DIMENSION SELECTION CRITERIA FOR PREDICTOR ENVELOPES IN UNIVARIATE REGRESSION

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

BY

SAM BALES

DR. MUNNI BEGUM – ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

MAY 2020

Acknowledgements

First, I would like to thank my thesis advisor Dr. Munni Begum, as well as Dr. Drew Lazar and Dr. Mayong Fan for serving on my committee. All three of them helped me greatly through this process, and I am thankful for their feedback and guidance. I would also like to thank Kofi Adgrani for introducing me to envelope methods.

I would also like to thank my parents Joe and Kelli Bales for supporting me throughout my college career. I would also like to thank my fiancee Addie Linegar for her advice and emotional support.

Contents

1. Introduction

Statistical models are becoming more important in a variety of fields from journalism to chemistry. One of the most commonly used multivariate statistical techniques is Ordinary Least Squares (OLS) regression. The general form of the model in matrix notation is:

$$Y = \mu + \mathbf{X}\beta + \varepsilon \quad (1).$$

Where $Y \in R^n$, $\mathbf{X} \in R^{n\,x\,p}$, $\beta \in R^p$ and $\varepsilon \in R^n$ is a normal random vector with zero mean and variance matrix $\Sigma = \sigma^2 I \in R^{n\,x\,n}$. This model is used widely in many areas of academic research and industry, and for good reason. OLS is highly flexible, easy to interpret, and a linear function is often a sufficient approximation for relationships between independent and dependent variables. Also, as errors often represent the sum of independent factors not included in the model, by the central limit theorem they are often normally distributed.

Several attempts have been made to improve the predictive accuracy and inferential capacity of OLS, ranging from simple mathematical transformations of the response and predictors to altering the objective function. These improvements have been undertaken for good reason: improving the performance of a widely used technique can reap substantial benefits. However, before new methodologies can be presented, the meaning of an "improvement" must be defined. To arrive at an appropriate definition, consultation of statistical theory is necessary.

Unbiasedness, Efficiency, and Sufficiency are three properties of estimators that underlie modern statistical theory. Unbiasedness asserts that the average value of an estimator–or statistic–will tend to the value of the estimated population parameter as the sample size approaches infinity. Efficiency relates to the variability of an estimator, and all else being equal, one prefers an estimator that has less variability (standard error). Sufficiency suggests a statistic should extract all relevant information about the parameter contained in observations.

Thus, an “improvement” in estimators should improve one of these three properties, among a number of others.

According to the Gauss-Markov theorem, an OLS estimator is the Best Linear Unbiased Estimator (BLUE) if a set of assumptions are satisfied, and is often a sufficient statistic. Thus, improvements in ordinary least square estimators are focused on increasing the efficiency of the estimator, which has positive practical implications. Estimators with smaller standard errors often lead to decreased error rates of statistical tests, translating into more robust discoveries by decreasing false positives (type I error) and improving the chance of detecting true relationships (power), all else remaining equal. Lower error rates translate into better business decisions, safer product development, and more robust scientific conclusions. Greater power translates into shorter decision lags, less costly experiments, and the potential for greater discovery.

Several techniques have been created that increase the efficiency of OLS in particular situations such as Ridge regression, Principal Component Regression (PCR), Partial Least Squares (PLS) and LASSO regression. However, these techniques can add unnecessary complexity outside of their intended use cases, which can complicate interpretability and affect inference. For example, PCR transforms predictors into a combination of orthogonal vectors from a Singular Value Decomposition of the sample covariance matrix. These orthogonal predictors reduce dimension and eliminate collinearity but also reduce standard errors of the coefficients. However, the effect of a single covariate becomes difficult to untangle since the estimated coefficients represent the marginal impact of a combination of covariates. Similar issues arise with Ridge regression, PLS, and LASSO.

Keeping these limitations in mind, a new technique called envelope methods has the potential to improve many of the commonly used multivariate statistical techniques, including OLS regression, without the limitations mentioned above. However, it is important to note that envelope methods are not a panacea, and this nascent methodology has many areas of active research. A review of the existing literature will show relevant applications and relevant research questions.

2. Literature Review

The literature review is divided into two sections. The first section will provide background information and the mathematics necessary for envelopes as well as the predictor envelope model, likelihood function, and estimation procedures. The second section details many of the extensions of envelope methods to commonly used techniques. Relevant notation needed for the presentation of envelope methods are included at the beginning of this section.

Notation and Reducing Subspaces

The following notations will be used throughout the paper:

I Let $\mathbb{R}^{a\,x\,b}$ denote the space of all real $a\,x\,b$ matrices, where a and b are positive integers

II Let span($\mathbf{A}$) denote the span of the columns of $\mathbf{A} \in \mathbb{R}^{axb}$

III Let $\mathbf{X} \in \mathbb{R}^{n\,x\,p}$ denote a matrix of p predictors with n observations

IV Let $\mathbf{S} \in \mathbb{R}^{p\,x\,p}$ denote a symmetric, positive-definite matrix

V Let $\mathbf{M} \in \mathbb{R}^{p\,x\,u}$ denote a semiorthogonal basis matrix such that $\mathbf{M}^T\mathbf{M} = \mathbf{I}$

VI For a subspace $\mathcal{R} \in \mathbb{R}^p$, let $\mathbf{S}\mathcal{R}$ denote the space of all vectors $\mathbf{S}x$ as x runs through $\mathcal{R}$

VII The projection onto the subspace $\mathcal{R}$ in the inner product of $\mathbf{S}$ is given as

$$P_{(\mathcal{R}(S))} = \mathbf{M}(\mathbf{M}^T\mathbf{S}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{S}$$

when $\mathbf{M}$ is a basis matrix for $\mathcal{R}$

VIII Let $\mathcal{R}^{\perp}$ denote the orthogonal complement of $\mathcal{R}$ and let $\mathbf{Q}_{(\mathcal{R}(\mathbf{S}))} = \mathbf{I}_p - \mathbf{P}_{(\mathcal{R}(\mathbf{S}))}$ such that $\mathbf{P}_{(\mathcal{R}^{\perp}(\mathbf{S}))} = \mathbf{Q}_{\mathcal{R}(\mathbf{S})}$.

The introduction of envelopes relies on the concept of reducing subspaces. Formally, a subspace $\mathcal{R} \in \mathbb{R}^p$ is said to be a reducing subspace of $\mathbf{S}$ if $\mathbf{S}\mathcal{R} \subseteq \mathcal{R}$ and $\mathbf{S}\mathcal{R}^{\perp} \subseteq \mathcal{R}^{\perp}$. Conventionally, If $\mathcal{R}$ is a reducing subspace of $\mathbf{S}$, $\mathcal{R}$ is said to reduce $\mathbf{S}$. A reducing subspace can be thought of intuitively in the following manner: $\mathbf{S}$ maps any vector in the reducing subspace back into the reducing subspace, and any vector in $\mathcal{R}$ will not be mapped into $\mathcal{R}^{\perp}$. So, the image of $\mathbf{S}$ can effectively be partitioned into two spaces: $\mathbf{S}\mathcal{R}$ and $\mathbf{S}\mathcal{R}^{\perp}$. The following definition states this notion in greater generality.

Definition (Cook, Helland and Su, 2013). A subspace $\mathcal{R} \subseteq \mathbb{R}^p$ is a reducing subspace of $\mathbf{M}$ if $\mathcal{R}$ decomposes $\mathbf{M}$ as $\mathbf{M} = \mathbf{P}_{\mathcal{R}}\mathbf{S}\mathbf{P}_{\mathcal{R}} + \mathbf{Q}_{\mathcal{R}}\mathbf{S}\mathbf{Q}_{\mathcal{R}}$.

Thus, by using a semiorthogonal basis matrix $\mathbf{M}$ for $\mathcal{R}$, $\mathbf{S}$ can be partitioned into the sum of two independent matrices using the projection operators defined in VII & VIII above and $\mathbf{P}_{(\mathcal{R}(\mathbf{S}))}$ is shortened to $\mathbf{P}_{\mathcal{R}}$. Partitioning symmetric matrices into the sum of two independent matrices is the defining feature of envelope methods.

Literature on Envelopes

The development of envelopes is founded on the notion of sufficient dimension reduction, which extends the concept of sufficiency for a statistic to functions of predictors in multivariate models. To illustrate, suppose $\mathbf{X}$ in (1) is replaced by a function $\mathrm{R}(\mathbf{X})$ that contains all the relevant information in $\mathbf{X}$ that is related to Y. Substantial gains can arise when $\mathrm{R}(\mathbf{X})$ is lower dimension than $\mathbf{X}$, and in situations where this occurs, it is called sufficient dimension reduction (Adragni and Cook, 2009).

The concept of sufficient dimension reduction was used by Cook, Li and Chiaromonte (2007), who presented a method for regression without inverting the predictor covariance matrix. This methodology requires estimating a central subspace of a $p\,x\,u$ matrix η such that $Y \perp X|X\eta$. When this condition is satisfied, $X\eta$ contains all of the relevant information about Y contained in X, and thus the regression of Y on the lower dimensional space $X\eta$ is informationally equivalent to the regression of Y on X–which is a *sufficient* dimension reduction (Cook et al., 2007). The advantages of this technique are multifold: it can facilitate regressions in which n does not dominate p, allow visualization of regressions in lower dimensions, and collinearity and base inference on a smaller subset of predictors (Cook et al., 2007). Most importantly, this methodology established that the regression can be based on a subspace of $\Sigma_{\mathbf{X}}$ rather than the entire covariance matrix.

This approach of basing the regression on a subspace of a covariance matrix was generalized by Cook, Li and Chiaromonte (2010), who proposed envelope methods for multivariate regression. In contrast to the methods presented by Cook et. al. (2007) which estimated a subspace of $\Sigma_{\mathbf{X}}$, the methods introduced by Cook et. al. (2010) focused on reducing $\Sigma_{\mathbf{Y}}$ where $Y \in \mathbb{R}^{n\,x\,r}$. By estimating the intersection of all invariant reducing subspaces $\mathcal{S} \in \mathbb{R}^r$ of $\Sigma_{\mathbf{Y}}$ called the envelope of Σ_Y, the authors were able to effectively reduce the dimension of Y without losing information. Moreover, if the regression is effectively based on $Y^* \in \mathbb{R}^{n\,x\,u}$ where $u < r$, then envelope methods can lead to substantial gains in efficiency for multivariate regression (Cook et. al., 2010).

Envelope methods were further extended to predictor envelopes by Cook, Helland, and Su (2013) which reduces Σ_X instead of reducing Σ_Y. To rigorously introduce predictor envelopes, some mathematical conditions must first be established. First, assume a reducing subspace $\mathcal{R}$ of $\mathbb{R}^p$ such that:

i $Q_{\mathcal{R}}\mathbf{X}$ is uncorrelated with $P_{\mathcal{R}}\mathbf{X}$ and

ii Y is uncorrelated with $Q_{\mathcal{R}}\mathbf{X}$ given $P_{\mathcal{R}}\mathbf{X}$.

Condition (i) asserts the two subspaces must be independent of each other, and condition (ii) requires $Q_{\mathcal{R}}\mathbf{X}$ to be unrelated to Y given the information contained in $P_{\mathcal{R}}\mathbf{X}$. For any R

with properties (i) and (ii), $Q_{\mathcal{R}}\mathbf{X}$ is said to be linearly immaterial to the regression and $P_{\mathcal{R}}\mathbf{X}$ is said to contain all of the information available about β in $\mathbf{X}$. This effectively partitions the predictor covariance matrix into a two matrices: one that contains variation material to the regression ($P_{\mathcal{R}}\mathbf{X}$) and one that contains immaterial variation ($Q_{\mathcal{R}}\mathbf{X}$). The following proposition establishes the algebraic properties that allow for the construction of envelopes.

Proposition (Cook et al. 2013) Assuming model (1), the condition (i) that corr($P_{\mathcal{R}}\mathbf{X}$, $Q_{\mathcal{R}}\mathbf{X}$) = 0 is algebraically equivalent to

a $\Sigma_{\mathbf{X}}\mathcal{R} \subseteq \mathcal{R}$ and $\Sigma_{\mathbf{X}}\mathcal{R}^{\perp} \subseteq \mathcal{R}^{\perp}$

When (a) holds, $\mathcal{R}$ is said to be a reducing subspace of $\Sigma_{\mathbf{X}}$. Condition (ii) that $corr(y, \mathbf{Q}_{\mathcal{R}}\mathbf{X}|\mathbf{P}_{\mathcal{R}}\mathbf{X}) = 0$ is algebraically equivalent to

b $span(\beta) \subseteq \mathcal{R}$

The smallest $\mathcal{R}$ satisfying (a) and (b) is called the Σ_X envelope of $span(\beta)$ and is denoted as $\mathcal{E}_{(\Sigma_X)}\{span(\beta)\}$. To rewrite (1) as an envelope model, let $d = dim\mathcal{E}_{(\Sigma_{\mathbf{X}})span(\beta)}$, let $\Sigma_{\mathbf{X}Y} = cov(\mathbf{X}, Y)$ and let $\Gamma \in R^{p\,x\,d}$ be a semi-orthogonal basis matrix for $\mathcal{E}_{\mathbf{X}}(\beta)$. With a known Γ, (1) can be rewritten as:

$$Y = \mu + \mathbf{X}\Gamma\alpha + \varepsilon \quad (2)$$

where

$$\Sigma_{\mathbf{X}} = \Gamma^T\Omega\Gamma + \Gamma_0^T\Omega_0\Gamma_0$$

and $\Omega = \Gamma^T\Sigma_{\mathbf{X}}\Gamma$ and $\Omega_0 = \Gamma_0^T\Sigma_{\mathbf{X}}\Gamma_0$. As mentioned above, a reduction in the predictor space leads to a partitioning of the predictor covariance matrix into sub-matrices that contain material and immaterial variation. The parameter $\alpha = (\Gamma^T\Gamma)^{-1}\Gamma^T\Sigma_{\mathbf{X}} \in R^d$ contains the coordinates of β relative to Γ (Cook et al., 2013). The coefficient vector in (2) is given by

$$\hat{\beta}_{\mathcal{E}} \equiv \Gamma\alpha = \Gamma(\Gamma^T\Sigma_{\mathbf{X}}\Gamma)^{-1}\Gamma^T\Sigma_{\mathbf{X}Y} = P_{(\mathcal{E}(\Sigma_{\mathbf{X}Y}))} = \beta.$$

The coefficient vector in (2) does not depend on the chosen Γ, and $\mathcal{E}_{\Sigma_{\mathbf{X}}}(span(\beta))$ is a parameter that must be estimated (Cook, Helland and Su, 2013). $\mathcal{E}_{\Sigma_{\mathbf{X}}}(span(\beta))$ lives in a set of d dimensional subspaces of R^p, which is a Grassmann manifold denoted as $\mathcal{G}(d,p)$. The likelihood function for Γ, the orthogonal basis matrix for $\mathcal{E}_{\Sigma_{\mathbf{X}}}(\beta)$, is given by:

$$J(\Gamma) = log|\Gamma^T\mathbf{S}_{\mathbf{X}}\Gamma| + log|\Gamma^T(\mathbf{S}_{\mathbf{X}} - \mathbf{S}_{\mathbf{X}Y}\mathbf{S}_{\mathbf{X}Y}^T/s_Y^2)^{-1}\Gamma| \quad (3)$$

The estimated semi-orthogonal basis matrix can be expressed in terms of (3) as:

$$\hat{\Gamma} = arg\,min\{J(\Gamma)\}.$$

Estimation of Γ can be performed using likelihood methods, the SIMPLS algorithm, or new methods proposed by Cook and Zhang (2016) and Cook and Zhang (2018). The most common envelope estimation procedure is the SIMPLS algorithm proposed by Cook, Forzani, and Su (2016), since likelihood methods require non-convex optimization and are computationally expensive. Instead of the non-convex optimization, the SIMPLS algorithm sequentially fixes the rows of $\hat{\Gamma}$ and minimizes (3). This estimation procedure is implemented in the **R** package **Renvlp** by Minji Lee and Zhihua Su, which provides a diverse suite of tools to most of the established envelope methods.

Additional algorithms designed to improve envelope estimation have been suggested. For example, the envelope coordinate descent (ECD) and envelope component screening (ECS) algorithms have been proposed by Cook and Zhang, (2018). The ECD stabilizes and speeds up the algorithm proposed by Cook and Zhang (2016) without decreasing accuracy, and the ECS provides and efficient way to reduce the dimension of the estimation task without losing information. However, the most recent algorithm has yet to be implemented in widely available software packages.

Implementation of envelope methods presents additional complexity. Most importantly, the size of the envelope, d, must be estimated a priori, and the success of the technique hinges on choosing an appropriate dimension. Selecting too small of an envelope will produce biased estimates, while estimating too large of an envelope will prevent the modeler from taking full advantage of the gains in efficiency (Cook, 2018). Cook (2018) proposed using holdout samples, cross-validation, or information criteria to estimate the dimension of predictor envelopes, but the performance of these techniques have not been examined in detail.

Envelope methods have been extended in several unique ways to increase the usefulness of the technique. Su and Cook (2012) introduced inner envelopes, which assume the entire vector Y is material to the estimation of β. Inner envelopes can offer gains in multivariate regression when the rank of the coefficient matrix β is equal to r. With this, one can reap efficiency gains in multivariate regression without requiring dimension reduction. Furthermore, Su and Cook (2011) introduced partial envelopes, which allow a subset of the predictors to be enveloped. Partial envelopes are useful when additional variables are included in the model that are not of particular interest such as control variables. Simultaneous envelopes for multivariate regression were proposed by Cook and Zhang (2013), which simultaneously reduces X and Y.

Extensions to many common statistical techniques have also been established. Envelope methods were applied to generalized linear models, including logistic, Poisson, and Cox regression through the development of a generalized enveloping model (Cook and Zhang, 2014). Khare, Pal and Su, (2017) applied Bayesian methods to envelope models by specifying a Bingham distribution on the set of orthogonal basis matrices Γ. The authors noted the posterior distribution can be approximated with a Gibb's sampler. Envelope methods have also been extend to matrix variate regression (Cook and Ding, 2017).

Research Questions

Some developments are unique to envelopes and extend their usefulness in special cases, while other developments generalize envelope methods to existing statistical methods such as generalized linear models. The estimation procedures of envelopes have also been improved by the development of faster and more efficient algorithms. However, Several unanswered practical questions about predictor envelopes still remain.

First, situations in which envelopes will offer advantages over OLS have not been explored in detail. Second, methods for selecting the dimension of the envelope have not be investigated in detail. Thus, the first objective of this paper is to determine when envelopes offer advantages over OLS. The second objective is to examine the performance of Likelihood Ratio Testing (LRT), Akaike Information Criteria (AIC), and Bayesian Information Criteria (BIC) which are implemented in the **Renvlp** package for selecting the predictor envelope dimension.

3. Methodology

The research methods can be divided into two main parts: a simulation study intended to test envelopes under a variety of controlled data conditions, and an application using Fama-Macbeth Regression, a commonly used estimation procedure in the finance literature. The simulation study was divided into two parts: the first phase simulated predictors from a multivariate normal distribution to principally test the effects of multicollinearity, sample size, and number of predictors. This phase was designed to test the aspects of the data that commonly complicate OLS. The second phase investigated the performance of the dimension selection criteria when some predictors were non-normally distributed. Predictors from distributions other than the normal, and the second phase was designed to compensate for this limitation in the first phase.

To rigorously test the performance of each method, a fully factorial design was employed in each phase. A detailed list of the simulation factors is included in the subsections below. Overall, each simulated dataset was fit with each of the following four models:

$$Y = \mu + \mathbf{X}\Gamma_{AIC}\alpha + \varepsilon \quad (4)$$

$$Y = \mu + \mathbf{X}\Gamma_{BIC}\alpha + \varepsilon \quad (5)$$

$$Y = \mu + \mathbf{X}\Gamma_{LRT}\alpha + \varepsilon \quad (6)$$

$$Y = \mu + \mathbf{X}\beta + \varepsilon \quad (7)$$

Where the subscript on Γ indicates which dimension selection criterion was used. Computing the information criteria and the likelihood ratio test was performed using the likelihood function expressed in equation (3), and the expressions for computing each selection criteria are:

$$AIC = 2p - 2J(\Gamma, p)$$

$$BIC = 2ln(p) - 2J(\Gamma, p)$$

$$LRT = 2(J(\Gamma, p) - J(\Gamma, (p-1))$$

All likelihood ratio tests for selecting the envelope dimension were performed with the default significance level of 0.01, and the specific simulation conditions in each phase are outlined below.

Phase 1

Sample size, number of predictors, correlation among predictors, variation of predictors, and distribution of ε were manipulated in the first phase. 1000 datasets under 72 different data conditions were simulated and 4 different models were estimated for each dataset. To mitigate the effect of choosing a specific value for a factor, some factors were simulated from a uniform distribution, and the upper and lower limits are specified in Table 1. For example, in the first arm of the study, σ_X^2 was simulated from a uniform distribution using one set of upper and lower limits of 1/10 and 4 respectively.

Table 1: Phase 1 Simulation Factors

Sample Size	Number of Predictors (% of Sample Size)	Predictor Correlation (Absolute Value)	Predictor Variance	Residual Distribution
25	10%	0 to 0.3	1/10 to 4	Normal(0, 100)
75	30%	0.3 to 0.7	4 to 15	Normal(0, 400)
25		0.7 to 1		

Other inputs that may affect the simulations were randomly generated to average out their affects. Specifically, μ_X was simulated with upper and lower limits of 5 and 15, respectively, and β was simulated with upper and lower limits of -2 and 2, respectively. To have an appropriately structured correlation structure, Y was constructed as the inner product of β and X plus the random vector specified in the Table 1. The number of predictors was computed as a percentage of the sample size and rounded up to the nearest integer to facilitate simultaneous modifications of the sample size and number of predictors.

Phase 2

The second phase examined the performance of each dimension selection technique when some predictor variables follow non-normal distributions. Nine total predictors were simulated in this phase: five from a multivariate normal and one predictor from the gamma, beta, logistic, and t distributions. Collinearity and variance were both modified for the five normal random variables, representing a special case of the first phase. The variances of each of the non-normal predictors were modified individually constituting the main difference between many simulation arms, and the parameters of the distributions are outlined in Table 2. Similar to the first phase, the number of predictors for a dataset was computed as a percentage of the sample size and rounded up to the nearest integer, and inputs that may affect the simulations were generated from uniform distributions with the same upper and lower limits specified in phase 1.

Table 2: Phase 2 Simulation Factors

Predictor Correlation (Absolute Value)	Residual Distribution	Non-Normal Distribution	Low Variance	High Variance
0 to 0.3	Normal(0, 100)	Gamma	scale = 1 & shape = 2	shape = 5 & scale = 10
0.3 to 0.7	Normal(0, 400)	Beta	scale = 1 & shape = 2	shape = 5 & scale = 10
0.7 to 1		Logistic	1	10
		t	df = 20	df = 1

The following is a discussion to clarify the study design and simulation procedure. First, assume the factors in the first row of Table 1. Thus 1000 datasets will be simulated with a sample size of 25 and 3 normally distributed predictors. Let x_{ij} indicate the i^{th} predictor in the j^{th} dataset, where $i = 1, 2, 3$ and $j = 1,\ 2,\ ...,\ 1000$. For the j^{th} dataset, $X_j \in \mathbb{R}^{25\,x\,3} \sim N(M_j, \Sigma_j)$ where $\mu_{ij} \sim Unif(5, 15)$, $\sigma_{ij} \sim Unif(1/10, 4)$ and $\rho_{ij} \sim Unif(0,\ 0.3)$, $i \neq j$. Furthermore, $\beta_j \sim Unif(-2, 2)$ and $Y = X_j\beta_j + \varepsilon_j$ where $\varepsilon_j \in \mathbb{R}^{25} \sim MVN(0, 100I_{25})$. Models (4) - (7) were fit to the j^{th} dataset to obtain $\hat{\beta}_{AIC}$, $\hat{\beta}_{BIC}$, $\hat{\beta}_{LRT}$ and $\hat{\beta}_{OLS}$. After the coefficients were estimated, the difference between the "true" population parameter and the estimated coefficients for each model were retained. For the first arm of the study, a matrix $R \in \mathbb{R}^{1000\,x\,4}$ was retained consisting of all $D_j \in \mathbb{R}^4$, $j = 1, ..., 1000$ where

$$D_j = \begin{bmatrix} \beta_j - \hat{\beta}_{AIC} \\ \beta_j - \hat{\beta}_{BIC} \\ \beta_j - \hat{\beta}_{LRT} \\ \beta_j - \hat{\beta}_{OLS} \end{bmatrix}.$$

Subsequent arms of the study were executed by repeating the above process with the next level of a factor while keeping all other factors from the previous trial constant. So, for the 2nd simulation round, datasets of size 75 were simulated with 8 predictors (10% of the sample size) without changing collinearity, number of predictors, or variance. Sequencing through all possible combinations of factors in Table 1 yields 72 different simulation scenarios, and this 72 D matrices were generated in phase 1. The same simulation procedure and study design were implemented in phase 2, which resulted in 96 D matrices. Results from both phases are presented in detail below.

Evaluation Criteria

Three criteria were used to evaluate the parameter estimates from models (4) - (7): mean-squared error (MSE), bias, and standard error (SE). The evaluation criteria for a arm of the simulation run are given in (8) - (10).

$$Bias = \left(\frac{1}{p}\right)\left(\frac{1}{n_{sim}}\right)\sum_{i=1}^{p}\sum_{j=1}^{n_{sim}}(\hat{\beta_{ij}} - \beta_{ij}) \quad (8)$$

$$Variance = \left(\frac{1}{p}\right)\left(\frac{1}{n_{sim}-1}\right)\sum_{i=1}^{p}\sum_{i=1}^{n_{sim}}(\hat{\beta_{ij}} - \bar{\beta_{ij}})^2 \quad (9)$$

$$MSE = \left(\frac{1}{p}\right)\left(\frac{1}{n_{sim}}\right)\sum_{i=1}^{p}\sum_{i=1}^{n_{sim}}(\hat{\beta_{ij}} - \beta_{ij})^2 \quad (10)$$

Note β_{ij} is the simulated population parameter, $\hat{\beta}_{ij}$ is the estimated parameter, and p is the number of parameters. So, these metrics are averaged across all datasets and parameters for an arm of the study.

4. Results

Phase 1

MSE is the preferred method of comparison among the methodologies, since it incorporates both bias and variance. Table 1 demonstrates that the techniques are unbiased on average in the first phase, as represented by the column of zeros and the equality of MSE and Variance since $MSE = Bias^2 + Variance$. So, MSE effectively captures the main differences among the techniques and is used in most tables and figures presented below.

Table 3: Phase 1 Overall Performance

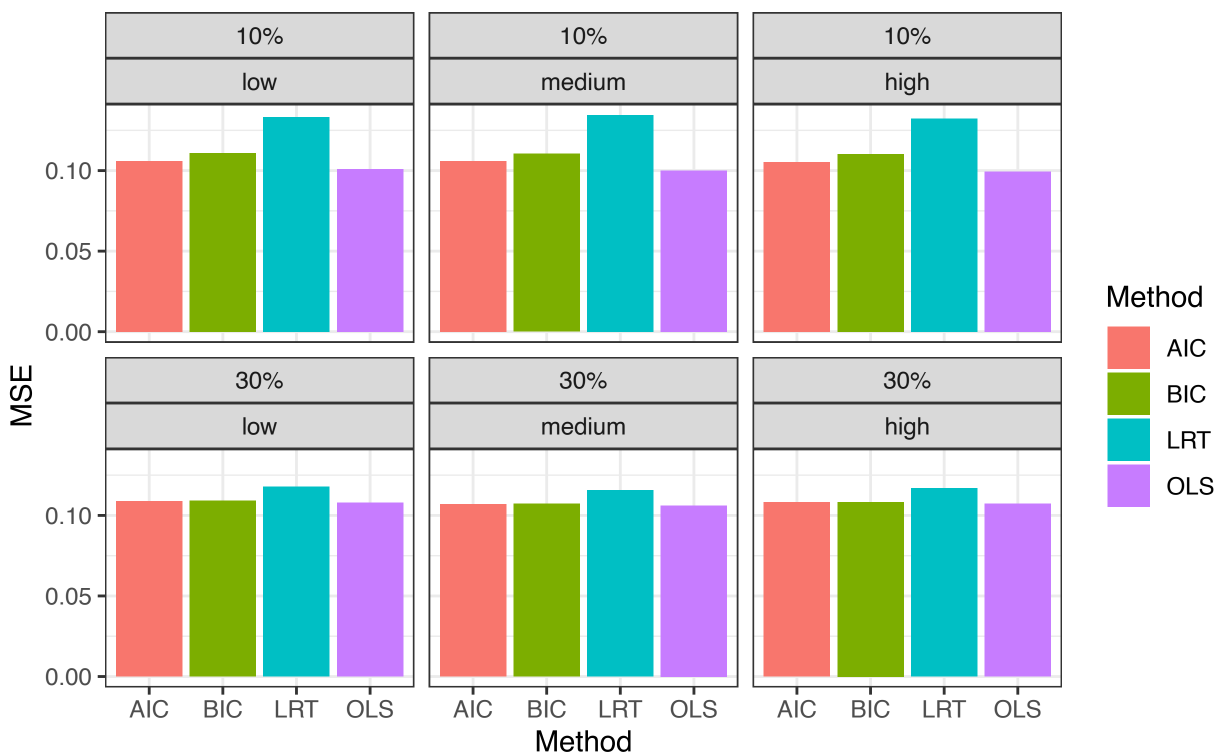
Method	MSE	Bias	Variance
AIC	1.060	0	1.060
BIC	0.965	0	0.965
LRT	0.639	0	0.639
OLS	1.163	0	1.163

Identifying when envelopes do not offer advantages over OLS is important, since envelopes methods are more computationally expensive and technically complicated. Thus, scenarios where envelopes do not outperform OLS are presented first. Figure 1 shows the MSE of each technique averaged over all sample sizes with high predictor variance and low response variance. The columns of the plot grid correspond to different levels of collinearity, and rows correspond to the number of predictors. In these scenarios, envelopes actually perform worse than OLS, and the degree of underperformance is a function of how small an envelope dimension the methods tend to estimate. LRT is the most conservative method, and it substantially underperforms the other three methods.

Also note that collinearity and the number of predictors have little impact on the relative performance of the different techniques, which can be seen in Table 4. For example, with an envelope model specified with BIC and a sample size of 25, changing the collinearity of the predictors only results in a change of 0.223. Changing the sample size from 25 to 250 for the same model decreases MSE by 4.079. It can be seen in Figure 1 that changing the number of predictors also has little effect on performance. These results suggest collinearity and the number of predictors do not substantially influence the relative performance of envelopes and OLS.

Figure 1

High Predictor Variance, Low Response Variation



Sample size does influence the relative performance of envelopes, but it is predicated on the other features of the data. Figure 2 illustrates the interaction between sample size and response variation. Rows correspond to the amount of response variation and columns correspond to the sample size. The bottom row of the figure indicates envelopes do not outperform OLS when there is little variability in the response. However, envelopes have

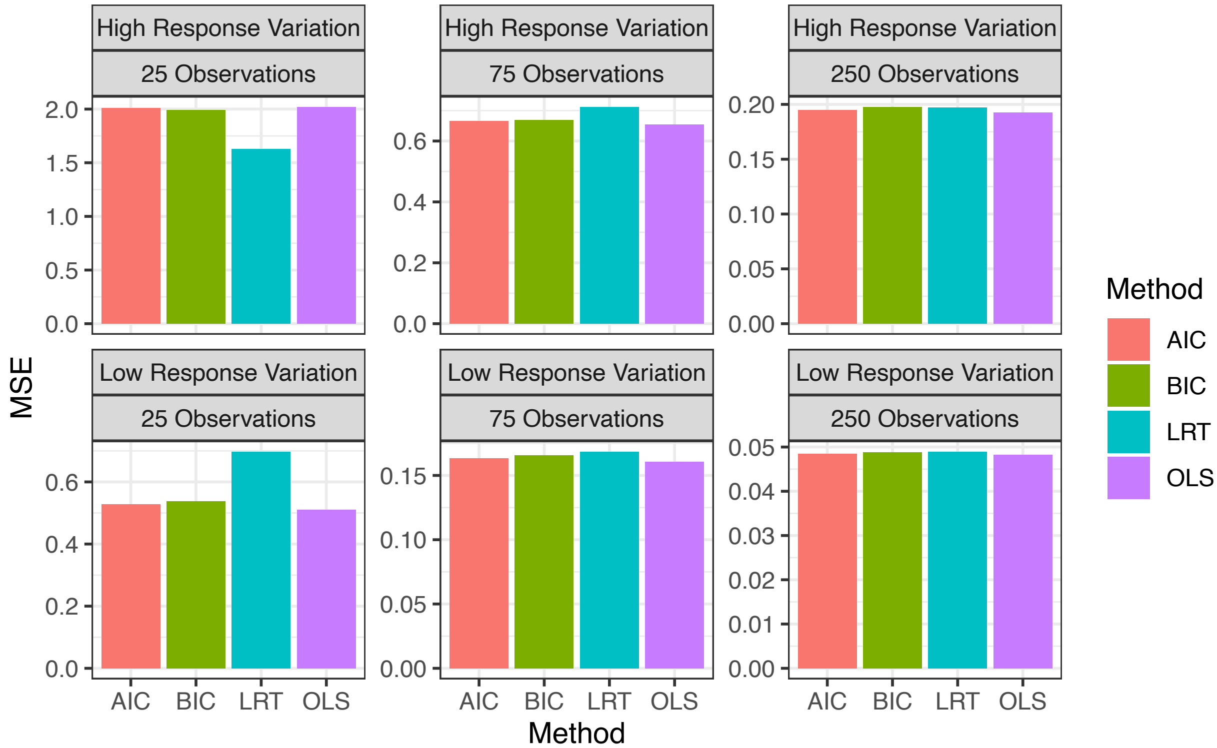
Table 4: MSE Over Collinearity and Sample Size

Correlation	25	75	250
AIC			
high	4.929	1.624	0.501
low	4.709	1.632	0.511
medium	4.873	1.676	0.510
BIC			
high	4.542	1.432	0.463
low	4.319	1.460	0.468
medium	4.468	1.510	0.471
LRT			
high	1.966	1.039	0.386
low	1.953	1.039	0.384
medium	1.823	1.104	0.384
OLS			
high	5.671	1.791	0.530
low	5.395	1.786	0.539
medium	5.489	1.832	0.536

a lower MSE with a sample size of 25 and high response variation. Also not that the envelope model specified with LRT outperformed all other methods in this scenario, but underperformed in all other panels of the plot. This reaffirms the finding from above that LRT amplifies the effects of envelopes and results in the best or worst performance given the characteristics of the data.

Figure 2

Effects of Response Variation and Sample Size with High Predictor Variability



Variability in the predictors is the most important factor for envelope outperformance. Figure 3 shows the MSE of each technique faceted by sample size and response variation when the variation in the predictors is low. Rows correspond to response variation and columns correspond to sample size. Envelopes outperform OLS in all cases, and LRT performs the best among all dimension selection criteria.

Table 5 demonstrates the degree of outperformance across levels of predictor variance. With high predictor variance, the difference in MSE between OLS and a LRT specified model is 0.187. Under low predictor variance, the difference in MSE between these two models is 12.405. Sample size does temper the difference in performance, but these figures demonstrate the power of envelopes.

Figure 3

Effects of Low Predictor Variation Across Sample Sizes and Response Variation

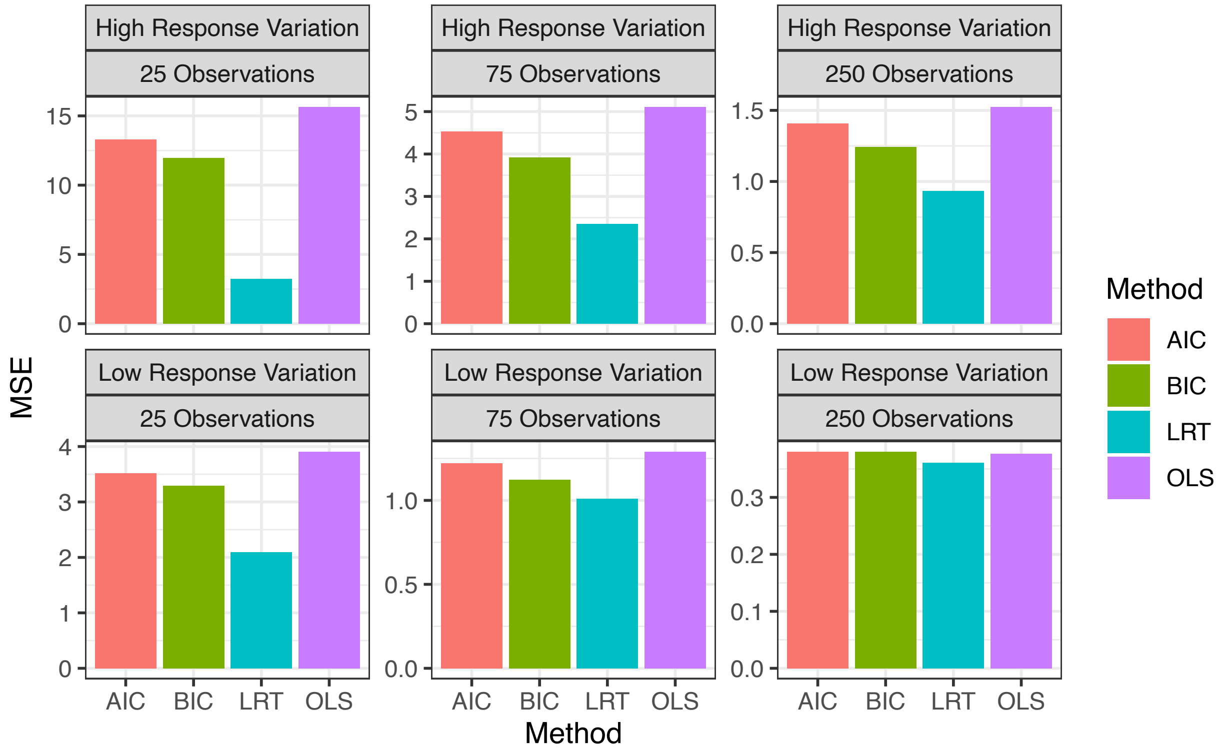


Table 5: High Response Variation by Sample Size and Predictor Variance

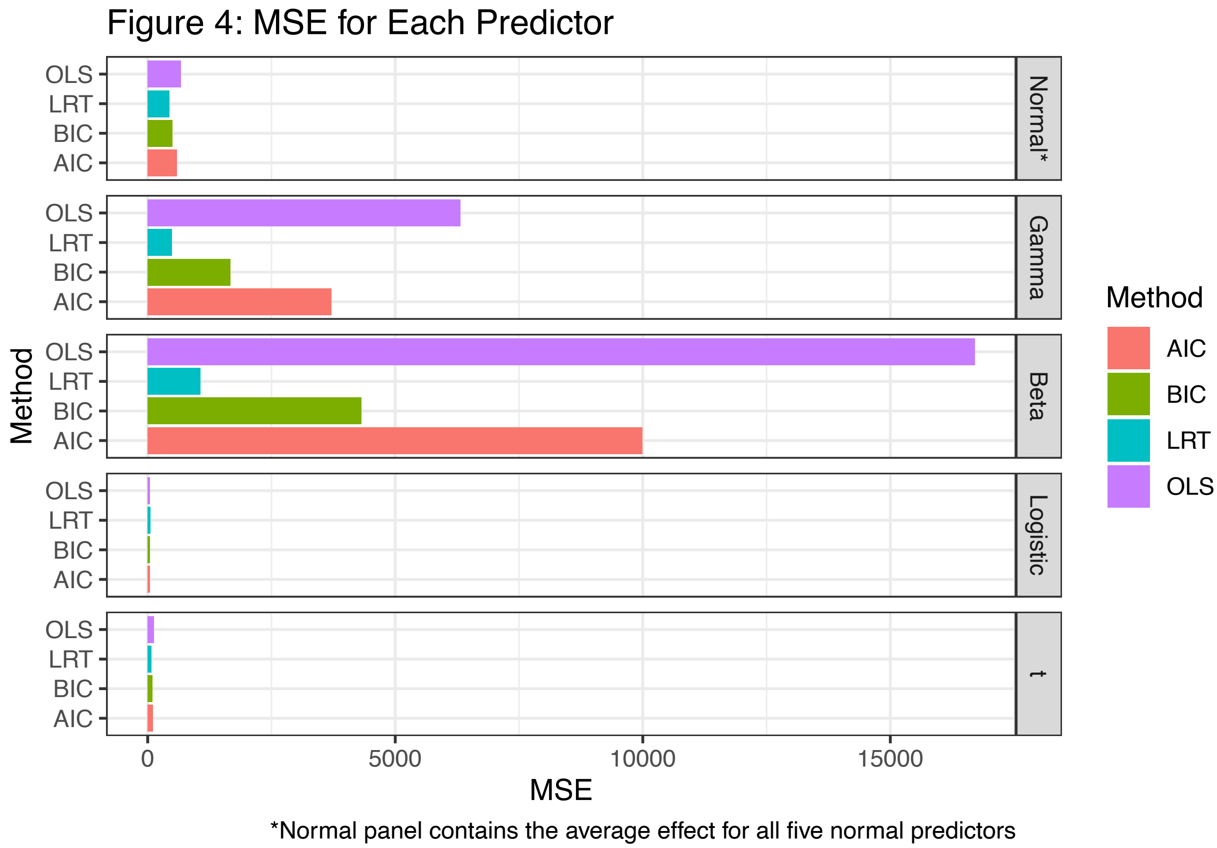
Method	25	75	250
High Predictor Variance			
AIC	0.529	0.163	0.048
BIC	0.538	0.166	0.049
LRT	0.697	0.168	0.049
OLS	0.510	0.161	0.048
Low Predictor Variance			
AIC	13.289	4.525	1.406
BIC	11.950	3.911	1.242
LRT	3.231	2.353	0.932
OLS	15.636	5.107	1.523

The best dimension selection criteria for normally distributed predictors is dependent on the structure of the data, with the most important factor being the variation in the predictors. Low variability in the predictors makes envelopes more efficient than OLS. As seen from the above figures, LRT testing is the best dimension selection criteria when the variability in the predictors is low. High variability in the response also leads to gains for envelopes, but their outperformance is tempered by sample size. Generally speaking, envelopes will outperform in small samples when the variability in the response is high. Most interestingly, the number of predictors and collinearity minimally impact the performance of envelopes. Thus, the relative variability of the response and predictors is important to consider when the data are normally distributed.

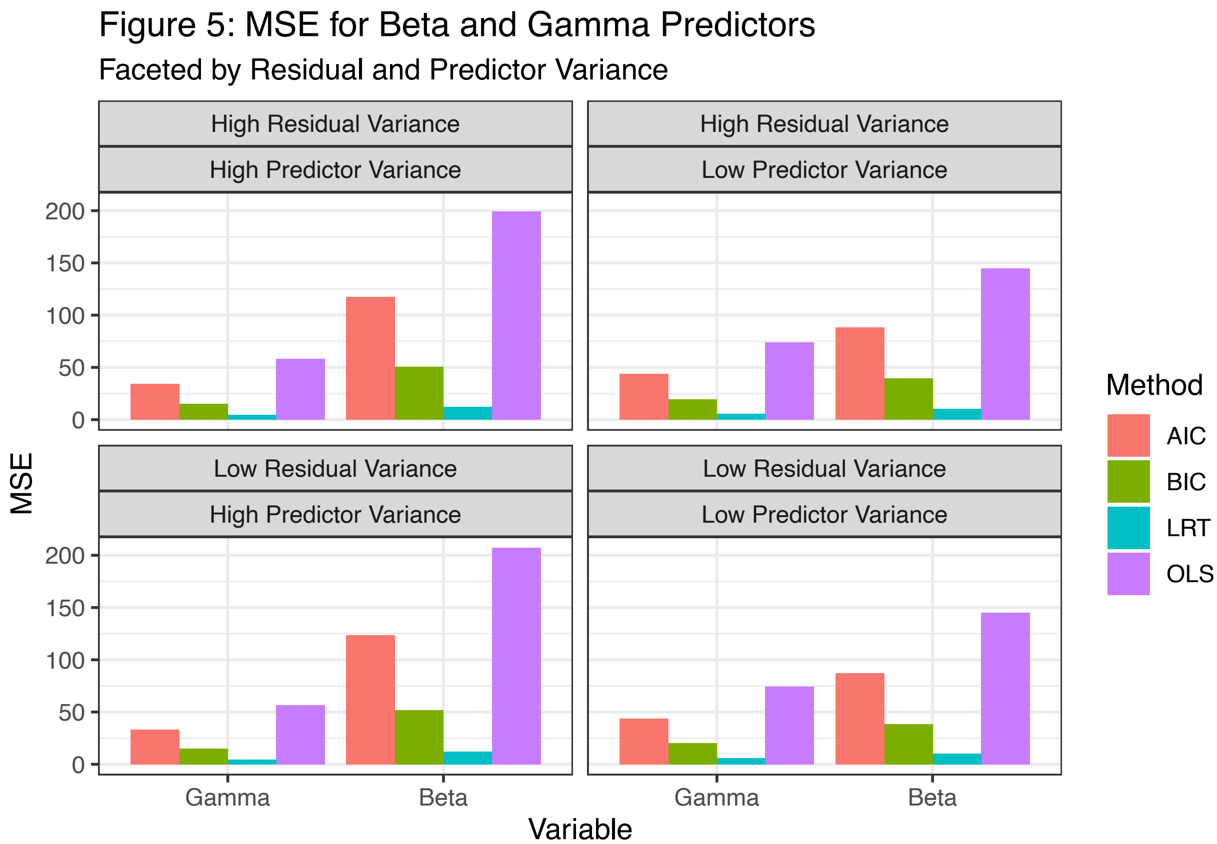
Phase 2

Although approximately normal predictors are common, predictors that are not approximately normal are often encountered in applications. Phase 2 is designed to test the performance of the three dimension selection criteria when there are non-normal predictors. To provide continuity from the first phase, five normally distributed data were included to test the effects of collinearity, sample size, and number of predictors. Four additional predictors from the gamma, beta, logistic, and t distributions were included to test the performance of envelopes on non-normally distributed predictors. The following scenario presents how these distributions could be combined in an analysis: change in assets (normal), change in operating cash flow (t), time since previous earnings beat (gamma), debt to assets ratio (beta), market capitalization (logistic) are used to predict the market return in the next year (normal).

Similar to Phase 1, MSE was the main method used to compare the performance of the techniques since the methods have negligible bias. Figure 4 shows the overall MSE of each dimension selection technique during this phase for each type of predictor distribution. From the fixed scaling on the x-axis, one can see that the MSE for the gamma and beta predictors is significantly larger than the MSE for the other predictors. Also note the difference in MSE among the four methodologies. Envelopes outperform OLS, and similar to the findings in phase 1, the degree of outperformance is a function of the conservatism of the technique, with LRT performing the best, followed by BIC and AIC.



To put precise figures on the performance, the MSE for the beta distribution using LRT is 11.13, whereas the MSE for BIC, AIC, and OLS is 45.02, 104.10, and 174.01 respectively. These results persist even while changing the two most important factors from phase 1: residual variance and predictor variance. Figure 5 shows the outperformance for the gamma and beta predictors for all levels of these factors.



The outperformance for the predictor from the beta distribution is likely a result of the constrained support of the distribution. Since a beta random variable can only take values between 0 and 1, the predictors likely behave similar to predictors with low variance in phase 1. Moreover, it appears the shape of the distribution also has an effect, given the outperformance for the gamma predictor. The gamma distribution is also constrained below by 0 and has "heavier" tails when compared to the normal distribution. Also note that the MSE for the gamma and beta distributions increases when the variance of the distribution increases. This is a reversal of the trend observed in phase 1, and is likely a result of the changing shape of the distribution.

The presence of predictors from the beta and gamma distributions does not ensure envelopes outperform OLS for all predictors in the model. Figure 6 presents the MSE for the predictors from the Normal, logistic, and t distributions. Table 6 contains the MSE for all variables grouped by response variation. From the figure and table, it can be seen that

Table 6: MSE for Normal, Logistic, and t Predictors

Variable	Residual_Variance	AIC	BIC	LRT	OLS
Normal*	high	1.217	1.027	0.904	1.392
Normal*	low	1.225	1.037	0.907	1.409
Logistic	high	0.456	0.473	0.527	0.440
Logistic	low	0.448	0.475	0.523	0.429
t	high	1.522	1.292	1.045	1.706
t	low	0.776	0.669	0.562	0.878

* Averaged across all 5 normal predictors

envelopes outperform OLS for normally distributed predictors when the response variation is high and predictor variation is low. However, envelopes do not outperform OLS for the predictor from the logistic distribution. This suggest the dispersion of the predictor may be able to offset the effect of skew. Most notably, envelopes outperform OLS under all scenarios for the predictor from the t distribution.

Figure 6: MSE for Normal, Logistic and t Predictors

Faceted by Response and Predictor Variation

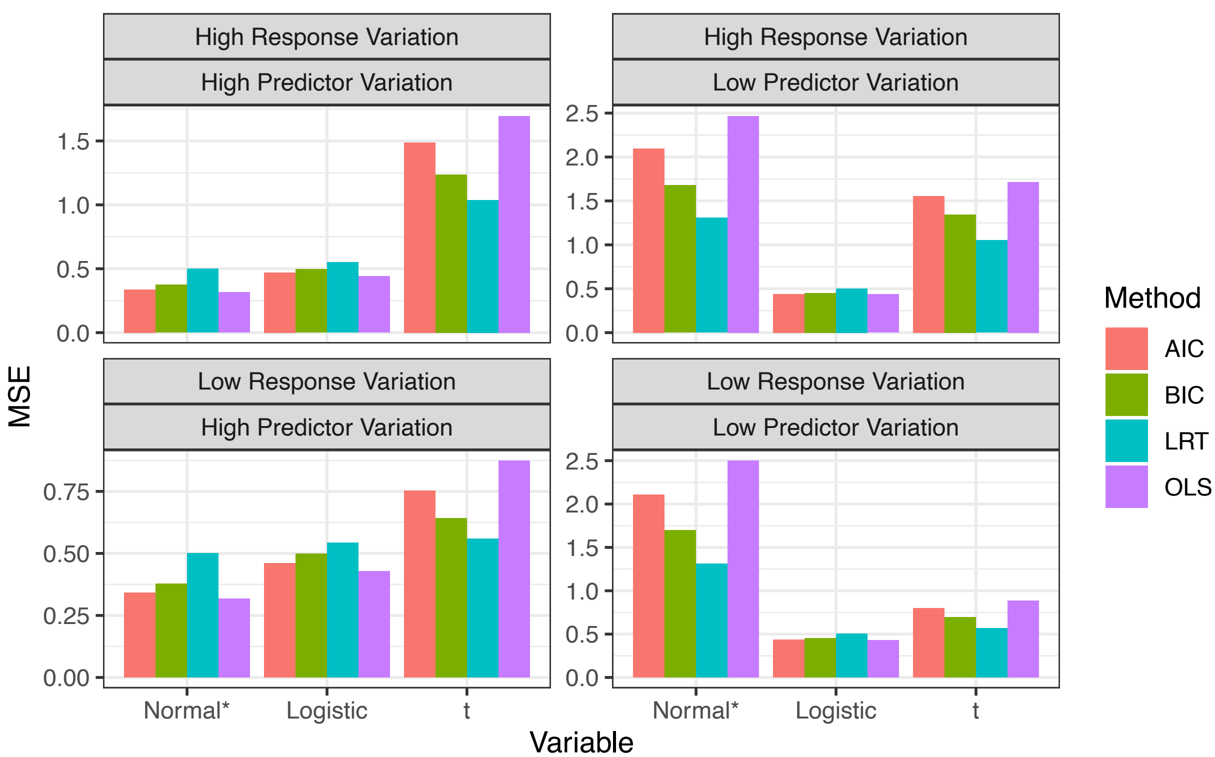


Figure 7 presents the MSE for predictors from $t_{(df=1)}$ and $t_{(df=20)}$ distributions under both levels of residual variation mentioned in Table 2. Note that envelopes outperform in all scenarios, and LRT has the lowest MSE. This serves to temper the results from phase 1, as the histogram, boxplots, and kernel density plots of a t distributed random variable with 20 degrees of freedom looks very similar to a normal distribution. So, routine exploratory analysis techniques may not be precise enough to indicate which dimension selection technique will be the best. Table 6 presents the MSE of each technique for the logistic predictor under all levels of predictor and response variation. Note that envelopes still do not have an advantage over OLS, and that LRT is the worst performer.

Figure 7: t Distribution

Faceted by Degrees of Freedom and Response Variation

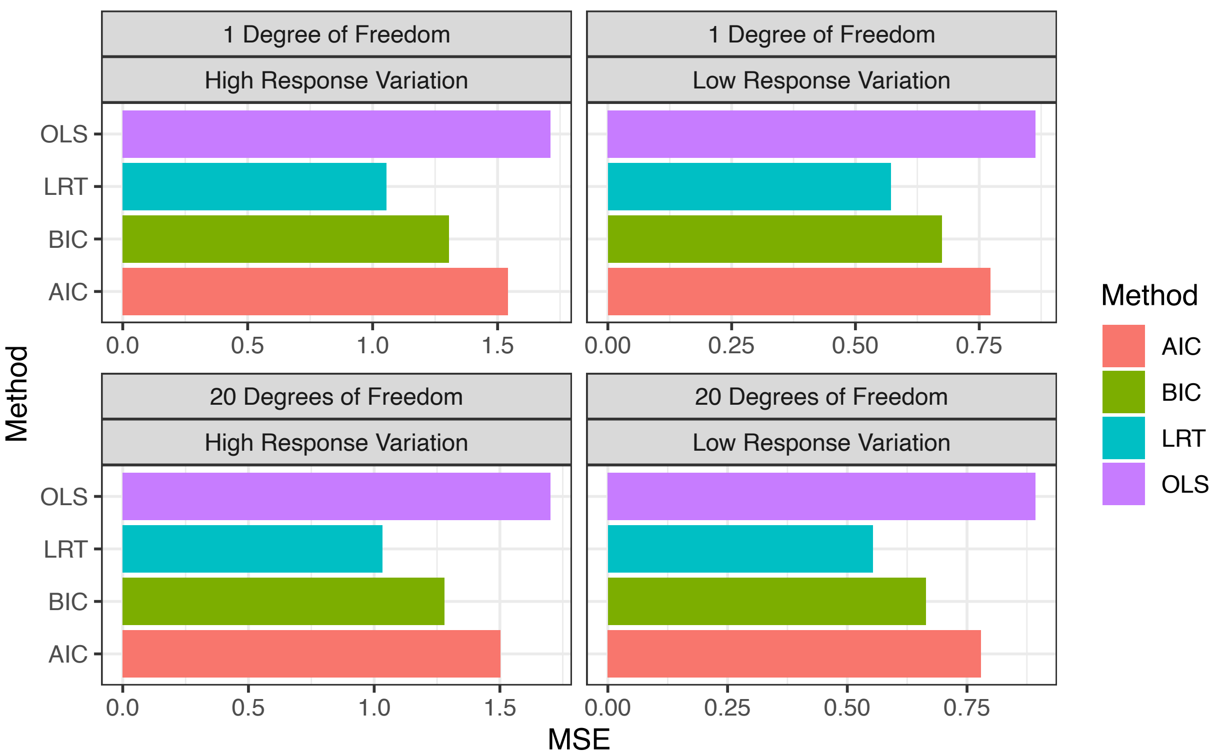


Table 7: Logistic Predictor

Logistic_Variance	Residual_Variance	AIC	BIC	LRT	OLS
high	high	0.453	0.471	0.526	0.438
high	low	0.455	0.483	0.526	0.434
low	high	0.460	0.476	0.527	0.442
low	low	0.441	0.467	0.520	0.424

Discussion and Recommendations

Phase 1 simulation study results indicate envelopes outperform OLS when there is either (1) a small sample size, (2) high response variation or (3) low variation in the predictors. Correlation among predictors and the number of predictors have small effects and do not result in gains in envelopes over OLS. Furthermore, when envelopes outperform OLS, LRT is the best dimension selection criterion.These findings are corroborated with the results from Phase 2, where envelopes had a substantially lower MSE for predictors form a beta, gamma, to t distribution. Similar to phase 1, LRT test is the best dimension selection criterion when envelopes offer advantages over OLS..

When predictors follow a normal distribution, the two most important factors are the variability in the repose and the variability of the predictors. It is unlikely that a practitioner will encounter data that follows the exact conditions examined in the simulations, so the identification of the appropriate dimension selection criterion should be done a priori via an exploratory analysis. A simple exploratory procedure to determine when envelopes can offer advantages over OLS is outlined below and demonstrated in the following application.

First, sample size is an important factor to consider and small sample sizes ($n \leq 25$) suggest envelopes can offer advantages. The variability and shape of the predictor distributions should be analyzed second through summary statistics and univariate plots. Low variability and skewed distributions indicate envelopes may offer advantages. Next, the mean squared error from an OLS regression (SSE_{OLS}) can be used as a rough measurement of the conditional variance of the response. Large values for SSE_{OLS} indicate envelopes will be the better option.

5. Application

Overview

To demonstrate the power of predictor envelopes and illustrate the exploratory process, predictor envelopes will be used to estimate the risk premia in the Fama-French 5 factor model using Fama-Macbeth Regression (FMR). The Fama-French 5 factor models aim to predict the return of securities based on five financial factors, which are: the excess return of the equity market over the risk-free-rate, and the difference in returns of diversified portfolios sorted on book-value, size, operating profitability, and investment practices. These five factors can be combined into the following Five Factor model:

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2 SML_t + \beta_3 HML_t + \beta_5 CMA_t + \beta_6 RMW_t + \varepsilon_{it} \quad (8)$$

- R_{it} is the return of the i^{th} stock in period
- R_{ft} is the risk free rate in period
- $R_{Mt} - R_{ft}$ is the excess return of the equity market over the risk free rate in period
- SML_t is the average excess return of portfolios of small stocks over portfolios of large stocks, measured by market capitalization
- HML_t is the average excess return of portfolios of low market-to-book stocks over portfolios of high market-to-book stocks in period
- CMA_t is the average excess return of portfolios of conservative investment firms over portfolios of aggressive investment firms in period
- RMW_t is the average excess return of portfolios of firms with robust operating profit margins over portfolios of firms with weak operating profit margins in period

- α_{it} captures the excess return of the security in time

- ε_{it} is Normal(0, $\sigma^2 I$)

FMR is a widely used econometric procedure used to estimate the risk premia for financial assets, and is the main technique used to estimate the loadings on each factor of interest. This requires a two-step estimation procedure. The first step is a time series regression of each security's return on the five factors. The second step is a cross-section regression of all stocks in a time period on the estimated coefficients from the first step. To present this FMR formally, let $R \in \mathbb{R}^{txs}$ contain the returns of s securities over t periods, and let $F \in \mathbb{R}^{txf}$ contain the f factors of interest over t periods. The first step involves estimating s time series regressions:

$$R_t - R_{ft} = \alpha_t + \beta_1(R_{Mt} - R_{ft}) + \beta_2 SML_t + \beta_3 HML_t + \beta_5 CMA_t + \beta_6 RMW_t + \varepsilon_t$$

Each β vector from all time-series regressions is retained, and let $B \in \mathbb{R}^{sxf}$ contain these estimates. The second step uses B as the predictor matrix for t cross-section regressions of asset returns. The coefficient matrix λ is estimated in the following manner:

$$R_i - R_f = \alpha_i + \lambda_1 \beta_{(R_M - R_f)} + \lambda_2 \beta_{SML} + \lambda_3 \beta_{HML} + \lambda_4 \beta_{CMA} + \lambda_5 \beta_{RMW} + \varepsilon_i.$$

The $\lambda_k' s$ are aggregated in $\Lambda \in \mathbb{R}^{txf}$ matrix. The loading on each factor is the average λ_j over time, and are calculated as $\text{Loading}_j = \frac{1}{t}\sum_{k=1}^{t} \lambda_{kj}$.

Data

Monthly returns for the S&P 500 and monthly factor data were collected from June 1987 to August 2019. Factor data was collected from Kenneth R. French's website [1], which performs the portfolio sorting and construction. Monthly return data was collected using the Bloomberg Professional service. After eliminating securities that did not have returns running the full length of the sample period, 389 security returns were used in the analysis.

Estimation

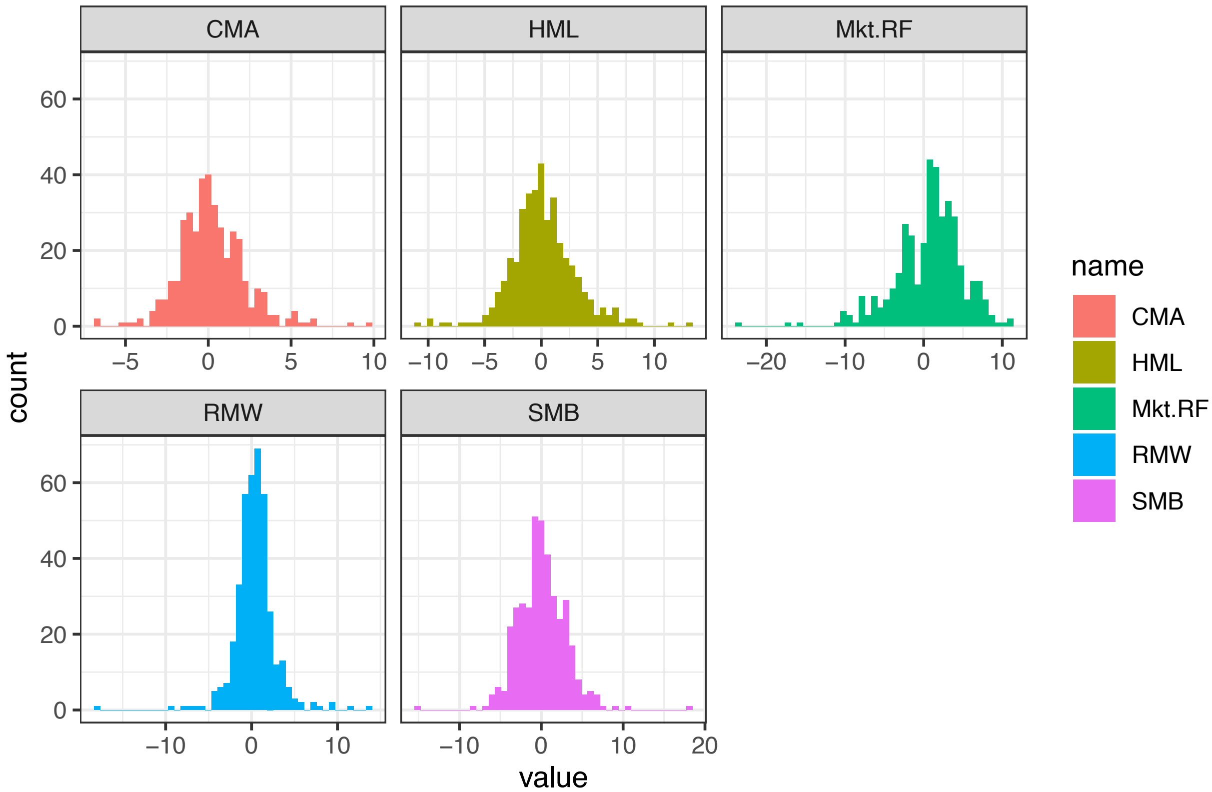
Predictor envelopes are likely to improve estimates in this procedure for several reasons. First, Table 7 shows the standard deviation of the five factors as well as the mean squared error of the residual return for Apple, Inc. from an OLS regression. There is relatively little variation in the predictors, indicating a predictor envelope may be beneficial. Furthermore, the standard deviation of the residuals is relatively large, also indicating envelopes may offer advantages. Phase 2 of the simulations demonstrated the distribution of the predictors impacts how well envelopes perform. Figure 8 shows histograms of the five factors. These plots show some factors have heavy tails, and some variables such as $(R_{Mt} - R_{ft})$ and HML are skewed. This further indicates envelopes will outperform OLS during this modeling task. The sample size, number of predictors, correlation matrix do not further suggest envelopes over OLS, but are still important to consider.

Table 8: Factor Standard Deviations

Mkt.RF	SMB	HML	RMW	CMA	AAPL SD
4.355	2.993	2.916	2.494	2.014	12.417

[1] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Figure 8: Fama–French Factor Histograms



After the preliminary analysis, the coefficients from the first stage of FMR can be estimated. The **xenv** function in the **Renvlp** package returns the ratio of the estimated standard errors from envelopes and OLS. The average ratio for each factor employing all dimension selection techniques is shown in Table 9. All dimension selection criteria have similar gains in efficiency over OLS, but LRT has greater gains in efficiency. Also note that the largest improvement comes for CMA, which has the lowest standard deviation as identified in Table 8. Gains in efficiency at this stage have a multiplicative effect for the second phase, as the estimated coefficients will now be used as predictors. This allows for the propagation of estimation errors throughout the analysis.

Table 9: Stage 1 OLS to Envelope Standard Error Ratios

	Mkt.RF	SMB	HML	RMW	CMA
AIC	1.296	1.355	1.651	1.522	2.605
BIC	1.389	1.504	1.798	1.660	2.983
LRT	1.400	1.507	1.810	1.661	2.994

Before beginning phase 2, another preliminary analysis of the predictors is in order. Table 10 shows the standard deviation for the estimated betas for each factor and dimension selection criteria. This table shows there is very low variation in the estimated betas, indicating envelopes will offer substantial advantages. The standard error of the residuals from an OLS model for the first year of cross-sectional returns is approximately 7.60, indicating high variation in the response, further justifying envelopes. Figure 9 shows kernel density plots of the betas for each variable from the OLS fit. These plots show the estimated betas have heavy tails, and slightly skewed distributions, further suggesting envelopes will outperform.

Table 10: Standard Deviation of Stage 1 Betas

	OLS	AIC	BIC	LRT
Mkt.RF	0.114	0.108	0.089	0.086
SMB	0.193	0.205	0.169	0.166
HML	0.271	0.232	0.202	0.200
RMW	0.252	0.212	0.177	0.177
CMA	0.299	0.253	0.180	0.177

Figure 9: OLS Beta Histograms

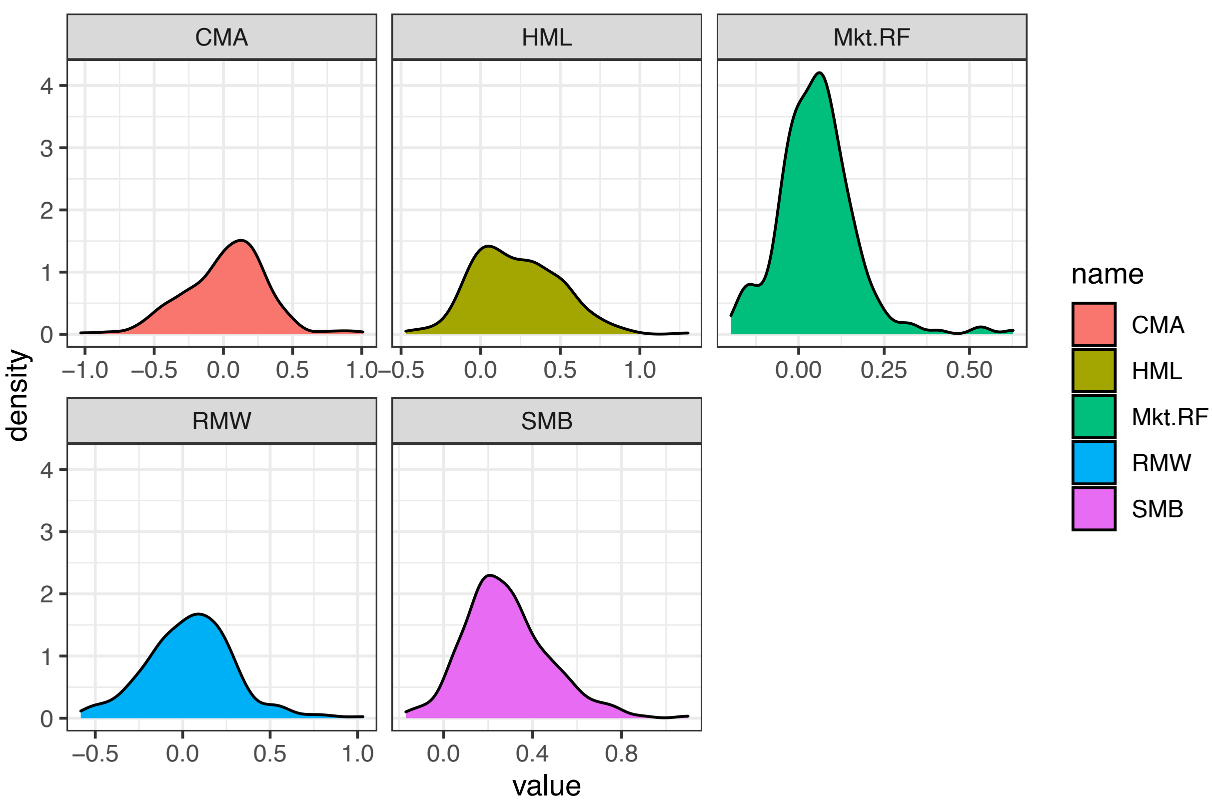


Table 11 contains the standard error ratios from the second phase of the Fama-Macbeth procedure. Envelopes are substantially more efficient for all of the factors, and LRT produces the biggest gains in efficiency. OLS standard errors for the $\beta_{(R_M - R_f)}$ variable are 50 times larger than the envelope standard errors. The results from the simulation study suggest envelopes have comparable levels of bias to that of OLS, indicating this a pure gain in precision.

Table 11: Stage 2 Average OLS to Envelope Standard Error Ratio

	Mkt.RF	SMB	HML	RMW	CMA
AIC	13.008	2.211	3.082	2.452	2.416
BIC	40.954	3.758	3.879	5.076	4.986
LRT	51.107	4.024	4.146	5.356	5.309

The estimated loadings can be found in Table 12. The most important takeaway is the difference in magnitude and signs of the loadings resulting from OLS and envelope methods. Interestingly, the loadings from the envelope methods align more closely to those estimated by Fama and French (2015), despite using substantially fewer observations than the authors used. This demonstrates the power of predictor envelopes to improve a widely used estimation technique. By adopting a new estimation procedure, practitioners have the potential to reach conclusions that otherwise may not be feasible with the available data and resources.

Table 12: Loadings

	Mkt.RF	SMB	HML	RMW	CMA
OLS	0.451	0.311	0.406	0.077	0.233
AIC	1.029	0.329	0.141	-0.057	0.064
BIC	1.678	0.858	-0.163	-0.261	-0.334
LRT	1.701	1.036	-0.339	-0.361	-0.333

6. Summary

When predictors are normally distributed, the most important factors are the variance of the predictors and residuals. When there is low variability in the predictors and high variability in the response, envelopes can offer significant decreases in MSE over OLS. Envelopes can also offer advantages in small sample sizes with high response variance. In all these scenarios with normally distributed predictors, Likelihood Ratio Testing results in the lowest MSE.

Envelopes can result in large decreases in MSE when some predictors do not follow a normal distribution. The greatest gains result when predictors follow a beta or gamma distribution, and reductions in MSE still occur for predictors from the t distribution. Envelopes continue to offer gains for these predictors despite changes in the variability of the response and the variability in the predictors. Likelihood Ratio Testing is also the best dimension selection criterion when predictors are non-normal.

Predictor envelopes have the potential to improve the estimation accuracy of existing research methods. When applying predictor envelopes to Fama-Macbeth regression, envelope standard errors were 50 times smaller than OLS for some variables. This allowed a smaller sample size and shorter observation period to arrive at similar estimates from previous research. Moreover, the simple exploratory analysis aided in intuition as to why envelopes result in dramatic reductions in estimated standard errors.

References

Adragni, K., and R. D. Cook. 2009. "Sufficient Dimension Reduction and Prediction in Regression." *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, November, 4385–4485.

Cook, Li, R. D., and F. Chiaromonte. 2010. "Envelope Models for Parsimonious and Efficient Multivariate Linear Regression." *Statistica Sinica* 20: 927–1010.

Cook, R. D. 2018. *INTRODUCTION to Envelope Models and Methods: Dimension Reduction for Efficient Estimation in Multivariate Statistics*. WILEY-BLACKWELL.

Cook, R. D., B. Li, and F. Chiaromonte. 2007. "Dimension Reduction in Regression Without Matrix Inversion." *Biometrika* 94 (3): 569–84.

Cook, R. D., and X. Zhang. 2015a. "Simultaneous Envelopes for Multivariate Linear Regression." *Technometrics* 57 (1): 11–25.

———. 2015b. "Foundations for Envelope Models and Methods." *Journal of the American Statistical Association* 110 (510): 599–611.

———. 2016. "Algorithms for Envelope Estimation." *Journal of Computational and Graphical Statistics*, no. 1: 284–300.

Ding, S., and R. D. Cook. 2017. "Matrix Variate Regressions and Envelope Models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80 (2): 387–408.

Fama, E. F., and K. R. French. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116 (1): 1–22.

Khare, K., S. Pal, and Z. Su. 2017. "A Bayesian Approach for Envelope Models." *The Annals of Statistics* 45 (1): 196–222.

R. D. Cook, I. S. Helland, and Z. Su. 2013. "Envelopes and Partial Least Squares Regression."

Royal Statistical Society, no. 5: 851–77.

Su, Z., and R. D. Cook. 2011. "Partial Envelopes for Efficient Estimation in Multivariate Linear Regression." *Biometrika* 98 (1): 133–46.

———. 2012. "Inner Envelopes: Efficient Estimation in Multivariate Linear Regression." *Biometrika* 99 (3): 687–702.

Wang, L., and S. Ding. 2018. "Vector Autoregression and Envelope Model." *Stat* 7 (1).

Zhang, X., and R. D. Cook. 2018. "Fast Envelope Algorithms." *Statistica Sinica*.

Appendix

Phase 1 Supplemental Tables

Table 13: Effects of Response Variation

	High Response Variation			Low Response Variation		
Method	MSE	Bias	Variance	MSE	Bias	Variance
AIC	1.674	0.000	1.675	0.446	-0.001	0.446
BIC	1.503	0.002	1.503	0.428	-0.001	0.428
LRT	0.905	0.000	0.905	0.372	-0.001	0.372
OLS	1.861	0.000	1.861	0.465	-0.001	0.465

Table 14: Effects of Sample Size

Sample_size	Method	Bias	MSE	Variance
25	AIC	0.002	4.837	4.837
	BIC	0.004	4.443	4.443
	LRT	0.002	1.914	1.914
	OLS	0.006	5.518	5.519
75	AIC	0.001	1.644	1.644
	BIC	0.002	1.467	1.467
	LRT	-0.001	1.060	1.060
	OLS	0.000	1.803	1.803
250	AIC	-0.001	0.508	0.508
	BIC	-0.001	0.467	0.467
	LRT	-0.001	0.385	0.385
	OLS	-0.001	0.535	0.535

Table 15: Effects of Predictor Variance

Predictor_Variance	Method	MSE	Bias	Variance
high	AIC	0.266	0.001	0.266
	BIC	0.268	0.001	0.268
	LRT	0.265	0.001	0.265
	OLS	0.264	0.001	0.264
low	AIC	1.854	-0.002	1.854
	BIC	1.663	0.000	1.663
	LRT	1.012	-0.002	1.012
	OLS	2.062	-0.002	2.062

Table 16: Effect of Collinearity

Method	Correlation	Bias	MSE	Variance
AIC	low	0.000	1.051	1.052
	medium	0.000	1.072	1.071
	high	-0.001	1.058	1.058
BIC	low	0.001	0.956	0.956
	medium	0.001	0.979	0.979
	high	-0.001	0.962	0.962
LRT	low	-0.001	0.636	0.636
	medium	-0.001	0.641	0.641
	high	0.001	0.639	0.639
OLS	low	0.000	1.153	1.154
	medium	0.000	1.168	1.168
	high	-0.001	1.167	1.167

Phase 2 Supplemental Tables

Table 17: Effects of Response Variation

Response_Variation	Variable	MSE	Bias	Variance
high	Normal*	1.135	-0.001	1.135
	Gamma	31.757	-0.028	31.755
	Beta	82.668	0.011	82.672
	Logistic	0.474	-0.002	0.474
	t	1.391	-0.001	1.391
low	Normal*	1.145	-0.001	1.145
	Gamma	31.694	0.013	31.697
	Beta	84.467	-0.050	84.474
	Logistic	0.469	0.001	0.469
	t	0.721	0.002	0.722

Table 18: Performance for the Gamma Predictor

Predictor_Variance	Method	MSE	Bias	Variance
high	AIC	33.589	-0.010	33.570
	BIC	14.973	-0.001	14.967
	LRT	4.451	0.000	4.451
	OLS	57.394	-0.040	57.368
low	AIC	43.729	-0.004	43.754
	BIC	19.797	0.001	19.803
	LRT	5.655	0.005	5.655
	OLS	74.217	-0.011	74.238

Table 19: Performance for the Beta Predictor

Predictor_Variance	Method	MSE	Bias	Variance
high	AIC	120.363	-0.065	120.352
	BIC	51.142	-0.032	51.149
	LRT	12.124	-0.003	12.124
	OLS	203.168	-0.035	203.189
low	AIC	87.839	-0.005	87.846
	BIC	38.904	-0.027	38.899
	LRT	10.147	-0.023	10.149
	OLS	144.854	0.035	144.874

Table 20: Performance for the Logistic Predictor

Predictor_Variance	Method	MSE	Bias	Variance
high	AIC	0.454	-0.002	0.453
	BIC	0.477	-0.001	0.477
	LRT	0.526	-0.004	0.526
	OLS	0.436	-0.003	0.436
low	AIC	0.451	0.004	0.451
	BIC	0.471	0.002	0.472
	LRT	0.524	0.004	0.523
	OLS	0.433	0.001	0.433

Table 21: Performance for the t Predictor

Degrees_of_Freedom	Method	MSE	Bias	Variance
1	AIC	1.156	0.004	1.157
	BIC	0.990	0.003	0.990
	LRT	0.814	0.003	0.814
	OLS	1.288	0.003	1.288
20	AIC	1.141	-0.007	1.141
	BIC	0.972	0.001	0.972
	LRT	0.793	0.005	0.793
	OLS	1.297	-0.008	1.297