ERASMUS UNIVERSITEIT ROTTERDAM

Erasmus University Rotterdam

Erasmus School of Economics

Econometrics and Operational Research

Bachelor Thesis: FEB63007

July 4, 2021

On the decorrelated score test for sparse high-dimensional linear models with FDR control

Author:

SIPKE DOM *509386*

Supervisor:

DR. A.J. KONING

Second assessor:

M.F.O. WELZ

Abstract

Advancements in data mining have increased interest in sparse high-dimensional models in fields of statistics, biology, and economics. However, conventional hypothesis testing methods often do not work when the number of covariates exceeds the number of observations. As an alternative testing method, Ning and Liu (2017) introduce the decorrelated score (DScore) test for generic penalized M-estimators. This thesis investigates the performance of this DScore test in two types of linear models: linear regressions and generalized linear models (GLMs). Extending the framework of Ning and Liu (2017), we also consider multiple testing, as it is often not pre-specified which covariates are of interest. For the performance evaluation of the DScore test, this thesis simulates its Type I errors, power, and false discovery rates (FDRs) and then applies the test on two real-world data sets. The main findings of this thesis are that, in general, the DScore test in linear regression models yields accurate Type I errors and high power. Conversely, this performance is less reliable in GLMs with a small sample size. Moreover, the DScore test fails to consistently control the FDRs, even after applying conventional FDR adjustment methods to its p-values.

Keywords: False Discovery Rate (FDR), Generalized Linear Models (GLMs), High-dimensional Inference, Hypothesis Testing, Lasso, Multiple Testing, Penalized Estimator, Sparsity.

Contents

List of Tables

List of Figures

1 Introduction

Advancements in data mining have led to a spike of interest in statistical models that deal with high dimensionality (Van de Geer et al., 2014). In these high-dimensional models, the number of covariates, d, is often larger than the number of observations, n, invalidating inference in many existing statistical models. Consider a standard linear regression, for example:

$$\mathbf{Y} = \mathbf{Q}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}, \tag{1.1}$$

where $\mathbf{Y} = (Y_1, ..., Y_n)^T \in \mathbb{R}^n$ contains the dependent variables, $\mathbf{Q} = (\boldsymbol{Q}_1, ..., \boldsymbol{Q}_n)^T \in \mathbb{R}^{n\times d}$ is the regressor matrix, $\boldsymbol{\beta}^* = (\beta_1^*, ..., \beta_d^*)^T \in \mathbb{R}^d$ is the coefficient vector, and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^T \in \mathbb{R}^n$ are the error terms. If $d > n$, standard ordinary least squares (OLS) faces the problem of multicollinearity as $\text{rank}(\mathbf{Q}) < d$ and $\mathbf{Q}\mathbf{Q}^T$ is a singular matrix (Heij et al., 2004). As an alternative method, Ning and Liu (2017) use penalized M-estimators for estimating the high-dimensional parameter $\boldsymbol{\beta}^*$:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\Omega}{\text{argmin}}\, \ell(\boldsymbol{\beta}) + P_\lambda(\boldsymbol{\beta}), \tag{1.2}$$

where Ω is the parameter space, $\ell(\boldsymbol{\beta})$ is a loss function (e.g., the negative log-likelihood) and $P_\lambda(\boldsymbol{\beta})$ is a penalty function with tuning parameter λ. The penalty function ensures the sparsity of the model by allowing for a limited number of nonzero estimates.

Most studies on high-dimensional models, such as Fonti and Belitser (2017), focus solely on these nonzero estimates and omit discussion about the statistical significance of these estimates. In many applications, however, it is of interest to test this significance. Hence, we consider the partition $\boldsymbol{\beta}^* = (\theta^*, \boldsymbol{\gamma}^{*T})^T$ and denote the null hypothesis as $H_0 : \theta^* = 0$, in which θ^* is the true parameter of interest. With hypothesis tests, a second problem arises, as many widely-used test statistics do not work in sparse high-dimensional models when $d > n$. Rao's score test statistic of Rao (1948), for example, is no longer asymptotically normally distributed under H_0 (Ning and Liu, 2017). Therefore, Ning and Liu (2017) propose a decorrelated score (DScore) test statistic for penalized M-estimators, which is an extension of Rao's score test statistic that asymptotically follows the normal distribution under H_0. According to Ning and Liu (2017), the DScore test is applicable to a variety of models, including linear regressions, logistic regressions, Poisson regressions, Gaussian graphical models, and additive hazard models. Specifically, this thesis focuses on two types of linear models: linear regressions and a class of GLMs that includes logistic and Poisson regressions.

This research replicates some derivations of Ning and Liu (2017) concerning the DScore test and its asymptotic properties with additional details. Moreover, we aim to investigate whether these asymptotic properties hold in finite samples before applying the test to real-world data. Therefore, the first research question of this thesis is as follows:

> **RQ1:** *How is the finite sample performance of the DScore test in linear models?*

This thesis attempts to answer this question by performing simulations of the Type I errors and power of the DScore test in linear regressions and two GLM specifications: logistic regressions and Poisson regressions. Here, we consider a smaller sample size and find that the DScore test still performs well in linear regression models. Conversely, the test seems to have a less reliable small sample performance in the GLMs.

Furthermore, researchers may want to test multiple hypotheses, as they often do not know beforehand which covariates are of interest. This thesis denotes these hypotheses as $H_{0,j} : \beta_j^* = 0$ for $j = 1, \ldots, m$. Multiple testing is useful in many statistical applications, such as feature selection with a large set of features, and it poses additional challenges compared to single hypothesis testing. A naive approach is to test the m hypotheses and reject them if their corresponding p-value is lower than some pre-specified significance level. However, when performing 10,000 tests with a 5% significance level, the expected number of false rejections[1] of the null hypotheses is already 500. Benjamini and Hochberg (1995) define the false discovery rate (FDR) as the expected number of false rejections out of all rejections, and they created the Benjamini-Hochberg (BH) procedure, which remains a widely-used method to control the FDR. Because Ning and Liu (2017) omit multiple testing from their discussion, this thesis expands on their work by evaluating the performance of the DScore test after applying one of four multiple testing correction methods, including the BH procedure.

To our current knowledge, this is the first study that investigates the multiple testing ability of the DScore test. We attempt to fill this gap in the literature by attempting to answer the second research question:

> **RQ2:** *How can we control the FDRs in linear models when testing multiple hypotheses with the DScore test?*

In this case, we simulate the FDRs of the DScore test under the same conditions as the Type I errors. Our results indicate that the DScore test does not consistently control the FDRs in small samples, even after applying our proposed FDR control techniques.

[1]False rejections are Type I errors: erroneously rejecting true null hypotheses. Other terms for false rejections are 'false discoveries' and 'false positives'.

After performing the simulations, this thesis shows two real-world data multiple testing applications of the DScore test. The first is a macro-economic application on the Freddie Mac House Price Index, which uses the linear regression specification, and the second an application in genetics with data from Singh et al. (2002) on prostate cancer gene expressions, for which this thesis uses the logistic regression specification.

1.1 Related work

The previously-mentioned penalized M-estimators work with a great variety of penalty functions. These penalty functions can either be convex or non-convex. A popular example of a convex penalty is the Lasso penalty that was introduced by Tibshirani (1996):

$$P_\lambda(\boldsymbol{\beta}) = \lambda \left\|\boldsymbol{\beta}\right\|_1, \tag{1.3}$$

which restricts the L_1-norm of $\boldsymbol{\beta}$ to be less than some tuning parameter $\lambda \in \mathbb{R}$. Examples of other penalties are the L_2 penalty, also called the ridge penalty, and non-convex penalties like the SCAD penalty of Fan and Li (2001) and the MCP penalty of Zhang (2010).

Other than Ning and Liu (2017), several studies proposed other hypothesis testing methods for sparse high-dimensional models in the last decade. Specifically for estimators that use the Lasso penalty, Javanmard and Montanari (2014) and Van de Geer et al. (2014) propose debiasing and desparsifying correction methods for (generalized) linear models. Another debiasing method is the ridge projection method of Bühlmann (2013), which works with estimations via the ridge penalty instead of the Lasso. Another popular approach is sample splitting (Shah and Samworth, 2013), but Neykov et al. (2018) argue that such methods face an inevitable efficiency loss.

Furthermore, a related concept to the FDR is the familywise error rate (FWER), which is the probability of making any Type I error. The FWER concept predates the FDR and equals the FDR when all null hypotheses are true. The most popular method for FWER control is the Bonferroni correction (Bonferroni, 1936). Another example is the Holm FWER correction procedure, which Van de Geer et al. (2014) use in their high-dimensional inference study (Holm, 1979). However, Benjamini and Hochberg (1995) note that FWER correction techniques yield a lower power compared to FDR correction techniques, such as the BH procedure. Although this thesis focuses on FDRs rather than FWERs, it also evaluates the FDR control ability of the Bonferroni and Holm procedures.

1.2 Outline of this thesis

The remainder of this thesis is structured as follows. First, Section 2 contains an elaborate discussion on the DScore test with descriptions of the models and some necessary derivations. Section 3 starts with the set-up of the simulation study and follows with the simulation results. Next, Section 4 includes the applications of the DScore test on the house price index and prostate cancer data-sets. Section 5 concludes.

2 Methodology

Section 2.1 contains the derivation of the general test statistic and its asymptotic properties. Next, Section 2.2 and 2.3 discuss the linear regression case and the GLM case, respectively. Last, Section 2.4 introduces four p-value adjustment techniques for FDR control. This thesis follows the notation and general framework of Ning and Liu (2017). Appendix A contains an overview of the mathematical notation and abbreviations used in this thesis.

2.1 General test statistic

Ning and Liu (2017) propose the DScore function that forms the basis of the hypothesis tests:

$$S(\theta, \boldsymbol{\gamma}) = \nabla_\theta \ell\left(\theta, \boldsymbol{\gamma}\right) - \mathbf{w}^T \nabla_{\boldsymbol{\gamma}} \ell\left(\theta, \boldsymbol{\gamma}\right) \text{ with } \mathbf{w}^T = \mathbf{I}_{\theta\boldsymbol{\gamma}} \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}, \tag{2.1}$$

where $\mathbf{I}_{\theta\boldsymbol{\gamma}}$ and $\mathbf{L}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ are partitions of the Fischer information matrix $\mathbf{I}$:

$$\mathbf{I} = \begin{pmatrix} I_{\theta\theta} & \mathbf{I}_{\theta\boldsymbol{\gamma}} \\ \mathbf{I}_{\boldsymbol{\gamma}\theta} & \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \end{pmatrix} := \mathbb{E}\left(\nabla^2 \ell(\boldsymbol{\beta})\right). \tag{2.2}$$

Furthermore, this thesis uses Algorithm 1 for the estimation of $\hat{S}(\theta, \hat{\boldsymbol{\gamma}})$, which closely resembles[2] the algorithm of Ning and Liu (2017).

[2] Apart from the penalized M-estimator approach that Algorithm 1 uses in step 2, Ning and Liu (2017) propose the Dantzig estimator $\tilde{\mathbf{w}}$ as an alternative estimator for $\mathbf{w}$ with tuning parameter λ':

$$\tilde{\mathbf{w}} = \text{argmin}\, \|\mathbf{w}\|_1 \quad \text{s.t.} \quad \left\| \nabla^2_{\theta\boldsymbol{\gamma}} \ell(\hat{\boldsymbol{\beta}}) - \mathbf{w}^T \nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(\hat{\boldsymbol{\beta}}) \right\|_\infty \leq \lambda'.$$

Algorithm 1 DScore function estimation.

Input: Loss function $\ell(\theta, \boldsymbol{\gamma})$, penalty function $P(\cdot)$ and tuning parameters λ and λ' .

1. Estimate $\hat{\boldsymbol{\beta}} = (\hat{\theta}, \hat{\boldsymbol{\gamma}})$ with Eq. (1.2).

2. Estimate $\mathbf{w}$ with the penalized M-estimator $\hat{\mathbf{w}}$ with penalty function $Q_{\lambda'}(\mathbf{w})$:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2n} \sum_{i=1}^{n} \left\{ \mathbf{w}^T \nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell_i(\hat{\boldsymbol{\beta}}) \mathbf{w} - 2\mathbf{w}^T \nabla^2_{\boldsymbol{\gamma}\theta} \ell_i(\hat{\boldsymbol{\beta}}) \right\} + Q_{\lambda'}(\mathbf{w}). \tag{2.3}$$

3. Calculate the estimated DScore function:

$$\hat{S}(\theta, \hat{\boldsymbol{\gamma}}) = \nabla_\theta \ell(\theta, \hat{\boldsymbol{\gamma}}) - \hat{\mathbf{w}}^T \nabla_{\boldsymbol{\gamma}} \ell(\theta, \hat{\boldsymbol{\gamma}}). \tag{2.4}$$

return $\hat{S}(\theta, \hat{\boldsymbol{\gamma}})$

The name 'decorrelated score function' comes from the notion that $S(\theta, \boldsymbol{\gamma})$ is uncorrelated with the nuisance gradient $\nabla_{\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma})$. This decorrelation is achieved by $\mathbf{w}^T \nabla_{\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma})$, which is the sparse projection of $\nabla_\theta \ell(\theta, \boldsymbol{\gamma})$ on the nuisance score space $\mathcal{N} = \operatorname{span}\{\nabla_{\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma})\}$. To illustrate the importance of this decorrelation, we compare the DScore function with Rao's score function $\nabla_\theta \ell(\theta, \boldsymbol{\gamma})$ (Rao, 1948). For Rao's score function, it holds that $n^{1/2} \nabla_\theta \ell(0, \tilde{\boldsymbol{\gamma}}) \rightsquigarrow N(0, \sigma)$ when $d << n$ for some estimator $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}^*$. The first-order Taylor expansion under $H_0: \theta^* = 0$ shows that this asymptotic normality does not hold when $d > n$ and $\boldsymbol{\beta}^* = (\theta^*, \boldsymbol{\gamma}^{*T})^T$ is sparse:

$$n^{1/2} \nabla_\theta \ell(0, \tilde{\boldsymbol{\gamma}}) = \underbrace{n^{1/2} \nabla_\theta \ell(0, \boldsymbol{\gamma}^*)}_{T_1} + \underbrace{n^{1/2} \nabla^2_{\theta\boldsymbol{\gamma}} \ell(0, \boldsymbol{\gamma}^*)(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)}_{T_2} + Rem, \tag{2.5}$$

where Rem is the remainder of the Taylor expansion. While T_1 is asymptotically normally distributed, this is not always the case for T_2 and Rem due to the sparsity of $\boldsymbol{\beta}^*$ (Ning and Liu, 2017). For the DScore function under H_0, the first order Taylor expansion becomes

$$\begin{aligned} n^{1/2} S(0, \tilde{\boldsymbol{\gamma}}) &= n^{1/2} S(0, \boldsymbol{\gamma}^*) + n^{1/2} \nabla_{\boldsymbol{\gamma}} S(0, \boldsymbol{\gamma}^*)(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) + Rem \\ &= n^{1/2} S(0, \boldsymbol{\gamma}^*) + n^{1/2} \nabla_{\boldsymbol{\gamma}} \left\{ \nabla_\theta \ell(0, \boldsymbol{\gamma}^*) - \mathbf{w}^T \nabla_{\boldsymbol{\gamma}} \ell(0, \boldsymbol{\gamma}^*) \right\} (\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) + Rem \\ &= n^{1/2} S(0, \boldsymbol{\gamma}^*) + n^{1/2} \left\{ \nabla^2_{\theta\boldsymbol{\gamma}} \ell(0, \boldsymbol{\gamma}^*) - \mathbf{I}_{\theta\boldsymbol{\gamma}} \mathbf{I}^{-1}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(0, \boldsymbol{\gamma}^*) \right\} (\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) + Rem \\ &= \underbrace{n^{1/2} S(0, \boldsymbol{\gamma}^*)}_{T_1} + \\ &\underbrace{n^{1/2} \left\{ \nabla^2_{\theta\boldsymbol{\gamma}} \ell(0, \boldsymbol{\gamma}^*) - \mathbb{E}[\nabla^2_{\theta\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma})] \mathbb{E}[\nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma})]^{-1} \nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(0, \boldsymbol{\gamma}^*) \right\} (\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)}_{T_2} + Rem. \end{aligned} \tag{2.6}$$

From the last line of this equation, we can see that the middle term of T_2 converges to 0. Generalizing this finding for Rem removes the asymptotic distribution problems of Rao's score function.

Using the estimated DScore function, the test statistic under H_0 is as follows:

$$\hat{U}_n = n^{1/2}\hat{S}(0, \hat{\boldsymbol{\gamma}})/\sqrt{\hat{\sigma}_S}. \tag{2.7}$$

Here, $\hat{\sigma}_S$ is a consistent estimator of $\sigma_S^* = (1, -\mathbf{w}^{*T})\boldsymbol{\Sigma}^*(1, -\mathbf{w}^{*T})^T$ with $\boldsymbol{\Sigma}^* = \lim_{n\to\infty} \text{Var}(n^{1/2}\nabla\ell(\boldsymbol{\beta}^*))$.

For the derivation of the asymptotic distribution of $\hat{U}_n$, Ning and Liu (2017) make four general assumptions that models must validate. Appendix B.1 shows these assumptions (Assumptions B.1-B.4). Subsequently, Ning and Liu (2017) derive Theorem 2.1, which shows the asymptotic normality of $\hat{U}_n$. Appendix C.1 contains an elaborate version of their proof of this theorem. Using Theorem 2.1, we can reject $H_0 : \theta^* = 0$ when n is sufficiently large and

$$\left|\hat{U}_n\right| > \Phi^{-1}(1 - \alpha/2), \tag{2.8}$$

where $\alpha \in [0, 1]$ is the significance level.

Theorem 2.1. *Let $\eta_1(n)$ and $\eta_2(n)$ be sequences that converge to 0 when $n \to \infty$. If Assumptions B.1-B.4 hold and $(\eta_1(n) + \eta_1(n))\sqrt{\log d} = o(1)$, then*

$$n^{1/2}\hat{S}(0, \hat{\boldsymbol{\gamma}})\sigma_S^{*-1/2} \rightsquigarrow N(0, 1), \tag{2.9}$$

and for any $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \left|\mathbb{P}\left(\hat{U}_n \leq t\right) - \Phi(t)\right| = 0. \tag{2.10}$$

2.2 Linear regression case

Consider a linear regression in the form

$$\begin{aligned} Y_i &= \boldsymbol{\beta}^T\boldsymbol{Q}_i + \varepsilon_i \\ &= \theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i + \varepsilon_i, \end{aligned} \tag{2.11}$$

where $\boldsymbol{Q}_i = (Z_i, \boldsymbol{X}_i^T)^T \in \mathbb{R}^d$ and $\boldsymbol{\beta} = (\theta, \boldsymbol{\gamma}^T)^T \in \mathbb{R}^d$ for $i = 1, ..., n$. This thesis assumes that the error terms ε_i satisfy $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{E}(\varepsilon_i^2) = \sigma^2$. Also, we assume for the first part of the derivation that $\sigma^2 > 0$ is known. Besides, Appendix B.2 shows an additional assumption that Ning and Liu (2017) make for the derivations of the DScore test statistic in the linear regression case.

Like Ning and Liu (2017), this thesis uses the negative Gaussian quasi-log-likelihood as the loss function $\ell(\cdot)$ for linear regressions:

$$\ell(\theta, \boldsymbol{\gamma}) = (2n\sigma^2)^{-1} \sum_{i=1}^{n} \left(Y_i - \theta Z_i - \boldsymbol{\gamma}^T \boldsymbol{X}_i\right)^2. \tag{2.12}$$

Appendix C.2 shows the formulae and derivations of the partial first and second order gradients of this loss function. The resulting formula for the DScore function for linear regressions is as follows:

$$\begin{aligned} S(\theta, \boldsymbol{\gamma}) &= \nabla_\theta \ell(\theta, \boldsymbol{\gamma}) - \mathbf{w}^T \nabla_{\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma}) \\ &= -\frac{1}{\sigma^2 n} \sum_{i=1}^{n} \left(Y_i - \theta Z_i - \boldsymbol{\gamma}^T \boldsymbol{X}_i\right) \left(Z_i - \mathbf{w}^T \boldsymbol{X}_i\right) \text{ with} \end{aligned} \tag{2.13}$$

$$\mathbf{w} = \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1} \mathbf{I}_{\boldsymbol{\gamma}\theta} = \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i^T\right)^{-1} \mathbb{E}\left(Z_i \boldsymbol{X}_i\right). \tag{2.14}$$

Under $H_0 : \theta^* = 0$, the estimated DScore function becomes

$$\hat{S}(0, \hat{\boldsymbol{\gamma}}) = -\frac{1}{\sigma^2 n} \sum_{i=1}^{n} \left(Y_i - \hat{\boldsymbol{\gamma}}^T \boldsymbol{X}_i\right) \left(Z_i - \hat{\mathbf{w}}^T \boldsymbol{X}_i\right), \tag{2.15}$$

in which $\hat{\boldsymbol{\gamma}}$ is the nuisance component of $\hat{\boldsymbol{\beta}}$, and this thesis estimates $\hat{\mathbf{w}}$ with the Lasso penalty:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(Z_i - \mathbf{w}^T \boldsymbol{X}_i\right)^2 + \lambda' \|\mathbf{w}\|_1 \right\}. \tag{2.16}$$

Furthermore, Appendix C.3 contains the derivations of the (partial) Fisher information matrices $\mathbf{I}^*$ and $I^*_{\theta|\boldsymbol{\gamma}}$, which we estimate as

$$\hat{\mathbf{I}} = \frac{1}{\sigma^2 n} \sum_{i=1}^{n} \boldsymbol{Q}_i \boldsymbol{Q}_i^T, \text{ and} \tag{2.17}$$

$$\hat{I}_{\theta|\boldsymbol{\gamma}} = \frac{1}{\sigma^2} \left\{ \frac{1}{n} \sum_{i=1}^{n} Z_i^2 - \hat{\mathbf{w}}^T \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i Z_i \right) \right\}. \tag{2.18}$$

Finally, note that the information identity $\boldsymbol{\Sigma}^* = \mathbf{I}^*$ holds for negative log-likelihoods (Mykland, 1999). Following this identity, we derive that $\sigma_S^* = I^*_{\theta|\boldsymbol{\gamma}}$ in Appendix C.4.

Using Eqs. (2.15) and (2.18), the resulting test statistic under the null hypothesis is

$$\begin{aligned} \hat{U}_n &= n^{1/2} \hat{S}(0, \hat{\boldsymbol{\gamma}}) \hat{I}_{\theta|\boldsymbol{\gamma}}^{-1/2} \\ &= -\frac{1}{\sigma n^{1/2}} \sum_{i=1}^{n} \left(Y_i - \hat{\boldsymbol{\gamma}}^T \boldsymbol{X}_i\right) \left(Z_i - \hat{\mathbf{w}}^T \boldsymbol{X}_i\right) \left(H_Z - \hat{\mathbf{w}}^T \boldsymbol{H}_{XZ}\right)^{-1/2}, \end{aligned} \tag{2.19}$$

where $H_Z = \frac{1}{n}\sum_{i=1}^n Z_i^2$ and $\boldsymbol{H}_{XZ} = \frac{1}{n}\sum_{i=1}^n Z_i \boldsymbol{X}_i$. Moreover, Ning and Liu (2017) show that Corollary 2.1.1 follows from Theorem 2.1 and that $\hat{U}_n$ is, thus, asymptotically normally distributed under $H_0 : \theta^* = 0$ for the linear regression case.

Corollary 2.1.1. *Let* $S = \operatorname{supp}(\boldsymbol{\beta}^*)$ *and* $S' = \operatorname{supp}(\mathbf{w}^*)$ *satisfy* $|S| = s^*$ *and* $|S'| = s'$. *If Assumption B.5 holds,* $n^{-1/2}\max(s', s^*)\log d = o(1)$ *and* $\lambda \asymp \lambda' \asymp \sqrt{\frac{\log d}{n}}$, *then under* $H_0 : \theta^* = 0$,

$$\lim_{n\to\infty}\left|\mathbb{P}\left(\hat{U}_n \le t\right) - \Phi(t)\right| = 0 \text{ for any } t \in \mathbb{R}. \tag{2.20}$$

As σ^2 is often unknown in practice, this research uses the consistent estimator $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\boldsymbol{\beta}}^T \boldsymbol{Q}_i)^2$ to estimate σ^2. Ning and Liu (2017) show that Corollary 2.1.1 still holds if $\sigma^2 \ge C$ for some C, after replacing σ with $\hat{\sigma}$ in Eq. (2.19).

2.3 Generalized linear model (GLM) case

This thesis considers the class of generalized linear models (GLMs) with negative log-likelihoods in the following form

$$\ell(\theta,\gamma) = -\frac{1}{n}\sum_{i=1}^n \frac{1}{a(\phi)}\left\{Y_i\left(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i\right) - b\left(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i\right)\right\}, \tag{2.21}$$

where $a(\cdot)$ and $b(\cdot)$ are two known functions. Appendix C.5 is a supplement to Eq. (2.21) with derivations of all relevant (partial) gradients, and Appendix C.6 shows the derivations of the theoretical quantities $\mathbf{I}^*$ and $I^*_{\theta|\boldsymbol{\gamma}}$. Like Ning and Liu (2017), this thesis assumes $a(\phi) = 1$ in the following derivations. Using this assumption, the negative log-likelihood in Eq. (2.21) leads to the DScore function for GLMs:

$$\begin{aligned} S(\theta,\boldsymbol{\gamma}) &= \nabla_\theta \ell(\theta,\boldsymbol{\gamma}) - \mathbf{w}^T \nabla_{\boldsymbol{\gamma}} \ell(\theta,\boldsymbol{\gamma}) \\ &= -\frac{1}{n}\sum_{i=1}^n \left(Y_i - b'\left(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i\right)\right)\left(Z_i - \mathbf{w}^T\boldsymbol{X}_i\right). \end{aligned} \tag{2.22}$$

Furthermore, the formula for $\mathbf{w}$ is

$$\mathbf{w} = \mathbf{I}_{\boldsymbol{\gamma\gamma}}^{-1}\mathbf{I}_{\boldsymbol{\gamma}\theta} = \mathbb{E}(b''(\boldsymbol{\beta}^T\boldsymbol{Q}_i)\boldsymbol{X}_i\boldsymbol{X}_i^T)^{-1}\mathbb{E}(b''(\boldsymbol{\beta}^T\boldsymbol{Q}_i)\boldsymbol{X}_i Z_i), \tag{2.23}$$

which we estimate with

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}}\left\{\frac{1}{2n}\sum_{i=1}^n b''(\hat{\boldsymbol{\beta}}^T\boldsymbol{Q}_i)\left(Z_i - \mathbf{w}^T\boldsymbol{X}_i\right)^2 + \lambda'\|\mathbf{w}\|_1\right\}. \tag{2.24}$$

Similar as in the linear regression case, this thesis estimates the DScore funcion and information matrices under $H_0 : \theta^* = 0$ as

$$\hat{S}(0, \hat{\boldsymbol{\gamma}}) = -\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - b'\left(\hat{\boldsymbol{\gamma}}^T \boldsymbol{X}_i\right)\right)\left(Z_i - \widehat{\mathbf{w}}^T \boldsymbol{X}_i\right), \tag{2.25}$$

$$\widehat{\mathbf{I}} = \frac{1}{n}\sum_{i=1}^{n} b''\left(\widehat{\boldsymbol{\beta}}^T \boldsymbol{Q_i}\right)\boldsymbol{Q}_i \boldsymbol{Q}_i^T, \text{ and} \tag{2.26}$$

$$\widehat{I}_{\theta|\boldsymbol{\gamma}} = \frac{1}{n}\sum_{i=1}^{n} b''\left(\widehat{\boldsymbol{\beta}}^T \boldsymbol{Q}_i\right) Z_i^2 - \widehat{\mathbf{w}}^T\left(\frac{1}{n}\sum_{i=1}^{n} b''\left(\hat{\boldsymbol{\beta}}^T \boldsymbol{Q}_i\right)\boldsymbol{X}_i Z_i\right). \tag{2.27}$$

For σ_S^* in the GLM case, the same derivation holds as in the linear regression case (Appendix C.4). Thus, again we have $\sigma_S^* = I_{\theta|\boldsymbol{\gamma}}^*$ and we can estimate the test statistic with

$$\widehat{U}_n = n^{1/2}\hat{S}(0, \hat{\boldsymbol{\gamma}})\widehat{I}_{\theta|\boldsymbol{\gamma}}^{-1/2}. \tag{2.28}$$

Appendix B.3 contains an additional assumption for GLMs, which ensures that Theorem 2.1 holds for all GLMs that validate this assumption. The corresponding corollary for the GLM case is Corollary 2.1.2. Two examples of GLMs that validate the assumptions of Corollary 2.1.2 are logistic regressions and Poisson regressions (Ning and Liu, 2017). Table 1 shows the specifications of these GLMs.

Corollary 2.1.2. *If Assumption B.6 holds,* $n^{-1/2}\max(s', s^*)\log d = o(1)$ *and* $\lambda \asymp \lambda' \asymp \sqrt{\frac{\log d}{n}}$, *then under* $H_0 : \theta^* = 0$ *for each* $t \in \mathbb{R}$,

$$\lim_{n\to\infty}\left|\mathbb{P}\left(\hat{U}_n \leq t\right) - \Phi(t)\right| = 0. \tag{2.29}$$

Table 1: GLM function specifications of logistic regressions and Poisson regressions.

Regression Type	$a(\phi)$	$b(x)$	$b'(x)$	$b''(x)$
Logistic	1	$\log(1 + \exp(x))$	$\frac{\exp(x)}{1+\exp(x)}$	$\frac{\exp(x)}{[1+\exp(x)]^2}$
Poisson	1	$\exp(x)$	$\exp(x)$	$\exp(x)$

2.4 False discovery rate (FDR) control

To the extent of our knowledge, this thesis is the first study that evaluates the multiple testing performance of the DScore test. Specifically, we test $H_{0,j} : \beta_j^* = 0$ for $j \in \text{supp}(\hat{\boldsymbol{\beta}})$, such that each test corresponds to a covariate with a nonzero parameter estimate: $|\text{supp}(\hat{\boldsymbol{\beta}})| = m$. Similar to Fang et al. (2020), this thesis defines the false discovery proportion (FDP) as the fraction of false rejections out of the total number of rejections for some pre-defined significance level $\alpha \in [0, 1]$:

$$\text{FDP}(\alpha) = \frac{\sum_{j \in (\text{supp}(\hat{\boldsymbol{\beta}}) \setminus \text{supp}(\boldsymbol{\beta}^*))} 1\,(p_j \leq \alpha)}{\max\left\{\sum_{j \in \text{supp}(\hat{\boldsymbol{\beta}})} 1\,(p_j \leq \alpha)\,, 1\right\}}, \tag{2.30}$$

of which the expected value is the false discovery rate (FDR):

$$\text{FDR}(\alpha) = \mathbb{E}[\text{FDP}(\alpha)]. \tag{2.31}$$

Furthermore, this thesis defines FDR control as controlling the FDRs to be lower than α, which is also referred to as the nominal FDR.

We consider four classic methods for adjusting the p-values for FDR control: (1) Bonferroni, (2) Holm, (3) Benjamini-Hochberg, and (4) Benjamini-Yekutieli (Bonferroni, 1936; Holm, 1979; Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). Algorithm 2 shows the procedure for calculating the adjusted p-values and rejecting the null hypotheses.

Algorithm 2 FDR control procedure.

Input: p-values $(p_1, ..., p_m)$, significance level α (=nominal FDR).

1. Sort the p-values such that $p_{(1)} \leq ... \leq p_{(m)}$.
2. Apply one of the following p-value adjustment methods[3]:
   - Bonferroni: For all $i \in \{1, ..., m\}$ calculate $p_{(i)}^{\text{B}} = \min\left(mp_{(i)}, 1\right)$.
   - Holm: For all $i \in \{1, ..., m\}$ calculate $p_{(i)}^{\text{H}} = \min\left(\left(\max_{j \leq i}(m - j + 1)p_{(j)}\right), 1\right)$.
   - Benjamini-Hochberg (BH): For all $i \in \{1, ..., m\}$ calculate $p_{(i)}^{\text{BH}} = \min\left(\left(\min_{j \geq i} \frac{m}{j} p_{(j)}\right), 1\right)$.
   - Benjamini-Yekutieli (BY): For all $i \in \{1, ..., m\}$ calculate $p_{(i)}^{\text{BY}} = \min\left(\left(\min_{j \geq i} \frac{m\,x(m)}{j} p_{(j)}\right), 1\right)$ with $x(m) = \sum_{i=1}^{m} 1/i$.
3. Reject hypothesis $H_{(i)}$ corresponding to $p_{(i)}$ if $p_{(i)}^t < \alpha$ for $t \in \{B, H, BH, BY\}$.

As discussed in Section 1.1, the Bonferroni and Holm methods are FWER correction techniques. Therefore, we can expect that these methods are conservative when it comes to controlling the

[3] The study of Yekutieli and Benjamini (1999) contains an overview of the p-value adjustment formulae of the Bonferroni, Holm, and BH methods. The formula of the BY method follows from the BH formula.

FDRs. The BH method is the original FDR correction method, which is more powerful than classic FWER correction methods according to Benjamini and Hochberg (1995) and is still widely used. The BY procedure is an extension of the BH method. Narum (2006) scrutinized the performance of the Bonferroni, BH, and BY methods. He finds that the BY method yields more power than the Bonferroni method but less power than the BH method. On the other hand, he finds that the BY method is better at controlling Type I errors than the BH method. Although FDR control methods, such as BH and BY, assume independence of the individual tests, Yekutieli and Benjamini (1999) state that most of these methods control the FDR in cases of dependency with some efficiency loss.

3 Simulation study

3.1 Set-up

The simulation set-up of this thesis is similar to the one of Ning and Liu (2017). This thesis performs 1000 simulations to assess the performance of the DScore test in linear regressions and GLMs. We believe this number of simulations is sufficiently large to approximate the distribution of the simulated Type I errors, power, and FDRs with the normal distribution via the central limit theorem. This assumption allows us to create confidence bands for the power plots and to perform two-tailed and right-tailed Z-tests to evaluate the empirical Type I errors and FDRs, respectively[4].

For both models, this thesis assumes a sample size of $n = 50$ and $d \in \{25, 50, 100\}$ covariates. This set-up differs from Ning and Liu (2017), as they use a larger sample size of $n = 200$. The smaller sample size allows us to see whether smaller samples can reflect the asymptotic properties of the test statistics. Furthermore, we select the tuning parameters λ and λ' of Algorithm 1 using 5-fold cross-validations and assume a 5% significance level. This thesis uses RStudio version 4.0.3 for the programming in R. Particularly, we use the R-package `glmnet` of Hastie and Qian (2016) for the estimation of Eqs. (1.2) and (2.3) with the Lasso penalty. The author of this thesis can supply the full programming codes of the simulations and real-world data applications upon request.

[4]For some significance level $\alpha \in [0, 1]$, this thesis considers the empirical Type I errors and FDRs to be accurate if we cannot reject the null-hypotheses $H_{0,1} : \mathbb{P}(\text{TI}) = \alpha$, where $\mathbb{P}(\text{TI})$ denotes the probability of a Type I error, and $H_{0,2} : \text{FDR} \leq \alpha$, respectively. Here, we choose to refrain from any multiple testing adjustments, as this would artificially increase the number of non-rejections.

3.1.1 Linear regression set-up

For a linear regression in the form $Y_i = \boldsymbol{\beta}^{*T}\boldsymbol{Q}_i + \varepsilon_i$, the simulations generate $\varepsilon_i \sim N(0,1)$ and $\boldsymbol{Q}_i \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{Q}})$, such that $\Sigma_{\boldsymbol{Q},jk} = \rho^{|j-k|}$ with Toeplitz parameter ρ. Like Ning and Liu (2017), the simulations consider $\rho \in \{0.25, 0.4, 0.6, 0.75\}$, and $\text{supp}(\boldsymbol{\gamma}^*) = s \in \{2,3\}$. The two settings for generating the nonzero elements of $\boldsymbol{\gamma}^*$ are:

1. Dirac setting: all nonzero elements are set to 1.
2. Uniform setting: all nonzero elements are independently drawn from the uniform distribution on the interval $[0,2]$.

Without loss of generality, we set $\beta_1^* = \theta^* = 0$ and test $H_0 : \beta_1^* = 0$ for the Type I error simulation. Next, this research compares the simulated Type I errors of the DScore test with the simulation results of two other testing methods[5]. The first is the debiasing method (SSLasso) of Javanmard and Montanari (2014) and the second is the ridge projection method (Ridge-Pro) of Bühlmann (2013). The test statistics corresponding to these methods asymptotically follow the normal distribution. For the computation of the SSLasso test statistic, we use the publicly available code[6] of Javanmard and Montanari (2014) and the `hdi`-package in R of Dezeure et al. (2015) for the Ridge-Pro test statistic.

Next, this thesis compares the power of the three methods by regenerating β_1^* on the interval $[0,1]$. Due to time constraints, we only consider the setting with $d = 25$, $\rho = 0.25$ and the Dirac approach for generating nonzero elements of $\boldsymbol{\gamma}^*$.

Last, this research performs FDR simulations of the DScore test following the same settings as the Type I error simulations. Here, we assume that each of the hypotheses concerns a covariate with a nonzero parameter estimate. Per iteration, the number of tests, thus, equals the number of nonzero estimates. The simulations compute the empirical FDRs for the raw p-values and the p-values after applying one of the four methods in Section 2.4. This thesis uses the build-in `stats`-package in R for the implementation of the p-value adjustment methods (R Core Team, 2020).

3.1.2 GLM set-up

In the GLM case, this research simulates logistic regressions and Poisson regressions. The generative processes of the covariates $\boldsymbol{Q}_i$ and the nonzero elements of $\boldsymbol{\gamma}^*$ are the same as in the linear regression

[5]For accurate comparisons, the simulations generate paired data, on which the three tests are applied.

[6]The SSLasso code is available at `https://web.stanford.edu/~montanar/sslasso/code.html`. Last accessed on July 4, 2021.

case. For the dependent variables Y_i, the generative processes are as follows:

- Logistic regression: $Y_i \sim Binom(1,p)$ with $p = \frac{1}{1+\exp(-\boldsymbol{\beta}^{*T}\boldsymbol{Q}_i)}$.
- Poisson regression: $Y_i \sim Poisson(\mu)$ with $\mu = \exp(\boldsymbol{\beta}^{*T}\boldsymbol{Q}_i)$.

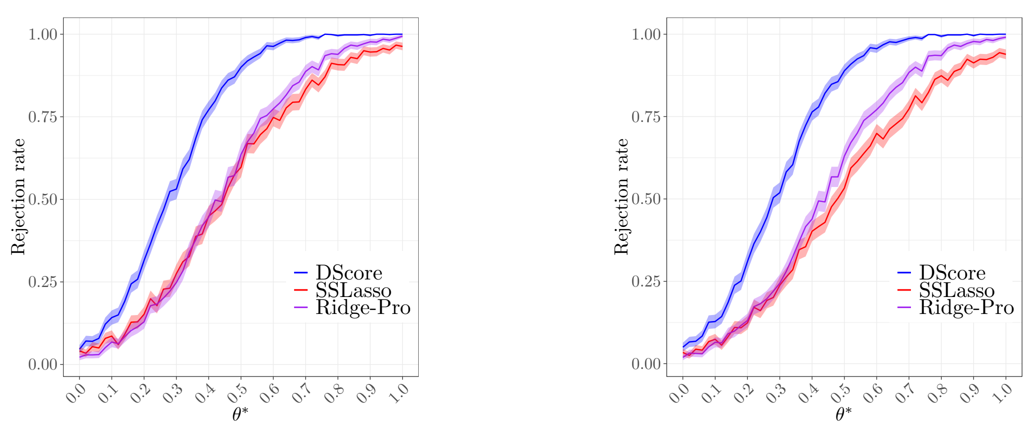
Furthermore, the simulations of the Type I errors and empirical FDRs follow the same setting as in the linear regression case. The power simulations use $\beta_1^* \in [0,2]$ for the logistic and Poisson regressions and compare them with the power of the linear regression over the same interval of β_1^*.

3.2 Simulation results

3.2.1 Linear regression results

Table 2 displays the simulated Type I errors of the three methods in the linear regression set-up. The results indicate that the Type I errors of the DScore test do not significantly differ from the significance level $\alpha = 5\%$ in all cases, excluding five slightly larger Type I errors when $d = 100$. Exceptions aside, the SSLasso method generates Type I errors that are larger than the significance level when $d \in \{50, 100\}$. For this method, the Type I errors only seem close to 5% for $s = 2$ and $d = 25$. These results differ from the findings of Ning and Liu (2017), as they found accurate Type I errors of the SSLasso method in all cases. Therefore, the smaller sample size used in our approach

Figure 1: Power of the test corresponding to $H_0 : \theta^* = 0$ in the linear regression case.



(a) $s = 2$. **(b)** $s = 3$.

Notes: DScore=Decorrelated Score method, SSlasso=Debiasing method, Ridge-Pro=Ridge projection method. The simulations assume $d = 25$, $\rho = 0.25$, and $\theta^* \in [0,1]$. The shaded areas around the curves indicate 95% confidence bands (based on a normal distribution approximation).

Table 2: Average simulated Type I errors (%) for the linear regression case.

			$\rho = 0.25$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.75$	
	s	d	Dirac	Unif	Dirac	Unif	Dirac	Unif	Dirac	Unif
DScore	2	25	4.6	4.5	4.9	4.9	5.2	4.4	5.5	5.3
	2	50	5.8	6.1	6.1	5.3	4.0	3.9	4.9	4.6
	2	100	4.8	5.0	6.7	7.1	7.0	6.3	6.9	6.8
	3	25	5.0	4.7	5.5	5.3	4.6	4.7	5.7	5.1
	3	50	5.3	5.2	6.1	6.2	4.5	4.7	5.1	4.7
	3	100	4.8	5.0	6.2	6.0	6.3	6.1	6.5	6.5
SSLasso	2	25	4.1	3.9	3.8	4.3	3.4	4.5	4.7	5.0
	2	50	6.9	7.0	8.2	7.8	6.8	6.5	9.9	8.7
	2	100	7.7	7.4	10.2	9.4	12.5	12.2	14.5	13.4
	3	25	3.4	3.0	2.9	3.8	3.2	3.8	3.3	4.1
	3	50	6.4	6.8	8.2	7.8	7.0	6.5	10.3	9.3
	3	100	7.0	6.8	10.0	9.1	12.3	11.8	14.0	13.0
Ridge-Pro	2	25	3.4	3.6	3.1	3.8	4.5	3.7	3.5	4.2
	2	50	2.9	2.5	2.3	3.5	3.2	2.7	4.8	4.3
	2	100	0.8	1.2	1.7	0.8	2.0	1.9	1.8	2.1
	3	25	3.6	3.4	3.1	3.4	4.4	2.3	3.9	3.7
	3	50	3.3	2.6	2.7	4.0	3.0	4.6	4.6	5.4
	3	100	0.8	0.5	1.5	0.4	1.5	1.7	1.6	2.6

Notes: Underlined values indicate that we cannot reject that the Type I error equals 5% using a two-tailed Z-test with a 5% significance level. We generate nonzero parameters to be equal to 1 (Dirac) or draw them uniformly on the interval [0,2] (Unif). Parameters: d=number of covariates, s = supp($\boldsymbol{\gamma}^*$), and ρ =Toeplitz parameter for the covariance matrix of the covariates. Tests: Decorrelated score test (DScore), Debiasing method (SSLasso), Ridge projection method (Ridge-Pro).

seems to less accurately reflect the asymptotic properties of this test statistic. Lastly, the Ridge-Pro method generates lower Type I errors than the significance level in almost all cases.

Fig. 1 shows two plots of the power of the three tests for $d = 25$ and $s \in \{2, 3\}$. In line with the findings of Ning and Liu (2017), the DScore test is the most powerful out of the three tests. Furthermore, inspections of the confidence bands of the SSLasso and Ridge-Pro methods indicate no significant difference in power when θ^* is small. For larger values of θ^*, the SSLasso method yields the lowest power. Note that the power curves corresponding to the Ridge-Pro method start at a lower rejection rate than the other two methods. This notion is as expected based on the Type I error results in Table 2. However, as the power curves of the Ridge-Pro method lie well outside the confidence bands of the DScore method, we believe the conclusion of a larger power for the latter method to hold.

Next, Table 3 shows the empirical FDRs of the linear regression simulations. Although all FDR correction methods decrease the empirical FDRs, the results indicate that none of the proposed FDR adjustment methods consistently control the FDRs when applied to the DScore test, especially when $\rho \in \{0.25, 0.4\}$. Nevertheless, the Bonferroni and BY methods cause the greatest empirical FDR reduction, followed by the Holm method. The BH method consistently causes the lowest FDR reduction, which is in line with the findings of Narum (2006). Furthermore, the larger FDR reduction of the Bonferroni and Holm methods compared to the BH method is as expected, as these two methods were designed to control the FWER, a more stringent condition than FDR control.

3.2.2 GLM results

Table 4 contains the average Type I errors for the GLM simulations of the logistic and Poisson regressions[7]. Although they do not differ significantly from 5% in most cases, the DScore method produces reliable errors less often in the two GLMs than in the linear regression case. Again, this result differs from Ning and Liu (2017), likely due to the smaller sample size.

Furthermore, Fig. 2 shows the power curves based on the simulations of the two GLMs and the linear regression. The DScore test statistic noticeably has a lower power in the logistic regression case than the other two models. The power curve of the Poisson regression case closely resembles the linear regression case, only with a slightly higher power when θ^* is small.

Finally, Appendix D contains the results of the FDR simulations in the logistic regression case and the Poisson regression case. These results are similar to the results in the linear regression case and draw the same conclusions.

[7] The Lasso method in the `glmnet`-package returns an empty model when the tuning parameters do not converge in the cross-validation stage. Due to the frequent occurrence of these convergence issues for logistic and Poisson regressions, this thesis re-simulates each of the instances that do not lead to converging tuning parameters. Hence, the number of simulations for these models still equals 1000 at the expense of longer runtimes. Also, this may cause potential violations of the independence of the simulated data if (some of) the convergence issues are related to specific groups of data.

Table 3: Empirical FDRs (%) of the DScore test for the linear regression case with a 5% nominal FDR.

			$\rho = 0.25$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.75$	
s	d	Adjustment	Dirac	Unif	Dirac	Unif	Dirac	Unif	Dirac	Unif
2	25	Raw	26.6	28.1	21.9	23.6	17.8	19.2	15.3	17.7
2	25	Bonferroni	6.0	7.9	4.9	6.1	3.6	4.8	3.4	3.9
2	25	Holm	9.8	11.5	8.1	8.8	5.9	6.7	5.7	5.7
2	25	BH	16.4	17.8	12.8	14.4	9.4	10.1	8.3	8.2
2	25	BY	7.3	8.9	5.9	6.4	4.2	4.8	3.7	3.8
2	50	Raw	37.8	39.0	32.5	33.7	23.2	26.4	17.6	20.8
2	50	Bonferroni	8.3	11.2	7.0	8.8	4.1	5.8	3.3	4.4
2	50	Holm	13.4	15.3	12.0	13.4	8.1	8.5	5.0	5.9
2	50	BH	23.3	23.5	19.9	20.2	12.9	12.8	8.1	9.1
2	50	BY	10.1	12.2	8.7	9.6	5.4	5.5	3.8	4.2
2	100	Raw	47.4	47.9	39.2	40.7	29.5	31.5	20.4	23.1
2	100	Bonferroni	13.1	16.5	9.7	11.5	6.0	7.5	3.8	4.1
2	100	Holm	20.0	21.1	16.2	16.3	10.1	10.8	5.9	6.1
2	100	BH	32.2	31.9	25.6	25.4	17.0	16.3	10.3	10.3
2	100	BY	16.7	18.3	12.5	13.2	7.6	7.8	4.3	4.1
3	25	Raw	20.9	22.7	17.8	19.4	13.5	15.4	11.7	13.8
3	25	Bonferroni	3.9	5.8	3.2	4.2	2.1	3.4	2.4	3.1
3	25	Holm	6.7	8.1	5.6	6.8	4.1	5.2	4.1	4.7
3	25	BH	13.2	14.3	10.9	11.5	7.4	8.3	6.9	7.6
3	25	BY	5.4	6.6	4.5	4.9	3.1	3.9	3.3	3.6
3	50	Raw	31.6	33.4	26.5	28.8	18.6	21.0	13.8	16.9
3	50	Bonferroni	5.2	7.7	4.2	5.7	2.6	3.6	2.1	2.9
3	50	Holm	9.5	12.0	7.9	9.3	5.0	6.0	3.7	4.2
3	50	BH	19.2	21.2	15.4	17.2	9.9	10.8	7.0	7.1
3	50	BY	8.0	9.7	6.0	7.3	3.8	4.2	2.8	3.1
3	100	Raw	41.5	43.9	34.2	36.7	23.1	25.8	16.0	18.9
3	100	Bonferroni	8.2	11.8	6.1	7.9	3.3	4.5	2.5	3.1
3	100	Holm	13.5	16.5	10.8	12.6	7.1	7.6	4.6	5.1
3	100	BH	27.3	28.9	21.7	22.0	13.1	14.5	8.3	8.7
3	100	BY	11.8	14.7	9.1	10.2	5.3	5.5	3.5	3.6

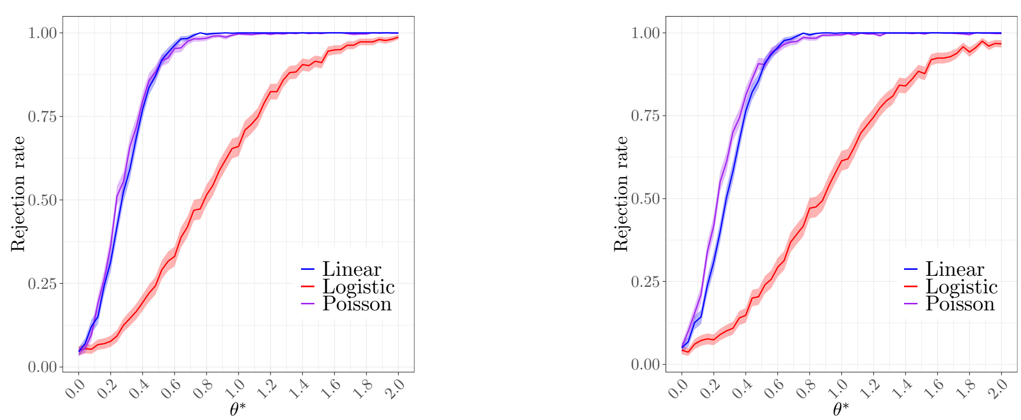
Notes: Underlined values indicate that we cannot reject that the FDR is lower than 5% using a right-tailed Z-test with a 5% significance level. We generate nonzero parameters to be equal to 1 (Dirac) or draw them uniformly in the interval [0,2] (Unif). Parameters: d=number of covariates, $s = \text{supp}(\boldsymbol{\gamma}^*)$, and ρ =Toeplitz parameter for the covariance matrix of the covariates. Adjustment methods: unadjusted p-values (Raw), Bonferroni, Holm, Benjamini-Hochberg (BH), and Benjamini-Yekutieli (BY).

Table 4: Average simulated Type I errors (%) for the GLM case (logistic regression and Poisson regression).

	s	d	$\rho = 0.25$ Dirac	$\rho = 0.25$ Unif	$\rho = 0.4$ Dirac	$\rho = 0.4$ Unif	$\rho = 0.6$ Dirac	$\rho = 0.6$ Unif	$\rho = 0.75$ Dirac	$\rho = 0.75$ Unif
Logistic	2	25	4.4	3.9	6.2	4.4	6.2	6.1	7.7	7.4
	2	50	4.4	3.3	5.1	3.7	5.2	4.6	6.4	5.4
	2	100	3.4	2.8	3.5	3.3	5.2	4.7	5.1	4.3
	3	25	4.6	4.6	6.3	4.6	5.7	4.8	5.6	5.0
	3	50	3.6	2.5	4.6	3.3	4.5	3.8	6.6	5.1
	3	100	1.9	2.8	3.0	3.0	3.5	3.5	4.6	4.3
Poisson	2	25	4.6	5.6	4.8	4.6	5.4	4.4	5.6	3.7
	2	50	3.3	4.2	3.8	3.2	4.1	3.9	4.8	3.5
	2	100	2.2	2.5	2.5	3.8	2.8	3.6	4.5	4.4
	3	25	5.2	6.2	5.9	5.6	7.3	5.1	4.2	3.9
	3	50	4.6	4.0	3.4	4.5	4.4	4.0	4.6	4.2
	3	100	2.1	2.7	2.7	5.0	4.2	4.6	4.4	5.7

Notes: Underlined values indicate that we cannot reject that the Type I error equals 5% using a two-tailed Z-test with a 5% significance level. We generate nonzero parameters to be equal to 1 (Dirac) or draw them uniformly on the interval [0,2] (Unif). Parameters: d=number of covariates, $s = \text{supp}(\boldsymbol{\gamma}^*)$, and ρ =Toeplitz parameter for the covariance matrix of the covariates.

Figure 2: Power of the test corresponding to $H_0 : \theta^* = 0$ in the linear, logistic and Poisson regression simulations.



(a) $s = 2$.

(b) $s = 3$.

Notes: The simulations assume $d = 25$, $\rho = 0.25$, and $\theta^* \in [0, 2]$. The shaded areas around the curves indicate 95% confidence bands (based on a normal distribution approximation).

4 Real-world applications

4.1 Linear regression application: Freddie Mac House Price Index

The first real-world data application of the DScore test relates to the macro-economy. Specifically, this thesis performs a linear regression using monthly seasonally-adjusted house price index (HPI) data of Freddie Mac (2021). This thesis investigates four sample periods:

1. January 1975 to December 1990: 16 years (192 observations).
2. January 1975 to December 2000: 26 years (312 observations).
3. January 1975 to December 2010: 36 years (432 observations).
4. January 1975 to December 2020: 46 years (552 observations).

Furthermore, we take the log-difference of the HPIs to measure the relative change in house prices. The regression equation is as follows:

$$HPI_{US,t} = \sum_{i=1}^{382} \beta_i HPI_{MSA,i,t} + \varepsilon_t \text{ for } t \in \{1, ..., \mathcal{T} - 1\}, \tag{4.1}$$

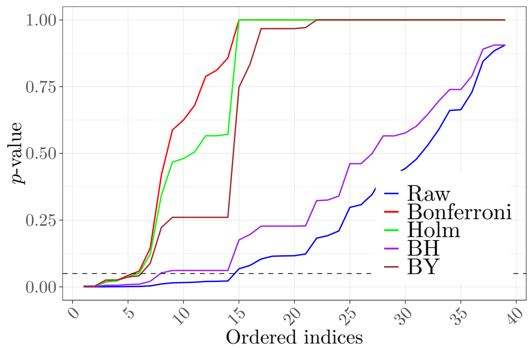
where $HPI_{US,t}$ is the log-differenced US HPI, $HPI_{MSA,i,t}$ is the log-differenced HPI corresponding to one of 382 metropolitan statistical areas (MSAs) in the US, and $\mathcal{T}$ is the sample size. Again, we use the Lasso penalty in Eq. (1.2) to estimate the coefficients. Next, this thesis tests the null hypotheses $H_{0,j}$: $\beta_j = 0$ for $j \in \text{supp}(\hat{\boldsymbol{\beta}})$, such that $\text{supp}(\hat{\boldsymbol{\beta}})$ is the set of nonzero estimates. Afterward, this thesis uses the four methods of Section 2.4 to correct for the FDRs.

Fig. 3 shows plots of the ordered p-values of the four samples before and after the adjustment with one of the FDR control methods. These plots show a consistent order of the adjustment methods over the four samples: the Bonferroni method is the most conservative, followed by the Holm method and the BY method. In line with our simulation results, the BH method decreases the number of significant p-values the least. For additional information, Appendix E contains an overview of the MSAs that correspond to these significant p-values in the four samples.

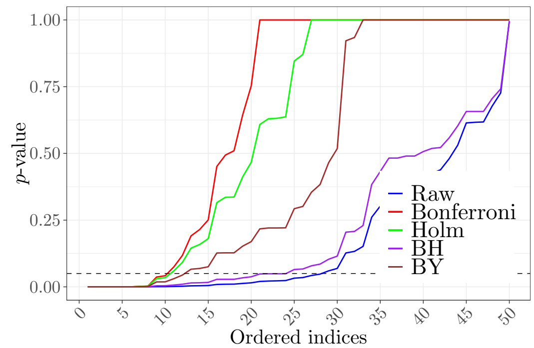
4.2 GLM application: Prostate cancer study of Singh et al. (2002)

Next, this thesis presents an application of the DScore test for a logistic regression in the field of genetics. In particular, we use the same gene expression data-set as Fonti and Belitser (2017), who investigated feature selection in high-dimensional GLMs using the Lasso estimator. The data-set originates from the study of Singh et al. (2002) and is available in the `sda`-package in R (Ahdesmäki

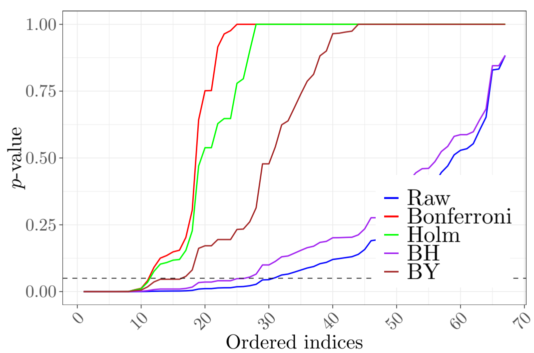
Figure 3: The ordered p-values of the house price index application.



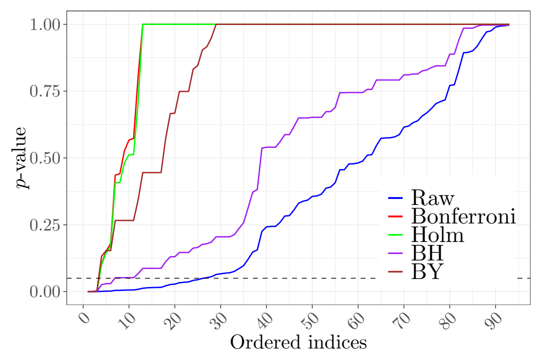
(a) January 1975-December 1990 (16 years).



(b) January 1975-December 2000 (26 years).



(c) January 1975-December 2010 (36 years).



(d) January 1975-December 2020 (46 years).

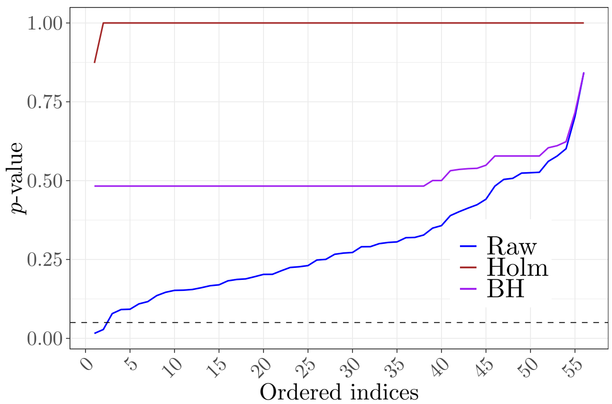
Notes: The plots include the raw p-values and the p-values after adjustment with the FDR correction methods of (1) Bonferroni, (2) Holm, (3) Benjamini-Hochberg (BH), and (4) Benjamini-Yekutieli (BY). The horizontal black dotted line corresponds with a 5% significance level.

and Strimmer, 2010). It consists of 102 samples with gene expression data of 6033 genes. Of these 102 samples, 52 are from individuals diagnozed with prostate cancer and 50 from healthy individuals. Hence, we define the dependent variable of the logistic regression as

$$Y_i = \begin{cases} 1, & \text{if individual } i \text{ has prostate cancer.} \\ 0, & \text{if individual } i \text{ is healthy.} \end{cases} \tag{4.2}$$

After regressing the dependent variable on the 6033 gene expressions, this thesis performs the DScore test over all coefficients with nonzero estimates. We find that the only gene expressions to be significant on a 5% level correspond to genes 610 and 914. Fonti and Belitser (2017) find a set of 10 relevant features (here: gene expressions), which includes these two genes. However, we must note that Fonti and Belitser (2017) do not perform significance tests and instead determine whether features are relevant by checking if their estimates are nonzero.

Figure 4: The ordered p-values of the prostate cancer gene expression application.



Notes: The plot includes the raw p-values and the p-values after adjustment with the Holm and Benjamini-Hochberg (BH) corrections. The horizontal black dotted line corresponds with a 5% significance level.

Applying each of the four FDR control methods yields no results that are significant on a 5% level. Similar to the HPI application example, Fig. 4 shows the ordered p-values before and after applying the FDR correction methods. We omit the Bonferroni and BY adjustments from this plot as they closely overlap with the adjusted p-values of the Holm method. A possible explanation of the findings of only two significant features before and no significant after FDR correction is the lower power of the DScore test for logistic regressions (see Fig. 2).

5 Discussion

5.1 Conclusion

This thesis investigates the DScore test of Ning and Liu (2017) in linear regressions and GLMs. In particular, we assess the small sample performance of the test and evaluate its multiple testing ability, which we believe has not been done before. Recall the research questions of this thesis:

RQ1: *How is the finite sample performance of the DScore test in linear models?*

RQ2: *How can we control the FDRs in linear models when testing multiple hypotheses with the DScore test?*

Concerning RQ1, this thesis finds that the DScore test generally has consistent Type I errors and more power than the Ridge-Pro and SSLasso methods when applied to linear regression models.

Conversely, the DScore test in GLMs with a small number of observations is less reliable than in linear regressions due to less consistent Type I errors and a lower power in logistic regressions. This finding differs from Ning and Liu (2017), who find a good performance for logistic regressions and Poisson regressions in a setting with more observations.

Regarding RQ2, this research finds that none of the four classic multiple testing correction methods consistently control the FDRs of the DScore test in linear regressions and GLMs. However, they still cause a substantial decrease in the empirical FDRs. Of the methods that this thesis investigates, the p-value adjustments of the Bonferroni and BY methods are the most conservative, while the BH method is the least conservative.

The findings of this thesis implicate that the DScore test is, in general, an efficient and powerful method for testing single hypotheses in sparse high-dimensional linear regression models. However, applying the DScore test in GLMs is not guaranteed to produce accurate results in small samples. Moreover, the test may not accurately control the FDRs when conducting multiple tests even after correction with conventional FDR control methods. Therefore, we believe that researchers who want to perform multiple testing with the DScore test should interpret the resulting p-values with caution.

5.2 Limitations and research recommendations

A limitation of this thesis relates to the simulated power curves of different tests. Because their power simulations do not always correspond to the same empirical Type I errors, this thesis cannot make any bold claims about differences in power of tests with roughly similar power curves. Also, we only consider two cases due to time constraints. Asserting whether one method consistently has a high power requires power analysis of other cases, such as the 48 cases considered for the simulations of the empirical Type I errors and FDRs.

Furthermore, the GLM simulations in this research re-simulate the data of iterations, in which the tuning parameters do not converge in the cross-validation stage. As inspections of potential relations between the data corresponding to these issues are costly, this thesis omits them from the analysis. This creates the problem of potential dependency structures in the simulated data, which can influence the results.

Several interesting avenues for future research exist. First, one can perform our simulation study for a range of sample sizes. Doing this for the same GLMs as in our study or others, like exponential regressions, may add additional support to our finding of a less reliable small sample performance of the DScore test in GLMs.

Also, it may be interesting to consider alternative FDR control methods. For example, the study of Fang et al. (2020) uses the FDR correction method of Storey (2002), which is more powerful than the BH method. It may be interesting to see whether this approach leads to less conservative results in the simulations and real-world data applications of our research.

6 Acknowledgement

I want to thank Dr. A.J. Koning for supervising my Bachelor thesis. He provided me with valuable insights and suggestions, while he was monitoring my progress.

Bibliography

Ahdesmäki, M. and Strimmer, K. (2010). Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. *The Annals of Applied Statistics*, pages 503–519.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, pages 1165–1188.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62.

Bühlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212 – 1242.

Dezeure, R., Bühlmann, P., Meier, L., and Meinshausen, N. (2015). High-dimensional inference: confidence intervals, p-values and R-software hdi. *Statistical Science*, pages 533–558.

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

Fang, E. X., Ning, Y., Li, R., et al. (2020). Test of significance for high-dimensional longitudinal data. *Annals of Statistics*, 48(5):2622–2645.

Fonti, V. and Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, 30:1–25.

Freddie Mac (2021). Freddie Mac House Price Index. Data retrieved from http://www.freddiemac.com/research/indices/house-price-index.page on June 12, 2021.

Hastie, T. and Qian, J. (2016). Glmnet vignette. Retrieved from https://mran.microsoft.com/snapshot/2018-04-27/web/packages/glmnet/vignettes/glmnet_beta.pdf on June 18, 2021.

Heij, C., Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K., et al. (2004). *Econometric methods with applications in business and economics*. Oxford University Press.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, pages 65–70.

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

Mykland, P. A. (1999). Bartlett identities and large deviations in likelihood theory. *The Annals of Statistics*, 27(3):1105–1117.

Narum, S. R. (2006). Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics*, 7(5):783–787.

Neykov, M., Ning, Y., Liu, J. S., and Liu, H. (2018). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statistical Science*, 33(3):427–443.

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

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Stats package information available at https://prs.ism.ac.jp/~nakama/Rjp/stats-manual.pdf (Last accessed: July 4, 2021).

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 44, pages 50–57. Cambridge University Press.

Shah, R. D. and Samworth, R. J. (2013). Variable selection with error control: Another look at stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):55–80.

Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1(2):203–209.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498.

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., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

Yekutieli, D. and Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82(1-2):171–196.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942.

Appendix A Glossary of the notation and abbreviations

This thesis follows the mathematical notation of Ning and Liu (2017). We denote vectors in bold italics and matrices in bold roman, e.g., $\boldsymbol{X}$ denotes a vector and $\mathbf{X}$ a matrix. Furthermore, Tables A.1 and A.2 contain an overview of respectively the mathematical symbols and operators and the abbreviations that this thesis uses.

Table A.1: Mathematical symbols and operators used in this thesis with their definitions.

Notation	Definition
$\text{supp}(\cdot)$	Set containing the nonzero elements of a vector: $\text{supp}(\mathbf{v}) = \{j : v_j \neq 0\}$
$\|\mathbf{v}\|_q$, $1 \leq q \leq \infty$	The L_q-norm of $\mathbf{v}$: $\|\mathbf{v}\|_q = \left(\sum_{i=1}^{d} \|v_i\|^q\right)^{1/q}$
$a_n = o(b_n)$	$\lim_{n\to\infty} \frac{a_n}{b_n} = 0$
$a_n = \mathcal{O}(b_n)$	$a_n \leq Cb_n$ for some $C > 0$ and n sufficiently large
$X_n = o_{\mathbb{P}}(b_n)$	$\lim_{n\to\infty} \frac{X_n}{b_n} = 0$
$X_n = \mathcal{O}_{\mathbb{P}}(b_n)$	For any $\epsilon > 0$, we have $P(X_n \leq Cb_n) \geq 1 - \epsilon)$ for some $C > 0$ and n sufficiently large
$a_n \asymp b_n$	$C \leq \frac{a_n}{b_n} \leq C'$ for some $C, C' > 0$
$a_n \lesssim b_n$	$a_n \leq Cb_n$ for some $C > 0$
$\lambda_{min}(\mathbf{M})$ and $\lambda_{max}(\mathbf{M})$	The minimum and maximum eigenvalue of $\mathbf{M}$
$\nabla f(\cdot)$	Gradient of $f(\cdot)$
$\nabla_S f(\cdot)$	Gradient of $f(\cdot)$ with respect to S
$X_n \rightsquigarrow X$	X_n converges weakly to X
$\|Y\|_{\psi 2}$	The sub-Gaussian norm of Y: $\sup_{p\geq 1} p^{-1/2} (\mathbb{E}\|Y\|^p)^{1/p}$

Notes: The vector $\boldsymbol{v}$ is defined as $\mathbf{v} = (v_1, ..., v_d)^T \in \mathbb{R}^d$. $\mathbf{M} \in \mathbb{R}^{n\times d}$ is a matrix. a_n and b_n are positive sequences. X_n is a sequence of random numbers. Y is a random variable.

Table A.2: Abbreviations used in this thesis with their descriptions.

Abbreviation	Description
BH	Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995)
BY	Benjamini-Yekutieli procedure (Benjamini and Yekutieli, 2001)
DScore	Decorrelated score test (Ning and Liu, 2017)
FDP	False discovery proportion
FDR	False discovery rate
FWER	Family-wise error rate
GLM	Generalized linear model
HPI	House price index
MSA	Metropolitan statistical area
Ridge-Pro	Ridge projection test (Bühlmann, 2013)
SSLasso	Debiasing test (Javanmard and Montanari, 2014)

Appendix B Assumptions

B.1 General assumptions

Ning and Liu (2017) make Assumptions B.1-B.4 that loss functions must validate before Theorem 2.1 can hold.

Assumption B.1 (Consistency conditions for initial parameter estimation)**.** *For some sequences $\eta_1(n)$ and $\eta_2(n)$ that converge to 0 when $n \to \infty$, it holds that*

$$\lim_{n\to\infty} \mathbb{P}\left(\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1 \lesssim \eta_1(n)\right) = 1 \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}\left(\|\hat{\mathbf{w}} - \mathbf{w}^*\|_1 \lesssim \eta_2(n)\right) = 1. \tag{B.1}$$

Intuition: $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{w}}$ are consistent estimators of $\boldsymbol{\beta}^*$ and $\mathbf{w}^*$.

Assumption B.2 (Concentration of the gradient and Hessian)**.** *For $\mathbf{v}^* = (1, -\mathbf{w}^{*T})^T$, we assume*

$$\begin{aligned}
&\|\nabla \ell\left(\beta^*\right)\|_\infty = \mathcal{O}_{\mathbb{P}}(\sqrt{\log d/n}), \text{ and}\\
&\left\|\mathbf{v}^{*T}\nabla^2 \ell\left(\boldsymbol{\beta}^*\right) - \mathbb{E}_{\beta^*}\left(\mathbf{v}^{*T}\nabla^2\ell\left(\boldsymbol{\beta}^*\right)\right)\right\|_\infty = \mathcal{O}_{\mathbb{P}}(\sqrt{\log d/n}).
\end{aligned} \tag{B.2}$$

Assumption B.3 (Local smoothness on the loss function)**.** *Let $\hat{\boldsymbol{\beta}}_0 = (0, \hat{\boldsymbol{\gamma}}^T)^T$, $\hat{\mathbf{v}} = (1, -\hat{\mathbf{w}}^T)^T$, and $\mathbf{v}^* = (1, -\mathbf{w}^{*T})^T$. For $\check{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_0$ and $\check{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$, we assume*

$$\begin{aligned}
&\mathbf{v}^{*T}\left\{\nabla\ell(\check{\boldsymbol{\beta}}) - \nabla\ell(\boldsymbol{\beta}^*) - \nabla^2\ell(\boldsymbol{\beta}^*)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right\} = o_{\mathbb{P}}\left(n^{-1/2}\right), \text{ and}\\
&(\hat{\mathbf{v}} - \mathbf{v}^*)^T(\nabla\ell(\check{\boldsymbol{\beta}}) - \nabla\ell(\boldsymbol{\beta}^*)) = o_{\mathbb{P}}\left(n^{-1/2}\right).
\end{aligned} \tag{B.3}$$

Assumption B.4 (Central limit theorem for the score function)**.** *Assume that*

$$\sqrt{n}\mathbf{v}^{*T}\nabla\ell(\boldsymbol{\beta}^*)/\sqrt{\sigma_S^*} \rightsquigarrow N(0,1), \tag{B.4}$$

where $\sigma_S^* = \mathbf{v}^{*T}\boldsymbol{\Sigma}^*\mathbf{v}^*$ and $\sigma_S^* \geq C$ for some constant $C > 0$.

B.2 Additional assumption for linear regressions

Ning and Liu (2017) make Assumption B.5 for the derivation of the DScore test statistic in the linear regression case.

Assumption B.5 (Additional assumption for linear regressions). *Ning and Liu (2017) assume the following for the derivation of the linear regression DScore test statistic:*

1. $\|\varepsilon_i\|_{\psi 2} \leq C$ *for some constant* C.

2. $2\kappa \leq \lambda_{min}(\mathbb{E}(\boldsymbol{Q}_i\boldsymbol{Q}_i^T)) \leq \lambda_{max}(\mathbb{E}(\boldsymbol{Q}_i\boldsymbol{Q}_i^T)) \leq 2/\kappa$ *for some constant* κ.

3. $\boldsymbol{Q}_i$ *is a sub-Gaussian vector.*[8]

B.3 Additional assumption for GLMs

Ning and Liu (2017) make Assumption B.6 for the derivation of the DScore test statistic in the GLM case.

Assumption B.6 (Additional assumption for GLMs). *Ning and Liu (2017) assume the following five conditions for the derivation of DScore test statistic for GLMs:*

1. $\lambda_{\min}(\mathbf{I}^*) \geqslant \kappa^2$ *for some constant* $\kappa > 0$.

2. $S = \operatorname{supp}(\boldsymbol{\beta}^*)$ *and* $S' = \operatorname{supp}(\mathbf{w}^*)$ *satisfy* $|S| = s^*$ *and* $|S'| = s'$.

3. $\|\boldsymbol{Q}_i\|_\infty \leqslant K, \left|\mathbf{w}^{*T}\boldsymbol{X}_i\right| \leqslant K$, *for some constant* K, *and* $\left|Y_i - b'\left(\boldsymbol{Q}_i^T\boldsymbol{\beta}^*\right)\right|$ *is sub-exponential.*[9]

4. $\boldsymbol{Q}_i^T\boldsymbol{\beta}^* \in [K_1, K_2]$ *for* $K_2 > K_1$ *and* $K_1, K_2 \in \mathbb{R}$.

5. $\forall t \in [K_1 - \epsilon, K_2 + \epsilon]$ *with some constant* $\epsilon > 0$ *and a sequence* t_1 *satisfying* $|t_1 - t| = o(1)$, *it holds that* $0 < b''(t) \leqslant C$ *and* $|b''(t_1) - b''(t)| \leqslant C|t_1 - t|\, b''(t)$ *for some constant* $C > 0$.

---

[8] A random variable v is sub-Gaussian if $\mathbb{P}(|v| > t) \leq \exp(1 - t^2/K^2)$ for some $K > 0$ and for all t.

[9] A random variable v is sub-exponential if $\mathbb{P}(|v| > t) \leq \exp(1 - t/K)$ for some $K > 0$ and for all t.

Appendix C Additional proofs and derivations

This section contains additional proofs and derivations of the parts that the DScore function calculations require or aid in the derivation of the asymptotic properties of the test statistic.

C.1 Proof of Theorem 2.1

This subsection shows an elaborate version of the proof of Theorem 2.1 from Ning and Liu (2017).

Proof. Let $\hat{\boldsymbol{\beta}}_0 = (0, \hat{\boldsymbol{\gamma}}^T)^T$, $\hat{\mathbf{v}} = (1, -\hat{\mathbf{w}}^T)^T$, and $\mathbf{v}^* = (1, -\mathbf{w}^{*T})^T$.The following holds:

$$
\begin{aligned}
n^{1/2}\left|\hat{S}(\hat{\boldsymbol{\beta}}_0) - S\left(\boldsymbol{\beta}^*\right)\right| &= n^{1/2}\left|\hat{\mathbf{v}}^T \nabla \ell\left(\hat{\boldsymbol{\beta}}_0\right) - \mathbf{v}^{*T}\nabla \ell\left(\boldsymbol{\beta}^*\right)\right| \\
&= n^{1/2}\left|\hat{\mathbf{v}}^T \nabla \ell\left(\hat{\boldsymbol{\beta}}_0\right) - \mathbf{v}^{*T}\nabla \ell\left(\boldsymbol{\beta}^*\right) + \mathbf{v}^{*T}\nabla \ell\left(\hat{\boldsymbol{\beta}}_0\right) - \mathbf{v}^{*T}\nabla \ell\left(\hat{\boldsymbol{\beta}}_0\right)\right| \\
&\leq n^{1/2}\left|\mathbf{v}^{*T}\left\{\nabla \ell\left(\hat{\boldsymbol{\beta}}_0\right) - \nabla \ell\left(\boldsymbol{\beta}^*\right)\right\}\right| + n^{1/2}\left|\left(\hat{\mathbf{v}} - \mathbf{v}^*\right)^T \nabla \ell\left(\hat{\boldsymbol{\beta}}_0\right)\right| \\
&:= I_1 + I_2.
\end{aligned}
\tag{C.1}
$$

From Assumption B.3, we know that

$$
\begin{aligned}
|I_1| &= n^{1/2}\left|\mathbf{v}^{*T}\left\{\nabla \ell\left(\hat{\boldsymbol{\beta}}_0\right) - \nabla \ell\left(\boldsymbol{\beta}^*\right)\right\}\right| \\
&\leq n^{1/2}\left|\mathbf{v}^{*T}\nabla^2 \ell\left(\boldsymbol{\beta}^*\right)\left(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^*\right)\right| + o_{\mathbb{P}}(1) \\
&\leq n^{1/2}\left\|\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^*\right\|_1 \left\|\nabla^2_{\theta\boldsymbol{\gamma}} \ell\left(\boldsymbol{\beta}^*\right) - \mathbf{w}^{*T}\nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell\left(\boldsymbol{\beta}^*\right)\right\|_\infty + o_{\mathbb{P}}(1), \text{ and that}
\end{aligned}
\tag{C.2}
$$

$$
\begin{aligned}
|I_2| &\leq n^{1/2}\left|\left(\hat{\mathbf{v}} - \mathbf{v}^*\right)^T \nabla \ell\left(\boldsymbol{\beta}^*\right)\right| + o_{\mathbb{P}}(1) \\
&\leq n^{1/2}\left\|\hat{\mathbf{v}} - \mathbf{v}^*\right\|_1 \left\|\nabla \ell\left(\boldsymbol{\beta}^*\right)\right\|_\infty + o_{\mathbb{P}}(1).
\end{aligned}
\tag{C.3}
$$

Furthermore, Assumptions B.1 and B.2 show that

$$
\begin{cases}
|I_1| \lesssim \eta_1(n)\sqrt{\log d} + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1), \text{ and} \\
|I_2| \lesssim \eta_2(n)\sqrt{\log d} + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).
\end{cases}
\tag{C.4}
$$

Combining the bounds of I_1 and I_2 with Eq. (C.1), we get

$$
n^{1/2}\left|\hat{S}(\hat{\boldsymbol{\beta}}_0) - S(\boldsymbol{\beta}^*)\right| = o_{\mathbb{P}}(1).
\tag{C.5}
$$

The final step is combining Eq. (C.5) with the notion that $n^{1/2} S\left(\boldsymbol{\beta}^*\right) \sigma_s^{*-1/2} \rightsquigarrow N(0,1)$ with $\sigma_S^* \geq C$ (Assumption B.4) into

$$
n^{1/2}\left|\hat{S}(0, \hat{\boldsymbol{\gamma}})\sigma_s^{*-1/2} - S\left(0, \boldsymbol{\gamma}^*\right)\sigma_s^{*-1/2}\right| = o_{\mathbb{P}}(1).
\tag{C.6}
$$

Now, applying Slutsky's theorem achieves the desired result. □

C.2 Supplement to Eq. (2.12)

Eq. (C.7) shows the derivations of respectively $\nabla_{\theta}\ell(\theta,\boldsymbol{\gamma})$ and $\nabla_{\boldsymbol{\gamma}}\ell(\theta,\boldsymbol{\gamma})$, which are inputs for the DScore function of the linear regression case.

$$\begin{aligned}
\nabla_{\theta}\ell(\theta,\boldsymbol{\gamma}) &= \nabla_{\theta}\left[(2n\sigma^2)^{-1}\sum_{i=1}^{n}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\mathbf{X}_i\right)^2\right] \\
&= (2n\sigma^2)^{-1}\sum_{i=1}^{n}\nabla_{\theta}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)^2 \\
&= (2n\sigma^2)^{-1}\sum_{i=1}^{n}2\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)(-Z_i) \\
&= -\frac{1}{n\sigma^2}\sum_{i=1}^{n}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)Z_i, \text{ and} \\
\nabla_{\boldsymbol{\gamma}}\ell(\theta,\boldsymbol{\gamma}) &= \nabla_{\boldsymbol{\gamma}}\left[(2n\sigma^2)^{-1}\sum_{i=1}^{n}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\mathbf{X}_i\right)^2\right] \\
&= (2n\sigma^2)^{-1}\sum_{i=1}^{n}\nabla_{\boldsymbol{\gamma}}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)^2 \\
&= (2n\sigma^2)^{-1}\sum_{i=1}^{n}2\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)(-\boldsymbol{X}_i) \\
&= -\frac{1}{n\sigma^2}\sum_{i=1}^{n}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)\boldsymbol{X}_i.
\end{aligned} \tag{C.7}$$

In addition, Eq. (C.8) derives the partial second-order gradients, which the formula of $\mathbf{w}$ requires, for example.

$$\begin{aligned}
\nabla^2_{\boldsymbol{\gamma}\theta}\ell(\theta,\boldsymbol{\gamma}) &= \nabla_{\theta}[\nabla_{\boldsymbol{\gamma}}\ell(\theta,\boldsymbol{\gamma})] \\
&= \nabla_{\theta}[-\frac{1}{n\sigma^2}\sum_{i=1}^{n}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)\boldsymbol{X}_i] \\
&= \frac{1}{n\sigma^2}\sum_{i=1}^{n}Z_i\boldsymbol{X}_i, \text{ and} \\
\nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\ell(\theta,\boldsymbol{\gamma}) &= \nabla_{\boldsymbol{\gamma}}[\nabla_{\boldsymbol{\gamma}}\ell(\theta,\boldsymbol{\gamma})] \\
&= \nabla_{\theta}[-\frac{1}{n\sigma^2}\sum_{i=1}^{n}\left(Y_i-\theta Z_i-\boldsymbol{\gamma}^T\boldsymbol{X}_i\right)\boldsymbol{X}_i] \\
&= \frac{1}{n\sigma^2}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^T.
\end{aligned} \tag{C.8}$$

C.3 Derivations of (partial) information matrices $\mathbf{I}^*$ and $I^*_{\theta|\boldsymbol{\gamma}}$ in the linear regression case

Eq. (C.9) shows the Fischer information matrix $\mathbf{I}^*$ and partial Fischer information matrix $I^*_{\theta|\boldsymbol{\gamma}}$ for the linear regression case.

$$
\begin{aligned}
\mathbf{I}^* &= \mathbb{E}\left(\nabla^2 \ell\left(\boldsymbol{\beta}^*\right)\right) \\
&= \mathbb{E}\left(\nabla\left[-\frac{1}{n\sigma^2}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{\beta}^T \boldsymbol{Q}_i\right)\boldsymbol{Q}_i^T\right]\right) \\
&= \mathbb{E}\left(\frac{1}{\sigma^2 n}\sum_{i=1}^{n}\boldsymbol{Q}_i\boldsymbol{Q}_i^T\right) \\
&= \frac{1}{\sigma^2 n}\sum_{i=1}^{n}\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{Q}_i^T\right) \\
&= \frac{1}{\sigma^2}\mathbb{E}\left(\boldsymbol{Q}_i\boldsymbol{Q}_i^T\right), \text{ and} \\
I^*_{\theta|\boldsymbol{\gamma}} &= I^*_{\theta\theta} - \mathbf{I}^*_{\theta\boldsymbol{\gamma}}\mathbf{I}^{*-1}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{I}^*_{\boldsymbol{\gamma}\theta} \\
&= \mathbb{E}\left(\nabla^2_{\theta\theta}\ell\left(\boldsymbol{\beta}^*\right)\right) - \mathbb{E}\left(\nabla^2_{\theta\boldsymbol{\gamma}}\ell\left(\boldsymbol{\beta}^*\right)\right)\mathbb{E}\left(\nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\ell\left(\boldsymbol{\beta}^*\right)\right)^{-1}\mathbb{E}\left(\nabla^2_{\boldsymbol{\gamma}\theta}\ell\left(\boldsymbol{\beta}^*\right)\right) \\
&= \frac{1}{\sigma^2}\left\{\mathbb{E}\left(Z_i^2\right) - \mathbb{E}\left(Z_i\boldsymbol{X}_i^T\right)\mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\mathbb{E}\left(\boldsymbol{X}_i Z_i\right)\right\} \\
&= \frac{1}{\sigma^2}\left\{\mathbb{E}\left(Z_i^2\right) - \mathbf{w}^T\mathbb{E}\left(\boldsymbol{X}_i Z_i\right)\right\}.
\end{aligned}
\tag{C.9}
$$

C.4 Derivation of σ^*_S in linear regressions and GLMs

Using the identity $\boldsymbol{\Sigma}^* = \mathbf{I}^*$ of Mykland (1999), we derive σ^*_S of Eq. (2.7) as

$$
\begin{aligned}
\sigma^*_S &= \mathbf{v}^{*T}\boldsymbol{\Sigma}^*\mathbf{v}^* \\
&= (1, -\mathbf{I}^*_{\theta\boldsymbol{\gamma}}{\mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}}}^{-1})\mathbf{I}^*(1, -\mathbf{I}^*_{\theta\boldsymbol{\gamma}}\mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}})^T \\
&= (1, -\mathbf{I}^*_{\theta\boldsymbol{\gamma}}{\mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}}}^{-1})\begin{pmatrix} I^*_{\theta\theta} & \mathbf{I}^*_{\theta\boldsymbol{\gamma}} \\ \mathbf{I}^*_{\boldsymbol{\gamma}\theta} & \mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\end{pmatrix}\begin{pmatrix} 1 \\ -{\mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}}}^{-1}\mathbf{I}^*_{\boldsymbol{\gamma}\theta}\end{pmatrix} \\
&= (I^*_{\theta\theta} - \mathbf{I}^*_{\theta\boldsymbol{\gamma}}{\mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}}}^{-1}\mathbf{I}^*_{\boldsymbol{\gamma}\theta}, \mathbf{0}^T)\begin{pmatrix} 1 \\ -{\mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}}}^{-1}\mathbf{I}^*_{\boldsymbol{\gamma}\theta}\end{pmatrix} \\
&= I^*_{\theta\theta} - \mathbf{I}^*_{\theta\boldsymbol{\gamma}}{\mathbf{I}^*_{\boldsymbol{\gamma}\boldsymbol{\gamma}}}^{-1}\mathbf{I}^*_{\boldsymbol{\gamma}\theta} := I^*_{\theta|\boldsymbol{\gamma}}.
\end{aligned}
\tag{C.10}
$$

C.5 Supplement to Eq. (2.21)

Eqs. (C.11) and (C.12) show the first and second order gradients of the loglikelihood function in Eq. (2.21).

$$
\begin{aligned}
\nabla_\theta \ell(\theta, \boldsymbol{\gamma}) &= \nabla_\theta[-\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\left(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i\right) - b\left(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i\right)\right\}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left\{Y_iZ_i - b'(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)Z_i\right\}. \\
\nabla_{\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma}) &= \nabla_{\boldsymbol{\gamma}}[-\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\left(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i\right) - b\left(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i\right)\right\}] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\boldsymbol{X}_i - b'(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\boldsymbol{X}_i\right\},
\end{aligned} \tag{C.11}
$$

$$
\begin{aligned}
\nabla^2_{\boldsymbol{\gamma}\theta} \ell(\theta, \boldsymbol{\gamma}) &= \nabla_\theta[-\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\boldsymbol{X}_i - b'(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\boldsymbol{X}_i\right\}] \\
&= \frac{1}{n}\sum_{i=1}^{n} b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)Z_i\boldsymbol{X}_i, \\
\nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(\theta, \boldsymbol{\gamma}) &= \nabla_{\boldsymbol{\gamma}}[-\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\boldsymbol{X}_i - b'(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\boldsymbol{X}_i\right\}] \\
&= \frac{1}{n}\sum_{i=1}^{n} b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\boldsymbol{X}_i\boldsymbol{X}_i^T,
\end{aligned} \tag{C.12}
$$

C.6 Derivations of (partial) information matrices $\mathbf{I}^*$ and $I^*_{\theta|\boldsymbol{\gamma}}$ in the GLM case

Eq. (C.13) shows the derivations of the true Fisher information and partial information matrices ($\mathbf{I}^*$ and $I^*_{\theta|\boldsymbol{\gamma}}$) in the GLM case.

$$
\begin{aligned}
\mathbf{I}^* &= \mathbb{E}\left(\nabla^2 \ell\left(\boldsymbol{\beta}^*\right)\right) \\
&= \mathbb{E}\left(\nabla\left[-\frac{1}{n}\sum_{i=1}^{n}\left(Y_i\boldsymbol{Q}_i - b'(\boldsymbol{\beta}^T\boldsymbol{Q}_i)\,\boldsymbol{Q}_i\right]\right)\right. \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} b''(\boldsymbol{\beta}^T\boldsymbol{Q}_i)\boldsymbol{Q}_i\boldsymbol{Q}_i^T\right) \\
&= \mathbb{E}\left(b''(\boldsymbol{\beta}^T\boldsymbol{Q}_i)\boldsymbol{Q}_i\boldsymbol{Q}_i^T\right), \text{ and} \\
I^*_{\theta|\boldsymbol{\gamma}} &= I^*_{\theta\theta} - \mathbf{I}^*_{\theta\boldsymbol{\gamma}}\mathbf{I}^{*-1}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{I}^*_{\boldsymbol{\gamma}\theta} \\
&= \mathbb{E}\left(\nabla^2_{\theta\theta}\ell\left(\boldsymbol{\beta}^*\right)\right) - \mathbb{E}\left(\nabla^2_{\theta\boldsymbol{\gamma}}\ell\left(\boldsymbol{\beta}^*\right)\right)\mathbb{E}\left(\nabla^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\ell\left(\boldsymbol{\beta}^*\right)\right)^{-1}\mathbb{E}\left(\nabla^2_{\boldsymbol{\gamma}\theta}\ell\left(\boldsymbol{\beta}^*\right)\right) \\
&= \mathbb{E}\left(b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)Z_i^2\right) - \{\mathbb{E}\left(Z_i\boldsymbol{X}_i^T b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\right) \\
&\quad \mathbb{E}\left(b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\mathbb{E}\left(b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\boldsymbol{X}_iZ_i\right)\} \\
&= \mathbb{E}\left(b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)Z_i^2\right) - \mathbf{w}^T\mathbb{E}\left(b''(\theta Z_i + \boldsymbol{\gamma}^T\boldsymbol{X}_i)\boldsymbol{X}_iZ_i\right).
\end{aligned} \tag{C.13}
$$

Appendix D Empirical FDR simulations in the GLM case

Tables D.1 and D.2 display the empirical FDRs of the logistic regression and the Poisson regression simulations respectively (on the next page).

Table D.1: Empirical FDRs (%) of the DScore test for the logistic regression case with a 5% nominal FDR.

			$\rho = 0.25$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.75$	
s	d	Adjustment	Dirac	Unif	Dirac	Unif	Dirac	Unif	Dirac	Unif
2	25	Raw	50.1	48.7	28.6	34.6	22.8	24.2	19.6	23.9
2	25	Bonferroni	17.0	16.7	9.1	13.5	8.8	8.8	9.2	10.0
2	25	Holm	18.1	17.8	10.0	15.0	9.3	9.5	9.6	10.6
2	25	BH	25.0	25.4	14.3	18.5	11.1	12.2	11.2	12.5
2	25	BY	13.2	13.7	6.9	11.0	6.4	7.6	6.7	7.2
2	50	Raw	63.1	60.6	42.0	41.7	25.5	26.7	22.5	23.9
2	50	Bonferroni	20.8	19.4	12.7	10.6	7.6	7.8	8.1	8.0
2	50	Holm	21.4	20.7	13.8	12.2	8.5	8.8	8.3	8.4
2	50	BH	28.3	28.7	18.4	17.8	10.6	10.5	9.7	10.0
2	50	BY	15.1	15.0	9.5	9.3	<u>5.4</u>	<u>6.1</u>	<u>5.9</u>	<u>6.1</u>
2	100	Raw	72.9	70.8	51.4	52.2	30.8	33.8	24.7	25.9
2	100	Bonferroni	25.1	19.8	12.8	16.2	6.4	9.0	<u>6.0</u>	6.4
2	100	Holm	25.8	21.1	13.8	18.1	7.2	10.0	6.3	6.7
2	100	BH	37.0	32.1	20.4	25.3	9.3	13.4	8.1	7.8
2	100	BY	15.5	14.5	8.3	13.0	<u>4.8</u>	6.5	<u>3.8</u>	<u>3.5</u>
3	25	Raw	46.8	43.1	29.7	27.6	18.2	16.2	16.7	14.6
3	25	Bonferroni	16.8	14.4	9.5	8.6	6.5	<u>6.2</u>	<u>6.2</u>	<u>4.7</u>
3	25	Holm	17.3	15.0	10.2	9.5	7.0	6.8	<u>6.2</u>	<u>5.1</u>
3	25	BH	21.5	20.4	13.5	13.4	8.9	8.2	7.7	6.5
3	25	BY	12.6	12.0	7.3	7.2	<u>5.3</u>	<u>5.2</u>	<u>5.4</u>	<u>4.5</u>
3	50	Raw	60.1	57.3	37.5	36.7	20.2	18.9	15.4	16.9
3	50	Bonferroni	19.5	14.9	11.1	9.0	<u>5.9</u>	<u>5.3</u>	<u>4.5</u>	<u>5.7</u>
3	50	Holm	19.9	15.7	11.6	9.9	6.4	<u>5.9</u>	<u>4.8</u>	6.3
3	50	BH	25.6	22.6	15.1	14.9	8.4	7.8	<u>5.2</u>	7.3
3	50	BY	12.6	11.8	8.2	6.4	<u>4.0</u>	<u>4.0</u>	<u>2.8</u>	<u>4.1</u>
3	100	Raw	70.6	65.9	47.1	48.2	21.9	23.1	16.3	16.5
3	100	Bonferroni	20.1	14.4	10.8	9.8	<u>4.0</u>	<u>4.9</u>	<u>4.0</u>	<u>4.5</u>
3	100	Holm	20.4	15.1	11.5	10.8	<u>4.6</u>	<u>5.4</u>	<u>4.1</u>	<u>4.7</u>
3	100	BH	27.5	24.9	16.3	16.9	6.4	8.3	<u>4.9</u>	<u>5.9</u>
3	100	BY	13.3	10.2	7.4	6.9	<u>2.9</u>	<u>3.5</u>	<u>3.2</u>	<u>3.3</u>

Notes: <u>Underlined</u> values indicate that we cannot reject that the FDR is lower than 5% using a right-tailed Z-test with a 5% significance level. We generate nonzero parameters to be equal to 1 (Dirac) or draw them uniformly in the interval [0,2] (Unif). Parameters: d=number of covariates, $s = \text{supp}(\boldsymbol{\gamma}^*)$, and ρ =Toeplitz parameter for the covariance matrix of the covariates. Adjustment methods: unadjusted p-values (Raw), Bonferroni, Holm, Benjamini-Hochberg (BH), and Benjamini-Yekutieli (BY).

Table D.2: Empirical FDRs (%) of the DScore test for the Poisson regression case with a 5% nominal FDR.

			$\rho = 0.25$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.75$	
s	d	Adjustment	Dirac	Unif	Dirac	Unif	Dirac	Unif	Dirac	Unif
2	25	Raw	43.1	40.0	30.8	29.8	18.0	19.2	13.0	15.7
2	25	Bonferroni	12.3	11.4	8.1	9.1	<u>5.1</u>	<u>6.2</u>	<u>4.1</u>	<u>4.9</u>
2	25	Holm	13.3	12.7	9.5	10.4	<u>6.2</u>	7.5	<u>4.5</u>	<u>5.3</u>
2	25	BH	21.3	19.3	15.5	14.7	9.5	9.3	6.6	6.8
2	25	BY	11.8	10.8	8.0	8.6	<u>4.7</u>	<u>5.3</u>	<u>4.1</u>	<u>3.5</u>
2	50	Raw	57.9	51.3	41.0	38.3	21.5	21.4	12.1	13.3
2	50	Bonferroni	18.2	15.8	12.2	11.8	<u>4.6</u>	<u>5.4</u>	<u>2.8</u>	<u>3.7</u>
2	50	Holm	19.5	17.2	12.9	12.7	<u>5.6</u>	6.4	<u>3.3</u>	<u>4.3</u>
2	50	BH	30.3	26.3	19.4	18.6	8.4	8.5	<u>4.8</u>	<u>5.3</u>
2	50	BY	17.0	15.3	11.3	10.9	<u>4.6</u>	<u>4.7</u>	<u>2.7</u>	<u>2.7</u>
2	100	Raw	64.6	57.0	45.8	41.9	20.0	20.8	9.3	12.9
2	100	Bonferroni	16.2	15.7	9.9	10.5	<u>3.2</u>	<u>4.4</u>	<u>1.7</u>	<u>3.1</u>
2	100	Holm	17.1	16.6	10.9	11.3	<u>4.0</u>	<u>5.4</u>	<u>2.2</u>	<u>3.6</u>
2	100	BH	29.2	28.3	17.8	17.8	7.0	7.1	<u>3.3</u>	<u>5.0</u>
2	100	BY	13.6	13.8	8.5	8.5	<u>2.8</u>	<u>3.6</u>	<u>1.6</u>	<u>2.3</u>
3	25	Raw	40.2	38.2	27.9	26.0	14.6	14.7	9.0	10.7
3	25	Bonferroni	15.3	13.4	10.1	9.4	<u>5.0</u>	<u>5.1</u>	<u>3.9</u>	<u>3.7</u>
3	25	Holm	16.6	14.5	11.2	10.2	<u>6.2</u>	<u>6.0</u>	<u>4.3</u>	<u>4.1</u>
3	25	BH	25.3	22.9	17.0	15.2	8.9	8.3	<u>6.0</u>	<u>5.7</u>
3	25	BY	16.4	14.0	10.8	9.5	<u>5.7</u>	<u>5.2</u>	<u>3.9</u>	<u>3.4</u>
3	50	Raw	54.7	51.9	38.5	36.2	16.7	16.9	8.4	9.5
3	50	Bonferroni	20.6	20.8	12.9	12.3	<u>4.7</u>	<u>4.2</u>	<u>2.2</u>	<u>2.7</u>
3	50	Holm	21.3	21.6	14.0	13.3	<u>5.4</u>	<u>4.8</u>	<u>2.5</u>	<u>3.2</u>
3	50	BH	33.0	31.6	20.8	19.9	8.6	8.0	<u>4.1</u>	<u>4.2</u>
3	50	BY	21.3	20.7	13.4	12.0	<u>4.9</u>	<u>4.0</u>	<u>1.9</u>	<u>2.4</u>
3	100	Raw	58.6	55.1	38.8	37.3	12.8	15.2	<u>4.7</u>	7.4
3	100	Bonferroni	19.2	21.2	9.9	13.4	<u>2.9</u>	<u>4.4</u>	<u>1.4</u>	<u>2.2</u>
3	100	Holm	20.0	22.2	10.7	14.5	<u>3.5</u>	<u>5.0</u>	<u>1.9</u>	<u>2.7</u>
3	100	BH	32.8	33.6	19.6	20.4	6.3	7.6	<u>3.0</u>	<u>3.4</u>
3	100	BY	19.4	21.4	10.3	13.7	<u>3.3</u>	<u>4.3</u>	<u>1.6</u>	<u>2.3</u>

Notes: <u>Underlined</u> values indicate that we cannot reject that the FDR is lower than 5% using a right-tailed Z-test with a 5% significance level. We generate nonzero parameters to be equal to 1 (Dirac) or draw them uniformly in the interval [0,2] (Unif). Parameters: d=number of covariates, s = supp($\boldsymbol{\gamma}^*$), and ρ =Toeplitz parameter for the covariance matrix of the covariates. Adjustment methods: unadjusted p-values (Raw), Bonferroni, Holm, Benjamini-Hochberg (BH), and Benjamini-Yekutieli (BY).

Appendix E Additional HPI application results

Table E.1 shows the MSAs with raw p-values that are significant on a 5%-level. The superscripts in this table indicate whether the MSA still has a significant p-value after FDR control.

Table E.1: Additional house price index results: MSAs with significant raw p-values in the four different samples.

16-year sample	
Alexandria, LA	Rochester, NY
Charlotte-Concord-Gastonia, NC-SC	Salinas, CA^{cd}
Chicago-Naperville-Elgin, IL-IN-WI	San Francisco-Oakland-Hayward, CA
Lewiston, ID-WA	Seattle-Tacoma-Bellevue, WA^{c}
Napa, CA^{abcd}	Texarkana, TX-AR^{abcd}
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	The Villages, FL^{abcd}
Prescott, AZ^{abcd}	Washington-Arlington-Alexandria, DC-VA-MD-WV^{abcd}
26-year sample	
Alexandria, LA^{abcd}	Napa, CA^{c}
Atlanta-Sandy Springs-Roswell, GA^{c}	Orlando-Kissimmee-Sanford, FL^{c}
Boston-Cambridge-Newton, MA-NH^{abcd}	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD^{abcd}
Charlotte-Concord-Gastonia, NC-SC^{c}	Prescott, AZ
Chicago-Naperville-Elgin, IL-IN-WI^{cd}	Salisbury, MD-DE^{cd}
Cincinnati, OH-KY-IN^{c}	Santa Cruz-Watsonville, CA^{abcd}
Decatur, AL^{c}	Santa Maria-Santa Barbara, CA^{abcd}
Indianapolis-Carmel-Anderson, IN	Seattle-Tacoma-Bellevue, WA^{abcd}
Lewiston, ID-WA^{abcd}	Springfield, MA^{c}
Los Angeles-Long Beach-Anaheim, CA^{abcd}	Syracuse, NY^{c}
Madera, CA^{c}	Texarkana, TX-AR^{abcd}
Madison, WI^{c}	The Villages, FL^{abcd}
Minneapolis-St. Paul-Bloomington, MN-WI	Vineland-Bridgeton, NJ^{c}
Missoula, MT^{c}	Waco, TX
36-year sample	
Alexandria, LA^{cd}	Orlando-Kissimmee-Sanford, FL^{c}
Asheville, NC	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD^{abcd}
Boston-Cambridge-Newton, MA-NH^{abcd}	Portland-Vancouver-Hillsboro, OR-WA^{abcd}
Chattanooga, TN-GA^{cd}	Riverside-San Bernardino-Ontario, CA^{abcd}
Chicago-Naperville-Elgin, IL-IN-WI^{abcd}	Salisbury, MD-DE^{c}
Colorado Springs, CO^{cd}	San Francisco-Oakland-Hayward, CA^{abcd}
Des Moines-West Des Moines, IA^{c}	Santa Cruz-Watsonville, CA
Greenville, NC^{c}	Seattle-Tacoma-Bellevue, WA^{abcd}

Las Vegas-Henderson-Paradise, NV[cd]	South Bend-Mishawaka, IN-MI
Lewiston, ID-WA[c]	Syracuse, NY[c]
Los Angeles-Long Beach-Anaheim, CA[abcd]	Texarkana, TX-AR[abcd]
Madera, CA[abcd]	Tucson, AZ[abcd]
Miami-Fort Lauderdale-West Palm Beach, FL	Vallejo-Fairfield, CA[c]
New York-Newark-Jersey City, NY-NJ-PA[cd]	Vineland-Bridgeton, NJ[c]
North Port-Sarasota-Bradenton, FL[c]	Wichita Falls, TX[c]
46-year sample	
Alexandria, LA	Naples-Immokalee-Marco Island, FL
Allentown-Bethlehem-Easton, PA-NJ	Oxnard-Thousand Oaks-Ventura, CA[c]
Charlottesville, VA	Portland-Vancouver-Hillsboro, OR-WA
Chico, CA[abcd]	Redding, CA[c]
Denver-Aurora-Lakewood, CO[abcd]	Riverside-San Bernardino-Ontario, CA
Detroit-Warren-Dearborn, MI	St. Joseph, MO-KS
Fayetteville-Springdale-Rogers, AR-MO	Salinas, CA[c]
Grand Rapids-Wyoming, MI	Seattle-Tacoma-Bellevue, WA
Gulfport-Biloxi-Pascagoula, MS	Sebring, FL
Lancaster, PA	Texarkana, TX-AR
Lawton, OK	Vallejo-Fairfield, CA[abcd]
Los Angeles-Long Beach-Anaheim, CA	Vineland-Bridgeton, NJ
Minneapolis-St. Paul-Bloomington, MN-WI	Youngstown-Warren-Boardman, OH-PA

Note: The superscripts indicate 5%-significance after applying correction methods of [a]= Bonferroni, [b]=Holm, [c]=Benjamini-Hochberg, [d]=Benjamini-Yekutieli.