

A New Test for Large Dimensional Regression Coefficients

June Luo¹, Yi-Jun Zuo²

¹*Department of Mathematical Sciences, Clemson University, Clemson, USA*

²*Department of Statistics and Probability, Michigan State University, Lansing, USA*

E-mail: jluo@clemson.edu

Received August 19, 2011; revised September 22, 2011; accepted September 30, 2011

Abstract

In the article, hypothesis test for coefficients in high dimensional regression models is considered. I develop simultaneous test statistic for the hypothesis test in both linear and partial linear models. The derived test is designed for growing p and fixed n where the conventional F -test is no longer appropriate. The asymptotic distribution of the proposed test statistic under the null hypothesis is obtained.

Keywords: High Dimension, Ridge Regression, Hypothesis Test, Partial Linear Model, Asymptotic

1. Introduction

Some high dimensional data, such as gene expression datasets in microarray, exhibits the property that the number of covariates greatly exceeds the sample size. The discovery of “large p , small n ” paradigm brings challenges to many traditional statistical methods, and thus the asymptotic properties of various estimators when p goes to infinity much faster than n have been discussed (see [1-3]). Reference [1] considered uniform convergence for a large number of marginal discrepancy measures targeted on univariate distributions, means and medians. Reference [3] proposed a two sample test on high dimensional means. Both of these aforementioned articles considered testing under “large p , small n ” without a regression structure, which the present article concentrates on.

Zhong and Chen in [4] proposed a test statistic for testing the regression coefficients in linear models when $p/n \rightarrow \rho$ in $(0,1)$. As 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 50 due to limitation for replications. The fact results in p going to infinity and thus I think the consideration of p going to infinity and n remains constant is more practical.

Covariate selection for high dimensional linear regression has received considerable attention in recent years. Penalizing methods are alternatives to the traditional least squares estimator for shrinkage estimation as in [5, 6]. Shao and Chow in [7] proposed a variable screening method using ridge estimators as both p and n go to infinity. In contrast to the assumptions in the literature, I

consider “large p , fixed n ” setting in linear models for variable selection. Testing hypothesis on the regression coefficients is critical in determining the effects of covariates on certain outcome variable. Motivated by the latest need in biology to identify significant sets of genes, rather than individual gene, I aim at developing simultaneous tests for coefficients in linear regression models.

The partial linear models have been extensively studied. They have a wide range of applications, from statistics to biomedical sciences. In these models, some of the relations are believed to be of certain parametric form while others are not easily parameterized. Several approaches have been developed to construct estimators. A profile likelihood approach was used in [8,9]. In this article, I apply a difference based estimation method in the partial linear models. The method of taking differences to eliminate the effect of the unknown nonparametric component has been used in both nonparametric and semiparametric settings. Rice in [10] first introduced a differencing estimator of the residual variance. Horowitz and Spokoiny in [11] used the differencing method to test between a parametric model and a nonparametric alternative. After taking the differences to eliminate the bias induced from the nonparametric term, I concentrate attention on estimating the linear component and then formulate the test statistic for testing the linear components.

The article begins with the conventional F -test. I will then discuss the efficiency of ridge estimator and propose a new test statistic for “large p , fixed n ” setting. The asymptotic distribution of the proposed test statistic under null hypothesis is established. Extensions to partial

linear models are then made in Section 3.

2. Test Statistics

Consider a linear regression model

$$Y = X\beta + \varepsilon \tag{1}$$

where $X = (X_1, X_2, \dots, X_n)'$ are independent and identically distributed observation matrix, covariates $X_{i1}, X_{i2}, \dots, X_{ip}$ are uncorrelated, $Y = (Y_1, Y_2, \dots, Y_n)'$ are independent responses, β is the $p \times 1$ vector of regression coefficients, and $\varepsilon \sim N(0, \sigma_p^2 I_n)$. I am interested in testing a high-dimensional hypothesis

$$H_0 : \beta = \beta_0 \text{ vs } H_1 : \beta \neq \beta_0 \tag{2}$$

for a specific β_0 in R^p .

2.1. F-Test for “Large n , Small p ”

I will start from reviewing the F -test for hypothesis (2) by Rao in [12]. When we have a large sample size, we can use least squares method to estimate the coefficients. The least squares estimator is $\hat{\beta} = (X'X)^{-1}X'Y$. The conventional test for (2) is given by

$$F_{n,p} = \frac{(\hat{\beta} - \beta_0)' X'X(\hat{\beta} - \beta_0) / p}{(Y - X\hat{\beta})'(Y - X\hat{\beta}) / (n-p)} \tag{3}$$

As proven in [12], under H_0 , $F_{n,p} \sim F_{p,n-p}$. Hence, an α -level F -test rejects H_0 if $F_{n,p} > F_{p,n-p;\alpha}$, the upper α -quantile of the $F_{p,n-p}$ distribution. The F statistic is a monotone function of the likelihood ratio statistic and is distributed as a noncentral F distribution under the alternative (see [13]).

2.2. A New Test Statistic for “Large p , Fixed n ”

I have seen a limitation with the F -test defined in Equation (3): it can not be applied to large p and small n . As more and more datasets exhibit larger dimension than sample size, we are in need to formulate a test statistic to suit the large “ p and small n ” paradigm. Because least squares estimator $\hat{\beta}$ is inappropriate when $p > n$, I modify the F -statistic in two aspects. One is to replace the least squares estimator with an appropriate estimator of β . The second is to find the asymptotic distribution of the new test statistic.

To overcome the singularity of $X'X$ when $p > n$ in model (1), consider using penalizing methods. The ridge estimator $\hat{\beta}$ of β in [14] is

$$\hat{\beta} = (X'X + h_p I_p)^{-1} X'Y \tag{4}$$

where h_p is the regularization parameter. Luo in [15] proved that the $\hat{\beta}$ in Equation (4) is mean squared er-

ror consistent of β under certain conditions. More recently, Luo in [16] proved the mean squared error consistency under less restrictive conditions. The assumptions and the results in [16] are given below.

Assumption A. $1/h_p = o(1)$. For sufficiently large p , there is a vector $b_{p \times 1}$ such that $\beta = X'Xb$. Furthermore, there exists a constant $\varepsilon > 0$ such that each component of $b_{p \times 1}$ is $O(1/p^{1+\varepsilon})$.

Assumption B. σ_p and h_p are chosen such that $p^{-\varepsilon} \cdot h_p = o(1)$ and $\sigma_p = o(h_p^{0.5})$.

It was proven in [16] that under the Assumption A and B, $\text{bias}(\hat{\beta}_j) = O(\sigma_p^2/h_p) = o(1)$ and $\text{var}(\hat{\beta}_i) = O(\sigma_p^2/h_p) = o(1)$. In this article, I will take the opportunity to explore more concise asymptotic results about $\hat{\beta}$ under Assumption A and B. Because $X'X$ can have at most n positive eigenvalues, without loss of generality, let λ_{ip} be the i th nonzero eigenvalue of $X'X$ and assume $\lambda_{ip} > 0$ for all $i = 1, 2, \dots, n$. Let $\Gamma = (\tau_{ij})_{p \times p}$ be an orthogonal matrix such that

$$X'X = \Gamma \begin{bmatrix} \Lambda_{n \times (p-n)} & O_{n \times (p-n)} \\ O_{(p-n) \times n} & O_{(p-n) \times (p-n)} \end{bmatrix} \Gamma'$$

where $\Lambda_{n \times n}$ is a diagonal matrix with elements λ_{ip} , $i = 1, 2, \dots, n$.

Theorem 1. Under Assumption A and B, given that the p covariates are uncorrelated, if h_p is chosen such that $p^{-\varepsilon/2} h_p / \sigma_p = o(1)$ and $\lambda_{ip} = o(h_p)$ for all $i = 1, 2, \dots, n$, I have

$$\frac{h_p}{\sigma_p} (\hat{\beta} - \beta) \rightarrow_n N(0, \text{diag}(X'X))$$

where $\text{diag}(X'X)$ means the diagonal matrix with diagonal elements of $X'X$.

Proof.

Because the random error ϵ is multivariate normal, $\hat{\beta}$ is a multivariate normal.

$$\begin{aligned} \text{cov}(\hat{\beta} - \beta) &= (X'X + h_p I_p)^{-1} \cdot X' \sigma_p^2 X (X'X + h_p I_p)^{-1} \\ &= \frac{\sigma_p^2}{h_p} \left(\frac{X'X}{h_p} + I_p \right)^{-1} - \left[\left(\frac{X'X}{h_p} + I_p \right)^{-1} \left(\frac{X'X}{h_p} + I_p \right)^{-1} \right] \\ &= \frac{\sigma_p^2}{h_p} [\Gamma A \cdot \Gamma' - \Gamma A \cdot \Gamma' \cdot \Gamma A \cdot \Gamma'] \\ &= \frac{\sigma_p^2}{h_p} [\Gamma (A - A^2) \Gamma'] \end{aligned}$$

where $A = (\frac{X'X}{h_p} + I_p)^{-1}$ is a diagonal matrix with $i = 1, 2, \dots, n$ as first n diagonal elements, and the rest $(p - n)$ diagonal elements all equal to 1. So

$$\text{var}(\hat{\beta}_j) = \frac{\sigma_p^2}{h_p^2} \sum_{i=1}^n \tau_{ij}^2 \frac{h_p^2 \lambda_{ip}}{(h_p + \lambda_{ip})^2}$$

for all $j = 1, 2, \dots, p$.

Under the assumption $\lambda_{ip} = o(h_p)$ for all $i = 1, 2, \dots, n$,

$$\lim_{p \rightarrow \infty} \frac{h_p^2}{(h_p + \lambda_{ip})^2} = 1 \text{ for all } i = 1, 2, \dots, n.$$

Notice that $\sum_{i=1}^n \tau_{ij}^2 \lambda_{ip}$ is the j th diagonal element of $X'X$, we have

$$\lim_{p \rightarrow \infty} \frac{h_p^2}{\sigma_p^2} \text{var}(\hat{\beta}_j) = \text{diag}(X'X)_j \text{ for all } j = 1, 2, \dots, p.$$

where $\text{diag}(X'X)_j$ means the j th diagonal element of $X'X$. Given that the p covariates are uncorrelated, I conclude

$$\lim_{p \rightarrow \infty} \frac{h_p^2}{\sigma_p^2} \text{cov}(\hat{\beta}) = \text{diag}(X'X) \tag{5}$$

As in [16], the bias($\hat{\beta}_j$) = $O(p^{-\epsilon} h_p) = o(1)$, the assumption that $p^{-\epsilon/2} h_p / \sigma_p = o(1)$ guarantees bias($\hat{\beta}_j$) $h_p / \sigma_p = o(1)$. Along with result in (5), that completes the proof for Theorem 1.

Now I can modify the $F_{n,p}$ to a test statistic for ‘‘large p , fixed n ’’ paradigm. Define

$$G_{n,p} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{h_p^2 (\hat{\beta} - \beta_0)' (\text{diag}(X'X))^{-1} (\hat{\beta} - \beta_0) / p} \tag{6}$$

Under assumption **A** and **B**, as $p \rightarrow \infty$, $\hat{\beta}$ in Equation (4) is mean squared error consistent of β which implies $\hat{\beta}$ converges in probability to β . Apply the continuous mapping theorem,

$$\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma_p^2} \text{ converges in probability to } \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma_p^2} = \frac{\epsilon'\epsilon}{\sigma_p^2}$$

which is χ_n^2 distribution. Under H_0 as $p \rightarrow \infty$, by Theorem 1,

$$\frac{h_p^2}{\sigma_p^2} (\hat{\beta} - \beta_0)' (\text{diag}(X'X))^{-1} (\hat{\beta} - \beta_0)$$

converges to χ_p^2 distribution. By Law of large numbers,

$$\frac{h_p^2}{\sigma_p^2} (\hat{\beta} - \beta_0)' (\text{diag}(X'X))^{-1} (\hat{\beta} - \beta_0) / p$$

converges to 1. Apply the Slutsky’s theorem, I conclude that under H_0 , test statistic $G_{n,p}$ in Equation (6) converges

in distribution to χ_n^2 as $p \rightarrow \infty$ and n is a fixed constant. Hence, an α -level $G_{n,p}$ statistic rejects H_0 if $G_{n,p} > \chi_{n;\alpha}^2$, the upper α quantile of the χ_n^2 distribution.

3. Extension to Partial Linear Models

Partial linear models are more flexible than standard linear models. They can be a suitable choice when one suspects that the response Y linearly depends on X , but Y is nonlinearly related to Z . Consider a fix design version of the partial linear model which has the matrix form

$$Y = X\beta + f(Z) + \epsilon \tag{7}$$

where $Y = (Y_1, Y_2, \dots, Y_{n+1})'$, X is a $(n + 1) \times p$ matrix whose i th row is given by x_i , the p covariates of x_i are uncorrelated and $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_{n+1})'$ is normally distributed with a mean vector 0 and covariance matrix $\sigma_p^2 I_{n+1}$. Estimators of the linear component for $n > p$ situation have been discussed in [17-19]. The methods are not applicable for $p > n$, I propose the following procedure to obtain a statistic for hypothesis (2) in partial linear model (7). Assume the sequence $\{z_i\} \rightarrow c_0$ as $p \rightarrow \infty$, for all $i = 1, 2, \dots, n + 1$, where c_0 is a finite constant. The unknown function f is continuous at point c_0 . Consider

$$y_{i+1} - y_i = (x_{i+1} - x_i)\beta + f(z_{i+1}) - f(z_i) + \epsilon_{i+1} - \epsilon_i. \tag{8}$$

Since $z_i \rightarrow c_0$ for all $i = 1, 2, \dots, n + 1$, for any $\psi > 0$, there exists a large enough p value so that we have $\max_{1 \leq i \leq n+1} |z_i - c_0| < \psi$.

Function f is continuous at point c_0 , so for a large enough p , we have

$$\max_{1 \leq i \leq n+1} |f(z_i) - f(c_0)| < \psi,$$

which implies that for a finite n ,

$$\max_{1 \leq i \leq n} |f(z_{i+1}) - f(z_i)| = o(1) \tag{9}$$

Define a matrix

$$D = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 0 & \dots & 0 & 0 & 1 & -1 \end{pmatrix}_{n \times (n+1)}. \tag{10}$$

We now consider the matrix form of Equation (8), which is

$$DY = DX\beta + Df(Z) + D\epsilon. \tag{11}$$

Because of Equation (9), I can ignore the presence of nonparametric part in model (11). Thus, (11) becomes

$$DY = DX\beta + D\epsilon \tag{12}$$

where matrix D is given in (10). Luo in [20] examined the asymptotic distribution of ridge estimator of β in (12). Obviously the random errors $D\epsilon$ are not independent and thus the following procedure is crucial for the extension of previous results. Without loss of generality, assume sample size n is even. Define (see Equations (13) and (14)).

So Equation (12) becomes

$$D_1 Y = D_1 X \beta + D_1 \epsilon \tag{15}$$

and

$$D_2 Y = D_2 X \beta + D_2 \epsilon \tag{16}$$

Notice that $D_1 \epsilon \sim N(0, \sigma_p^2 I_{n/2})$ and $D_2 \epsilon \sim N(0, \sigma_p^2 I_{n/2})$. Now I can apply the results in Section 2.2 in model (15) and model (16). It follows that the two statistics for testing hypothesis (2) in model (15) and (16) are given by

$$G_{n,p}^1 = \frac{(D_1 Y - D_1 X \hat{\beta}_1)' (D_1 Y - D_1 X \hat{\beta}_1)}{h_p^2 (\hat{\beta}_1 - \beta_0)' \left(\text{diag}(X'D_1'D_1 X) \right)^{-1} (\hat{\beta}_1 - \beta_0) / P}$$

and

$$G_{n,p}^2 = \frac{(D_2 Y - D_2 X \hat{\beta}_2)' (D_2 Y - D_2 X \hat{\beta}_2)}{h_p^2 (\hat{\beta}_2 - \beta_0)' \left(\text{diag}(X'D_2'D_2 X) \right)^{-1} (\hat{\beta}_2 - \beta_0) / P}$$

where $\hat{\beta}_1 = (X'D_1'D_1 X + h_p I_p)^{-1} X'D_1'D_1 Y$ and $\hat{\beta}_2 = (X'D_2'D_2 X + h_p I_p)^{-1} X'D_2'D_2 Y$. When all assumptions for Theorem 1 hold, under H_0 , both $G_{n,p}^1$ and $G_{n,p}^2$ converge in distribution to χ_n^2 as $p \rightarrow \infty$. Hence, the decision rule is we reject H_0 if $\min(G_{n,p}^1, G_{n,p}^2) > \frac{\chi_n^2}{2}; \alpha$ and otherwise, fail to reject H_0 .

5. References

[1] M. Kosorok and S. Ma, "Marginal Asymptotics for the

'Large p , Small n ' Paradigm: With Applications to Microarray Data," *Annals of Statistics*, Vol. 35, No. 4, 2007, pp. 1456-1486. [doi:10.1214/009053606000001433](https://doi.org/10.1214/009053606000001433)

[2] J. Fan, P. Hall and Q. Yao, "To How Many Simultaneous Hypothesis Tests Can Normal Student's t or Bootstrap Calibrations Be Applied," *Journal of the American Statistical Association*, Vol. 102, No. 480, 2007, pp. 1282-1288. [doi:10.1198/016214507000000969](https://doi.org/10.1198/016214507000000969)

[3] S. Chen and Y. Qin, "A Two Sample Test for High Dimensional Data with Applications to Gene-Set Testing," *Annals of Statistics*, Vol. 38, No. 2, 2010, pp. 808-835. [doi:10.1214/09-AOS716](https://doi.org/10.1214/09-AOS716)

[4] P. Zhong and S. Chen, "Tests for High-Dimensional Regression Coefficients with Factorial Designs," *Journal of American Statistical Association*, Vol. 106, No. 493, 2011, pp. 260-274. [doi:10.1198/jasa.2011.tm10284](https://doi.org/10.1198/jasa.2011.tm10284)

[5] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B*, Vol. 58, No. 1, 1996, pp. 267-288.

[6] J. Fan and J. Lv, "Sure Independence Screening for Ultra-high Dimensional Feature Space (with Discussion)," *Journal of Royal Statistical Society*, Vol. 70, No. 5, 2008, pp. 849-911. [doi:10.1111/j.1467-9868.2008.00674.x](https://doi.org/10.1111/j.1467-9868.2008.00674.x)

[7] J. Shao and S. Chow, "Variable Screening in Predicting Clinical Outcome with High-Dimensional Microarrays," *Journal of Multivariate Analysis*, Vol. 98, No. 8, 2007, pp. 1529-1538. [doi:10.1016/j.jmva.2004.12.004](https://doi.org/10.1016/j.jmva.2004.12.004)

[8] R. Carroll, J. Fan, I. Gijbels and M. Wand, "Generalized Partially Linear Single-Index Models," *Journal of American Statistical Association*, Vol. 92, No. 438, 1997, pp. 477-489. [doi:10.2307/2965697](https://doi.org/10.2307/2965697)

[9] T. Severini and W. Wong, "Profile Likelihood and Conditionally Parametric Models," *Annals of Statistics*, Vol. 20, No. 4, 1992, pp. 1768-1802. [doi:10.1214/aos/1176348889](https://doi.org/10.1214/aos/1176348889)

[10] J. Rice, "Bandwidth Choice for Nonparametric Regression," *Annals of Statistics*, Vol. 12, No. 4, 1984, pp. 1215-1230. [doi:10.1214/aos/1176346788](https://doi.org/10.1214/aos/1176346788)

[11] J. Horowitz and V. Spokoiny, "An Adaptive Rate-optimal Test of a Parametric Mean-regression Model against a Nonparametric Alternative," *Econometrica*, Vol. 69, No. 3, 2001, pp. 599-631. [doi:10.1111/1468-0262.00207](https://doi.org/10.1111/1468-0262.00207)

[12] C. Rao, H. Toutenburg, and C. Heumann, "Linear Models and Generalizations," Springer, New York, 2008.

$$D_1 = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & \dots & 0 \\ \vdots & & & & & & \\ 0 & \dots & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix} \frac{n}{2} \times (n+1) \tag{13}$$

$$D_2 = \begin{pmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & \dots & 0 \\ \vdots & & & & & & \\ 0 & \dots & 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \frac{n}{2} \times (n+1). \tag{14}$$

- [13] T. Anderson, "An Introduction to Multivariate Statistical Analysis," Wiley, Hoboken, 2003.
- [14] A. Hoerl and R. Kennard, "Ridge Regression Biased Estimation for Nonorthogonal Problems," *Technometrics*, Vol. 12, No. 1, 1970, pp. 55-67. [doi:10.2307/1267351](https://doi.org/10.2307/1267351)
- [15] J. Luo, "The Discovery of Mean Square Error Consistency of Ridge Estimator," *Statistics and Probability Letters*, Vol. 80, No. 5, 2010, pp. 343-347. [doi:10.1016/j.spl.2009.11.008](https://doi.org/10.1016/j.spl.2009.11.008)
- [16] J. Luo, "Asymptotical Properties of Coefficient of Determination for Ridge Regression with Growing Dimensions," *Oriental Journal of Statistical Methods, Theory and Applications*, Vol. 1, No. 1, 2011, pp. 41-49.
- [17] L. Wang, L. Brown and T. Cai, "A Difference Based Approach to Semiparametric Partial Linear Model," *Electronic Journal of Statistics*, Vol. 5, 2011, pp. 619-641.
- [18] A. Yatchew, "An Elementary Estimator of the Partial Linear Model," *Economics Letters*, Vol. 57, No. 2, 1997, pp. 135-143. [doi:10.1016/S0165-1765\(97\)00218-8](https://doi.org/10.1016/S0165-1765(97)00218-8)
- [19] A. Yatchew, "Scale Economies in Electricity Distribution: A Semiparametric Analysis," *Journal of Applied Economics*, Vol. 15, No. 2, 2000, pp. 187-210.
- [20] J. Luo, "Asymptotic Efficiency of Ridge Estimator in Linear and Semiparametric Linear Models," *Statistics and Probability Letters*, Vol. 82, No. 1, 2011, pp.58-62. [doi:10.1016/j.spl.2011.08.018](https://doi.org/10.1016/j.spl.2011.08.018)