sense as a small SNR is much in tune with the local alternatives.

Table 3 and 4 report the empirical powers and sizes of the proposed tests and the EB tests when p were much larger than n, which makes F-test unapplicable. We observe that the sizes of the proposed tests became closer to the nominal level 0.05 than Table 1 and 2. This is also confirmed by the null distributions plots in Figure 2. The power of the proposed test were increased quite rapidly as the SNR was increased. In contrast, the EB test suffered from rather severe size distortion for all cases considered. At the meanwhile, the power of the EB test endured very low power when $T = 10$. This alarming performance may be due to the fact that its justification as in Goeman *et al.* (2009) was made for p being fixed while $n \to \infty$.

Considering that the proposed test is an asymptotic test, we plotted in Figure 2 the kernel density estimates for the standardized test statistics of proposed test under H_0 for $T = 10$ and compared them with the standard normal distribution. It shows that the null distribution was quite closer to that of $N(0,1)$, which confirmed the asymptotic null distribution of the standardized test statistic given in Theorem 3. There was some right skewness when p is less than n. However, as p was increased, this skewness was largely reduced when p was increased.

The second set of the simulations were designed to understand performance of the proposed test under the factorial designs. We simulated a two-factor balanced design with two levels for each factor:

$$Y_{ijk} = \alpha_{ij} + X'_{ijk}\beta + \varepsilon_{ijk}, \quad k = 1, 2, \cdots, n_c \tag{6.3}$$

where $c = 2(i-1) + j$ and $i, j = 1, 2$, corresponding to (i, j)-th cell and the parameters $(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}) = (1, 3, 3, 4)$. The sparsity set-ups for β were the same to those for simple random designs used in (6.1). Within each cell, independent and identically distributed p-dimensional X_{ijk} were generated from the moving average model (6.2) with $T = T_c$, where T_c equals to $10, 15, 20$ and 25 for $c = 1, 2, 3, 4$ respectively. Using the different T values was to generate different dependence structure in Σ. We assigned the $n_c = 20$ and 30 in all cells, and three values of p: $100, 150$ and 200. The simulation results for the proposed test are summarized in Table 5. We observe that the sizes of the proposed test were satisfactorily around 0.05. The power of the test increased as the SNR_f, the factorial design version of SNR, was increased. When the sample size was increased from 20 to 30, we observed significant increase in the power under all

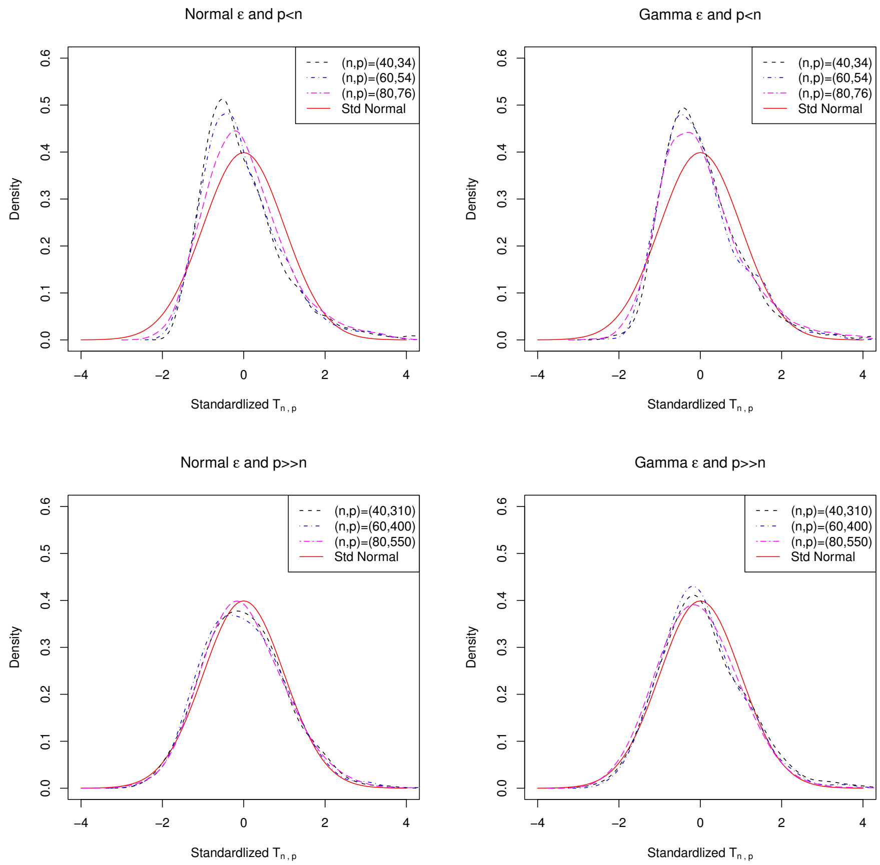

Figure 2: The null distributions of standardized $T_{n,p}$.

settings.

7. ASSOCIATION TEST FOR GENE-SETS

We applied the proposed test for association between gene-sets and certain clinical outcomes in a randomized factorial design experiment applied to 24 six-month-old Yolkshire gilts. The gilts were genotyped according to the melanocortin-4 receptor gene, 12 of them with D298 and the other with N298. Two diet treatments were randomly assigned to the 12 gilts in each genotype. One treatment is ad libitum (no restrictions) in the amount of feed consumed; the other is fasting. More details of the experiment could be found at Lkhagvadorj *et al.*(2009). The genotypes and the diet treatments were the two factors in the factorial experiments. The purpose of our study was to identify associations between gene-sets and triiodothyronine (T_3) measurement, a vital thyroid hormone that increases the metabolic rate, protein synthesis and stimulates breakdown of cholesterol.

The gene expression values were obtained for 24,123 genes in liver and adipose tissues, as well as measurements of T_3 in the blood on each gilt. Gene sets are defined by Gene Ontology (GO term) (The Gene Ontology Consortium, 2000), which classifies genes into different sets according to their biological functions among three broad categories: cellular component, molecular function and biological process. The data-set contained 6176 GO terms. Our objective is to find the GO terms which are significantly correlated with T_3 after accounting for the design factors.

Let i, j, k be indices for treatment, genotype and observations, respectively. For instance, Y_{ijk} denote the T_3 measurement for the $k-$th gilt in the $i-$th treatment with $j-$th genotype, and X^g_{ijk} be the corresponding p_g-dimension gene expressions for the $g-$th GO term. We consider the following four models corresponding to four types of designs:

$$
\begin{aligned}
&\text{Design I:} && Y_k = \alpha + X_k^{g'} \beta^g + \varepsilon_k^g, \quad k = 1, \cdots, 24;\\
&\text{Design II:} && Y_{ik} = \alpha + \mu_i + X_{ik}^{g'} \beta^g + \varepsilon_{ik}^g, \quad k = 1, \cdots, 12;\\
&\text{Design III:} && Y_{jk} = \alpha + \tau_j + X_{jk}^{g'} \beta^g + \varepsilon_{jk}^g, \quad k = 1, \cdots, 12;\\
&\text{Design IV:} && Y_{ijk} = \alpha + \mu_i + \tau_j + \mu\tau_{ij} + X_{ijk}^{g'} \beta^g + \varepsilon_{ijk}^g, \quad k = 1, \cdots, 6
\end{aligned}
$$

for $i = 1, 2$, $j = 1, 2$ and $g = 1, \cdots, G$ where $G = 6176$ is the total number of the GO terms, μ_i stand for diet treatment effects, τ_j for genotype effects and $\mu\tau_{ij}$ represent the interaction between treatment and genotype. For each GO term, we tested for

$$H_0 : \beta^g = 0 \quad vs \quad H_1 : \beta^g \neq 0.$$

Among the 6176 GO terms, the dimension p_g of the gene-sets ranged from 1 to 5158, and many of the gene-sets shared common genes. Hence, there were both high dimensionality and multiplicity. We applied the proposed high dimensional test for $p_g \geq 5$ and the F-test for $p_g < 5$. Without confusion, we call this combination of the proposed high dimensional test and F-test as the proposed test in this Section. For comparison purposes, the Empirical Bayes test was also carried out.

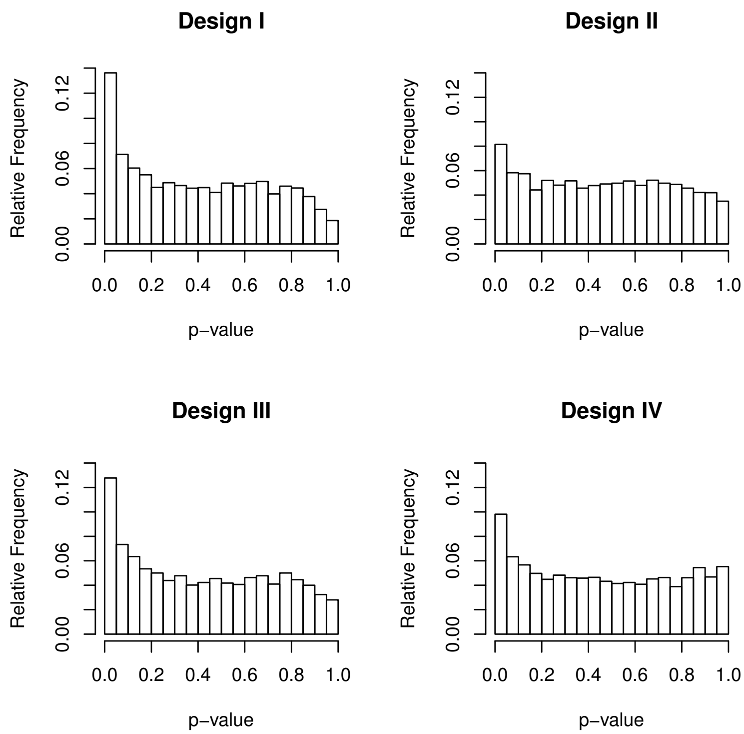

Figure 3: Histograms of the p-values on all GO terms using the proposed tests.

Figure 3 and 4 display histograms of p-values of the proposed tests and the EB tests under the four designs (I-IV) for all the gene-sets, respectively. Both Figures 3 and 4 show that the histograms for Designs I and III were very similar, so were the histograms of Designs II and IV. This was confirmed by Figure 5 where we plots the histograms for the differences in the p-values from the proposed tests. We observed that the p-values from Design I and III had higher portion of small p-values than those under Design II and IV. These features show that the form of design is important and it is necessary to account for different designs into the analysis.

By controlling the false discover rate (FDR) for the p-values from the proposed tests at 5%, 129, 23, 51 and 40 GO terms were declared statistically significant under designs I-IV respectively. We list in Table 6 significant GO terms identified by the proposed tests under at least three designs, together with their p-values and dimensions. They include GO terms that significant under all four

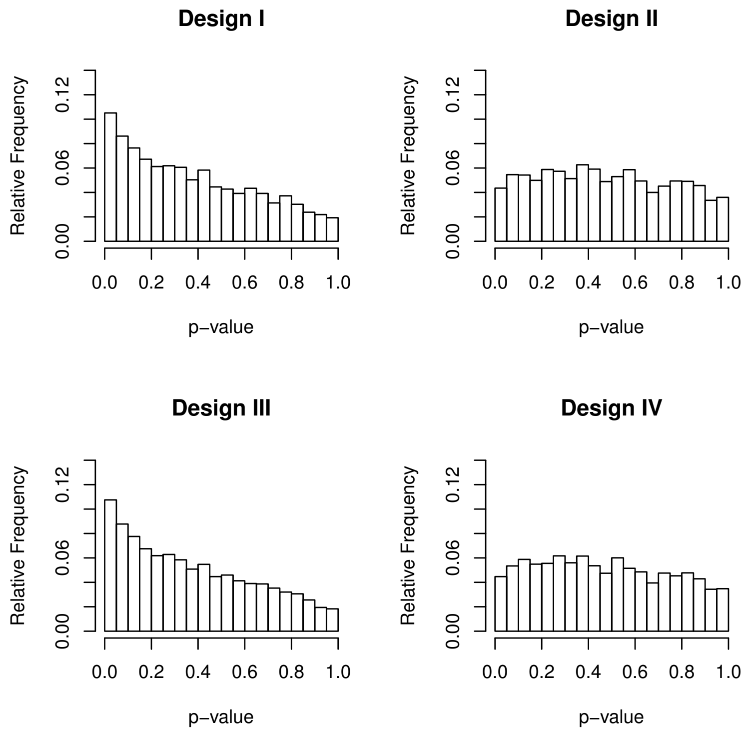

Figure 4: Histograms of the p-values on all GO terms
using Empirical Bayes (EB) tests.

designs: GO:0005086, GO:0007528 and GO:0032012. GO:0005086 is related to the molecular function, which stimulates the exchange of guanyl nucleotides associated with the GTPase ARF. GO:0007528 belongs to the biological process category. Its role in the progression of the neuromuscular junction over time, whose association with T_3 was discovered by other authors including Kawa and Obata (1982). GO:0032012 also belongs to the biological process, which was also found significant by the EB test.

The EB tests detected one significant GO term for each design: GO:0032012 for Designs I and III, and GO:0004731 for Designs II and IV. They were all among the significant GO terms discovered by the proposed tests. That the EB test detected quite few gene-sets is not entirely unexpected as our simulation has shown it tends to have relative low power.

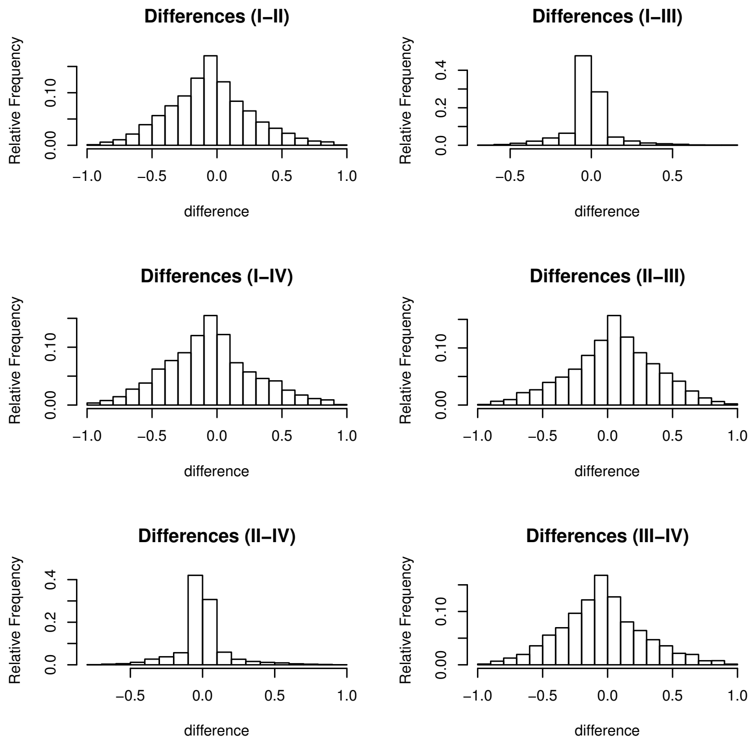

Figure 5: Differences in the p-values among Designs I-IV.

APPENDIX: TECHNICAL DETAILS.

In this appendix, we give technical proofs for the results we presented in Sections 2 and 4. We will use $\delta_\beta = \beta - \beta_0$ through the Appendix.

Proof of Theorem 1

Let $\gamma_0 = (\alpha, \beta_0^\tau)^\tau$. By plugging in the least square estimate $\hat{\gamma}$, we could write the F-statistics in (2.3) as

$$G_{n,p} = \frac{(Y - U\gamma_0)' P_{Au} (Y - U\gamma_0)/p}{Y'(I_n - P_U)Y/(n-p-1)}$$

where $P_{Au} = U(U'U)^{-1}A'(A(U'U)^{-1}A')^{-1}A(U'U)^{-1}U'$, $P_U = U(U'U)^{-1}U'$ and $P_{\mathbf{1}} = \mathbf{1}\mathbf{1}'/n$ be the projection matrices of $U(U'U)^{-1}A'$, U and $\mathbf{1}$ respectively. By applying the matrix inverse formula on $(U'U)^{-1}$, $U(U'U)^{-1}A' = (I - P_{\mathbf{1}})X\{X'(I - P_{\mathbf{1}})X\}^{-1}$. It then follows that $P_{Au} = (I - P_{\mathbf{1}})X(X'(I - P_{\mathbf{1}})X)^{-1}X'(I - P_{\mathbf{1}})$.

Since $P_{Au}(I - P_U) = 0$, the numerator and the denominator of $G_{n,p}$ are independent, and P_{Au} is an idempotent matrix with rank p. We may write

$$\frac{p}{n-p-1}G_{n,p} \stackrel{d}{=} \frac{\{Q\varepsilon + Q(U(\gamma-\gamma_0))\}' diag(\mathbf{1}_p', \mathbf{0}_{n-p}')\{Q\varepsilon + Q(U(\gamma-\gamma_0))\}}{\mathbf{z_1}'\mathbf{z_1}},$$

where $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)' \sim N(0, I_n)$ and $\mathbf{z_1} \sim N(0, I_{n-p-1})$ are independent random variables, and Q is an orthogonal matrix such that $P_{Au} = Q' diag(\mathbf{1}_p', \mathbf{0}_{n-p}')Q$. Here $\stackrel{d}{=}$ means the two random vectors on either side have the same distribution. Write $Q = (Q_1, Q_2, \cdots, Q_n)'$. Note that $Q\varepsilon \stackrel{d}{=} \varepsilon$. Furthermore, write $pG_{n,p}/(n-p-1)$ as

$$\frac{p}{n-p-1}G_{n,p} \stackrel{d}{=} \sum_{i=1}^{p}\{\varepsilon_i^2 + 2\varepsilon_i Q_i' X\delta_\beta\}/\mathbf{z_1}'\mathbf{z_1} + \delta_\beta' X' P_{Au} X\delta_\beta/\mathbf{z_1}'\mathbf{z_1} \quad \text{(A.1)}$$

where $X'P_{Au}X = X'(I-P_{\mathbf{1}})X = \Gamma Z'(I-P_{\mathbf{1}})Z\Gamma'$ and $Z = (Z_1, \cdots, Z_n)'$.

For the numerator of the (A.1), we can show that under Model (2.4), $E\{\delta_\beta' X' P_{Au} X\delta_\beta\} = (n-1)\delta_\beta'\Sigma\delta_\beta$. It is easy to see that $E\{\sum_{i=1}^p \varepsilon_i Q_i' X\delta_\beta\} = 0$ and

$$\text{Var}\{\sum_{i=1}^{p}\varepsilon_i Q_i' X\delta_\beta\} = (n-1)\sigma^2\delta_\beta'\Sigma\delta_\beta. \quad \text{(A.2)}$$

It can be shown that

$$\text{Var}\{\delta_\beta' X' P_{Au} X\delta_\beta\} = 2(n-1)(\delta_\beta'\Sigma\delta_\beta)^2 + (n+2+1/n)\Delta tr(A_1 \circ A_1). \quad \text{(A.3)}$$

Direct calculation shows that $E\big(\frac{1}{\mathbf{z_1}'\mathbf{z_1}}\big) = 1/(n-p-3)$ and $E\big(\frac{1}{\mathbf{z_1}'\mathbf{z_1}}\big)^2 = 1/\{(n-p-3)(n-p-5)\}$. (A.2) implies that $\sum_{i=1}^p \varepsilon_i Q_i' X\delta_\beta/\mathbf{z_1}'\mathbf{z_1} = O_p\{\frac{1}{\sqrt{n}}\sqrt{\delta_\beta'\Sigma\delta_\beta}\}$ and note that $E(X'P_{Au}X) = (n-1)\Sigma$. Then (A.3) yields

$$\frac{\delta_\beta' X' P_{Au} X\delta_\beta}{\mathbf{z_1}'\mathbf{z_1}} = \frac{\delta_\beta'\Sigma\delta_\beta}{1-\rho} + O_p\{\tfrac{1}{\sqrt{n}}\delta_\beta'\Sigma\delta_\beta\}.$$

If $\delta_\beta'\Sigma\delta_\beta = o(1)$, then

$$\frac{p}{n-p-1}G_{n,p} \stackrel{d}{=} \sum_{i=1}^{p}\frac{\varepsilon_i^2}{\mathbf{z_1}'\mathbf{z_1}} + \frac{\delta_\beta'\Sigma\delta_\beta}{1-\rho} + o_p(n^{-1/2}).$$

From Bai and Saranadasa (1996),

$$\frac{p}{n-p-1}F_{p,\, n-p-1;\alpha} = \frac{\rho_n}{1-\rho_n} + \sqrt{\frac{2\rho}{(1-\rho)^3 n}}z_\alpha + o(n^{-1/2}),$$

where z_α is the α quantile of N(0,1) and it can be shown

$$\sqrt{\frac{(1-\rho)^3 n}{2\rho}}\left(\sum_{i=1}^{p}\frac{\varepsilon_i^2}{\mathbf{z_1}'\mathbf{z_1}}-\frac{\rho_n}{1-\rho_n}\right)\xrightarrow{d} N(0,1).$$

Therefore the power of the F-test is

$$\Omega_F(\|\beta-\beta_0\|)=P\left(\frac{p}{n-p-1}G_{n,p}>\frac{p}{n-p-1}F_{p,\,n-p-1;\alpha}\right)$$
$$=P\Big\{\sqrt{\frac{(1-\rho)^3 n}{2\rho}}\left(\sum_{i=1}^{p}\frac{\varepsilon_i^2}{\mathbf{z_1}'\mathbf{z_1}}-\frac{\rho_n}{1-\rho_n}\right)>z_\alpha-\sqrt{\frac{(1-\rho)^3 n}{2\rho}}\frac{\delta_\beta'\Sigma\delta_\beta}{1-\rho}+o_p(1)\Big\}$$
$$=\Phi\left(-z_\alpha+\sqrt{\frac{(1-\rho)n}{2\rho}}\delta_\beta'\Sigma\delta_\beta\right)+o(1). \qquad \square$$

Proof of Theorem 2

It is straightforward to show that $E(T_{n,p})=\|\Sigma\delta_\beta\|^2$. To derive $\mathrm{Var}(\mathrm{T_{n,p}})$ we need to derive the variance of h_1,h_2,h_3 and h and then apply the variance decomposition given in (3.2).

Let $A_0=\Gamma'\Gamma$, $A_1=\Gamma'\delta_\beta\delta_\beta'\Gamma$, $A_2=\Gamma'\Sigma\delta_\beta\delta_\beta'\Sigma\Gamma$, $A_3=\Gamma'\Sigma\Gamma$ and $B_i=\delta_\beta'\Sigma^i\delta_\beta$. It can be shown that

$$\zeta_1=\frac{1}{4}B_1B_3+\frac{1}{4}\sigma^2B_3+\frac{1}{4}B_2^2+\frac{1}{4}\Delta tr(A_1\circ A_2). \tag{A.4}$$

We can also show that

$$\zeta_2=\frac{1}{36}\Big\{\sigma^4tr(\Sigma^2)+21B_2^2+22B_1B_3+22\sigma^2B_3+B_1^2tr(\Sigma^2)+2\sigma^2tr(\Sigma^2)B_1$$
$$+2\Delta\{B_1+\sigma^2\}tr(A_1\circ A_3)+20\Delta tr(A_1\circ A_2)+\Delta^2tr\{A_0diag(A_1)\}^2\Big\}, \tag{A.5}$$

As $\zeta_4\geq\zeta_3$, we first derive ζ_4. It may be shown that

$$\zeta_4=\tfrac{1}{2}\sigma^4tr(\Sigma^2)+B_2^2+\tfrac{8}{3}B_1B_3+\tfrac{8}{3}\sigma^2B_3+\tfrac{11}{24}B_1^2tr(\Sigma^2)+\sigma^2tr(\Sigma^2)B_1 \tag{A.6}$$
$$+\tfrac{11}{24}\{B_1+\sigma^2\}\Delta tr(A_1\circ A_3)+2\Delta tr(A_1\circ A_2)+\tfrac{1}{6}\Delta^2tr\{A_0diag(A_1)\}^2.$$

Note that (A.5) and (A.6) show that ζ_2 and ζ_4 are both the linear combination of $tr(\Sigma^2),B_2^2,B_1B_3,B_3,B_1^2tr(\Sigma^2),B_1tr(\Sigma^2),(B_1+\sigma^2)tr(A_1\circ A_3),tr(A_1\circ A_2)$ and $tr\{A_0diag(A_1)\}^2$. So it implies that ζ_2 and ζ_4 are of the same order. By Proposition 1, ζ_2, ζ_3 and ζ_4 are of the same order. Hence, the third and fourth term in the Hoeffding decomposition are all of smaller order. Thus

$\mathrm{Var}(\mathrm{T_{n,p}}) = \left\{\frac{16}{\mathrm{n}}\zeta_1 + \frac{72}{\mathrm{n(n-1)}}\zeta_2\right\}\{1 + \mathrm{o}(1)\}$. Substituting ζ_1 and ζ_2, the results in Theorem 2 follow. $\square$

The following two inequalities will be useful in the proof of Theorem 3. By the Cauchy-Schwarz inequality together with (A.4) and (A.5), we have

$$\zeta_1 \leq \{(\frac{1}{2} + \frac{1}{4}\Delta)B_1 + \frac{1}{4}\sigma^2\}B_3, \tag{A.7}$$

$$\zeta_2 \leq \frac{1}{36}\Big\{[\sigma^2 + (\Delta+1)B_1]^2 tr(\Sigma^2) + [22\sigma^2 + (43+20\Delta)B_1]B_3\Big\}. \tag{A.8}$$

Proof of Theorem 3

Let

$$\widehat{T}_{n,p} - \|\Sigma\delta_\beta\|^2 = \frac{12}{n(n-1)} \sum_{1\leq i_1<i_2\leq n} \tilde{h}_2(W_{i_1}, W_{i_2}) \tag{A.9}$$

be the projection of $T_{n,p}$. We can decompose $T_{n,p} - \|\Sigma\delta_\beta\|^2 = \widehat{T}_{n,p} - \|\Sigma\delta_\beta\|^2 + (T_{n,p} - \widehat{T}_{n,p})$, where $T_{n,p} - \widehat{T}_{n,p}$ can still be written as a U-statistics with kernel

$$H(W_1, W_2, W_3, W_4) = \tilde{h}(W_1, W_2, W_3, W_4) - \sum_{1\leq i_1<i_2\leq 4} \tilde{h}_2(W_{i_1}, W_{i_2}). \tag{A.10}$$

The projections of H are $H_1(w_1) = -2\tilde{h}_1(w_1)$, $H_2(w_1, w_2) = -2\sum_{i=1}^2 \tilde{h}_1(w_i)$ and $H_3(w_1, w_2, w_3) = \tilde{h}_3(w_1, w_2, w_3) - \sum_{i=1}^3 \tilde{h}_1(w_i) - \sum_{1\leq i<j\leq 3} \tilde{h}_2(w_i, w_j)$. Thus if the null hypothesis or the local alternatives conditions (4.2) hold, $\mathrm{Var(h_1)} = \mathrm{o(n^{-1}}\zeta_2)$. By Hoeffding's variance formula, $\mathrm{Var}(\widehat{\mathrm{T}}_{\mathrm{n,p}}) = \mathrm{O(n^{-2}}\zeta_2)$ and $\mathrm{Var}(\mathrm{T_{n,p}} - \widehat{\mathrm{T}}_{\mathrm{n,p}}) = \mathrm{o(n^{-2}}\zeta_2)$. Here we used the fact that ζ_2, ζ_3 and ζ_4 are of the same order as we have shown in Theorem 2. Thus,

$$\frac{T_{n,p} - \|\Sigma\delta_\beta\|^2}{\sqrt{\mathrm{Var}(\widehat{\mathrm{T}}_{\mathrm{n,p}})}} = \frac{\widehat{T}_{n,p} - \|\Sigma\delta_\beta\|^2}{\sqrt{\mathrm{Var}(\widehat{\mathrm{T}}_{\mathrm{n,p}})}} + o_p(1).$$

Hence we only need to show that

$$\frac{\widehat{T}_{n,p} - \|\Sigma\delta_\beta\|^2}{\sqrt{\mathrm{Var}(\widehat{\mathrm{T}}_{\mathrm{n,p}})}} \xrightarrow{d} N(0,1). \tag{A.11}$$

From (A.9), $\widehat{T}_{n,p} - \|\Sigma\delta_\beta\|^2 = \widehat{T}_{n,p}^{(1)} + \widehat{T}_{n,p}^{(2)}$ where

$$\widehat{T}_{n,p}^{(1)} = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \Big\{[\delta_\beta'(X_i - X_j) + (\varepsilon_i - \varepsilon_j)](X_i - X_j)'\Sigma\delta_\beta$$
$$+ [\delta_\beta'(X_iX_i' + \Sigma) + \varepsilon_i X_i'](X_jX_j' + \Sigma)\delta_\beta + \varepsilon_j X_j'(X_iX_i' + \Sigma)\delta_\beta\Big\} - 6\|\Sigma\delta_\beta\|^2$$

and $\widehat{T}^{(2)}_{n,p} = \binom{n}{2}^{-1}\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j X_i'X_j$. Under the assumptions of this theorem and following (A.7) and (A.8), $\mathrm{Var}(\widehat{\mathrm{T}}_{\mathrm{n,p}}) = \mathrm{Var}(\widehat{\mathrm{T}}^{(2)}_{\mathrm{n,p}})\{1+\mathrm{o}(1)\}$ and $\widehat{T}^{(1)}_{n,p}/\sqrt{\mathrm{Var}(\widehat{\mathrm{T}}_{\mathrm{n,p}})} = o_p(1)$. To prove the theorem, we only need to show

$$\widehat{T}^{(2)}_{n,p}/\sqrt{\mathrm{Var}(\widehat{\mathrm{T}}^{(2)}_{\mathrm{n,p}})} = \sqrt{\binom{n}{2}}\widehat{T}^{(2)}_{n,p}/\sqrt{\sigma^4 tr(\Sigma^2)} \xrightarrow{d} N(0,1). \tag{A.12}$$

Now write $\widetilde{T}_{nk} = \sqrt{\binom{n}{2}}\widehat{T}^{(2)}_{n,p} = \sum_{i=2}^{k} Z_{ni}$ and $\widetilde{T}_{nn} = \widetilde{T}_{n,p}$, where $Z_{ni} = \sum_{j=1}^{i-1}\varepsilon_i\varepsilon_j X_i'X_j/\sqrt{\binom{n}{2}}$. Let $\mathscr{F}_i = \sigma\left\{\binom{X_1}{\varepsilon_1},\cdots,\binom{X_i}{\varepsilon_i}\right\}$ be the $\sigma-$field generated by $\{(X_j^\tau,\varepsilon_j),\ j\le i\}$. It is easy to see that $E(Z_{ni}|\mathscr{F}_{i-1}) = 0$ and it follows that $\{\widetilde{T}_{nk},\mathscr{F}_k : 2\le k\le n\}$ is a zero mean martingale. Let $v_{ni} = E(Z_{ni}^2|\mathscr{F}_{i-1}),\ 2\le i\le n$ and $V_n = \sum_{i=2}^{n} v_{ni}$. The central limit theorem will hold (Hall and Heyde, 1980) if we can show

$$\frac{V_n}{\mathrm{Var}(\widetilde{\mathrm{T}}_{\mathrm{n,p}})} \xrightarrow{p} 1 \tag{A.13}$$

and for any $\epsilon > 0$

$$\sum_{i=1}^{n}\sigma^{-4}tr^{-1}(\Sigma^2)E\{Z_{ni}^2 I(|Z_{ni}| > \epsilon\sigma^2\sqrt{tr(\Sigma^2)})|\mathscr{F}_{i-1}\} \xrightarrow{p} 0. \tag{A.14}$$

It can be shown that $v_{ni} = \binom{n}{2}^{-1}\sigma^2\Big\{\sum_{j=1}^{i-1}\varepsilon_j^2 X_j'\Sigma X_j + 2\sum_{1\le j<k<i}\varepsilon_j\varepsilon_k X_j'\Sigma X_k\Big\}$ and

$$\begin{aligned}\frac{V_n}{\mathrm{Var}(\widetilde{\mathrm{T}}_{\mathrm{n,p}})} &= \frac{1}{\binom{n}{2}^2 tr(\Sigma^2)\sigma^2}\Big\{\sum_{j=1}^{n-1} j\varepsilon_j^2 X_j'\Sigma X_j + 2\sum_{1\le j<k\le n}\varepsilon_j\varepsilon_k X_j'\Sigma X_k\Big\}\\ &= C_{n1} + C_{n2},\ \text{say.}\end{aligned}$$

We know that $E(C_{n1}) = 1$ and

$$\mathrm{Var}(\mathrm{C}_{\mathrm{n1}}) = \frac{1}{\binom{n}{2}^4 tr^2(\Sigma^2)\sigma^4}E\Big\{\sum_{j=1}^{n-1} j^2(\varepsilon_j^4(X_j'\Sigma X_j)^2 - tr^2(\Sigma^2)\sigma^4)\Big\}.$$

As $tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}$ implies $E\{(X_j'\Sigma X_j)^2\} = o(n)tr^2(\Sigma^2)$. Hence, $\mathrm{Var}(\mathrm{C}_{\mathrm{n1}}) \to 0$ and $C_{n1} \xrightarrow{p} 1$. Similarly, $E(C_{n2}) = 0$ and

$$\mathrm{Var}(\mathrm{C}_{\mathrm{n2}}) = \frac{4}{\binom{n}{2}^4}\Big\{\sum_{i=3}^{n}\binom{i}{2} + \sum_{i=3}^{n-1}(n-i)\binom{i}{2}\Big\}\frac{tr(\Sigma^4)}{tr^2(\Sigma^2)}.$$

Thus, $tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}$ implies $C_{n2} \xrightarrow{p} 0$. In summary, (A.13) holds.

It remains to show (A.14). Since

$$E\{Z_{ni}^2 I(|Z_{ni}| > \epsilon\sigma^2\sqrt{tr(\Sigma^2)})|\mathscr{F}_{i-1}\} \le E(Z_{ni}^4|\mathscr{F}_{i-1})/(\epsilon^2\sigma^4 tr(\Sigma^2)),$$

by the law of large numbers, we only need to prove that

$$\sum_{i=1}^{n} E(Z_{ni}^4) = o\{\sigma^4 tr^2(\Sigma^2)\}. \tag{A.15}$$

Let $\kappa_4 = E(\varepsilon^4)$ which is assumed to be finite. Then

$$\sum_{i=1}^{n} E(Z_{ni}^4) \le \binom{n}{2}^{-1} \kappa_4^2 \Big(3tr^2(\Sigma^2) + (6 + 6\Delta + \Delta^2) tr(\Sigma^4)\Big)$$
$$+ \binom{n}{2}^{-2} \frac{1}{3}(n^3 - 3n^2 + 2n)\kappa_4 \sigma^4 \left(tr^2(\Sigma^2) + (2+\Delta) tr(\Sigma^4)\right).$$

Under the assumption that $tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}$, (A.15) follows immediately. This completes the proof. $\square$

Power under fixed alternative (4.7)

In this part, we consider two scenarios of fixed alternatives under (4.7) mentioned in Section 4. One is

$$\delta_\beta' \Sigma^3 \delta_\beta = o\left\{\frac{1}{n}\delta_\beta' \Sigma \delta_\beta tr(\Sigma^2)\right\}, \tag{A.16}$$

which complements (4.2). If $\delta_\beta' \Sigma \delta_\beta$ is truly bounded, (A.16) implies $\delta_\beta' \Sigma^3 \delta_\beta = o\left\{\frac{1}{n} tr(\Sigma^2)\right\}$ which mimics the second part of (4.2).

A complement to both (4.2) and (A.16) is

$$\frac{1}{n}\delta_\beta' \Sigma \delta_\beta tr(\Sigma^2) = o\left\{\delta_\beta' \Sigma^3 \delta_\beta\right\}. \tag{A.17}$$

If $\delta_\beta' \Sigma \delta_\beta$ is bounded, (A.17) implies $\frac{1}{n} tr(\Sigma^2) = o\left\{\delta_\beta' \Sigma^3 \delta_\beta\right\}$, which prescribes a larger discrepancies between β and β_0. Without causing much confusion, we call both (A.16) and (A.17) under (4.7) as fixed alternatives.

To quantify the asymptotic power, we define

$$\begin{aligned}
\sigma_{A_1}^2 &= 2\sigma^4 tr(\Sigma^2) + 2B_1^2 tr(\Sigma^2) + 4\sigma^2 tr(\Sigma^2) B_1 + 4\Delta(B_1 + \sigma^2) tr(A_1 \circ A_3) \\
&\quad + 2\Delta^2 tr\{(A_0 diag(A_1))^2\} \quad \text{and} \\
\sigma_{A_2}^2 &= (B_1 + \sigma^2) B_3 + B_2^2 + \Delta tr(A_1 \circ A_2).
\end{aligned}$$

We note that $\sigma_{A_1}^2$ is part of the variance of M_{n2}, where we only keep the leading order terms under (A.16) and $\sigma_{A_2}^2$ is the same as ζ_1, the variance of M_{n1} up to a constant.

Theorem A *Assume Model (2.4), Conditions (2.8) and (4.7), then (i) under the first fixed alternatives (A.16)*

$$\frac{n}{\sigma_{A_1}}(T_{n,p} - \|\Sigma\delta_\beta\|^2) \xrightarrow{d} N(0,1); \tag{A.18}$$

(ii) under the second fixed alternatives (A.17)

$$\frac{\sqrt{n}}{\sigma_{A_2}}(T_{n,p} - \|\Sigma\delta_\beta\|^2) \xrightarrow{d} N(0,1). \tag{A.19}$$

The proof of theorem A is contained in a longer version of this paper. The theorem implies that the asymptotic power of the test under the first fixed alternatives (A.16) is

$$\Omega_{H_1}(\|\delta_\beta\|) \doteq \Phi\left(-\frac{\sqrt{2tr(\Sigma^2)}\sigma^2 z_\alpha}{\sigma_{A_1}} + \frac{n\|\Sigma\delta_\beta\|^2}{\sigma_{A_1}}\right). \tag{A.20}$$

Since B_1 is not $o(1)$ and $\sigma_{A_1}^2 > 2B_1^2 tr(\Sigma^2)$, the first term $\sqrt{2tr(\Sigma^2)}\sigma^2 z_\alpha/\sigma_{A_1} < \sigma^2 z_\alpha/B_1$ is always bounded from infinity. In particular, if B_1 diverges to ∞, the first term converges to 0. Hence, the test attains at least 50% power in this case. If $n\|\Sigma\delta_\beta\|^2/\sigma_{A_1} \to \infty$, the power converges to 1.

The asymptotic power under the second fixed alternatives (A.17) is

$$\Omega_{H_2}(\|\delta_\beta\|) \doteq \Phi\left(-\frac{\sqrt{2tr(\Sigma^2)}\sigma^2 z_\alpha}{\sqrt{(n-1)\sigma_{A_2}^2}} + \frac{\sqrt{n}\|\Sigma\delta_\beta\|^2}{\sigma_{A_2}}\right).$$

As (A.17) implies $\frac{1}{n}tr(\Sigma^2)/\sigma_{A_2}^2 = o(1)$, the proposed test is consistent as long as

$$\sqrt{n}\|\Sigma\delta_\beta\|^2/\sigma_{A_2} \to \infty. \tag{A.21}$$

Even if $\sqrt{n}\|\Sigma\delta_\beta\|^2/\sigma_{A_2}$ does not converge to ∞, the power is still at least 50% asymptotically. The power of the test under the fixed alternatives attains at least 50% power is assuring and it can be shown that the proposed test is more powerful under two fixed alternatives than the local alternative if all the eigenvalues are of the same order. It is also the reason that we call the two alternatives in (A.16) and (A.17) as fixed alternatives. It may be shown that a sufficient condition for (A.21) is $\lambda_p/\lambda_1 = o(n)$.

REFERENCES

Ahmad, M., Werner, C. and Brunner E. (2008), Analysis of high-dimensional repeated measures designs: The one sample case, *Computational Statistics and Data Analysis*, **53** 416-427.

Akritas, M. and Arnold, S. (2000), Asymptotics for Analysis of Variance When the Number of Levels Is Large, *Journal of the American Statistical association*, **95**, 212-226.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley.

Bathke, A. and Harrar, S. (2008), Nonparametric Methods in Multivariate Factorial Designs for Large Number of Factor Levels, *Journal of Statistical Planning and Inference*, **138**, 588-610.

Bathke, A. and Lankowski, D. (2005), Rank Procedures for a Large Number of Treatments, *Journal of Statistical Planning and Inference*, **133**, 223-238.

Brunner, E., Dette, H. and Munk, A. (1997), Box-Type Approximations in Nonparametric Factorial Designs, *Journal of the American Statistical association*, **92**, 1494-1502.

Bai, Z. and Saranadasa, H. (1996), Effect of high dimension: by an example of a two sample problem, *Statistica Sinica,* **6**, 311-329.

Boos, D. D. and Brownie, C. (1995), ANOVA and rank tests when the number of treatments is large, *Statistics and Probability Letters*, **23**, 183-191.

Chen, S. X. and Qin, Y. (2010), A Two Sample Test for High Dimensional Data with Applications to Gene-set Testing, *The Annals of Statistics*, **38**, 808-835.

Chen, S. X., Zhang, L.X. and Zhong, P.S. (2010), Tests for High Dimensional Covariance Matrices, *Journal of the American Statistical Association*, **109**, 810-819.

Candes, E. and Tao, T. (2007), The Dantzig selector: Statistical estimation when p is much larger than n, *The Annals of Statistics,* **35**, 2313-2351.

Efron, B. and Tibshirani, R. (2007), On testing the significance of sets of genes, *The Annals of Applied Statistics*, **1**, 107-129.

Fan, J. and Hall, P. and Yao, Q. (2007), To how many simultaneous hypothesis tests can normal student's t or bootstrap calibrations be applied, *Journal of* *the American Statistical Association,* **102**, 1282-1288.

Fan, J. and Li, R. (2001), Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association,* **96**, 1348-1360.

Fan, J. and Lv, J. (2008), Sure independence screening for ultra-high dimensional feature space (with discussion), *Journal of Royal Statistical Society, Series B*, **70,** 849-911.

Frank, I. E. and Friedman, J. H. (1993), A statistical view of some chemometrics regression tools (with discussion), *Technometrics,* **35**, 109-148.

The Gene Ontology Consortium. (2000), Gene ontology: tool for the unification of biology, *Nature Genetics*, **25**, 25-29.

Goeman, J., Finos, L. and van Houwelingen, J. C. (2009), Testing against a high dimensional alternative in the generalized linear model: asymptotic alpha-level control, *Manuscript.*

Hall, P. and Heyde, C.C. (1980), *Martingale limit theory and its application,* New York : Academic Press.

Harrar, S. and Bathke, A. (2008), Nonparametric methods for unbalanced multivariate data and many factor levels, *Journal of Multivariate Analysis*, **99**, 1635-1664.

Hoeffding, W. (1948), A class of statistics with asymptotically normal distribution, *The Annals of Mathematical Statistics,* **19**, 293-325.

Huang, J., Wang, D. and Zhang, C.H. (2005), A Two-Way Semilinear Model for Normalization and Analysis of cDNA Microarray Data, *Journal of the American Statistical Association,* **100**, 814-829.

Kawa, K. and Obata, K. (1982), Altered developmental changes of neuromuscular junction in hypo- and hyperthyroid rats, *The Journal of Physiology*, **329**, 143-161.

Kosorok, M. R. and Ma, S. (2007), Marginal asymptotics for the "large p, small n" paradigm: With applications to microarray data, *The Annals of Statistics,* **35**, 1456-1486.

Lee, A. J. (1990), *U-statistics: theory and practice.* Marcel Dekker.

Lkhagvadorj, S., Qu, L., Cai, W., Couture, O., Barb, C., Hausman, G., Nettleton, D., Anderson, L., Dekkers, J. and Tuggle, C. (2009), Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant, *Physiol. Genomeics*, **38**, 98-111, 2009.

Newton, M., Quintana, F., Den Boon, J., Sengupta, S. and Ahlquist, P. (2007), Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis, *The Annals of Applied Statistics*, **1**, 85-106.

Portnoy, S. (1984), Asymptotic Behavior of the M-estimators of p-regression parameters when p^2/n is large: Consistency. *The Annals of Statistics*, **12**, 1298-1309.

Portnoy, S. (1985), Asymptotic Behavior of the M-estimators of p-regression parameters when p^2/n is large: Normal approximation. *The Annals of Statistics*, **13**, 1403-1417.

Rao, C. R., Touteburg, H., Shalabh and Heumann, C. (2008), *Linear models and generalizations*, Springer, New York.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, *Proceedings of the National Academy of Sciences,* **102** 15545-15550.

Tibshirani, R. (1996), Regression shrinkage and selection via the Lasso, *Journal of Royal Statistical Society, Series B,* **58**, 267-288.

Wang H.S. (2009), Forward regression for ultra-high dimensional variable screening, *Journal of the American Statistical Association,* **104**, 1512-1524.

Wang, H. and Akritas, M. (2009). Rank Tests in Heteroscedastic Multi-Way HANOVA, *Journal of Nonparametric Statistics*, **21**, 663-681

Zhang, C.H. and Huang, J. (2008), The sparsity and bias of the Lasso selection in high-dimensional linear regression, *The Annals of Statistics,* **36**, 1567-1594.

Table 1: Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p\times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p\times 1}$ at significant level 5% for normal residual.[a]

			$T = 10$				$T = 20$		
(n, p)	$\|\beta\|^2$	SNR	F-test	EB	New	SNR	F-test	EB	New
(a) Non-sparse case									
(40, 34)	0.00 (size)	0.00	0.05	0.04	0.06	0.00	0.05	0.04	0.07
	0.02	0.96	0.16	0.19	0.26	4.31	0.19	0.65	0.71
	0.04	1.92	0.31	0.36	0.44	8.62	0.35	0.90	0.93
	0.06	2.89	0.41	0.48	0.57	12.94	0.51	0.97	0.98
(60, 54)	0.00 (size)	0.00	0.05	0.03	0.06	0.00	0.05	0.04	0.06
	0.02	1.48	0.21	0.26	0.34	8.19	0.28	0.92	0.95
	0.04	2.95	0.43	0.53	0.62	16.38	0.53	1.00	1.00
	0.06	4.44	0.62	0.70	0.80	24.57	0.72	1.00	1.00
(80, 76)	0.00 (size)	0.00	0.06	0.03	0.06	0.00	0.04	0.04	0.06
	0.02	1.25	0.19	0.24	0.33	6.19	0.25	0.87	0.91
	0.04	2.51	0.34	0.48	0.56	12.39	0.41	0.99	1.00
	0.06	3.76	0.52	0.68	0.77	18.58	0.56	1.00	1.00
(b) Sparse case									
(40, 34)	0.02	0.59	0.08	0.12	0.18	1.41	0.09	0.25	0.32
	0.04	1.19	0.12	0.19	0.27	2.82	0.15	0.43	0.52
	0.06	1.78	0.17	0.29	0.38	4.23	0.20	0.60	0.68
(60, 54)	0.02	0.81	0.09	0.14	0.22	2.22	0.09	0.42	0.50
	0.04	1.63	0.13	0.26	0.36	4.45	0.18	0.68	0.76
	0.06	2.44	0.18	0.40	0.50	6.68	0.22	0.85	0.90
(80, 76)	0.02	0.62	0.07	0.11	0.17	1.67	0.09	0.34	0.42
	0.04	1.25	0.10	0.22	0.33	3.35	0.11	0.57	0.67
	0.06	1.87	0.13	0.32	0.44	5.03	0.16	0.80	0.87

[a]The standard error of power entries is bounded by 0.016 calculated based on 1000 simulations. SNR (signal-to-noise ratio) is $n\|\Sigma\beta\|^2/\{\sqrt{2tr(\Sigma^2)}\sigma^2\}$.

Table 2: Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p\times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p\times 1}$ at significant level 5% for centralized gamma residual.

		$T = 10$				$T = 20$			
(n, p)	$\|\beta\|^2$	SNR	F-test	EB	New	SNR	F-test	EB	New
(a) Non-sparse case									
(40, 34)	0.00 (size)	0.00	0.04	0.04	0.05	0.00	0.05	0.04	0.06
	0.02	0.96	0.14	0.22	0.28	4.31	0.20	0.67	0.73
	0.04	1.92	0.30	0.36	0.45	8.62	0.35	0.88	0.92
	0.06	2.89	0.47	0.49	0.59	12.94	0.52	0.95	0.96
(60, 54)	0.00 (size)	0.00	0.06	0.03	0.06	0.00	0.05	0.04	0.06
	0.02	1.48	0.22	0.29	0.39	8.19	0.28	0.90	0.93
	0.04	2.95	0.46	0.55	0.63	16.38	0.53	0.99	0.99
	0.06	4.44	0.63	0.73	0.79	24.57	0.72	1.00	1.00
(80, 76)	0.00 (size)	0.00	0.04	0.03	0.06	0.00	0.05	0.04	0.06
	0.02	1.25	0.21	0.23	0.31	6.19	0.24	0.86	0.90
	0.04	2.51	0.38	0.48	0.58	12.39	0.41	0.98	0.98
	0.06	3.76	0.51	0.68	0.75	18.58	0.59	1.00	1.00
(b) Sparse case									
(40, 34)	0.02	0.59	0.07	0.13	0.20	1.41	0.09	0.26	0.35
	0.04	1.19	0.14	0.22	0.31	2.82	0.13	0.49	0.58
	0.06	1.78	0.15	0.29	0.40	4.23	0.21	0.62	0.70
(60, 54)	0.02	0.81	0.09	0.15	0.23	2.22	0.09	0.42	0.49
	0.04	1.63	0.11	0.30	0.40	4.45	0.17	0.69	0.76
	0.06	2.44	0.15	0.45	0.56	6.68	0.24	0.86	0.91
(80, 76)	0.02	0.62	0.06	0.11	0.18	1.67	0.08	0.37	0.43
	0.04	1.25	0.10	0.24	0.33	3.35	0.12	0.65	0.72
	0.06	1.87	0.12	0.35	0.48	5.03	0.14	0.77	0.84

Table 3: Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p\times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p\times 1}$ at significant level 5% for normal residual.

			$T = 10$			$T = 20$	
(n, p)	$\|\beta\|^2$	SNR	EB	New	SNR	EB	New
(a) Non-sparse case							
(40, 310)	0.00 (size)	0.00	0.00	0.06	0.00	0.02	0.06
	0.02	0.30	0.01	0.09	1.99	0.26	0.46
	0.04	0.61	0.01	0.15	3.99	0.47	0.68
	0.06	0.92	0.05	0.21	5.98	0.62	0.81
(60, 400)	0.00 (size)	0.00	0.01	0.05	0.00	0.01	0.05
	0.02	0.49	0.02	0.14	2.51	0.30	0.54
	0.04	0.98	0.05	0.23	5.03	0.63	0.82
	0.06	1.47	0.08	0.31	7.54	0.83	0.93
(80, 550)	0.00 (size)	0.00	0.00	0.05	0.00	0.02	0.06
	0.02	0.55	0.02	0.15	4.02	0.63	0.79
	0.04	1.11	0.08	0.29	8.05	0.91	0.96
	0.06	1.66	0.13	0.37	12.08	0.98	0.99
(b) Sparse case							
(40, 310)	0.02	0.16	0.01	0.08	0.58	0.05	0.15
	0.04	0.32	0.01	0.12	1.17	0.09	0.23
	0.06	0.48	0.01	0.11	1.75	0.12	0.30
(60, 400)	0.02	0.27	0.01	0.08	0.60	0.05	0.16
	0.04	0.54	0.02	0.14	1.21	0.09	0.25
	0.06	0.82	0.04	0.18	1.82	0.14	0.35
(80, 550)	0.02	0.35	0.02	0.10	1.05	0.11	0.24
	0.04	0.70	0.03	0.16	2.11	0.25	0.46
	0.06	1.05	0.05	0.25	3.17	0.38	0.58

Table 4: Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p\times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p\times 1}$ at significant level 5% for centralized gamma residual.

		$T = 10$			$T = 20$		
(n, p)	$\|\beta\|^2$	SNR	EB	New	SNR	EB	New
		(a) Non-sparse case					
(40, 310)	0.00 (size)	0.00	0.01	0.06	0.00	0.01	0.06
	0.02	0.30	0.01	0.12	1.99	0.24	0.45
	0.04	0.61	0.03	0.19	3.99	0.52	0.70
	0.06	0.92	0.05	0.24	5.98	0.69	0.83
(60, 400)	0.00 (size)	0.00	0.01	0.04	0.00	0.01	0.04
	0.02	0.49	0.02	0.13	2.51	0.35	0.57
	0.04	0.98	0.05	0.24	5.03	0.65	0.82
	0.06	1.47	0.10	0.36	7.54	0.82	0.93
(80, 550)	0.00 (size)	0.00	0.01	0.05	0.00	0.02	0.05
	0.02	0.55	0.03	0.16	4.02	0.67	0.82
	0.04	1.11	0.07	0.23	8.05	0.91	0.97
	0.06	1.66	0.16	0.40	12.08	0.97	0.99
		(a) Sparse case					
(40, 310)	0.02	0.16	0.01	0.08	0.58	0.05	0.16
	0.04	0.32	0.01	0.10	1.17	0.11	0.25
	0.06	0.48	0.02	0.14	1.75	0.14	0.33
(60, 400)	0.02	0.27	0.02	0.09	0.60	0.04	0.15
	0.04	0.54	0.02	0.12	1.21	0.10	0.25
	0.06	0.82	0.04	0.20	1.82	0.18	0.38
(80, 550)	0.02	0.35	0.01	0.10	1.05	0.10	0.24
	0.04	0.70	0.03	0.17	2.11	0.27	0.48
	0.06	1.05	0.06	0.25	3.17	0.39	0.60

Table 5: Empirical size and power of the proposed test for $H_0 : \beta = \mathbf{0}_{p\times 1}$ in a 2×2 factorial design with $n_1 = 20$ and $n_2 = 30$ replicates in each cell.[a]

		Non-sparse				Sparse			
p	$\|\beta\|^2$	SNR_f	n_1	SNR_f	n_2	SNR_f	n_1	SNR_f	n_2
(a) Normal residuals									
100	0.00 (size)	0.00	0.06	0.00	0.06	0.00	0.07	0.00	0.05
	0.02	3.05	0.65	4.58	0.85	0.70	0.20	1.06	0.26
	0.04	6.10	0.88	9.16	0.98	1.41	0.29	2.12	0.48
	0.06	9.16	0.96	13.74	1.00	2.12	0.44	3.18	0.65
150	0.00 (size)	0.00	0.06	0.00	0.06	0.00	0.05	0.00	0.06
	0.02	2.59	0.57	3.89	0.77	0.57	0.15	0.85	0.21
	0.04	5.18	0.84	7.78	0.97	1.14	0.28	1.71	0.39
	0.06	7.78	0.94	11.67	0.99	1.71	0.35	2.57	0.54
200	0.00 (size)	0.00	0.07	0.00	0.06	0.00	0.07	0.00	0.06
	0.02	2.28	0.50	3.43	0.73	0.49	0.14	0.73	0.18
	0.04	4.57	0.78	6.86	0.94	0.98	0.22	1.47	0.35
	0.06	6.86	0.89	10.29	0.99	1.47	0.31	2.21	0.48
(b) Gamma residuals									
100	0.00 (size)	0.00	0.07	0.00	0.05	0.00	0.07	0.00	0.06
	0.02	3.05	0.66	4.58	0.83	0.70	0.15	1.06	0.28
	0.04	6.10	0.86	9.16	0.97	1.41	0.31	2.12	0.48
	0.06	9.16	0.95	13.74	0.99	2.12	0.47	3.18	0.66
150	0.00 (size)	0.00	0.07	0.00	0.05	0.00	0.04	0.00	0.06
	0.02	2.59	0.57	3.89	0.78	0.57	0.16	0.85	0.22
	0.04	5.18	0.81	7.78	0.96	1.14	0.28	1.71	0.39
	0.06	7.78	0.93	11.67	0.99	1.71	0.37	2.57	0.57
200	0.00 (size)	0.00	0.05	0.00	0.06	0.00	0.06	0.00	0.05
	0.02	2.28	0.53	3.43	0.74	0.49	0.14	0.73	0.18
	0.04	4.57	0.77	6.86	0.93	0.98	0.24	1.47	0.32
	0.06	6.86	0.89	10.29	0.98	1.47	0.30	2.21	0.48

[a]The $\text{SNR}_f = n_c(\sum_c \|\Sigma_c\beta\|^2)/(\sigma^2\sqrt{\sum_c 2tr(\Sigma_c^2)})$.

Table 6: P-values of the GO terms which are significant under at least three designs using the proposed test, and their number of genes.

GO term	Design I	Design II	Design III	Design IV	No. of Genes
GO:0004115	3.253E-04	2.774E-06		1.992E-06	8
GO:0005086	2.345E-10	1.945E-05	7.220E-06	1.629E-05	14
GO:0005677	1.082E-04	3.102E-06		7.575E-05	5
GO:0006342	3.068E-04	3.444E-06		5.951E-05	5
GO:0007528	1.110E-16	7.922E-07	2.235E-08	3.203E-04	8
GO:0017136	1.082E-04	3.102E-06		7.575E-05	5
GO:0032012	0.000E-04	2.586E-06	2.746E-10	5.418E-06	12
GO:0050909	1.545E-09	3.842E-05	4.216E-05		5