Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis: International Econometrics and Operation Research

An Application of a General Theory to Sparse High Dimensional models

Author

Guangyao Zhou (474384)

Supervisor

Alex J. Koning

Second assessor

Max Welz

Date final version

July 4, 2021

Abstract

This paper explores the applicability of decorrelated score methods. I used three interest models with (non)convex penalty functions to find the degree of application of this method under the pointwise weak convergence assumptions. In a simulation, I find that decorrelated score methods have high applicability in a low dimension. However, the interest models show varying degrees of general application in high dimensions.

1 Introduction

With the development of information gathering techniques, the observed data with prominent amount characteristics in economics has increased rapidly and considerably impacted the traditional regression analysis of economics. It is a proven and refined big data technology that causes an observation value containing too much information. When performing regression analysis, the amount of explanatory variables is far more than the observation value. However, traditional regression methods, such as linear regression, usually use least squares to estimate parameters, which is more likely to cause overfitting problems and strict multicollinearity. Besides, in the case that the number of regressors d is far more extensive than the number of observations n, the coefficient vector β is rank deficient and bias (Javanmard and Montanari, 2014). Many high-dimensional models are applied to solve the above issues over the last decades. The high-dimensional model currently plays a crucial role in some economic research fields. It presents both wide ranges of application and considerable importance in the economics fields. The high-dimensional models, with a small number of significant explanatory variables s≪n, are to capture the main characteristics of regression models Belloni and Chernozhukov (2011). It is worth noting that the high dimensionality and a few significant explanatory variables form a high-dimensional sparse matrix of parameters. At present, the mainstream of the most pursued research progress of high-dimensional model is point estimation. Buehlmann (2006) and Greenshtein and Ritov (2004) focus on consistency for prediction; Oracle inequalities and parameter estimation, for example, are studied in Bunea et al. (2007) and Juditsky et al. (2012). In addition, some papers (Fan and Lv (2008), Meinshausen and Bühlmann (2006)) were to introduce a method concerning variable selection for high-dimensional models. Belloni et al. (2011) elaborates some examples considering the databases with many regressors, e.g., the American Housing Survey records. Moreover, the explosive growth and collection of panel data have prompted the large-scale use of high-dimensional models. Vector autoregression model is used with a high-dimensional database to make analysis and forecast based on Fan et al.

(2011).

A large part of theoretical research in high-dimensional models is conducting a statistical test and confidence regions for the assessment study. In some statistical theoretical studies, confidence intervals and hypothesis testing are the keys to exploring the reliability of models. For this reason, many econometric economists have proposed related testing methods. C.-H. Zhang and Zhang (2014), Voorman et al. (2014), and Javanmard and Montanari (2014) provide methods to explore the confidence intervals and hypothesis testing for the parameters of high-dimensional models. Athey et al. (2016) illustrates a general method to find the average treatment effect without exploring some treatment prerequisites for nuisance parameters for high dimensional models, which is a further study for statistic test.

To realize and apply the high-dimensional models, many researchers employ penalized regression and study an assessment work for these estimators, such as maximum likelihood (MLE) or maximum penalized likelihood estimator (MPLE). The penalty functions consist of a convex penalty and a non-convex penalty. Generally, we consider Lasso and Dantzig selectors as convex penalty functions, where both are approximately equivalent shown by Bickel et al. (2009), Lounici et al. (2008), and James et al. (2009). Inverting Karush– Kuhn–Tucker (KKT) conditions Van de Geer et al. (2014) are mentioned to give optimality properties under the assumption on high-dimensional sparse models. MCP and SCAD are considered as non-convex penalty. Admittedly, convex penalty functions solve high-dimensional sparse regression problems, especially lasso penalty function C.-H. Zhang and Huang (2008). However, it inevitably caused the bias of the estimators Liu et al. (2012). Hereby, some debiased methods are proposed to correct the bias of regression with convex functions, e.g. Zou (2006), T. Zhang et al. (2009), and Zhou (2009).

Ning, Liu, et al. (2017a) proposes a new method named decorrelated score methods to make a hypothesis test and find an optimal confident interval for a univariate interest parameter in a general framework. Besides, it develops a limiting distribution for the statistic test under the pointwise weak convergence. Then, a one-step estimator, following asymptotically normal distribution, is to construct an optimal confidence interval.

In this paper, I focus on sparse high dimensional modeling with the general decorrelated score methods proposed in Ning, Liu, et al. (2017a) and explores to which extend does this general theory applies in regression models. A research question arises about the applicability and generalization of decorrelated score methods for interest models and other available regression models. Does general application mean feasibility under multiple penalty functions, various regression models, characteristics of high-dimensional data sets, e.g., collinearity, and even different convergence conditions?. To solve them, the basic models, namely linear regression, logistic regression, and Poisson regression, are first performed with decorrelated score methods, lasso, and SCAD penalty functions to verify the general framework of decorrelated score methods. This study considers pointwise weak convergence as an asymptotic standard and characterizes the limiting behavior of decorrelated score functions. Afterward, a simulation procedure is to perform three interest models. Finally, the coverage rate of a confidence interval and statistical power of the local alternative hypothesis study the applicability of decorrelated score models.

Moreover, two applications in practice are conducted with linear regression and Poisson regression. A discussion and concerns about the simulation results are presented combined with mathematically theory to find the limitation, asymptotic property, doubts in the study and answer the research question in this paper. Finally, I give a general framework of decorrelated score function aimed at double interest parameters based on the asymptotic conditions and mathematical proof in Ning, Liu, et al. (2017a) and Ning, Liu, et al. (2017b). The simulation results imply that decorrelated score methods have high applicability and generalization in the interest models and other general models with $d = n$. This model can tolerate convex and non-convex penalty functions and a variety of regression models. However, in the case of $d \gg n$, different models show different applicability. Linear regression performs well, while Poisson regression does the opposite.

This paper is organized as follows. Section 2 describes the methodology and simulation settings. Afterward, Section 3 presents the simulation results of three interest models and gives two applications in the real world. The implication of the results is then discussed in Section 4. Lastly, section 4 illustrates some limitations and concludes the research questions.

2 Methodology

In this section, I demonstrate the application of score function concerning high dimensional models and the confidence regions for θ^*. S is defined as a multivariate random variable. Assume n is the sample size such that the database is a set $\{S_1, S_2, ..., S_n\}$. Let β denote a d dimensional parameter for this model and $\Pi = span(\beta)$. Let β^* denote the optimal estimated value of β. To explore a univariate unknown parameter, I divide the parameters into two parts where one part is the target parameter θ; The other part is regarded as nuisance parameters in a (d-1) dimensional parameter vector denoted as γ, such that $\beta = (\theta, \gamma^T)^T$. An asymptotic theory is introduced for the three interest models to achieve the limiting distribution of decorrelated score functions. I consider the linear regression, logistic regression, and Poisson regression with decorrelated score methods as interest models and make the power of statistical tests and coverage of confident regions with them. Finally, I implement a simulation study to find the coverage of the confidence and the power of the statistical test.

The estimated β is obtained by maximum penalized likelihood estimation (MPLE) for sparse high-dimensional models. The general form of MPLE is as the below equation.

$$\hat{\beta} = arg \min_{\beta \in \Pi} l(\beta) + P_{\lambda}(\beta) \tag{1}$$

where $l(\beta)$ is a loss function and $P_{\lambda}(\beta)$ is a penalty function with a tuning parameter λ. Besides, β is divided into two components: parameters of interest θ and nuisance parameter γ. In this study, $l(\beta)$ is the negative log-likelihood. Due to the wide range of selection of penalty, both convex penalty functions and non-convex penalty functions are applicable to estimate β. Therefore, lasso and SCAP are selected as candidates for penalty functions. The penalty functions are introduced as the following part.

- **The Least Absolute Shrinkage and Selection Operator** (Lasso) penalty function is a selection for maximum penalized likelihood estimator (MPLE). Tibshirani (1996) illustrated this penalty, which is a linear regression method that uses L1 regularization. The use of this L1 regularization makes the weight of some learned features achieve sparsity and feature selection. The equation of Lasso is that $p_{\lambda}(\beta) = \lambda \|\beta\|_1$.

- **Smoothly clipped absolute deviation** (SCAD) penalty function is to introduce a penalty to reduce bias in regression process Fan and Li (2001). The penalty function of this method is symmetric and non-convex, and can handle singular matrices to produce sparse solutions.

$$p_{\lambda}(\beta) = \int_0^{|\beta|} \{\lambda I(z \leq \lambda) + \frac{(a\lambda - z)_+}{(a-1)} I(z > \lambda)\} dz$$

Given the loss function $l(\beta)$, Fisher information in this study is defined as I, where $I_{\gamma\theta}$, $I_{\theta\theta}$, $I_{\gamma\gamma}$, and $I_{\theta\gamma}$ are the partitions of I. A general decorrelated score method was proposed to construct a general framework to make a score test for $\theta = 0$ and find an optimal confidence interval. The equation of this new type of score function is as the below.

$$S(\theta, \gamma) = \bigtriangledown_{\theta} l(\theta, \gamma) - w^T \bigtriangledown_{\gamma} l(\theta, \gamma) \tag{2}$$

where $w^T = I_{\theta\gamma} I_{\gamma\gamma}^{-1}$. In practice, β is estimated by MPLE while $\hat{w}$ is the estimation of Dantzig type estimator. In order to construct a valid and optimal confidence region with decorrelated score function for the interest parameter θ, this paper use one-step estimator to obtain $\tilde{\theta}$. Similar to one-step huber estimation (Bickel, 1975), one-step method are considered to resolve the multiple roots of $\hat{S}(\theta, \hat{\gamma}) = 0$. The general one-step estimator of penalized M-estimator $\hat{\theta}$ is $\tilde{\theta} = \hat{\theta} - \hat{S}(\hat{\beta})/\hat{I}_{\theta|\gamma}$, where $\hat{I}_{\theta|\gamma} = \bigtriangledown^2_{\theta\theta} l(\hat{\beta}) - \hat{w}^T \bigtriangledown^2_{\gamma\theta} l(\hat{\beta})$. Besides, the one-step estimator $\tilde{\theta}$ is asymptotic normal and semiparametrically efficient in this study.

Asymptotics

Pointwise weak convergence and uniform weak convergence are two important asymptotic theories to find the limiting distribution of decorrelated score functions under the null hypothesis. The convergence of the former is weaker than that of the latter. Ning, Liu, et al. (2017a) illustrate four conditions *Assumption* 3.1-3.4 to establish point-wise and weak convergence to give a limiting distribution of decorrelated score functions with linear regression, logistic regression, and Poisson regression, namely $\lim_{n\to\infty} |\mathbb{P}_{\beta^*}(\hat{U}_n \leq t) - \Phi(t)| = 0$. If the Corollary 4.1, 4.3, 4.4, or 4.5 hold in Ning, Liu, et al. (2017a), $\hat{U}_n$ is calculated by the below equation with the null hypothesis of $\theta = 0$.

$$\hat{U}_n = n^{1/2} \hat{S}(0, \hat{\gamma}) \hat{I}_{\theta|\gamma}^{-1/2} \quad (3)$$

It is highlighted that U_n is a density of the given n^{th} statistical series (score statistics series), which is a pivotal quantity. Besides, probability distribution of U_n does not rely on nuisance parameters as the decorrelated score function $S(\theta, \gamma)$ is uncorrelated with nuisance score functions $\bigtriangledown_\gamma l(\beta)$. Furthermore, Ning, Liu, et al. (2017a) provides another assumptions A.1-A.4 for uniform weak convergence. If the new conditions hold, new limiting behaviour arise: $\lim_{n\to\infty} \sup_{\beta^*\in\Pi_0} \sup_{t\in\mathbb{R}} |\mathbb{P}_{\beta^*}(\hat{U}_n \leq t) - \Phi(t)| = 0$, where Π_0 is a parameter space that $\Pi_0 = \{(0, \gamma) s.t. \|supp(\gamma)\| \leq supp(\beta^*), supp(\beta) \ll n\}$. Besides, the establishment of $\hat{U}_n$ under the uniform weak convergence gives the limiting behaviour for local power of alternative hypothesis. In this situation, θ is equal to $\tilde{C} n^{-\phi}$, where ϕ is set to 1/2. Based on Godambe (1991), the general estimated decorrelated score functions $\hat{S}(\theta, \hat{\gamma})$ is an approximately unbiased regarding to θ. In addition, one-step estimator $\tilde{\theta}$[1] is biased but approximately normal distributed. The optimal confidence interval of θ^* in a (1 - α) $\times$ 100% level with

[1] As the loss function is the negative log-likelihood, $I^*_{\theta|\gamma} = \sigma^*_s$, the variance of $\tilde{\theta}$ is identical to ${I^*_{\theta|\gamma}}^{-1}$.

decorrelated score functions, under the pointwise weak convergence, is constructed as the follows.

$$[\tilde{\theta} - n^{-1/2}\Phi^{-1}(1-\alpha/2)\hat{I}_{\theta|\gamma}^{-1/2}, \tilde{\theta} + n^{-1/2}\Phi^{-1}(1-\alpha/2)\hat{I}_{\theta|\gamma}^{-1/2}] \tag{4}$$

The proof of the above limiting behaviour is shown in Ning, Liu, et al. (2017a) and Ning, Liu, et al. (2017b).

Interest Models with Decorrelated Score Methods

In this part, the general framework of decorrelated score methods is applied in linear regression, logistic regression, and Poisson regression. Define $Q_i = (Z_i, X_i^T)^T$, a sub-Gaussian vector under the above assumptions, as the collection of all covariates for observation i, where $Z_i \in \mathbb{R}$, $X_i \in \mathbb{R}^{d-1}$. Let λ' denote a tuning parameter for the Dantzig selector (Candes, Tao, et al., 2007). The general estimation procedure of decorrelated score function is as follows: First, it is required to calculate a penalized M-estimators $\hat{\beta}$ with convex or nonconvex penalty functions in equation 1, where the estimation process is operated by cross-validations with tuning parameter λ. Next, employing a Dantzig type estimator $\hat{w}$ is to calculate an approximate value of w with the tuning parameter λ'. Finally, the decorrelated score function $\hat{S}(\theta, \hat{\gamma})$ is estimated as equation 2.

Linear Regression with Decorrelated Score Method

The general from of linear regression with decorrelated score methods is $Y_i = \theta^* Z_i + \gamma^T X_i + \varepsilon_i$, where ε is the error item satisfying $\mathbb{E}(\varepsilon_i) = 0$ and homoscedasticity assumption, namely $\mathbb{E}(\varepsilon_i^2) = \sigma^2$, for i = 1,2,...,n. Based on Ning, Liu, et al. (2017a), Bickel et al. (2009), and Javanmard and Montanari, 2014, the negative Guassian quasi log-likelihood is that $l(\beta) = (2n)^{-1}\sum_{i=1}^{n}(Y_i - \beta^T Q_i)^2$. Thus, the estimated decorrelated score function of linear regression with the known σ is as the below equation.

$$\hat{S}(0, \hat{\gamma}) = -\frac{1}{\sigma^2 n}\sum_{i=1}^{n}(Y_i - \gamma^T X_i)(Z_i - \hat{w}^T X_i)$$

where $\hat{w}$ is the estimator of Dantzig Selector, namely, $\hat{w} = argmin\|w\|_1$ such that

$$\left\| \frac{1}{n}\sum_{i=1}^{n} X_i(Z_i - w^T X_i) \right\|_\infty \leq \lambda'$$

In this case, the estimation of Fisher information is $\hat{I} = \frac{1}{\sigma^2 n}\sum_{i=1}^n Q_i Q_i^T$ and partial Fisher information is as the following equation.

$$\hat{I}_{\theta|\gamma} = \frac{1}{\sigma^2}\left\{\frac{1}{n}\sum_{i=1}^n Z_i^2 - \hat{w}^T\left(\frac{1}{n}\sum_{i=1}^n X_i Z_i\right)\right\} \tag{5}$$

The test statistic $\hat{U}_n$ under the null hypothesis in this model is that

$$\hat{U}_n = -\frac{1}{\sigma n^{1/2}}\sum_{i=1}^n (Y_i - \gamma^T X_i)(Z_i - \hat{w}^T X_i)\left\{\frac{1}{n}\sum_{i=1}^n Z_i^2 - \hat{w}^T\left(\frac{1}{n}\sum_{i=1}^n X_i Z_i\right)\right\}^{-1/2} \tag{6}$$

The error variance is unknown in many cases, but decorrelated score functions in linear regression are highly dependent on the standard deviation σ. Especially in practical applications, we usually estimate standard deviation $\hat{\sigma}$ of the regression error. With the estimated $\hat{\beta}$, $\hat{\sigma}$ can be obtained with the equation $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}^T Q_i)^2$. $\tilde{U}_n$ is that

$$\tilde{U}_n = -\frac{1}{\hat{\sigma} n^{1/2}}\sum_{i=1}^n (Y_i - \hat{\gamma^T} X_i)(Z_i - \hat{w}^T X_i)(H_Z - \hat{w} H_{XZ})^{-1/2} \tag{7}$$

where $H_Z = \frac{1}{n}\sum_{i=1}^n Z^2$ and $H_{XZ} = \frac{1}{n}\sum_{i=1}^n Z_i X_i$. The asymptotic properties show $\tilde{U}$ and $\hat{U}$ are uniformly asymptotically equivalent, where Corollary 4.3 in Ning, Liu, et al. (2017a) gives the asymptotic distribution of $\tilde{U}_n$ with the null hypothesis.

Logistic Regression with Decorrelated Score Method

As introduced before, logistic regression with decorrelated score function is the model of the interest. This model in this study is used for binary classification with high dimensional explanatory variables. Since the Logistic regression model is an extension of linear regression applied to the category, which implies that this regression model satisfies the assumptions of Corollary 4.1-4.3 in Ning, Liu, et al. (2017a), and its test statistical $\hat{U}_n$ approximation conforms to the standard normal distribution. The estimated decorrelated score function is as the below equation with the estimated Dantzig Selector $\hat{w}$ under the hypothesis of $\theta = 0$.

$$\hat{S}(0, \hat{\gamma}) = -\frac{1}{n}\sum_{i=1}^n (Y_i - \frac{exp(\hat{\gamma}^T X_i)}{1 + exp(\hat{\gamma}^T X_i)})(Z_i - \hat{w}^T X_i)$$

The Dantzig Selector $\hat{w}$ in this case is estimated with the objective function $\hat{w} = argmin\|w\|_1$ such that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \frac{exp(\hat{\beta}^T Q_i)}{(1 + exp(\hat{\beta}^T Q_i))^2} (Z_i - w^T X_i) X_i \right\|_\infty \leq \lambda'$$

The equation of test statistics $\hat{U}_n$ is identical to another two interest models. One of the distinct is partial Fisher information; See the below equation.

$$\hat{I}_{\theta|\gamma} = \frac{1}{n} \sum_{i=1}^{n} \frac{exp(\hat{\beta}^T Q_i)}{(1 + exp(\hat{\beta}^T Q_i))^2} Z_i (Z_i - w^T X_i) \tag{8}$$

The test statistic $\hat{U}_n$ here is that

$$\hat{U}_n = -\sum_{i=1}^{n} (Y_i - \frac{exp(\hat{\gamma}^T X_i)}{1 + exp(\hat{\gamma}^T X_i)})(Z_i - \hat{w}^T X_i)\{\sum_{i=1}^{n} \frac{exp(\hat{\beta}^T Q_i)}{(1 + exp(\hat{\beta}^T Q_i))^2} Z_i (Z_i - w^T X_i)\}^{-1/2} \tag{9}$$

Poisson Regression with Decorrelated Score Method

Another interest model in this study is Poisson regression, which is another generalized linear regression form. Likewise, I calculate the estimated decorrelated score function, confidence interval to make further simulation and test statistics. $\hat{U}_n$ of this model. The decorrelated score function is shown below with the hypothesis of $\theta = 0$.

$$\hat{S}(0, \hat{\gamma}) = -\frac{1}{n} \sum_{i=1}^{n} (Y_i - exp(\hat{\gamma} X_i))(Z_i - \hat{w}^T X_i)$$

where $\hat{w} = \text{argmin}\|w\|_1$, such that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} exp(\hat{\beta}^T Q_i)(Z_i - w^T X_i) X_i \right\|_\infty \leq \lambda'$$

In addition, the partial Fisher information in this model is that

$$\hat{I}_{\theta|\gamma} = \frac{1}{n} \sum_{i=1}^{n} exp(\hat{\beta}^T Q_i) Z_i (Z_i - w^T X_i) \tag{10}$$

Here, the test statistic $\hat{U}_n$ in this model is that

$$\hat{U}_n = \sum_{i=1}^{n} (Y_i - exp(\hat{\gamma} X_i))(Z_i - \hat{w}^T X_i)\{\sum_{i=1}^{n} exp(\hat{\beta}^T Q_i) Z_i (Z_i - w^T X_i)\}^{-1/2} \tag{11}$$

Simulation

In this simulation study, a derived confidence interval, power of statistical test with decorrelated score function are covered and performed. Besides, this procedure is based on the pointwise weak convergence to explore the limiting behavior under the null and alternative hypothesis. Moreover, the implementation of confident intervals' coverage involves the Markov Chain Monte Carlo (MCMC) simulation process. The interest models here consist of linear regression, logistic regression, and Poisson regression with decorrelated score methods.

Simulation Procedure

Throughout the simulation study, I first set the data generator process (DGP) of the covariates X: $n = 100$ independent and identical distribution samples with a multivariate Gaussian distribution $N_d(0, \Sigma)$, where $d = 100, 200, 500$ and Σ is a diagonal-constant matrix with $\Sigma_{ij} = \rho^{|i-j|}$. ρ has four potential values, namely, 0.25, 0.4, 0.6, and 0.75. The magnitude of ρ determines the strength of the collinearity of the data from DGP. For the true value of β^*, it satisfies $\|\beta^*\|_0 = s$, where $\beta_S = (1, ..., 1)$ is Dirac measure with $s = 3$. For the simulation process of linear regression, there is a standard Gaussian noise assumed in DGP, that is, $Y = X\beta^* + \varepsilon$, where ε is a $n \times 1$ vector following standard normal distribution. Regarding to generator of Y in logistic regression, Y is assumed to follow the binomial distribution with the probability of success on each trial is equal to $\frac{1}{1+exp(-X\beta^*)}$. In terms of Poisson regression, Y is following Poisson distribution with a vector of non-negative means $exp(X\beta^*)$.

This simulation study mainly studies several aims: finding the impact of different penalties and different degrees of high-dimensional d on coverage and powers of statistical tests. First, the three interest models are conducted with the $n = d$. Interest models take maximum penalized likelihood separately with lasso and SCAD penalty functions. Then, for the discussion and comparison of the two penalty equations, a penalty formula is selected to do a higher-dimensional simulation. It is emphasized that the selection of penalty items is not based on the simulation results that best meet the theoretical expectations. On the contrary, the selection of penalty functions should consider and weigh the theoretical expectations and the actual performance of the simulation. For example, the lasso is regarded as a penalty function that causes a relatively large bias by many researchers. Thus, it is more likely to produce rather unsafe or unreasonable simulation results during the simulation process. However, suppose the interest model with lasso does not have abnormal outcomes. In that case, it is acceptable to use lasso for higher-dimensional models since this paper aims to explore the generalization of the general theory of decorrelated score methods and tolerance to different parameters and penalties. SCAP is a non-convex penalty term. Typically, it is considered to be better than lasso in the process of debias. However, this does not mean that SCAP will be selected as a higher-dimensional simulation study, as described above. Then, the performance of the interest model in a higher dimension, that is, $d \gg n$, becomes the focus of further research. Here, the ratio of the dimensions d and n is 2 and 5.

In each simulation process, DGP first generates covariates. Then according to different models and settings, I calculate Y. Next, calculate penalized M-estimators and Dantzig type estimators to get $\hat{U}_n$ and part Fisher Information $\hat{I}_{|\hat{\gamma}}$, according to equation 6, 9, or 11[2]. Then, the confidence interval coverage and power of the statistical test are calculated. The hypothesis in this study is H_0 with $\beta = 0$ versus local alternative H_1 with $\beta_1 \neq 0$. The tuning parameters λ and λ' are corresponding to the cross-validated loss for the lasso or SCAD penalties of MPLE and the Dantzig Selector. The alternative values β_1 in this simulation study ranges from 0.05 to 0.55 to evaluate the power of the selected tests

There are 500 times replications for each simulation process. A global count variable records whether the actual value θ is contained in the estimated confident interval in each replication with a particular model. If it holds, the count variable will add 1, otherwise 0. Then, I can obtain the coverage by the calculation of the proportion of the containing times. For the power of the statistical test, with various values of ρ and interest models, I calculate the probability that rejects the null hypothesis of $\theta = 0$, where the alternative hypothesis H_a of a particular value between 0 and 0.55 is assumed true.

In practice, I exploit "cv.glmnet"[3] to realize the linear model via Lasso penalized maximum likelihood with k-fold cross-validation for λ. Likewise, "cv-gds"[4] is applied to make Dantzig Selector with the cross-validation to obtain an optimal λ. Then, we plugin this optimum into "gds"[5] and obtain w, which is the coeffient of Dantzig Selector. Other reference R packages are "MASS"[6], "ncvreg"[7], and "fastclime"[8]. The detailed tutorial is explained in the Appendix.

[2]The details of this process is introduces in Section Methodology
[3]cv.glmnet, R: https://cran.r-project.org/web/packages/glmnet/glmnet.pdf
[4]cv-gds, R: https://cran.r-project.org/web/packages/hdme/hdme.pdf
[5]*gds*, R: https://cran.r-project.org/web/packages/hdme/hdme.pdf
[6]R, https://cran.r-project.org/web/packages/MASS/MASS.pdf
[7]R,https://cran.r-project.org/web/packages/ncvreg/ncvreg.pdf
[8]R,https://cran.r-project.org/web/packages/fastclime/fastclime.pdf

Assessment and comparison

Once all simulation estimates are completed, I will evaluate the decorrelated score methods with the linear regression model, logistics regression model, and Poisson regression through coverage rate and power of statistics. For each type of model, a confidence interval of coverage is estimated to find whether (1-α) significance level. Due to the fixed simulation times, random and independent generation, and binary outcomes, binomial proportion confidence interval is employed to construct a reasonable region for the coverage proportions $\hat{p}$. Based on the central limit theorem, I approximated $\hat{p}$ to the coverage probability, which approximately follows a normal distribution Wallis (2013). **Clopper-Pearson exact method** is exploited to obtain a relatively accurate confidence interval. For the power of a statistical test, this paper will test whether rejection rates of null hypothesis will converge to 1 as the local alternative hypothesis β_1 increases.

3 Results

In this section, to explore the applicability and generalization of general theories to the interest model, I present the findings of simulation results and real-world applications with decorrelated score methods via the statistical tests and confidence intervals (CI) of coverage. For simulation results, this paper first presents the estimated coverage rates for interest models and corresponding CIs with different ρ and different penalties in Tables 1-2. Afterward, I consider the case of d$\gg$n, and the corresponding simulation is set to d=200,500>n=100 to study the performance of decorrelated score functions with interest models under high-dimensional data sets. The simulation results for the case d with lasso penalty are exhibited in Table 3-4. The abnormal results are marked in bold font. In addition, a noteworthy but normal value is marked in red. Figures 1-4 reveal the power of the given interest models with lasso and SCAP. Moreover, two databases applied in the real world are to find confidence regions of target parameters and give hypothesis tests in practice.

Findings for coverage of confidence interval with decorrelated score methods

Following the proposed procedure, the estimated coverage rates and corresponding confidence intervals are shown with interest models. Due to the insufficiency of normal approximation, especially for a large estimated coverage $\hat{p}$, Clopper-Pearson exact CI is employed to obtain more accurate and credible ranges of coverage rates.

Table 1: Coverage of CI with the decorrelated score method with Lasso Penalty and d=100 for the linear regression, logistics regression, and Poisson regression at 5% significance level

	$\rho = 0.25$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.75$
Linear Regression				
Coverage at 95% significant level	0.9400	0.9520	0.9480	0.9600
CI of Coverage via Clopper-Pearson exact	[0.9155,0.9592]	[0.9294,0.9690]	[0.9247,0.9658]	[0.9389,0.9754]
Logistics Regression				
Coverage at 95% significant level	0.9540	0.9620	0.9560	0.9660
CI of Coverage via Clopper-Pearson exact	[0.9318,0.9706]	[0.9413,0.9770]	[0.9341,0.9722]	[0.9461,0.9801]
Poisson Regression				
Coverage at 95% significant level	0.9360	0.9320	0.9300	0.9400
CI of Coverage via Clopper-Pearson exact	[0.9108,0.9558]	[0.9063,0.9525]	[0.9040,0.9508]	[0.9155,0.9592]

Table 2: Coverage of CI with the decorrelated score method with SCAP Penalty and d=100 for the linear regression, logistics regression, and Poisson regression at 5% significance level

	$\rho = 0.25$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.75$
Linear Regression				
Coverage of 95% significant level	0.9320	0.9380	0.9400	0.9300
CI of Coverage via Clopper-Pearson exact	[0.9063,0.9525]	[0.9133,0.9560]	[0.9155,0.9592]	[0.9040,0.9508]
Logistics Regression				
Coverage of 95% significant level	0.9620	0.9680	**0.9720**	0.9500
CI of Coverage via Clopper-Pearson exact	[0.9413,0.9770]	[0.9486,0.9816]	[**0.9535**,**0.9846**]	[0.9271,0.9674]
Poisson Regression				
Coverage of 95% significant level	0.9380	0.9360	0.9320	0.9340
CI of Coverage via Clopper-Pearson exact	[0.9133,0.9560]	[0.9108,0.9558]	[0.9065,0.9509]	[0.9088,0.9526]

For the three interest models, with the Lasso penalty, the estimation consequences with various simulation settings ρ are presented in Table 1. Likewise, the results with the SCAP penalty and the same simulation settings are shown in Table 2. The estimated coverage rates with linear regression and Poisson regression are all around 95%, and their confidence intervals all contain 95%. Undoubtedly, linear regression and Poisson regression with penalties perform well as expected when n=d. Thus, it shows that decorrelated score functions have strong applicability in linear regression and Poisson regression.

Moreover, it implies that these results demonstrate the versatility and generalization with penalty items. Both Lasso (convex penalty) and SCAP (non-convex penalty) can be applied to this case and obtain reliable simulation results. The left boundary of the confidence interval for coverage with $\rho = 0.6$ and SAPC penalty is larger than 0.95. The abnormal confidence interval raises questions about the general applicability of decorrelated score function. This result shows that a theoretically expected value of 95% coverage is far lower than the actual value of 95%. Although the excessively high real coverage value violates the expectation of the theoretical value, the overly high coverage rate is still a safe and effective embodiment. It implies that this model that deviates from the theoretical expectations still has a certain degree of feasibility and credibility. Decorrelated score methods are described as a general framework to solve high-dimensional problems, such as exploring the limiting distribution in a specific condition of a sample. It is highlighted that this "general form" is not only feasible for regression models but also for various collinearity and penalty terms. However, the selection of penalties will affect the performance of decorrelated score methods based on Tables 1-2. LASSO penalty is considered a relatively better choice with n=d. But this advantage is not apparent enough. In other words, the research limit is not enough to completely negate the non-convex penalty. Section 4 elaborates on the limitations of the study.

Table 3: Coverage of CI with the decorrelated score method with Lasso Penalty and d=200 for the linear regression, logistics regression, and Poisson regression at 5% significance level

	$\rho = 0.25$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.75$
Linear Regression				
Coverage of 95% significant level	0.9380	0.9580	0.9360	0.9580
CI of Coverage via Clopper-Pearson exact	[0.9131,0.9575]	[0.9365,0.9738]	[0.9108,0.9558]	[0.9365,0.9738]
Logistics Regression				
Coverage of 95% significant level	0.9500	0.9640	0.9680	0.9620
CI of Coverage via Clopper-Pearson exact	[0.9271,0.9674]	[0.437,0.9785]	[0.9486,0.9816]	[0.9413,0.9770]
Poisson Regression				
Coverage of 95% significant level	0.9480	0.9300	**0.9280**	**0.9260**
CI of Coverage via Clopper-Pearson exact	[0.9247,0.9658]	[0.9040,0.9508]	[**0.9017**,**0.9491**]	[**0.8994**,**0.9474**]

The consequences from Tables 3-4 reveal an interesting phenomenon. The coverage rate obtained by linear regression and logistic regression with decorrelated score functions is normal. Moreover, the corresponding CIs contain 95%, which means that with higher dimensions, decorrelated score methods are still suitable for linear regression and logistic regression from the perspective of coverage.

Table 4: Coverage of CI with the decorrelated score method with Lasso Penalty and d=500 for the linear regression, logistics regression, and Poisson regression at 5% significance level

	$\rho = 0.25$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.75$
Linear Regression				
Coverage of 95% significant level	0.9340	0.9460	0.9360	0.9460
CI of Coverage via Clopper-Pearson exact	[0.9086,0.9541]	[0.9224,0.9641]	[0.9108,0.9558]	[0.9224,0.9641]
Logistics Regression				
Coverage of 95% significant level	0.9680	0.9620	0.9660	0.9520
CI of Coverage via Clopper-Pearson exact	[0.9486,0.9816]	[0.9413,0.9770]	[0.9461,0.9801]	[0.9294,0.9690]
Poisson Regression				
Coverage of 95% significant level	0.9320	**0.9280**	**0.9260**	0.9300
CI of Coverage via Clopper-Pearson exact	[0.9063,0.9525]	**[0.9017,0.9491]**	**[0.8994,0.9474]**	[0.9040,0.9508]

However, some coverage rates corresponding to Poisson regression are in irregular confidence intervals, especially when ρs tend to be large. The right boundary of this abnormal confidence interval was less than 95%. Wrong and unreasonable confidence intervals negate the applicability of decorrelated score functions in Poisson regression to a certain extent. Regardless of case d=200 or d=500, decorrelated score methods cannot be fully applied to Poisson regression. Nevertheless, slight collinearity (when ρ=0.25) does not seem to affect the coverage negatively. However, there is a doubt here that when d = 500 and ρ = 0.75, the coverage of CI with decorrelated score method is at a normal level, which is expected to be abnormal. In section 4, I will discuss this phenomenon.

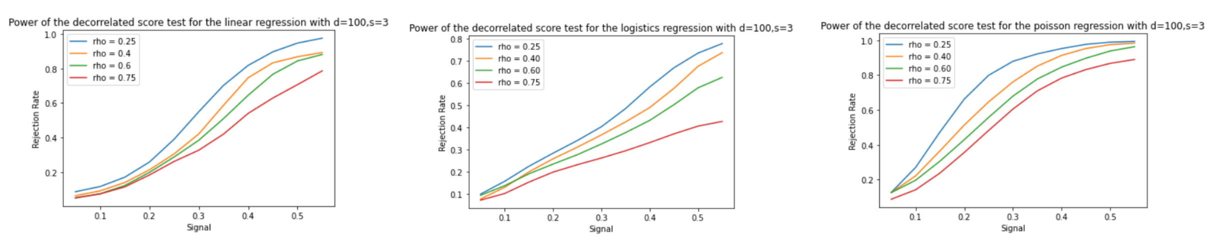

Figure 1: Power of the decorrelated score test for the linear regression (left panel), logistic regression (middle panel), and Poisson regression (right panel) with LASSO Penalty, n = 100, d = 100, s = 3

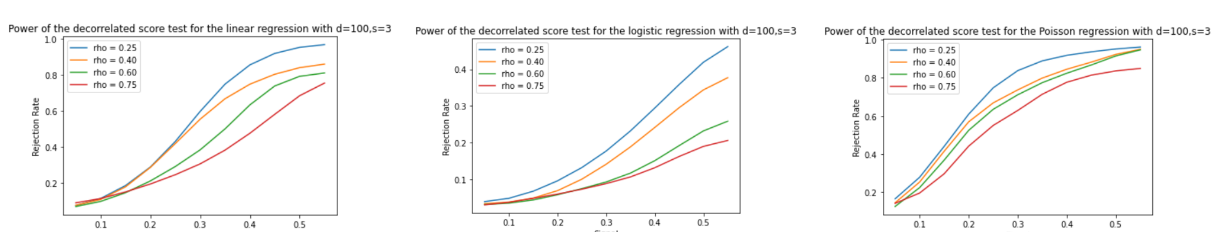

Figure 2: Power of the decorrelated score test for the linear regression (left panel), logistic regression (middle panel), and Poisson regression (right panel) with SCAP Penalty, n = 100, d = 100, s = 3

Figures 1-4 reveal the statistical power of the hypothesis test. Under the assumption of the true alternative hypothesis, the rejection rate of the null hypothesis is calculated to obtain statistical power. It is highlighted that with the increase of the local alternative hypothesis from 0 to 0.55, the probability of accepting the null hypothesis of $\theta = 0$ reduces for all the cases. Besides, multicollinearity negatively affects the rejection rates of the null hypothesis. ρ here represents the magnitude of the multicollinearity of the simulated data set and positive correlation. Strict multicollinearity (large ρ) restricts rejection rates.

On the contrary, weak collinearity promotes rejection rates of the null hypothesis to converge to 1. In Figure 1-2, rejection rates of interest models with different penalties present the same convergence trend, which shows that different types of penalty items have the approximately same impact on statistical power. Similarly, the performance of the interest model with high-dimensional data set in rejection rates displays the same convergence as in Figure 1-2. It is worth noting that the curve of logistic regression with decorrelated score methods shows lower rejection rates of the null hypothesis. A low statistical power implies a high risk of committing Type II errors.

The rejection rate of logistic regression with decorrelated score methods and high dimensional data sets is relatively insensitive to signals' increase. In summary, the application of decorrelated score functions in three interest models seems general and reliable and has good applicability and practicality if the dimension does not high.

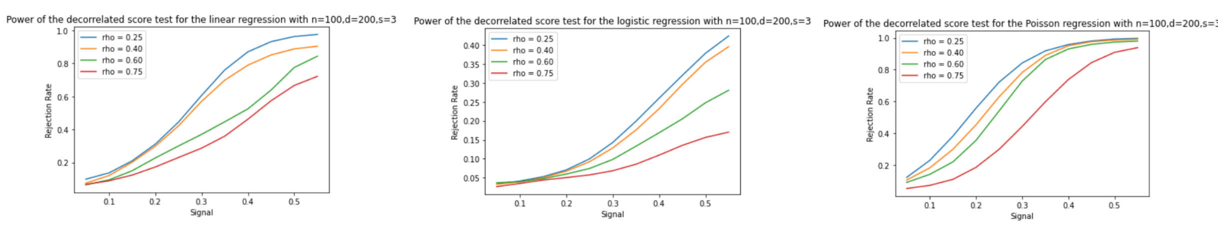

Figure 3: Power of the decorrelated score test for the linear regression (left panel), logistic regression (middle panel), and Poisson regression (right panel) with LASSO Penalty, n = 100, d = 200, s = 3

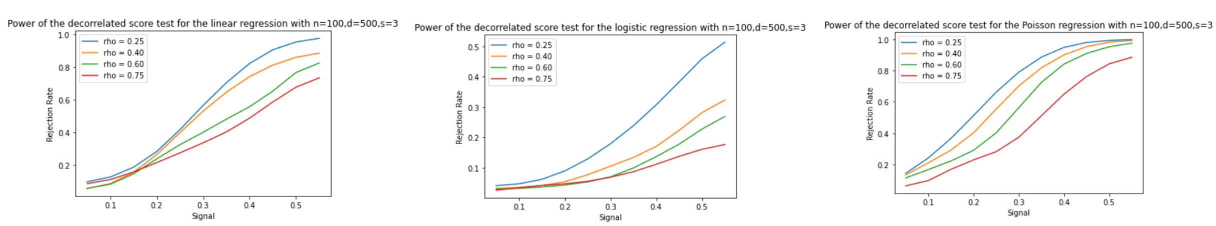

Figure 4: Power of the decorrelated score test for the linear regression (left panel), logistic regression (middle panel), and Poisson regression (right panel) with LASSO Penalty, n = 100, d = 500, s = 3

Real-World Data

The application with decorrelated score functions is divided into two parts regarding the real-number data sets and count data sets, which are investigated separately via linear and Poisson regression.

Application 1: Investment Portfolios

The first application is to use linear regression with a high-dimensional database. I consider high-dimensional investment portfolios formed on size (market equity) and investment. This data set is made available by "Portfolios of Investment" (2021). The data set is collected from Dec, 08, 2020 - Apr, 30, 2021 and contains $n = 100$ samples and $d = 100$ covariates corresponding to 100 investment portfolios. These portfolios for the given periods include all NYSE, AMEX, and NASDAQ stocks. The dependent variable $R_{i,t} - r_{F,t}$ is a time series of Apple Inc. (AAPL) stock returns at time t ("Apple Inc." (2021)) minus risk-free rate of return $r_{F,t}$. The linear regression model is as the below, where α_i is a constant item and ε_{it} is the error item.

$$R_{i,t} - r_{F,t} = \alpha_i + \Gamma^T X + \varepsilon_{it} \tag{12}$$

Portfolio "SMALL LoBM" in this study is the interest covariate. In this application, I make a null hypothesis that the "SMALL LoBM" parameter is equal to zero versus an alternative hypothesis of the wrong null. Γ is first estimated via equation 1. As the variance of residuals ε_{it} is unknown, this paper first calculates the estimated variance $\hat{\sigma}$ and plugin this estimator to obtain $\tilde{U}_n$ by equation 7. Next, I estimate the partial Fisher information via equation 5. The final step is to find the confident interval with decorrelated score methods. The result is that $|\tilde{U}_n| = |-0.384| \leq 1.96$, where confidence interval : $[-0.216, 0.114]$ with 5% significance level. It implies that interest covariate "SMALL LoBM" is not significant from zero. Besides, zero is contained in the estimated confidence interval, which gives the above hypothesis verification a verification.

Application 2: The number of Died Drivers in UK

As the second application, I consider a portfolios formed on size and operating data set from "Portfolios of Operating Profitability" (1972), which contains $n = 100$ samples and $d = 100$ covariates regarding to the intersections of size (market equity) and profitability. Dependent variable Y in this case is the number of death of drivers applied in U.K. from 01/1969 to 12/1972 (Harvey and

Durbin (1986)). A Poisson regression model is constructed with Y and the given covariates. Portfolio "ME1 OP2" is considered as interest explanatory variables. The null hypothesis is that the parameter of "SMALL LoBM" is equal to zero. By equation 10, 11, and confident interval formula 4, $|\hat{U}_n| = |-0.42| \leq 1.96$, which means the interest parameter is not significant from zero. Besides, the confident interval of interest parameter is that $[-0.113, 0.291]$ with 5% significance level.

4 Conclusion

I first elaborated on some concerns combined with simulation results and econometric theory. Subsequently, a summary of the simulation results was drawn to solve the research problem of this paper, that is, to explore the applicability of the general theory of decorrelated score methods in the interest model. Moreover, this paper discusses some shortcomings, constraints, and unresolved research problems. Finally, I introduced future research to further explore the general theory of decorrelated score methods and give a general calculation algorithm for double-dimensional interest parameters with decorrelated score methods.

Discussion

- **Concern 1: Limiting distribution of $\hat{U}_n$.** Ning, Liu, et al. (2017a) illustrates an approximate normal distribution for the statistics test $\hat{U}$ of decorrelated score functions. However, in the actual case, the test statistic generated $\hat{U}_n$ by the simulation occasionally has extreme values, which causes the tails to be fatter in the sample distribution. Heavy tails lead theoretically approximated normal distribution closer to t-distribution.

- **Concern 2: Abnormal simulation results of coverage**. In Section 3, some abnormal simulation results have raised questions about the applicability of decorrelated score methods and the exploration of factors that affect generalization. One speculation about the influencing factors is that too high a dimension may increase abnormal simulation coverage. A consistency condition about the parameter estimation introduced in Ning, Liu, et al. (2017a) is that $\|\hat{\beta} - \beta^*\|_1 = \emptyset_P(s^* \sqrt{log \frac{d}{n}})$, where $s^* = \|\beta^*\|_0$, $\emptyset_P(n)$ is a sequence converging to 0 with $n \to \infty$. However, with the increase of d and fixed n, the value of $\|\hat{\beta} - \beta^*\|$ increase as the grow of magnitude $\emptyset_P$. It implies that there is an obvious bias with the extremely high dimension. Likewise, the consistency of Dantzig selector has a identical magnitude. That is, $\|\hat{w} - w^*\|_1 = \emptyset_P(\|supp(w^*)\|^* \sqrt{log \frac{d}{n}})$. The bias of $\hat{w}$ is also positive affected by dimension.

This explains that too high dimensionality may cause too high bias, which may restrict the applicability of decorrelated score methods.

- **Concern 3: Selection of penalty items**. Convex and nonconvex penalties lead to distinct (de)biases of the penalized M-estimator $\hat{\theta}$. One-step estimator $\tilde{\theta}$ is bias to some extent. $\tilde{\theta}$ is difference of $\hat{\theta}$ and $\frac{\hat{S}(\beta)}{\hat{I}_{\theta|\gamma}}$. Based on theorem 3.2 in Ning, Liu, et al. (2017a), it can infer that $S(\beta^*) = \text{o}(1)$, where $S(\beta^*)$ approximately follows normal distribution with the mean of 0. Thus, there is not evidence to show $\mathbb{E}(\tilde{\theta}) = 0$ under the simulation settings.

- **Concern 4: Uniformly weak convergence**. In result 3, linear regression with decorrelated score method presents simulation results as the theoretical expectations. Besides, uniform convergence in Corollary 4.1 holds with Theorem A.1 Ning, Liu, et al. (2017b). However, no proof to show logistic regression and Poisson regression follow uniform weak convergence. Perhaps it is one of the reasons that simulation results of linear regression perform better than others if the two models did not satisfy uniform weak convergence.

- **Concern 5: Confusing simulation results marked in red**. In Table 4, the result is marked in red give a confusing conclusion. Coverage "0.93" denotes 35 times that true value is not contained in the estimated confident interval. This simulation result seems acceptable at a 95% significance level. However, if there are 36 times for abnormal situations, I will draw a different conclusion. Insufficient simulation times may cause inaccurate conclusions at some boundary points.

Summary of this study

The decorrelation score method is a new solution for the issues in high-dimensional models. Classical Rao's score methods with the Lasso penalty pose a challenge in untractable limiting distribution due to bias and sparsity. This newly proposed method presents its general application in estimation approaches and limiting behavior to solve the above problems. This study discusses the general application of the high-dimensional sparse model with decorrelated score functions and compares the performance of distinct penalty functions and various simulation settings. Linear regression, logistic regression, and Poisson regression are applied to test the applicability and generalization of decorrelated score methods. Coverage rates of confidence regions with the decorrelated score function and power of statistic test in section 3 show that this new method has extremely high applicability in linear regression, regardless of the degree of collinearity, selection of penalty functions, or the dimensionality of the data set. However, the applicability of decorrelation score methods in other models is limited. Logistic regression with decorrelated score methods is used in various simulation settings, but strict multicollinearity dramatically increases the risk of committing Type II errors in terms of power. Moreover, decorrelated score methods are restricted by some factors in the application of Poisson regression. Studies have shown that higher dimensions, such as d = 200 and 500, and multicollinearity may cause poor performance of decorrelated score methods in coverage rates for Poisson regression. It is undeniable that the general form of decorrelation score methods solves the regression problem of interest models in high-dimensional data sets but needs to avoid multicollinearity or too high dimensions, especially for Poisson regression. Besides, the choice of penalty functions also plays a significant role in model performance. In this study, two selected penalty functions have little impact on statistical power, but there is a difference in coverage rates, which may violate the expectations of statistical theory. The pointwise weak convergence is the core asymptotical standard for the limiting behavior of the new sore function, which requires all cases to satisfy its assumptions.

To sum up, in lower dimensions, such as $d = n$, three interest models usually perform well. However, in the high dimension, $d \gg n$, the decorrelated score method seems not to achieve high applicability in logistic regression and Poisson regression. In general, decorrelated score methods are widely applicable to convex penalty terms and non-convex penalty terms, as well as a variety of regression models, under the assumption of pointwise weak convergence but is restricted with too high dimension and strict multicollinearity.

Limitations

In this study, the simulation setting restricts the conclusions of this study: Five hundred times of replications of simulation may result in a puzzled simulation result. Concern 5 in subsection Discussion provides a corresponding example. Simulation replication times are a compromise between calculation time and simulation accuracy. Besides, there are still some questions that remained. First, can the general framework of decorrelated score functions be fully approved only by coverage and statistical power? Statistical power is usually related to the sample sizes, which is little related to test model performance. Hence, the test of the applicability is questioned by trying the above two statistical methods. Second, is the weak sparse matrix model applicable for this general theory? There is no reference to verify it. Dantzig Selector parameter w depends on β, but how much influence does this dependency have on this research? Third, is there a specific penalty function to significantly reduce the bias of estimated parameters, especially for the one-step estimator and decorrelated score functions?

Future Research

Future research here is further to explore the applicability and improvement of decorrelated score methods. One such point is that I can study whether uniform weak convergence, a more robust gradual theory, will improve the performance of the usual forms of decorrelation score methods. Previous theories have proved that this kind of convergence is applicable in the linear regression model, so if we are more stringent on the assumptions, the performance of Poisson regression may meet the theoretical expectations. In addition, another direction is to test the performance of decorrelated score methods to the multi-dimensional case.

Based on Ning, Liu, et al. (2017b), I can get an initial inference of the two-dimensional interest parameters with decorrelated score methods. Likewise, I assume there is a $\theta = (\theta_1, \theta_2)$, and γ here is the nuisance parameter whose dimension is $d_1 = d - 2$. The null hypothesis here become $H_0^2 : \theta^* = (0, 0)$. In this study, negative log-likelihood is still an estimation approach to simply formula inference. The general score function is similar to equation 2. The new score function is as the below:

$$\mathbf{S}_{\theta,\gamma)} = \bigtriangledown_\theta l(\theta, \gamma) - \mathbf{W}^T \bigtriangledown_\gamma l(\theta, \gamma)$$

where $\mathbf{W}^T = \mathbf{I}_{\theta\gamma}\mathbf{I}_{\gamma\gamma}^{-1} \in \mathbb{R}^{2\times d-2}$. And $\mathbf{W} = (\hat{W}_1, \hat{W}_2)$. $\mathbf{W}$ is estimated by the formula $\mathbf{W}_i = argmin\|w\|_1$, $i = 1, 2$. Similar to equation 2, $\hat{W}_i$ is estimated as the same approach as the case with a univariate interest parameter. $\hat{U}_n^2 = n[\hat{S}(0,\gamma)]^T \hat{I}_{\theta|\gamma}^{-1}[\hat{S}(0,\hat{\gamma})]$, which approximately follows χ_2^2, based on the theorem 3.1 and pointwise weak convergence in Ning, Liu, et al. (2017a). The partial Fisher Information is that $\mathbf{I}_{I\theta|\gamma} = \bigtriangledown_{\theta\theta}^2 l(\hat{\beta}) - \mathbf{W}^T \bigtriangledown_{\gamma\theta}^2 l(\hat{\beta})$. Finally, the confidence interval of one-step estimator $\tilde{\theta}$ is given by $[\tilde{\theta} - n^{-1/2}\phi(1-\alpha/2)(c^T \hat{I}_{\theta|\gamma} c)^{-1/2}, \tilde{\theta} + n^{-1/2}\phi(1-\alpha/2)(c^T \hat{I}_{\theta|\gamma} c)^{-1/2}]$, where c is constant vector.

References

Apple inc. (2021). https://finance.yahoo.com/quote/AAPL?p=AAPL&.tsrc=fin-srch (accessed: 30.04.2021)

Athey, S., Imbens, G. W., & Wager, S. (2016). Approximate residual balancing: De-biased inference of average treatment effects in high dimensions. *arXiv preprint arXiv:1604.07125.*

Belloni, A., & Chernozhukov, V. (2011). High dimensional sparse econometric models: An introduction. *Inverse problems and high-dimensional estimation* (pp. 121–156). Springer.

Belloni, A., Chernozhukov, V., & Hansen, C. (2011). Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*.

Bickel, P. J. (1975). One-step huber estimates in the linear model. *Journal of the American Statistical Association*, *70*(350), 428–434.

Bickel, P. J., Ritov, Y., Tsybakov, A. B., et al. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of statistics*, *37*(4), 1705–1732.

Buehlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics*, *34*(2), 559–583.

Bunea, F., Tsybakov, A., & Wegkamp, M. (2007). Sparsity oracle inequalities for the lasso. *Electronic journal of statistics*, *1*, 169–194.

Candes, E., Tao, T. et al. (2007). The dantzig selector: Statistical estimation when p is much larger than n. *Annals of statistics*, *35*(6), 2313–2351.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(5), 849–911.

Fan, J., Lv, J., & Qi, L. (2011). Sparse high-dimensional models in economics.

Godambe, V. P. (1991). Estimating functions: An overview. *Estimating functions*.

Greenshtein, E., & Ritov, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli*, *10*(6), 971–988.

Harvey, A. C., & Durbin, J. (1986). The effects of seat belt legislation on british road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society: Series A (General)*, *149*(3), 187–210.

James, G. M., Radchenko, P., & Lv, J. (2009). Dasso: Connections between the dantzig selector and lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(1), 127–142.

Javanmard, A., & Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, *15*(1), 2869–2909.

Juditsky, A., Karzan, F. K., Nemirovski, A., & Polyak, B. (2012). Accuracy guaranties for *backslash*ell 1 recovery of block-sparse signals. *The Annals of Statistics*, *40*(6), 3077–3107.

Liu, J., Wonka, P., & Ye, J. (2012). A multi-stage framework for dantzig selector and lasso. *Journal of Machine Learning Research*, *13*(4).

Lounici, K. et al. (2008). Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electronic Journal of statistics*, *2*, 90–102.

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, *34*(3), 1436–1462.

Ning, Y., Liu, H. et al. (2017a). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Annals of statistics*, *45*(1), 158–195.

Ning, Y., Liu, H. et al. (2017b). Supplement to "a general theory of hypothesis tests and confidence regions for sparse high dimensional models", DOI:10.1214/16–AOS1448SUPP.

Portfolios of investment. (2021). https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/tw_5_ports_me_inv.html (accessed: 30.04.2021)

Portfolios of operating profitability. (1972). https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_szme_op.html (accessed: 12.1972)

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Van de Geer, S., Bühlmann, P., Ritov, Y., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, *42*(3), 1166–1202.

Voorman, A., Shojaie, A., & Witten, D. (2014). Inference in high dimensions with the penalized score test. *arXiv preprint arXiv:1401.2678*.

Wallis, S. (2013). Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics*, *20*(3), 178–208.

Zhang, C.-H., & Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics*, *36*(4), 1567–1594.

Zhang, C.-H., & Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 217–242.

Zhang, T. et al. (2009). Some sharp performance bounds for least squares regression with l1 regularization. *The Annals of Statistics*, *37*(5A), 2109–2144.

Zhou, S. (2009). Thresholding procedures for high dimensional variable selection and statistical estimation. *Advances in Neural Information Processing Systems*, *22*, 2304–2312.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.

A Appendix: Tutorial of R code

```
# This is a tutorial for the simulation process.
#It is a general form and we can adjust some settings
#to get the coverage and the power of statistical test.
# d = 200 Poisson regression for rho = 0.25/0.4/0.6/0.75
print("start the code")
result = array(0, dim = c(500,1))
result_stat = array(0, dim = c(500, 1))
result_005 = array(0, dim = c(500, 1))
result_010 = array(0, dim = c(500, 1))
result_015 = array(0, dim = c(500, 1))
result_020 = array(0, dim = c(500, 1))
result_025 = array(0, dim = c(500, 1))
result_030 = array(0, dim = c(500, 1))
result_035 = array(0, dim = c(500, 1))
result_040 = array(0, dim = c(500, 1))
result_045 = array(0, dim = c(500, 1))
result_050 = array(0, dim = c(500, 1))
result_055 = array(0, dim = c(500, 1))
result_center = array(0, dim = c(500,1))
result_left = array(0, dim = c(500, 1))
result_right = array(0, dim = c(500, 1))
for (m in 1:500){
  d = 100 # 200, 500
  Sigma <- array(1,dim = c(d, d))
  rho_1 = 0.25 # 0.4 0.6 0.75
  # Generate Sigma
  for (i in 1:d){
    for (j in 1:d){
      Sigma[i,j]=rho_1^(abs(i-j))
    }
  }
  library(MASS)
  mean_<- array(0, dim = c(d,1))
  n = 100
```

```
x = mvrnorm(n,mean_,Sigma) # DGP
# Generate Y and beta
beta <- array(0, dim = c(d,1))
beta[2] = 1
beta[3] = 1
beta[4] = 1
#Linear regression
#y = x %*% beta
#Poisson regression
# y <- rpois(n, exp(x %*% beta))
#Logistic regression
y <- rbinom(n, 2, (1 + exp(-x %*% beta))^(-1))
library(glmnet)
library(ncvreg)
# You can choose "glmnet" for lasso or "ncvreg" for SCAD.
# You can choose a family type to make lienar, logistic, and Poisson regression
fitcv<-cv.glmnet(x, y, family="binomial", alpha=1)
beta_est = coef(fitcv, fitcv$lambda.min)
#Poisson regression with lasso or SCAD
# You can choose a family type to make lienar, logistic, and Poisson regression
#fitcv<-cv.ncvreg(x, y, penalty='SCAD',family="poisson" )
#fitcv<-ncvreg(x, y, penalty='SCAD',family="poisson")
#beta_est = coef(fitcv, fitcv$lambda.min)
library(hdme)
c = array(0, dim = c(500, 1))
# This part is to estimate a Dantzig Selector.
# For linear regression, we do not need to calculate x_,z_
#x_,z_ are to calculate the Dantzig Selector.
# For Poisson regression, x_,z_ is set based on partial Fisher information
for (i in 1:n ) {
  c[i,1] = sqrt(exp( x[i, 1:d]%*%beta_est[2:(d+1) ] )
  / (1 + exp( x[i, 1:d]%*%beta_est[2:(d+1) ] ) ))
}
c_ = sqrt(c)
x_ = x
for (i in n){
```

```
  x_[i, 1:d] = x_[i,1:d]*c[i,1]
}

z_ = x[1:n,1]
x_ = x[1:n, 2:d]
# Cross-validation
cv_fit <- cv_gds(x_, z_, family = "gaussian", no_lambda = 50, n_folds = 10)
fit_gds = gds(x_, z_, family = "gaussian", lambda = cv_fit$lambda_min)
gamma = coef(fit_gds)
#calculate sum of score
score=0
score_005 = 0
score_010 = 0
score_015 = 0
score_020 = 0
score_025 = 0
score_030 = 0
score_035 = 0
score_040 = 0
score_045 = 0
score_050 = 0
score_055 = 0
score_interval = 0
I=0
for (j_2 in (1:n)){
  #calculate wx
  wx=0
  for(j_3 in 1:dim(gamma)[1]){
    wx = wx + x_[j_2, gamma[j_3,1]]*gamma[(j_3),2]
    #print(x[j_2, gamma[j_3,1]])
    #print(gamma[j_3, 2])
  }
  if(dim(gamma)[1] == 0 ){
    wx = 0
  }
  score_interval = score_interval + (y[j_2] - (exp(x[j_2, 1:d]%*%beta_est[2:(d+1) ])
```

```
  /(1+exp(beta_est[1]+ x[j_2, 1:d]%*%beta_est[2:(d+1)])) ) )*(x[j_2, 1]-wx)
  score=score+ (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)])) ) )*(x[j_2, 1]-wx)
  score_005 = score_005 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.05 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.05*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_010 = score_010 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.10 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.10*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_015 = score_015 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.15 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.15*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_020 = score_020 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.20 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.20*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_025 = score_025 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.25 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.25*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_030 = score_030 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.30 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.30*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_035 = score_035 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.35 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.35*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_040 = score_040 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.40 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.40*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_045 = score_045 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.45 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.45*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_050 = score_050 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.50 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.50*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  score_055 = score_055 + (y[j_2] - (exp(x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.55 * x[j_2, 1])
  /(1+exp( x[j_2, 2:d]%*%beta_est[3:(d+1)] + 0.55*x[j_2, 1] ) ) ) )*(x[j_2, 1]-wx)
  # Partial Fisher Information
  I = I + (exp(x[j_2, 1:d]%*%beta_est[2:(d+1)])
  / (1+exp(x[j_2, 1:d]%*%beta_est[2:(d+1)]) )^2) *x[j_2,1]*(x[j_2,1] - wx)
}
I = abs(I)/n
score = (-1/n)*score
score_interval = score_interval*(-1/n)
U = sqrt(n)*score*I^(-0.5)
U_005 = sqrt(n)*score_005*abs(I)^(-0.5)*(-1/n)
U_010 = sqrt(n)*score_010*abs(I)^(-0.5)*(-1/n)
U_015 = sqrt(n)*score_015*abs(I)^(-0.5)*(-1/n)
```

```
  U_020 = sqrt(n)*score_020*abs(I)^(-0.5)*(-1/n)
  U_025 = sqrt(n)*score_025*abs(I)^(-0.5)*(-1/n)
  U_030 = sqrt(n)*score_030*abs(I)^(-0.5)*(-1/n)
  U_035 = sqrt(n)*score_035*abs(I)^(-0.5)*(-1/n)
  U_040 = sqrt(n)*score_040*abs(I)^(-0.5)*(-1/n)
  U_045 = sqrt(n)*score_045*abs(I)^(-0.5)*(-1/n)
  U_050 = sqrt(n)*score_050*abs(I)^(-0.5)*(-1/n)
  U_055 = sqrt(n)*score_055*abs(I)^(-0.5)*(-1/n)
  result[m,1] = U
  result_005[m,1] = U_005
  result_010[m,1] = U_010
  result_015[m,1] = U_015
  result_020[m,1] = U_020
  result_025[m,1] = U_025
  result_030[m,1] = U_030
  result_035[m,1] = U_035
  result_040[m,1] = U_040
  result_045[m,1] = U_045
  result_050[m,1] = U_050
  result_055[m,1] = U_055
  #One-step estimator
  theta_center = beta_est[2,1] + score_interval / I
  print(sqrt(1/n)*1.96*I^(-0.5))
  result_center[m,1] = theta_center
  result_left[m,1] = theta_center - sqrt(1/n)*1.96*I^(-0.5)
  result_right[m,1]= theta_center + sqrt(1/n)*1.96*I^(-0.5)
}
# Coverage rates
time = 0
for (i in 1:500){
  if(result_left[i,1] > 0 || result_right[i,1] < 0){
    time = time + 1
  }
}
# The following part is to calculate the power
num_1 = 0
```

```
for (i in 1:500){
  if( (abs(result_[i])) > 1.96){
    num_1 = num_1 + 1
  }
}
num_005 = 0
for (i in 1:500){
  if( abs(result_005_[i]) > 1.96 ){
    num_005 = num_005 + 1
  }
}
num_010 = 0
for (i in 1:500){
  if( abs(result_010[i]) > 1.96 ){
    num_010 = num_010 + 1
  }
}
num_015 = 0
for (i in 1:500){
  if( abs(result_015[i]) > 1.96 ){
    num_015 = num_015 + 1
  }
}
num_020 = 0
for (i in 1:500){
  if( abs(result_020[i]) > 1.96 ){
    num_020 = num_020 + 1
  }
}
num_025 = 0
for (i in 1:500){
  if( abs(result_025[i]) > 1.96 ){
    num_025 = num_025 + 1
  }
}
num_030 = 0
```

```
for (i in 1:500){
  if( abs(result_030[i]) > 1.96 ){
    num_030 = num_030 + 1
  }
}
num_035 = 0
for (i in 1:500){
  if( abs(result_35) > 1.96 ){
    num_035 = num_035 + 1
  }
}
num_040 = 0
for (i in 1:500){
  if( abs(result_040[i])> 1.96 ){
    num_040 = num_040 + 1
  }
}
num_045 = 0
for (i in 1:500){
  if( abs(result_045[i]) > 1.96 ){
    num_045 = num_045 + 1
  }
}
num_050 = 0
for (i in 1:500){
  if( abs(result_050[i]) > 1.96 ){
    num_050 = num_050 + 1
  }
}
num_055 = 0
for (i in 1:500){
  if( abs(result_055[i]) > 1.96 ){
    num_055 = num_055 + 1
  }
}
```

B Appendix: A Code for Plotting

```
# This part is to show how to draw a power plotting
import numpy as np
import matplotlib.pyplot as plt
from scipy.ndimage import gaussian_filter1d
#coverage 0.25 16 0.40 19 0.60 17 0.75 24
x=np.array([0.05, 0.1, 0.15, 0.20, 0.25
, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55])
y_25=np.array([19/500, 22/500, 27/500,
43/500, 61/500, 89/500, 115/500, 154/500, 189/500, 232/500, 271/500 ])
y_40=np.array([15/500, 17/500, 20/500,
23/500, 37/500, 54/500, 65/500, 80/500,
110/500, 143/500, 172/500 ])
y_60=np.array([14/500, 15/500, 18/500,
20/500, 26/500, 31/500, 47/500, 70/500, 84/500, 113/500, 146/500 ])
y_75=np.array([10/500, 18/500, 21/500,
23/500, 25/500, 35/500, 40/500, 56/500, 69/500, 81/500, 92/500 ] )
y_smoothed_1 = gaussian_filter1d(y_25, sigma=0.9)
y_smoothed_2 = gaussian_filter1d(y_40, sigma=0.9)
y_smoothed_4 = gaussian_filter1d(y_60, sigma=0.9)
y_smoothed_75 = gaussian_filter1d(y_75, sigma=0.9)
plt.plot(x, y_smoothed_1, label = "rho = 0.25" )
plt.plot(x, y_smoothed_2, label = "rho = 0.40" )
plt.plot(x, y_smoothed_4, label = "rho = 0.60" )
plt.plot(x, y_smoothed_75, label = "rho = 0.75")
plt.legend()
plt.title("Power of the decorrelated score test for the logistic regression with n=100,d=500,s=3")
plt.xlabel("Signal")
plt.ylabel("Rejection Rate")
plt.show()
```