The Effects of Errors in a Randomized Design Setting of Linear Regression

Terry Lay, Ayusha Upreti [1]

Abstract

This work will explore the nature of linear regression in a random design setting largely inspired by the theorems established by Daniel Hsu, Sham M. Kakade, and Tong Zhang in their paper *Random Design Analysis of Ridge Regression*. Under mild assumptions in the random design setting, they established explicit bounds to prediction error for regression estimators which revealed a close connection between the random and fixed design settings.

1. Problem Statement

The primary goal of many applications in data analytics is the ability to generate insights by transforming raw data into models. In regression, independent pairs of covariates and responses sampled from a population are hypothesized to have a linear relationship with random error terms.

In theory, treating covariates as deterministic yields convenient results because the covariance structure does not need to be estimated, since only responses are treated as random. In contrast, random design analysis is concerned with the predictive performance of a model on unseen data where the covariance structure is completely unknown.

A disadvantage of the fixed design setting is the inability to address out-of-sample prediction, whereas in a completely randomized experiment, causal statements and inferences about the general population are statistically valid.

Though there are multiple principal outcomes concluded in the original work, the primary focus here will be on the result that errors have minimal effect on the estimated covariance structure in a random design setting and when sample sizes are sufficiently large, the error effect is essentially second-order.

To provide deeper understanding of the core intuitions, the investigation of this conclusion will illustrate examples and implement empirical studies to demonstrate why various assumptions and conditions are necessary. Specifically, it hopes to address:

- How errors have an effect on the estimated covariance structure

- How the estimated covariance structure is impacted by errors if the linear assumption of the regression model is false

- How noise in a response influences the effect of errors

2. Preliminary Details

This section will outline some basic notation and review statistical concepts necessary to derive the main result.

2.1. Basic Notation

All vectors defined in this work belong to a finite dimensional Euclidean inner product space with inner product $\langle \cdot, \cdot \rangle$ defined as the dot product.

Definition 2.1. The spectral norm of a matrix A is:

$$||A|| = \sup_{\mathbf{v} \neq 0} \frac{||A\mathbf{v}||}{||\mathbf{v}||}$$

Definition 2.2. The Frobenius norm of a matrix A is:

$$||A||_F = \sqrt{tr(A^T A)}$$

Definition 2.3. The norm of a vector $\mathbf{x}$ with respect to a positive definite matrix A is:

$$||\mathbf{x}||_A = \sqrt{\mathbf{x}^T A \mathbf{x}}$$

Definition 2.4. Define the empirical expectation of a particular function as

$$\hat{\mathbb{E}}[f] = \frac{1}{n} \sum^n f(\mathbf{x}_i, y_i)$$

2.2. Linear Regression

Let the covariate $\mathbf{x}$ be a random vector, and the response y be a random variable, both with finite variances, i.e.,

$$\mathbb{E}[||\mathbf{x}||^2] < \infty, \qquad \mathbb{E}[y^2] < \infty$$

Definition 2.5. The second moment matrix of $\mathbf{x}$ is

$$\Sigma = \mathbb{E}[\mathbf{x}\mathbf{x}^T]$$

where $\mathbf{x}\mathbf{x}^T$ is the outer product of $\mathbf{x}$ with itself.

Proposition 2.6. *The eigenvectors of Σ are $\{\mathbf{v_j}\}$ such that they form an orthonormal basis.*

Proposition 2.7. *The eigenvalues of Σ are*

$$\lambda_j = \langle \boldsymbol{v}_j, \Sigma \boldsymbol{v}_j \rangle, \quad \lambda_j > 0$$

Note:

$$\begin{aligned}
\mathbb{E}[\langle \boldsymbol{v}_j, \boldsymbol{x} \rangle^2] &= \mathbb{E}[\langle \boldsymbol{v}_j, \boldsymbol{x} \rangle \langle \boldsymbol{v}_j, \boldsymbol{x} \rangle], \\
&= \mathbb{E}[\langle \boldsymbol{v}_j, \langle \boldsymbol{v}_j, \boldsymbol{x} \rangle \boldsymbol{x} \rangle], \\
&= \mathbb{E}[\langle \boldsymbol{v}_j, (\boldsymbol{x}\boldsymbol{x}^T) v_j \rangle], \\
&= \langle \boldsymbol{v}_j, \Sigma \boldsymbol{v}_j \rangle
\end{aligned}$$

Definition 2.8. Let β be the value that minimizes the mean squared error loss function.

$$\begin{aligned}
l(\beta) &= \mathbb{E}[(\langle \beta, \mathbf{v}_j \rangle - y)^2] \\
&= \min_w \mathbb{E}[(\langle w, \mathbf{v}_j \rangle - y)^2]
\end{aligned}$$

Definition 2.9. The excess MSE of an estimator w over the minimum estimator is given by:

$$\begin{aligned}
l(w) - l(\beta) &= ||w - \beta||_\Sigma^2 \\
&= \sqrt{\langle (w - \beta)\Sigma(w - \beta) \rangle}
\end{aligned}$$

2.3. Ridge and Ordinary Least Squares Estimators

Let $(\mathbf{x}_1, y_1)...(\mathbf{x}_n, y_n)$ be i.i.d samples from the distribution of the random variables $(\mathbf{x}, y)$

Definition 2.10. The estimate of the covariance matrix Σ is

$$\begin{aligned}
\hat{\Sigma} &= \hat{\mathbb{E}}[\mathbf{x}\mathbf{x}^T] \\
&= \frac{1}{n} \sum^n \mathbf{x}_i \mathbf{x}_i^T
\end{aligned}$$

where $\hat{\mathbb{E}}$ is the empirical expectation from Definition 2.4.

Definition 2.11. The ridge estimator $\hat{\beta}_\lambda$ with $\lambda \geq 0$ is the minimizer of the empirical MSE

$$\hat{\beta}_\lambda = \arg\min_w \hat{\mathbb{E}}[(\langle w, \mathbf{x} \rangle - y)^2] + \lambda ||w||^2$$

Remark 2.12. There is a special case of Definition 2.11 which occurs when $\lambda = 0$. This transforms the minimization problem to the regular OLS minimization problem: $\arg\min_w \hat{\mathbb{E}}[(\langle w, \mathbf{x} \rangle - y)^2]$.

3. Decomposition of the Data Model

The main theorem is concluded through a simple decomposition of the response model in terms of subgaussian noise and approximation error.

Definition 3.1. The response noise is

$$\begin{aligned}
noise(\mathbf{x}) &= y - \mathbb{E}[y|\mathbf{x}], \\
&= \epsilon(\mathbf{x})
\end{aligned}$$

Definition 3.2. The approximation error of β, from Definition 2.8, is

$$\begin{aligned}
approx(\mathbf{x}) &= \mathbb{E}[y|\mathbf{x}] - \langle \beta, \mathbf{x} \rangle, \\
&= \alpha(\mathbf{x})
\end{aligned}$$

From Definition 3.1 and Definition 3.2, the modelling equation can be written as

$$y = \mathbf{x}^T\beta + \alpha(\mathbf{x}) + \epsilon(\mathbf{x})$$

This decomposition is illustrated below in Figure 2 for a non-linear case and Figure 1 for a linear case, and the code to generate these plots can be found in Appendix A.3.
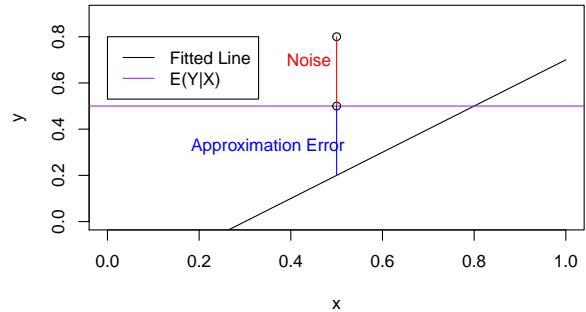

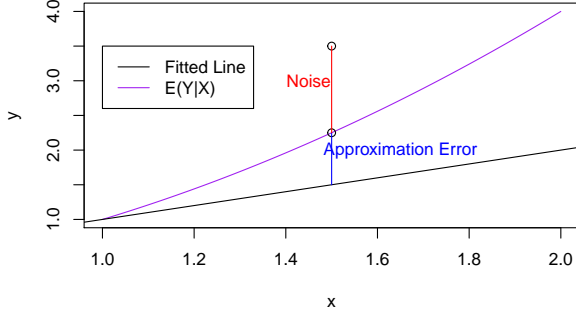*Figure 1.* Linear example of the modelling equation decomposition

Figure 2. Non-linear example of the modelling equation decomposition

The purpose of linear regression extends beyond modeling an assumed linear relationship between a response and predictor. In general, even in the presence of non-parametric components, linear regression can extract the parametric part of the surface of a response to derive inferences on parameters (Buja et al., 2014).

The key difference between the cases illustrated in Figure 1 and Figure 2 is that when the linear assumption is violated, the approximation error also considers the deviation from linearity in the decomposition and can be considered the non-linearity term (Buja et al., 2014).

Though not independent, the non-linearity term is uncorrelated with predictors since it is the population residual of the regression of $\mathbb{E}(Y|\mathbf{x})$ (Buja et al., 2014).

Lemma 3.3. *(Subgaussian Noise)* $\exists \sigma < \infty$ *where* $\rho_\lambda \geq 1$ *such that the following subgaussian noise bound occurs almost surely:*

$$\mathbb{E}[exp\{(\eta)noise(\boldsymbol{x})\}|\boldsymbol{x}] \leq exp\{\eta^2\sigma^2/2\}, \quad \forall \eta \in \mathbb{R}$$

This condition is satisfied in the case when the response noise is normally distributed, i.e., $\epsilon(\boldsymbol{x}) \sim N(0, \sigma^2)$ *(Hsu et al., 2014).*

Definition 3.4. The ridge estimator β_λ is the minimizer of the regularized MSE, i.e.,

$$\beta_\lambda = \arg\min_w \mathbb{E}[(\langle w, \mathbf{x}\rangle - y)^2] + \lambda||w||^2$$
$$= (\Sigma + \lambda I)^{-1}\mathbb{E}[\mathbf{x}y]$$

Definition 3.5. The approximation error of β_λ, from Definition 3.4, is

$$approx_\lambda(\mathbf{x}) = \mathbb{E}[y|\mathbf{x}] - \langle\beta_\lambda, \mathbf{x}\rangle$$
$$= \alpha_\lambda(\mathbf{x})$$

Lemma 3.6. *(Bounded Approximation Error)* $\exists b_\lambda < \infty$ *where* $b_\lambda \geq 1$ *such that the following bounded approximation error occurs almost surely:*

$$\frac{||(\Sigma + \lambda I)^{-1/2}\boldsymbol{x}\alpha_\lambda(\boldsymbol{x})||^2}{\sqrt{\mathbb{E}[||(\Sigma + \lambda I)^{-1/2}\boldsymbol{x}||^2]}} \leq \frac{||(\Sigma + \lambda I)^{-1/2}\boldsymbol{x}\alpha_\lambda(\boldsymbol{x})||^2}{\sqrt{d_{1,\lambda}}}$$
$$\leq b_\lambda$$

(Hsu et al., 2014).

3.1. Dimensions of the Covariate

The most effective dimensions of **x** depend on the eigenvalues of Σ, as defined in Proposition 2.7, and the ridge regularization parameter λ, as chosen when solving Definition 2.11.

Definition 3.7. The effective dimensions of **x** are

$$d_{p,\lambda} = \sum_j \left(\frac{\lambda_j}{\lambda_j + \lambda}\right)^p, \qquad p \in \{1, 2\}$$

Remark 3.8. $d_{2,\lambda} \leq d_{1,\lambda}$

Remark 3.9. In the special case that $\lambda \longrightarrow \infty$, define the dimensions as $\tilde{d}_{1,\lambda} = max\{d_{1,\lambda}, 1\}$

Definition 3.10. The main condition requires that the squared length of **x** is less than or equal to a constant factor greater than its expectation. Define the expectation of the λ-whitening mapping when $\lambda > 0$ as

$$\mathbb{E}[||(\Sigma + \lambda I)^{-1/2}x||^2] = tr[(\Sigma + \lambda I)^{-1/2}\Sigma(\Sigma + \lambda I)^{-1/2}]$$
$$= \sum_j \left(\frac{\lambda_j}{\lambda_j + \lambda}\right)$$
$$= d_{1,\lambda}$$

Lemma 3.11. *(Bounded Statistical Leverage)* $\exists$ *finite* $\rho_\lambda \geq 1$ *such that almost surely:*

$$\frac{||(\Sigma + \lambda I)^{-1/2}x||^2}{\sqrt{\mathbb{E}[||(\Sigma + \lambda I)^{-1/2}x||^2]}} \leq \frac{||(\Sigma + \lambda I)^{-1/2}x||^2}{\sqrt{d_{1,\lambda}}}$$
$$\leq \rho_\lambda$$

where the first inequality follows from Definition 3.10 (Hsu et al., 2014).

4. Random Design Setup for Ridge Regression

Recall the two previously defined ridge estimators.

- $\hat{\beta}_\lambda$ as the estimator for the empirical MSE from Definition 2.11

- β_λ as the estimator for the regularized MSE from Definition 3.4

Remark 4.1. It is important to note that in the case of a fixed design, $\hat{\beta}_\lambda$ is an unbiased estimator of β_λ, i.e.,

$$\mathbb{E}[\hat{\beta}_\lambda] = \beta_\lambda$$

This is not true in a random design setting, and using a biased estimator may lead to errors in the analysis. To prevent these errors, it is beneficial to use a different estimator. (Hsu et al., 2011)

Definition 4.2. Another estimator can be defined as the conditional expectation of $\hat{\beta}_\lambda$ given $\mathbf{x}_1, \mathbf{x}_2...\mathbf{x}_n$ is:

$$\bar{\beta}_\lambda = \mathbb{E}[\hat{\beta}_\lambda | \mathbf{x}_1, \mathbf{x}_2...\mathbf{x}_n]$$
$$= (\hat{\Sigma} + \lambda I)^{-1}\hat{\mathbb{E}}[\mathbf{x}\mathbb{E}[y|\mathbf{x}]]$$

Definition 4.3. Using Definition 2.9, define the (1) regularization bias, (2) random design bias, and (3) variance as:

$$\epsilon_{rg} = ||\beta_\lambda - \beta||_\Sigma^2 \tag{1}$$
$$\epsilon_{bs} = ||\bar{\beta}_\lambda - \beta_\lambda||_\Sigma^2 \tag{2}$$
$$\epsilon_{vs} = ||\hat{\beta}_\lambda - \bar{\beta}_\lambda||_\Sigma^2 \tag{3}$$

The main result provides strict upper bounds for the regularization bias, random design bias, and variance individually. These bounds can be used to create an upper bound for ridge regression in a random design setting.

Finding this ridge bound will require a generic quantity that bounds the excess MSE of the empirical MSE estimator $\hat{\beta}_\lambda$ over the minimum estimator β.

Using Definition 4.3, the following random design decomposition provides an upper bound on ridge regression in the randomized setting.

Proposition 4.4.

$$||\hat{\beta}_\lambda - \beta||_\Sigma^2 \leq \epsilon_{rg} + \epsilon_{bs} + \epsilon_{vr} + 2(\sqrt{\epsilon_{rg}\epsilon_{bs}} + \sqrt{\epsilon_{rg}\epsilon_{vr}}$$
$$+ \sqrt{\epsilon_{bs}\epsilon_{vr}}) \quad ,$$
$$\leq 3(\epsilon_{rg} + \epsilon_{bs} + \epsilon_{vr})$$

5. Main Result

The main result of the original paper concerns itself with the effect of errors on the covariance structure.

This result depends on the following assumptions:

- Assume Lemma 3.3, Lemma 3.6, Lemma 3.11 are true

- Assume that the sample size n has a lower bound of $n \geq 6\rho_\lambda^2 d_{1,\lambda}(log(\tilde{d}_{1,\lambda}) + t)$.

Theorem 5.1. *Choose a constant ridge parameter* $\lambda \geq 0$. *Choose* $t > max\{0, 2.6 - log(\tilde{d}_{1,\lambda})\}$.

Given these assumptions, the equations and inequalities below hold true with a probability of at least $1 - 4e^{-t}$:

1. **Spectral norm error of** $\hat{\Sigma} + \lambda I$

$$||(\Sigma + \lambda I)^{1/2}(\hat{\Sigma} + \lambda I)^{-1}(\Sigma + \lambda I)^{1/2}|| \leq (1 - \delta_s)^{-1}$$

where:

$$\delta_s = \sqrt{\frac{4\rho_\lambda^2 d_{1,\lambda}(log(\tilde{d}_{1,\lambda}) + t)}{n}}$$
$$+ \frac{2\rho_\lambda^2 d_{1,\lambda}(log(\tilde{d}_{1,\lambda}) + t)}{3n}$$

and $\delta_s \leq 0.93 < 1$

2. **Frobenius norm error of** $\hat{\Sigma}$

$$||(\Sigma + \lambda I)^{1/2}(\hat{\Sigma} + \Sigma)^{-1}(\Sigma + \lambda I)^{1/2}||_F \leq \delta_f \sqrt{d_{1,\lambda}}$$

where:

$$\delta_f = \sqrt{\frac{\rho_\lambda^2 d_{1,\lambda} - d_{1,\lambda}/d_{1,\lambda}}{n}}(1 + \sqrt{8t})$$
$$+ \frac{4\sqrt{\rho_\lambda^2 d_{1,\lambda} - d_{1,\lambda}/d_{1,\lambda}}t}{3n}$$

3. **Regularization effect**

$$\epsilon_{rg} \leq \sum_j \frac{\lambda_j}{(\frac{\lambda_j}{\lambda} + 1)^2}\beta_j^2$$

In the special OLS case when $\lambda = 0$, *we have* $\epsilon_{rg} = 0$

4. **The Effect of Bias due to Random Design**

$$\epsilon_{bs} \leq \frac{2}{(1 - \delta_s)^2}\left(\frac{\mathbb{E}[||(\Sigma + \lambda I)^{-1/2}(x\alpha_\lambda(x) - \lambda\beta_\lambda)||^2]}{n}\right.$$
$$(1 + \sqrt{8t})^2 + \frac{16(b_\lambda\sqrt{d_{1,\lambda}} + \sqrt{\epsilon_{rg}})^2 t^2}{9n^2}\right)$$
$$\leq \frac{4}{(1 - \delta_s)^2}\left(\frac{\rho_\lambda^2 d_{1,\lambda}\mathbb{E}[\alpha_\lambda(x)^2] + \epsilon_{rg}}{n}(1 + \sqrt{8t})^2\right.$$
$$\left. + \frac{(b_\lambda\sqrt{d_{1,\lambda}} + \sqrt{\epsilon_{rg}})^2 t^2}{n}\right)$$

5. *The Effect of Noise*

$$\epsilon_{vr} \leq \frac{\sigma^2(d_{2,\lambda} + \sqrt{d_{1,\lambda}d_{2,\lambda}}\delta_f)}{n(1-\delta_s)^2} + \frac{2\sigma^2 t}{n(1-\delta_s)}$$
$$+ \frac{2\sigma^2\sqrt{(d_{2,\lambda} + \sqrt{d_{1,\lambda}d_{2,\lambda}}\delta_f)t}}{n(1-\delta_s)^{3/2}}$$

(Hsu et al., 2014)

Note that for this theorem, the bound on n implies that n should be sufficiently large compared to some constant factor of the dimensions of the covariate x.

As previously stated, this result is used to provide upper bounds on $\epsilon_{vr}, \epsilon_{rg}, \epsilon_{bs}$, which will imply an upper bound on $||\hat{\beta}_\lambda - \beta||_{\hat{\Sigma}}^2$.

Proposition 5.2. *The bound on $||\hat{\beta}_\lambda - \beta||_{\hat{\Sigma}}^2$ using results 3, 4, and 5 from Theorem 5.1 is given by the following:*

$$||\hat{\beta}_\lambda - \beta||_{\hat{\Sigma}}^2 \leq ||\beta_\lambda - \beta||_{\hat{\Sigma}}^2$$
$$+ \mathcal{O}\Bigg(\frac{\rho_\lambda^2 d_{1,\lambda}\mathbb{E}[\alpha(x)^2]}{n}$$
$$+ \frac{(\rho_\lambda^2 d_{1,\lambda} + 1)||\beta_\lambda - \beta||_{\hat{\Sigma}}^2 + \sigma^2 d_{2,\lambda}}{n}\Bigg)$$

(Hsu et al., 2014)

5.1. Significance of Results

The Effects of Errors: Though these results appear rather complex, the simple implication is that the accuracy of $\hat{\Sigma}$ has a fairly mild effect on the bound.

Specifically, it appears through constant factors of the spectral and Frobenius norm errors from Theorem 5.1, where the terms are decreasing as sample size increases, therefore contributing to lower-order terms overall.

The Effect of Approximation Errors: The effect of approximation error ϵ_{bs} is an isolated term in Proposition 4.4. Its upper bound is the addition of the scaling of approximation errors depending on $\rho_\lambda d_1$, with some stochastic noise. This scaling term also contributes to lower-order terms overall and approaches 1 as $n \to \infty$

Significance: Based on the above remarks, the conclusion is that the ridge estimators behave similarly under fixed and random design settings, even though there is an influence of approximation error under random design.

This is because of the upper bound which tells us that the effect is second order and converges to a constant given sufficiently large sample sizes. This phenomenon will be explored in the proceeding section through a simple linear regression example.

6. Numerical Studies

The fixed design setting is restricted to observations made for controlled values of X, so there is no statistical validity to out-of-sample predictions. For experimenters, it may not always be feasible to control the predictors. In an observational study, certain groups may be more frequently observed than others, since the covariates follow their own natural distribution.

This introduces a bias or approximation error on top of random noise when attempting to model the relationship between predictors and responses. To further understand the effects of errors in a random design setting, where covariance structures are completely unknown, it is necessary to highlight its differences with a fixed design setting.

6.1. Simple Linear Regression in a Fixed vs Random Design Setting

For simplicity in the following example, consider a single explanatory variable for an ordinary least squares regression model, which is a special case of ridge regression where the regularization parameter $\lambda = 0$.

A population with predictors X and responses Y has the linear relationship $Y = \beta X + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

Assumptions:

- Assume in both cases that the noise term ϵ is normally distributed

- For fixed values of X: $\mathbb{E}(\epsilon) = 0$, $\text{var}(\epsilon) = \sigma^2$ and $\text{cov}(\epsilon_i, \epsilon_j) = 0$

- When the predictor is random, it can only be assumed conditionally that $\mathbb{E}(\epsilon|x) = 0$ and $\text{var}(\epsilon|x) = \sigma^2$ for particular outcomes of X
  - This is a key assumption in random design analysis because it ensures similar properties to a fixed design setting that would allow the covariates to be treated as if they were deterministic

(Hill et al., 2011)

The Fixed Design Setting: Treat X as deterministic, meaning values of X are controlled for in this particular experiment.

To simulate a fixed design setting, let the true value of the coefficient β be 1, and suppose a sequence of 100 numbers between -2 and 2 is fit to the response variable with added noise from the standard normal distribution.

The code for this section is found in Appendix A.1. Table 1, Figure 3, and Figure 4 summarize an instance of this linear regression model in a fixed design setting.

TERM	ESTIMATE	STD. ERROR	P. VALUE
(INTERCEPT)	0.1088874	0.0902675	0.2306156
EXPLANATORY	0.9888363	0.0773960	0.0000000

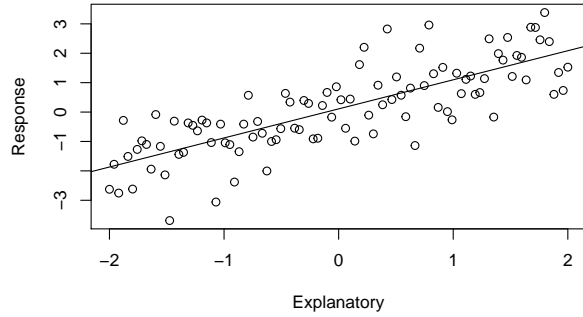*Table 1.* Linear regression model in a fixed design setting

TERM	ESTIMATE	STD. ERROR	P. VALUE
(INTERCEPT)	-0.0376926	0.0969873	0.6983896
EXPLANATORY	0.9989396	0.1077270	0.0000000

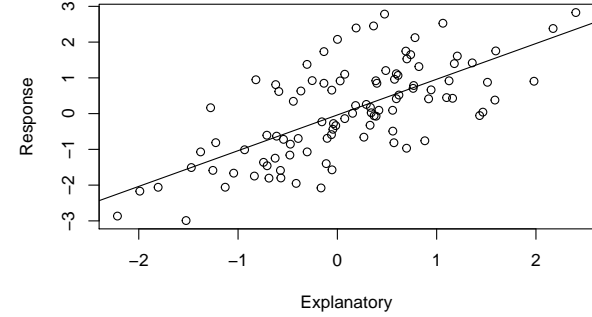*Table 2.* Linear regression model in a random design setting

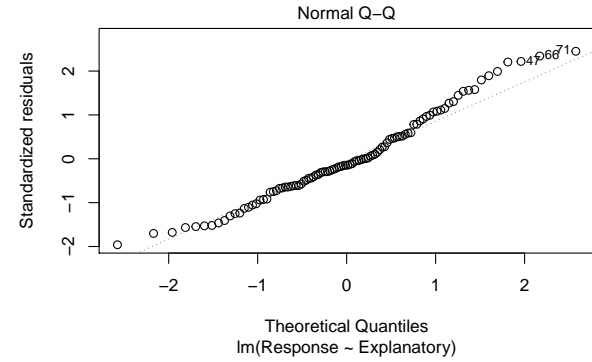*Figure 3.* Linear regression model in a fixed design setting

Figure 5. Linear regression model in a random design setting

Figure 4. Linear regression model normal quantile plot in a fixed design setting

Figure 6. Linear regression model normal quantile plot in a random design setting

The Random Design Setting: Treat X as random, meaning values of X are not controlled for and are determined jointly with Y and ϵ for each sample in the experiment.

To simulate a random design setting, suppose values of X are drawn from a standard normal distribution. Let the true value of the coefficient β also be 1, and fit the model with added noise from the standard normal distribution.

The code for this section is found in Appendix A.2. Table 2, Figure 5, and Figure 6 summarize an instance of this linear regression model in a random design setting.

With a sample size of 100, the estimated coefficient in both instances is statistically significant and very close to the true value 1 (from the tables, note that the p-value < 2e-16).

The plots vary slightly since X in the random design setting follows a normal distribution and is not fixed. This leads to more observations being made towards the center; however, the model was still able to accurately fit the relationship between the response and explanatory variables.

The normal quantile plots of the residuals indicate that the distribution of errors in both instances is relatively normal.

This is all evidence to suggest that the error effect is indeed second order for sufficiently large sample sizes. To further demonstrate this conclusion, Monte Carlo simulations can help explain the impact of the uncertainty between different samples.

6.2. Monte Carlo Simulations

The simulations explained above were repeated to obtain estimated values of β. The distribution of these estimates are shown below for the random design setting with increasing numbers of repeated sampling (N). The code for these simulations can be found in Appendix A.4.
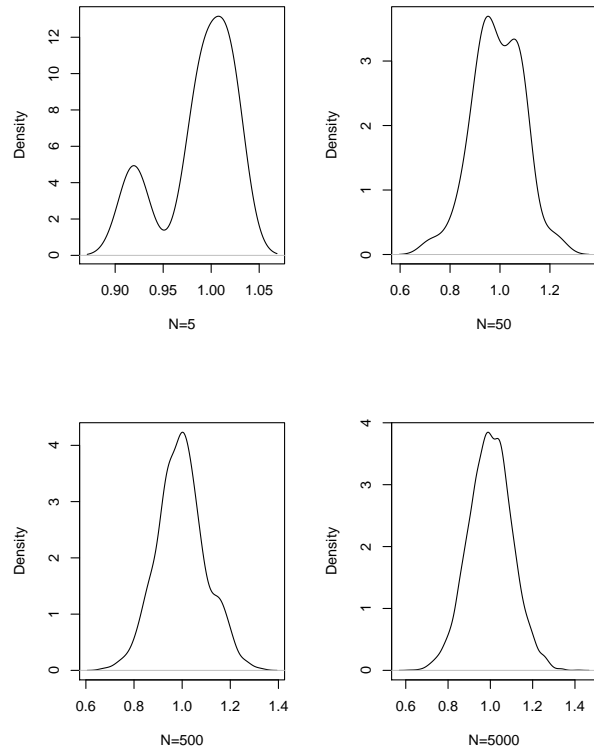

Figure 8. Distribution of Monte Carlo simulation estimations

Figure 7. Distribution of Monte Carlo simulation estimations

Figure 7 illustrates how the distribution of sample estimates of β converges to a normal distribution centered at the true value which is 1.

For further insight, the same simulation for both the fixed and random design setting was repeated 10 000 times to compare the distribution of the estimated regression coefficient.
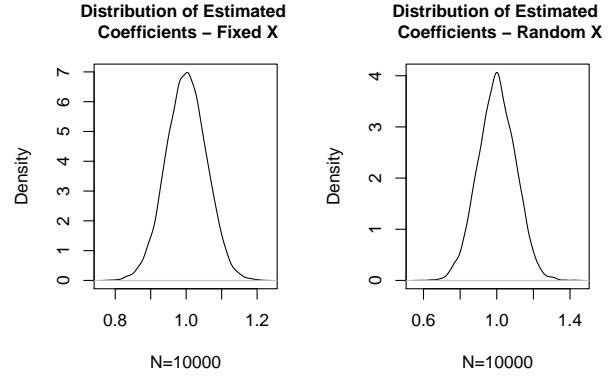
Based on the Monte Carlo simulations in Figure 8, the distributions of estimated values in both the fixed and random setting appear to follow a normal curve centered at the true value of β, which is 1.

This is evidence supporting the claim that the estimator behaves similarly under fixed and random designs, with the main difference being the lack of error in $\hat{\Sigma}$ under a fixed setting and the influence of bias and approximation error under random design.

This difference is highlighted by the variation seen in the distribution of $\hat{\beta}$, with there being a slightly larger spread of estimated values in the random setting. This illustrates how the scaling of the bounded statistical leverage $\rho_\lambda^2 d_{1,\lambda}$ with λ is crucial in controlling the effect of errors in random design settings when compared to fixed. (Hsu et al., 2014)

7. Limitations

One of the biggest limitations of Theorem 5.1 is the restriction on sample size because it will only hold if the sample size is sufficiently large, and will fail otherwise. The rest of the assumptions are not overly restrictive, as the numerical study has shown that scaling the bounded statistical leverage is necessary in order for the results to hold.

This paper analyzes a fixed effect model where the regression coefficients are assumed to be deterministic. In the case where effects may be random, however, the impact of errors may be more substantial so the results of the original work may not hold under these conditions.

8. Future Directions

A future research direction could be to perform similar analyses for Lasso regression, or ElasticNet regression. The original paper concludes that the effect of ridge regression errors are similar in random design vs fixed design settings,

but it may be beneficial to also understand the impact of different regularization terms on the regression function.

Another future research direction for this paper might interest itself in determining whether or not a random design setting behaves similar to a fixed design setting when both the predictors and regression slopes are random.

References

Buja, A., Berk, R., Brown, L., George, E., Pitkin, E., Traskin, M., Zhang, K., and Zhao, L. The Conspiracy of Random Predictors and Model Violations against Classical Inference in Regression. 2014. URL http://stat.wharton.upenn.edu/~buja/PAPERS/Notes-on-Wrong-Models-and-Random-X.pdf?fbclid=IwAR001YsBNBmc-MTgZKamHDZ1VlyQUgKWrxy0pgSh1RsVS38Mw0LmDowCSO8.

Hill, R. C., Griffiths, W. E., and Lim, G. C. *Principles of Econometrics, 4th Edition*. Wiley, 2011.

Hsu, D., Kakade, S. M., and Zhang, T. An Analysis of Random Design Linear Regression. 2011. URL https://arxiv.org/pdf/1106.2363v1.pdf.

Hsu, D., Kakade, S. M., and Zhang, T. Random Design Analysis of Ridge Regression. 2014. URL https://arxiv.org/pdf/1106.2363v2.pdf.

Rosset, S. and Tibshirani, R. J. From Fixed-X to Random-X Regression: Bias-Variance Decompositions, Covariance Penalties, and Prediction Error Estimation. 2017. URL https://www.stat.cmu.edu/~ryantibs/papers/randomx.pdf.

A. Appendix

A.1. The Fixed Design Setting Code

```r
library(tidyverse)
set.seed(1)
Explanatory <- seq(-2,2,length.out=100)
Response <- Explanatory + rnorm(length(Explanatory))
model <- lm(Response~Explanatory)
model_summary <- model %>% broom::tidy() %>% kable()

# PLOT
par(mfrow=c(1,2))
plot(Response~Explanatory)
abline(model)
plot(model, which = 2)
```

A.2. The Random Design Setting Code

```r
set.seed(1)
Explanatory <- rnorm(100)
Response <- Explanatory + rnorm(length(Explanatory))
model <- lm(Response~Explanatory)
model_summary <- model %>% broom::tidy() %>% kable()

# PLOT
par(mfrow=c(1,2))
plot(Response~Explanatory)
abline(model)
plot(model, which=2)
```

A.3. Decomposition Plot Code

```r
## Non Linear data
set.seed(1)
curve(x^2,from=1,to=2, xlab="x",ylab="y", col="purple")
abline(0,1)
points(c(1.5,1.5),c(3.5,1.5^2))
legend(1,3.5, legend=c("Fitted Line", "E(Y|X)"),
       col=c("black", "purple"), lty=1)
lines(c(1.5,1.5),c(1.5,1.5^2), col="blue")
lines(c(1.5,1.5),c(1.5^2,3.5), col="red")
text(1.65,2,label = "Approximation Error", col="blue")
text(1.45,3,label = "Noise",col="red")

## Linear data
curve(1*x,from=0,to=1, col="purple",ylab="y")
abline(h=0.2)
points(c(.5,.5),c(.5,0.8))
legend(0,0.8, legend=c("Fitted Line", "E(Y|X)"),
       col=c("black", "purple"), lty=1)
lines(c(.5,.5),c(0.2,0.5), col="blue")
lines(c(.5,.5),c(0.5,0.8), col="red")
text(.645,.3,label = "Approximation Error", col="blue")
text(.45,.7,label = "Noise",col="red")
```

A.4. Monte Carlo Simulations Code

```r
set.seed(1)
N <- 10000
X <- seq(-3,3,length.out=100)
beta_fixed <- vector()
for (i in 1:N){
  y <- X +rnorm(100)
  model <- lm(y~X)
  beta_fixed[i] <- model$coefficients[2]
}
beta_random <- vector()
for (i in 1:N){
  X <- rnorm(100)
  y <- X+rnorm(100)
  model <- lm(y~X)
  beta_random[i] <- model$coefficients[2]
}

par(mfrow=c(1,2))
plot(density(beta_fixed),cex.main=1, main="Distribution of Estimated
    Coefficients - Fixed X", xlab = "N=10000")
plot(density(beta_random),cex.main=1, main="Distribution of Estimated
    Coefficients - Random X", xlab = "N=10000")
```