A Comparison of Random Projection-Based Test Statistics in High Dimensions

Connor T. McNeill* Miles Woollacott*

December 10, 2024

1 Introduction

1.1 Motivation

In this report, we compare the random projection-based approach for three classical test statistics to a novel test statistic proposed by Changyu Liu and Huang (2024) in the high-dimensional setting. For Single Index Models (SIMs), recent literature has focused on random projection-based approaches for hypothesis testing in high dimensions. Using random projections, we transform $\boldsymbol{X}$ into $\boldsymbol{U}_k \in \mathbb{R}^{n \times k}$, where $k < n$. For the classical test statistics, we then compute them ordinarily, but using $\boldsymbol{U}_k$ in place of $\boldsymbol{X}$. Changyu Liu and Huang, 2024 performed a simulation in their introductory paper for their novel test statistic in high dimensions, and the other classical statistics, that is very similar to what we are investigating. In their simulation, they measure type I error under our null hypothesis which is when $\boldsymbol{\beta} = 0$, and empirical power under the alternative hypothesis $\boldsymbol{\beta} \neq 0$ for all four statistics. However, their simulation study only considers the logistic model for their data. Alkhalaf (2017) concluded that the Wald statistic may vary in power

*Department of Statistics, North Carolina State University

and Type I error based on nuisance factors. This means that the results presented by Changyu Liu and Huang (2024) may not reflect the performance of one or more test statistics in other settings. As such, we will study the performance a simulation study to extend the paper's original study from its supplemental material by looking at the Poisson model, and then compare our results to that found in the paper.

1.2 Model

Lanteri et al. (2022) defines a Single-Index Model (SIM) as "a statistical model for intrinsic regression where the responses are assumed to depend on a single yet unknown linear combination of the predictors, allowing to express the regression function as $\mathrm{E}(Y|X) = f(\boldsymbol{X}^T\boldsymbol{\beta}, \epsilon)$ for some unknown index vector ϵ and link function f." There exists similarities from SIMs to generalized linear models (GLMs), since both use link functions. However, there is a key difference: as mentioned by Lanteri et al. (2022), in SIMs the responses are assumed to depend on a single linear combination of the predictors, while GLMs allow for more complex relationships between multiple independent variables and the predictor.

Changyu Liu and Huang (2024) use the Logistic model (with a logit link), and for this report, we will consider the Poisson model, which has the form

$$Y_i|\boldsymbol{X_i} \overset{\text{ind}}{\sim} \text{Poisson}(\mu_i) \text{ where } \log \mu_i = \exp(\boldsymbol{X_i}^T\boldsymbol{\beta}).$$

As shown in the above model, we will be using the log link, which is also the canonical link.

2 Methodology

2.1 Random Projection Approach

Since we are working with a high-dimensional model, the first thing we need to handle is the high dimensions. The number of covariates p is larger than the sample size n, which means that we can run into the curse of dimensionality. As our number of covariates increases, we run into two key issues. First, our covariate matrix $\boldsymbol{X}$ will not be of full column rank. This means that $\boldsymbol{X}^T\boldsymbol{X}$ will not be invertible, meaning that we cannot obtain least-squares estimates for all of our model coefficients or compute our typical F-test statistic. Second, in the case of relatively high dimensions when $n > p$ and $p \approx n$, our model will overfit on the training data and likely perform poorly when applied to a testing dataset (James et al., 2021).

One approach to solving this problem is by performing dimension reduction on the covariate matrix. A common method of doing so is principal component analysis (PCA), in which we utilize the singular value decomposition (SVD) of the covariate matrix to create a matrix of principal components which are composed of the eigenvectors of $\boldsymbol{X}$. This process allows us to select a certain number of components and then use this lower k-dimensional matrix to fit models and compute test statistics. A similar approach called *random projections* allow us to perform dimension reduction as well. It is also computationally faster than PCA (Yang et al., 2021) which is quite important when considering that high-dimensional data may be quite large.

The following steps go through how we implemented uniform orthogonal random projections (Yang et al., 2021) to perform dimension reduction in order to be able to compute both the proposed test statistic and the classical test statistics as proposed by Changyu Liu and Huang (2024).

1. For a set projection ratio ρ, we must obtain $k = \lceil \rho n \rceil$ which is the projection dimension. This will be the number of columns of our random projected data.

2. Obtain the projection matrix $\boldsymbol{I} - \boldsymbol{P}_1$, where $\boldsymbol{P}_1 = \frac{1}{n}\mathbf{1}\mathbf{1}^T$.
3. For $d = 1, \ldots, D$, create a $p \times k$ matrix with random entries from a standard normal distribution. This gives us D random matrices that are all of the same dimensions.
4. Take the mean of each entry across the D matrices to obtain our $n \times p$ projection matrix. The random projection matrix $\boldsymbol{P}_k$ is equal to $p^{-1/2}$ multiplied to the obtained matrix.
5. Obtain our dimension-reduced matrix $\boldsymbol{U}_k = (\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{X}\boldsymbol{P}_k$, which is a $n \times k$ matrix.

Now that we have created a matrix $\boldsymbol{U}_k$ where the number of predictors $k < n$, we can utilize this matrix to compute our test statistics in place of where $\boldsymbol{X}$ would be typically. Theorem 4.1 of Changyu Liu and Huang, 2024 shows that utilizing this approach, we obtain the result that $\boldsymbol{U}_k^T\boldsymbol{U}_k$ is full rank with probability 1.

2.2 Proposed Test Statistic

Changyu Liu and Huang (2024) propose a test statistic that improves on the usual F-test statistic computed in regards to full-rank linear hypotheses. We utilize the random-projection approach described above to obtain our random-projected covariate matrix $\boldsymbol{U}_k$ and then utilize this matrix to form our test statistic. For the remainder of this paper, we will refer to their test statistic as the *random projection test*. The statistic takes the form

$$T_{RP} = \frac{\boldsymbol{y}^T\boldsymbol{H}_k\boldsymbol{y}/k}{\boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{P}_1 - \boldsymbol{H}_k)\boldsymbol{y}/(n-k-1)}$$

which is quite similar to the classical F-test with the exception of the denominator. Note that $\boldsymbol{H}_k = \boldsymbol{U}_k(\boldsymbol{U}_k^T\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^T$ is the hat matrix based on the random-projected covariate matrix.

Also, we can derive the asymptotic distribution of the test statistic. Assume that the number of predictors $p \gg n$ and that there is a constant projection ratio ρ such that $\frac{k}{n} \to p$ as $n \to \infty$. Under

that assumption and two other assumptions proved in Changyu Liu and Huang, 2024, including that we get the result that as $n \to \infty$, we have

$$\frac{T_{RP} - 1}{\sqrt{2/n\rho(1-\rho)}} \xrightarrow{d} \mathcal{N}(0,1).$$

This means that we would reject the null hypothesis when the quantity above is larger than z_α, the upper α-quantile of the standard normal distribution.

2.3 Classical Tests

The Wald, score and likelihood ratio tests are all likelihood-based tests that are asymptotically equivalent as they all converge to a chi-square distribution with the same degrees of freedom (Boos & Stefanski, 2013). However, this is true only under typical settings. Changyu Liu and Huang (2024) performed a simulation study that demonstrated the inadequacy of the asymptotic equivalency in relatively high dimensions (specifically, they chose $n = 1000$ varied the number of predictors p up to 200). However, when we are in a high-dimensional setting, the situation is even worse as our test statistics are not well-defined. As such, we will use the same random projection approach to perform dimension reduction as was done with the proposed test statistic, and then utilize the resulting randomly projected data of dimension k to construct our three tests.

In order to derive the classical tests, we first need to derive the likelihood function $\ell(\boldsymbol{\beta})$, score function $S(\boldsymbol{\beta})$, and the Fisher information matrix $I(\boldsymbol{\beta})$. All of these functions are derived in Appendix A. In general, for the hypothesis test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ vs. $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, our parameter is $\boldsymbol{\beta}$ and its maximum likelihood estimator is $\hat{\boldsymbol{\beta}}_n$. The Wald statistic is defined as

$$T_W = (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T \{nI(\hat{\boldsymbol{\beta}}_n)\}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0),$$

the score statistic is defined as

$$T_S = S(\boldsymbol{\beta}_0)\{nI(\boldsymbol{\beta}_0)\}^{-1}S(\boldsymbol{\beta}_0)^T,$$

and the likelihood ratio test is defined as

$$T_{LR} = 2\{\ell(\hat{\boldsymbol{\beta}}_n) - \ell(\boldsymbol{\beta}_0)\}.$$

The test statistics' explicit forms for the Poisson model are found in Appendix A.

While Changyu Liu and Huang (2024) mentions that the null distribution of our random projection-based classical test statistics will not really be a χ^2_k, for the purpose of the simulation study, we will utilize a $\chi^2_{\alpha,k}$ critical value to determine our rejection regions.

3 Simulation Study

3.1 Data Generation

One advantage of the test statistic is that it is able to perform decently when we have a great deal of sparsity (Changyu Liu & Huang, 2024). We will take this into consideration as we generate the data to be utilized in the simulation study. For our study, we utilize a high-dimensional setting where $n = 400$ and $p = 1000$. The data generating process is as follows:

1. Generate the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{ODO}^T$ based off of the specified sparsity in the setting of the simulation.
    - $\boldsymbol{D}$ is a diagonal matrix with the first $s = \lceil n^{0.8} \rceil$ entries equal to $w_i = 1$ and the remaining $n - s$ entries being equal to $w_i = (i - s)^4$.
    - $\boldsymbol{O}$ is an orthogonal matrix. First, we generated a blockwise diagonal matrix based off

of the specified sparsity that controls the number of blocks. Each submatrix is filled with random entries from a standard normal distribution. We add make the matrix symmetric by adding it to its transpose, and then use spectral decomposition to obtain the orthogonal matrix $\boldsymbol{O}$.

2. Generate the covariate matrix $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{Z}$ is a random matrix with entries generated from a standard normal distribution.

3. Generate the true values of $\boldsymbol{\beta} \in \mathbb{R}^p$ in one of two ways based off of the value of $\boldsymbol{\delta}$. We compute this as $\boldsymbol{\beta} = b\boldsymbol{\delta}/\sqrt{\boldsymbol{\delta}^T\boldsymbol{\Sigma}\boldsymbol{\delta}}$ for specified value b and vector $\boldsymbol{\delta}$.

   - $\boldsymbol{\delta}_1$ is a sparse vector with 10 non-zero values.
   - $\boldsymbol{\delta}_2$ is randomly selected from the span of the first 100 columns from the orthogonal $\boldsymbol{O}$ matrix. Since $\boldsymbol{O}$ is obtained from the `eigen` function in R, the columns are returned in a sorted order by the eigenvalues. We can interpret this result similar to the resulting matrix in PCA, where the first columns explain more of the data. This is why we take solely the first 100 columns and randomly select from the span of those columns.

4. Lastly, we randomly generate Y from our single-index Poisson model using $\boldsymbol{X}$ and $\boldsymbol{\beta}$.

3.2 Setup

This simulation study is relatively simple in terms of setup on purpose - due to working with high-dimensional data, the run time of a single setting of the simulation is quite high. In fact, if we had chosen to replicate the main simulation study of the paper, without a high performance computing cluster, it would likely take days to run each setting. As such, we chose to extend one of the simulation studies in the supplemental material because of the shorter runtime and the relevancy to content covered throughout the course.

We will use a simulation-based approach to compare the Type I errors and empirical power for all the four following test statistics:

- T_{RP}: the random projection test proposed by the paper designed to work in ultrahigh dimensions.
- T_W: the Wald statistic derived earlier utilizing the random-projected data.
- T_S: the score statistic derived earlier utilizing the random-projected data.
- T_{LR}: the likelihood ratio test derived earlier utilizing the random-projection data.

The workflow of the simulation study is the following. For each of the 1000 iterations, we first generate our data as described above. Then, we compute the 4 test statistics for each iteration. The Wald and likelihood ratio tests rely on the estimate of $\hat{\beta}_n$, so we will utilize maximum likelihood estimation to estimate those coefficients from our data. Lastly, we will utilize the resulting test statistics to compute Type I error rates and empirical powers for the global test $H_0 : \boldsymbol{\beta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$. Note that in the table below, since we have 1000 simulations, 3 significant digits is appropriate. In addition, due to the context of interpretation and the desire to compare our results to that of the paper, our simulation results table formatting will match that which can be found in the paper.

3.3 Results

Table 1 displays the simulation results for the Poisson model, and table 2 consists of the results that Changyu Liu and Huang (2024) obtained in the supplemental section of their paper. First, we discuss Table 1 on its own. The random projection test had the closest Type I error rate to $\alpha = 0.05$, but it struggled in power for when $b^2 = 0.1$. This means it needed the nonzero $\boldsymbol{\beta}$'s to be larger in magnitude in order to be more powerful. When this holds, the likelihood ratio test T_{LR} becomes the most powerful test overall, and performs similarly to the proposed random projection

$\boldsymbol{\beta}$ Setting	b^2	T_{RP}	T_{LR}	T_W	T_S	
$\mathbf{0}$	0	0.062	0.011	0.205	0.068	Type I Error
$\boldsymbol{\delta}_1$	0.1	0.534	0.533	0.928	0.802	Power
	0.2	0.940	0.974	0.999	0.995	Power
$\boldsymbol{\delta}_2$	0.1	0.525	0.516	0.910	0.806	Power
	0.2	0.929	0.972	1.000	0.997	Power

Table 1: Empirical Type I Error rates and Power for the Poisson model found in our simulation study.

$\boldsymbol{\beta}$ Setting	b^2	T_{RP}	T_{LR}	T_W	T_S	
$\mathbf{0}$	0	0.059	0.830	0.000	0.021	Type I Error
$\boldsymbol{\delta}_1$	0.4	0.469	0.993	0.000	0.227	Power
	0.8	0.831	0.973	0.026	0.581	Power
$\boldsymbol{\delta}_2$	0.4	0.480	0.987	0.003	0.214	Power
	0.8	0.830	0.981	0.019	0.613	Power

Table 2: Empirical Type I Error rates and Power for the Logistic model found in the supplemental material of Changyu Liu and Huang (2024).

test otherwise. The Wald test T_W cannot be considered to be a valid test here due to the high Type I error rate. The score test T_S performed the best overall, especially when $b^2 = 0.1$.

Now, we compare this previous discussion to what we can infer from Table 2. The biggest take-away is that the three classical test statistics performed much more differently when we switched from a Poisson to a Logistic model. T_W was the biggest change, going from having a high type I error rate to having a type I error rate of 0. As such, it is unusable for the Logistic model due to this along with the extremely low power. T_{LR} is invalid here for the Logistic model as its type I error rate is 0.830, far from the specified $\alpha = 0.05$. T_S is no longer the best statistic, as its power is lower than that of the proposed random projection test for each setting.

What about the proposed T_{RP} as we switched models? While the power decreased when moving

to the Logistic model, this decrease was far better than what happened with the score test T_S. The Type I error rate also stayed about the same when we switched the models, which contrasted with T_{LR} and T_W. So, while T_{RP} may not be the best model on a case-by-case basis, its overall consistency between the two different models leads to it being the most appealing choice out of the statistics we compared.

4 Conclusion

Throughout this report, we demonstrated that a random projection-based approach will allow us to conduct hypotheses tests in high-dimensional settings, even for test statistics that normally require $n > p$. However, the classical test statistics under the random projection setting can be erratic or misleading when comparing across different models, and so we need a more reliable statistic. We demonstrated in the simulation study that T_{RP}, as proposed by Changyu Liu and Huang, 2024, is generally effective and consistent across different types of models.

Future work may involve testing different types of models, in order to further demonstrate the stability of T_{RP}. We only compared the Poisson and Logistic models, which have both intuitive and simple link functions. However, there are popular two-parameter link functions, such as Negative Binomial or Gamma, that may be of interest. This will be more challenging to construct, since we would have to handle nuisance parameters. In addition, we only considered $n = 400$ and $p = 1000$ in this report, but comparing the simulation results for differing ratios $\frac{p}{n}$ may be of interest. Lastly, Changyu Liu and Huang, 2024 introduced other test statistics designed for high-dimensional settings in their article, most notably T_{GC}, proposed by Guo and Chen, 2016. The table we used to compare the Poisson model did not include T_{GC}, but seeing how T_{RP} compares to T_{GC} would be another question of interest.

5 References

Alkhalaf, A. A. (2017). *The impact of predictor variable(s) with skewed cell probabilities on the wald test in binary logistic regression* [Doctoral dissertation, University of British Columbia]. https://doi.org/10.14288/1.0343609

Boos, D. D., & Stefanski, L. A. (2013). *Essential Statistical Inference : Theory and Methods*. Springer New York. https://doi.org/10.1007/978-1-4614-4818-1

Changyu Liu, X. Z., & Huang, J. (2024). A random projection approach to hypothesis tests in high-dimensional single-index models. *Journal of the American Statistical Association*, *119*(546), 1008–1018. https://doi.org/10.1080/01621459.2022.2156350

Guo, B., & Chen, S. X. (2016). Tests for high dimensional generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *78*(5), 1079–1102. https://doi.org/10.1111/rssb.12152

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer US. https://doi.org/10.1007/978-1-0716-1418-1

Lanteri, A., Maggioni, M., & Vigogna, S. (2022). Conditional regression for single-index models. *Bernoulli*, *28*(4), 3051–3078. https://doi.org/10.3150/22-BEJ1482

Yang, F., Liu, S., Dobriban, E., & Woodruff, D. P. (2021). How to Reduce Dimension With PCA and Random Projections? [Conference Name: IEEE Transactions on Information Theory]. *IEEE Transactions on Information Theory*, *67*(12), 8154–8189. https://doi.org/10.1109/TIT.2021.3112821

A Derivation of Classical Tests

First we derive the log-likelihood function for our Poisson model as

$$\begin{aligned}\ell(\boldsymbol{\beta}; \boldsymbol{Y}, \boldsymbol{X}) &= \sum_{i=1}^{n} f(Y_i.\boldsymbol{X_i}; \boldsymbol{\beta}) \\ &= \sum_{i=1}^{n} \left\{ -\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}) + Y_i(\boldsymbol{X}_i^T \boldsymbol{\beta}) - \log Y_i! \right\}.\end{aligned}$$

Now that we have the log-likelihood function, we next need to find the score function.

$$\begin{aligned}S(\boldsymbol{\beta}; \boldsymbol{Y}, \boldsymbol{X}) &= \frac{\partial}{\partial \boldsymbol{\beta}^T} \ell(\boldsymbol{\beta}; \boldsymbol{Y}, \boldsymbol{X}) \\ &= \frac{\partial}{\partial \boldsymbol{\beta}^T} \sum_{i=1}^{n} \left\{ -\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}) + Y_i(\boldsymbol{X}_i^T \boldsymbol{\beta}) - \log Y_i! \right\} \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ -\exp(\boldsymbol{X}_i^T \boldsymbol{\beta}) + Y_i(\boldsymbol{X}_i^T \boldsymbol{\beta}) - \log Y_i! \right\} \\ &= \sum_{i=1}^{n} [Y_i - \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})] \boldsymbol{X}_i.\end{aligned}$$

Lastly, we can derive the fisher information matrix as

$$\begin{aligned}
I(\boldsymbol{\beta}) &= \frac{1}{n}\mathrm{E}\left[-\frac{\partial}{\partial\boldsymbol{\beta}}S(\boldsymbol{\beta};\boldsymbol{Y},\boldsymbol{X})\right] \\
&= \frac{1}{n}\mathrm{E}\left[-\frac{\partial}{\partial\boldsymbol{\beta}}\sum_{i=1}^{n}[Y_i-\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})]\boldsymbol{X}_i\right] \\
&= \frac{1}{n}\mathrm{E}\left[-\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\beta}}[Y_i-\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})]\boldsymbol{X}_i\right] \\
&= \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n}\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\boldsymbol{X}_i\boldsymbol{X}_i^T\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\boldsymbol{X}_i\boldsymbol{X}_i^T\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\boldsymbol{X}_i\boldsymbol{X}_i^T,
\end{aligned}$$

which can be rewritten as $I(\boldsymbol{\beta}) = \boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}/n$ where $\boldsymbol{V} = \mathrm{diag}\{\exp(\boldsymbol{X}_1^T\boldsymbol{\beta}),\ldots,\exp(\boldsymbol{X}_n^T\boldsymbol{\beta})\}$ (Boos & Stefanski, 2013, pp. 74–75).

As defined in the paper, for the hypothesis test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ vs. $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, our parameter is $\boldsymbol{\beta}$ and its maximum likelihood estimator is $\hat{\boldsymbol{\beta}}_n$. The Wald statistic can then be explicitly defined as

$$\begin{aligned}
T_W &= (\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)^T\{nI(\hat{\boldsymbol{\beta}}_n)\}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0) \\
&= (\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)^T\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0).
\end{aligned}$$

The score statistic can be explicitly defined as

$$\begin{aligned} T_S &= S(\boldsymbol{\beta}_0)\{nI(\boldsymbol{\beta}_0)\}^{-1}S(\boldsymbol{\beta}_0)^T \\ &= \left\{\sum_{i=1}^{n}[Y_i - \exp(\boldsymbol{X}_i^T\boldsymbol{\beta})]\boldsymbol{X}_i\right\}\left\{\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}\right\}^{-1}\left\{\sum_{i=1}^{n}[Y_i - \exp(\boldsymbol{X}_i^T\boldsymbol{\beta})]\boldsymbol{X}_i\right\}^T. \end{aligned}$$

The likelihood ratio test is explicitly defined as

$$\begin{aligned} T_{LR} &= 2\{\ell(\hat{\boldsymbol{\beta}}_n) - \ell(\boldsymbol{\beta}_0)\} \\ &= 2\sum_{i=1}^{n}\left\{-\exp(\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_n) + Y_i(\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_n) - \log Y_i!\right\} - \left\{-\exp(\boldsymbol{X}_i^T\boldsymbol{\beta}_0) + Y_i(\boldsymbol{X}_i^T\boldsymbol{\beta}_0) - \log Y_i!\right\} \\ &= 2\sum_{i=1}^{n}\left\{Y_i\boldsymbol{X}_i^T(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \exp(\boldsymbol{X}_i^T\boldsymbol{\beta}_0) - \exp(\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_n)\right\}. \end{aligned}$$