# General model-free weighted envelope estimation

Daniel J. Eck

*Department of Statistics, University of Illinois*

dje13@illinois.edu

**Abstract:** Envelope methodology is succinctly pitched as a class of procedures for increasing efficiency in multivariate analyses without altering traditional objectives [Cook, 2018, first sentence of page 1]. This description comes with the additional caveat that efficiency gains obtained by envelope methodology are mitigated by model selection volatility to an unknown degree. Recent strides to account for model selection volatility have been made on two fronts: 1) development of a weighted envelope estimator to account for this variability directly in the context of the multivariate linear regression model; 2) development of model selection criteria that facilitate consistent dimension selection for more general settings. We unify these two directions and provide weighted envelope estimators that directly account for the variability associated with model selection and are appropriate for general multivariate estimation settings. Our weighted estimation technique provides practitioners with robust and useful variance reduction in finite samples. Theoretical and empirical justification is given for our estimators and validity of a nonparametric bootstrap procedure for estimating their asymptotic variance are established. Simulation studies and a real data analysis support our claims and demonstrate the advantage of our weighted envelope estimator when model selection variability is present.

**Keywords**: Bootstrap smoothing, Dimension reduction, Model averaging, Model selection, Nonparametric bootstrap

## 1 Introduction

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an independent sample where $\theta \in \mathbb{R}^p$ is a target parameter that we want to estimate. Suppose that $\widetilde{\theta} = \widetilde{\theta}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is a $\sqrt{n}$-consistent and asymptotically normal estimator of $\theta$ with asymptotic variance $\Sigma > 0$ such that

$$\sqrt{n}\left(\widetilde{\theta} - \theta\right) \xrightarrow{d} N(0, \Sigma). \tag{1}$$

The goal of envelope methodology is to consistently estimate $\theta$ with as little variance as possible. This is achieved by exploiting a parametric link between $\theta$ and $\Sigma$ in which only a part of $\Sigma$ is relevant for the estimation of $\theta$ [Cook et al., 2010, Cook and Zhang, 2015, Cook, 2018]. Envelope methodology originated as a method to reduce the variability of a regression coefficient matrix $\beta$ in the multivariate linear regression model [Cook et al., 2010, Su and Cook, 2011, 2012, Cook and Su, 2013, Cook, 2018]. The key insight behind envelope methodology as a variance reduction tool was the observation that some linear combinations of the response vector may be invariant to changes in the predictors. Such linear combinations represent variability in the response vector that is not directly relevant to the estimation of $\beta$ and should be discarded. Cook and Zhang [2015] extended envelope methodology to the general setting where one only has a target parameter $\theta$, a $\sqrt{n}$ consistent and asymptotically normal estimator of $\theta$ as in (1), and a $\sqrt{n}$ consistent estimator $\widehat{\Sigma}$ of $\Sigma$.

In both the multivariate linear regression model and the general estimation framework (1), variance reduction obtained through envelope methodology arises from exploiting a subspace of the spectral structure of $\Sigma$ with dimension $u < p$ that contains span$(\theta)$. The dimension of the envelope space $u$ is unknown in practice. In many envelope modeling contexts one can estimate $u$ with information criteria, likelihood ratio tests, or cross-validation. Zhang and Mai [2018] proposed new model-free information criteria that can estimate $u$ consistently. With $u$ estimated, the variability of the envelope estimator is assessed via the bootstrap. However, most bootstrap implementations are conditional on $u = \hat{u}$ where $\hat{u}$ is the estimated dimension of the envelope space. These procedures ignore the variability associated with model selection. Eck and Cook [2017] provided a weighted envelope bootstrap to alleviate this problem in the context of the multivariate linear regression model. In this context the variability of the weighted envelope estimator was appreciably lower than that obtained by bootstrapping the multivariate linear regression model parameters as in Eck [2018]. Eck et al. [2020] provided a double bootstrap procedure for envelope estimation of expected Darwinian fitness from an aster model [Geyer et al., 2007, Shaw et al., 2008] which demonstrated useful variance reduction empirically. That being said, the theoretical motivations for each of these bootstrap procedures are not applicable for envelope estimation in general settings. The weights in Eck and Cook [2017] are constructed from the Bayesian Information Criterion (BIC) values of the multivariate linear regression model evaluated at all envelope estimators fit at dimension $u = 1$, $\ldots$, $p$. Model selection volatility is taken into account in Eck et al. [2020] by criteria that also require a likelihood.

In this paper we provide weighted envelope estimators that are appropriate for envelope estimation in the general setting (1). We will not require the existence of a likelihood function as in Eck and Cook [2017] and we do not require conditioning on an estimated envelope dimension to obtain inference. We then provide bootstrap procedures which estimate the variability of our weighted envelope estimators. More importantly, our bootstrap procedures target the variability of the envelope estimator at the true, unknown, dimension $u$ with an additional minor cost associated with the variability in the weights. As $n \to \infty$ this cost disappears. This is because the weight at the true unknown dimension $u$ converges to $1$ fast enough to not incorporate influences from other envelope dimensions. In finite samples, our weighted envelope estimators are robust to model misspecification associated with not knowing $u$. Our methodology unifies the separate methodologies proposed in Eck and Cook [2017] and Zhang and Mai [2018].

A double bootstrap procedure that incorporate variability in model selection was proposed by Efron [2014]. This procedure is applicable for exponential families, and it has been applied to envelope methodology within this context [Eck et al., 2020, 2018]. Neither Efron [2014] or Eck et al. [2020] provided asymptotic justification for the bootstrap procedures that are implemented within. We provide formal justification for the bootstrap procedures that are developed within. Moreover, our weighted envelope estimators are appropriate for a more general class of envelope models than either Efron [2014] or Eck et al. [2020] can claim. The weighted envelope estimation methodology that we develop in this paper extends to partial envelopes [Su and Cook, 2011], inner envelopes [Su and Cook, 2012], scaled envelopes [Cook and Su, 2013], predictor envelopes [Cook and Su, 2016], sparse response envelopes [Su et al., 2016], tensor response regression [Li and Zhang, 2017], matrix-variate response regression [Ding and Cook, 2018a,b], and envelope models with nonlinearity and heteroscedasticity [Zhang et al., 2020]. We now motivate envelope methodology and weighted estimation techniques.

# 2 Envelope preliminaries

We first provide the definition of a reducing subspace and an envelope.

**Definition 1** (Reducing subspace)**.** *A subspace $\mathcal{R} \subset \mathbb{R}^p$ is a reducing subspace of a matrix $M$ if $M\mathcal{R} \subset \mathcal{R}$ and $M\mathcal{R}^c \subset \mathcal{R}^c$ where $\mathcal{R}^c$ is the orthogonal complement of $\mathcal{R}$ relative to the usual inner product.*

A reducing subspace $\mathcal{R}$ of a matrix $M$ allows one to decompose $M$ as $M = P_{\mathcal{R}}MP_{\mathcal{R}} + Q_{\mathcal{R}}MQ_{\mathcal{R}}$ where $P_{\mathcal{R}}$ is the projection into $\mathcal{R}$ and $Q_{\mathcal{R}} = I - P_{\mathcal{R}}$. When the eigenvalues of $M$ are distinct, a reducing subspace is a direct sum of eigenspaces of $M$.

**Definition 2** (Envelope)**.** *The $M$ envelope of span($U$) is defined as the intersection of all reducing subspaces $\mathcal{R}$ of $M$ which satisfies span$(U) \subseteq \mathcal{R}$. The envelope subspace is denoted by $\mathcal{E}_M(U)$.*

The subspace $\mathcal{E}_M(U)$ is a targeted part of the spectral structure of $M$ which contains the span$(U)$. Let $u$ be the dimension of $\mathcal{E}_M(U)$, where $0 < u \leq p$. In practical settings, all envelope modeling quantities, including $u$, require estimation.

In general settings where the likelihood function is not known, Zhang and Mai [2018] proposed to estimate a semi-orthogonal basis matrix $\Gamma \in \mathbb{R}^{p \times u}$ of $\mathcal{E}_M(U)$ by minimizing the generic moment-based objective function:

$$J_n(\Gamma) = \log \mid \Gamma^T \widehat{M} \Gamma \mid + \log \mid \Gamma^T (\widehat{M} + \widehat{U})^{-1} \Gamma \mid, \tag{2}$$

where $\widehat{M}$ and $\widehat{U}$ are $\sqrt{n}$-consistent estimators of $M$ and $U$. We denote $\widehat{\Gamma}$ as the minimizer of (2). The motivations for (2) comes from its population counterpart, $J(\Gamma) = \log \mid \Gamma^T M \Gamma \mid + \log \mid \Gamma^T (M + U)^{-1} \Gamma \mid$, where Cook and Zhang [2016] showed that any $\Gamma$ which minimizes $J(\Gamma)$ must satisfy the envelope condition that span$(U) \subseteq$ span$(\Gamma)$. We thus take $\Gamma$ as a semi-orthogonal basis for the envelope space, where $\Gamma$ is also a basis for elements in a Grassmann manifold [Cook and Zhang, 2016]. We will therefore refer to the minimization of (12) as Full Grassmannian (FG) optimization.

Assuming that the true envelope dimension $u$ is supplied, the estimated projection $\widehat{\Gamma}\widehat{\Gamma}^T$ into the envelope space is $\sqrt{n}$-consistent [Cook and Zhang, 2016, Proposition 3]. This motivates the envelope estimator $\widehat{\theta}^{\text{FG}} = \widehat{\Gamma}\widehat{\Gamma}^T \widetilde{\theta}$ where $\widetilde{\theta}$ is an estimator satisfying (1). We will also consider envelope estimators constructed with a sequential one-dimensional (1D) coordinate optimization algorithm [Cook and Zhang, 2016, 2015], see Section 3.2 for details on the 1D algorithm. In the next section we motivate general weighted envelope estimation with respect to an envelope estimator $\hat{\theta}$ computed using FG or 1D optimizations, or any other valid envelope optimization routine [Cook et al., 2016, Lee and Su, 2019, Wang et al., 2019]. We will then discuss theoretical properties and bootstrap routines for weighted envelope estimators computed using FG or 1D optimizations.

We note that in this section we motivate $\hat{\theta}$ with the true dimension specified. Therefore, we will often write $\hat{\theta}_u$ to emphasize this point. We will also write $\hat{\theta}_k$ as the envelope estimator at dimension $k$.

# 3 Weighted envelope methodology

We introduce model-free weighted envelope estimation that offers a balance between variance reduction and model misspecification in finite samples. The weighted envelope estimators that we propose are of the form

$$\hat{\theta}_w = \sum_{k=1}^{p} w_k \hat{\theta}_k, \qquad w_k = f_k(\mathcal{I}_n(1), \ldots, \mathcal{I}_n(p)), \tag{3}$$

where $\hat{\theta}_k$ is an envelope estimator at dimension $k$, $\mathcal{I}_n(k)$ is an information criteria that assess the fit of envelope dimension $k$, and $f_k$ is a function of all information criteria at proposed dimension $k$. We note that we do not consider the case $u = 0$ in (3), this case corresponds to a degenerate problem in which $\theta$ is the zeros vector.

As is standard in model averaging, we require that $f_k$ and $\mathcal{I}_n(k)$ be chosen so that $w_k \geq 0$ and $\sum_k w_k = 1$. However, unlike typical model averaging contexts, $\hat{\theta}_k$ is consistent for all weight choices that satisfy $\sum_{k=u}^p w_k \to 1$ where $w_k \geq 0$ for $k \geq u$. Such weight choices induce a consistent estimator since the envelope estimator $\hat{\theta}_k$ is a consistent estimator for $\theta$ for all $k \geq u$. It is of course more desirable to select $\mathcal{I}_n(k)$ and $f_k$ so that $w_u \to 1$.

We specifically study two specific choices of $\mathcal{I}_n(\cdot)$ and one choice of $f_k$ so that $w_u \to 1$ at a fast enough rate to facilitate reliable estimation of the variability of $\hat{\theta}_u$, $u$ unknown, via a nonparametric bootstrap. The weights will be of the form

$$w_k = \frac{\exp\{-n\mathcal{I}_n(k)\}}{\sum_{j=1}^p \exp\{-n\mathcal{I}_n(j)\}}. \tag{4}$$

The choice of $f_k$ that yields the weights (4) is motivated by Eck and Cook [2017], where in that context $\mathcal{I}_n(k)$ is the BIC value of a multivariate linear regression model with the envelope estimator at dimension $k$ plugged in.

The two choices of $\mathcal{I}_n(k)$ that we study here facilitate weighted envelope estimation within the general envelope estimation context (1). The first choice of $\mathcal{I}_n(k)$, denoted $\mathcal{I}_n^{\text{FG}}(k)$ where the superscript FG denotes envelope estimation with respect to full Grassmannian optimization, allows for consistent estimation of the variability of $\hat{\theta}_u$ using the nonparametric bootstrap. This is achieved by setting $\mathcal{I}_n^{\text{FG}}(k) = J_n(\hat{\Gamma}) + \text{pen}(k)$ and showing that 1) $J_n(\hat{\Gamma})$ can be cast a quasi-likelihood objective function that is optimized via a full Grassmannian envelope optimization routine [Zhang and Mai, 2018]; 2) this partially optimized quasi-likelihood objective function can be cast as an M-estimation problem. The second choice of $\mathcal{I}_n(k)$, denoted $\mathcal{I}_n^{\text{1D}}(k)$, corresponds to optimization via the 1D algorithm, hence the 1D superscript. The choice $\mathcal{I}_n^{\text{1D}}(k)$ does not facilitate the same consistent estimation of the variability of $\hat{\theta}_u$. However, we demonstrate that the choice $\mathcal{I}_n^{\text{1D}}(k)$ allows for reliable estimation of the variability of the envelope estimator that is estimated using the 1D algorithm at the true $u$. Both of our simulations, and the simulations presented in Zhang and Mai [2018] find that the choice of $\mathcal{I}_n^{\text{1D}}(k)$ exhibits greater empirical variance reduction than the choice of $\mathcal{I}_n^{\text{FG}}(k)$. It is also noted that the 1D algorithm is more stable than FG optimization [Wang et al., 2019, Zeng et al., 2020].

### 3.1 Weighted envelope estimation via quasi-likelihood optimization

In this section we construct $\hat{\theta}_w$ in (3) using $\mathcal{I}_n^{\mathrm{FG}}(k)$ as the chosen the information criteria. At candidate dimension $k$, the minimizer $\widehat{\Gamma}_k \in \mathbb{R}^{p\times k}$ of the objective function (2) is the estimated basis of the envelope subspace. After obtaining $\widehat{\Gamma}_k$, the envelope estimator is $\widehat{\theta}_k^{\mathrm{FG}} = \widehat{\Gamma}_k\widehat{\Gamma}_k^T\widetilde{\theta}$. This envelope estimator is the original estimator projected into the estimated envelope subspace at dimension $k$. When the true dimension $u$ is known, then $\widehat{\theta}_u^{\mathrm{FG}}$ is $\sqrt{n}$-consistent and has been shown to have lower variability than $\widetilde{\theta}$ in finite samples [Cook et al., 2010, Cook and Zhang, 2015, Cook, 2018]. The weighted envelope estimator corresponding to $\mathcal{I}_n^{\mathrm{FG}}(k)$ is

$$\widehat{\theta}_w^{\mathrm{FG}} = \sum_{k=1}^{p} w_k^{\mathrm{FG}}\widehat{\theta}_k^{\mathrm{FG}}, \qquad w_k^{\mathrm{FG}} = \frac{\exp\left\{-n\mathcal{I}_n^{\mathrm{FG}}(k)\right\}}{\sum_{j=1}^{p}\exp\left\{-n\mathcal{I}_n^{\mathrm{FG}}(j)\right\}}, \tag{5}$$

In the remaining part of this section we construct $\mathcal{I}_n^{\mathrm{FG}}(k)$ and demonstrate how this information criteria choice supplements Section 4.1 to yield consistent estimation of the variability of $\widehat{\theta}_u^{\mathrm{FG}}$.

Zhang and Mai [2018] showed that optimization of $J_n$ in (2) is the same as optimization of a partially minimized quasi-likelihood function. Define this quasi-likelihood function as,

$$l_n(M,\theta) = \log \mid M \mid + \mathrm{tr}\left[M^{-1}\left\{\widehat{M} + (\widetilde{\theta}-\theta)(\widetilde{\theta}-\theta)^T\right\}\right], \tag{6}$$

and, for some candidate dimension $k = 1, \ldots, p$, define the constraint set for the minimization of (6) to be,

$$\mathcal{A}_k = \left\{(M,\theta) : M = \Gamma\Omega\Gamma^T + \Gamma_o\Omega_o\Gamma_o^T > 0,\ \theta = \Gamma\eta,\ \eta \in \mathbb{R}^k,\ (\Gamma,\Gamma_o)^T(\Gamma,\Gamma_o) = I_p\right\}. \tag{7}$$

Minimization of (6) over the constraint set (7) is the same as minimizing $J_n$ in (2).

**Lemma 1.** *[Zhang and Mai, 2018, Lemma 3.1]. The minimizer of $l_n(M,\theta)$ in* (6) *under the envelope parameterization* (7) *is $\widehat{M} = \widehat{\Gamma}\widehat{\Gamma}^T\widehat{\Omega}\widehat{\Gamma}\widehat{\Gamma}^T + \widehat{\Gamma}_o\widehat{\Gamma}_o^T\widehat{\Omega}_o\widehat{\Gamma}_o\widehat{\Gamma}_o^T$ and $\widehat{\theta} = \widehat{\Gamma}\widehat{\Gamma}^T\widetilde{\theta}$ where $\widehat{\Gamma}$ is the minimizer of the partially optimized objective function $l_n(\Gamma) = \min_{\Omega,\Omega_o,\eta} l_n(\Gamma,\Omega,\Omega_o,\eta) = J_n(\Gamma) + \log \mid \widehat{M} + \widehat{U} \mid + p$ where $\widehat{U} = \widetilde{\theta}\widetilde{\theta}^T$.*

We show that optimization of the quasi-likelihood $l_n(M,\theta)$ can be cast as an M-estimation problem when $\widehat{M} = n^{-1}\sum_{i=1}^{n} g(\mathbf{X}_i,\widetilde{\theta})$ provided that $\widetilde{\theta}$ is an optimal solution to some objective function, as in the same vein as Stefanski and Boos [2002, pg. 29]. The justification that a nonparametric bootstrap procedure will consistently estimate the variability of $\widehat{\theta}_u^{\mathrm{FG}}$ follows from bootstrap theory for M-estimators in Section 2 of Andrews [2002]. Therefore recasting (2) and (6) as an M-estimation problem is an important theoretical consideration for our proposed methodology. The requirement that $\widehat{M} = n^{-1}\sum_{i=1}^n g(\mathbf{X}_i, \widetilde{\theta})$ is mild and it holds in linear regression, maximum likelihood estimation, and M-estimation. Our parameterization of $\widehat{M}$ gives,

$$
\begin{aligned}
l_n(M,\theta) &= \log \mid M \mid + \mathrm{tr}\left[n^{-1}\sum_{j=i}^{n} M^{-1}\left\{g(\mathbf{X}_i,\widetilde{\theta}) + (\widetilde{\theta}-\theta)(\widetilde{\theta}-\theta)^T\right\}\right] \\
&= n^{-1}\sum_{i=1}^{n}\left(\log \mid M \mid + \mathrm{tr}\left[M^{-1}\left\{g(\mathbf{X}_i,\widetilde{\theta}) + (\widetilde{\theta}-\theta)(\widetilde{\theta}-\theta)^T\right\}\right]\right) \\
&= n^{-1}\sum_{i=1}^{n} h(\mathbf{X}_i,\theta,M).
\end{aligned}
$$

Lemma 1 then gives,

$$
l_n(\Gamma) = \min_{\Omega,\Omega_o,\eta} n^{-1}\sum_{i=1}^{n} h(\mathbf{X}_i,\Gamma,\Omega,\Omega_o,\eta) = J_n(\Gamma) + \log \mid \widehat{M} + \widehat{U} \mid + p, \tag{8}
$$

where the minimization takes place over $\mathcal{A}_k$. Now define

$$
\mathcal{I}_n^{\mathrm{FG}}(k) = J_n(\widehat{\Gamma}_k) + \frac{Ck\log(n)}{n}, \qquad (k = 1,\dots p), \tag{9}
$$

as in Zhang and Mai [2018] where $C > 0$ is a constant. The envelope dimension is selected as $\hat{u}_{\mathrm{FG}} = \mathrm{argmin}_{1\leq k\leq p}\mathcal{I}_n(k)$. Theorem 3.1 in Zhang and Mai [2018] showed that $\mathbb{P}(\hat{u}_{\mathrm{FG}} = u) \to 1$ as $n \to \infty$, provided that $C > 0$ and $\widehat{M}$ and $\widehat{U}$ are $\sqrt{n}$-consistent estimators of $M$ and $U$ respectively. We use $\mathcal{I}_n^{\mathrm{FG}}(k)$ in (9) to construct $\widehat{\theta}_w^{\mathrm{FG}}$ in (5). This construction yields consistent estimation of $\theta$ as seen in Section 3.3 and consistent estimation of the variability of $\widehat{\theta}_u^{\mathrm{FG}}$ through the combination of Theorem 1 and our formulation of $l_n(M,\theta)$ as an M-estimation problem. In Section 4.1 we develop a nonparametric bootstrap to consistently estimate the variability of $\widehat{\theta}_u^{\mathrm{FG}}$.

## 3.2 Weighted envelope estimation via the 1D algorithm

In this section we construct $\hat{\theta}_w$ in (3) where the information criteria $\mathcal{I}_n^{\mathrm{1D}}(k)$ is derived from the 1D algorithm. The 1D algorithm performs a sequence of optimizations that each return a basis vector of the envelope space (in the population) or a $\sqrt{n}$-consistent estimator of a basis vector for the envelope space (in finite-samples). The number of optimizations corresponds to the dimension of the envelope space and is provided by the user. The returned envelope estimator is $\widehat{\theta}_u^{\text{1D}} = \widehat{\Gamma}\widehat{\Gamma}\widetilde{\theta}$ where the estimated basis matrix $\widehat{\Gamma}$ is obtained from the 1D algorithm. The weighted envelope estimator corresponding to $\mathcal{I}_n^{\text{1D}}(k)$ is

$$\widehat{\theta}_w^{\text{1D}} = \sum_{k=1}^{p} w_k^{\text{1D}} \widehat{\theta}_k^{\text{1D}}, \qquad w_k^{\text{1D}} = \frac{\exp\left\{-n\mathcal{I}_n^{\text{1D}}(k)\right\}}{\sum_{j=1}^{p} \exp\left\{-n\mathcal{I}_n^{\text{1D}}(j)\right\}}. \tag{10}$$

We briefly state the 1D algorithm: Set $u_o \leq p-1$ to be the user inputted number of optimizations. For step $k = 0, \ldots, u_o$, let $g_k \in \mathbb{R}^p$ denote the $k$-th direction to be obtained by the 1D algorithm. Define $G_k = (g_1, \ldots, g_k)$, and $(G_k, G_{0k})$ to be an orthogonal basis for $\mathbb{R}^p$ and set initial value $g_o = G_0 = 0$. Define $M_k = G_{0k}^T M G_{0k}$, $U_k = G_{0k}^T U G_{0k}$, and the objective function after $k$ steps $\phi_k(v) = \log(v^T M_k v) + \log\{v^T (M_k + U_k)^{-1} v\}$. The $(k+1)$-th envelope direction is $g_{k+1} = G_{0k} v_{k+1}$ where $v_{k+1} = \text{argmax}_v \phi_k(v)$ subject to $v^T v = 1$.

Replacing $M$ and $U$ with $\sqrt{n}$-consistent estimators $\widehat{M}$ and $\widehat{U}$ yields $\sqrt{n}$-consistent estimates $\widehat{G}_k = (\widehat{g}_1, \ldots, \widehat{g}_k) \in \mathbb{R}^{p\times k}$, $k = 1, \ldots, p$ by optimizing $\phi_{k,n}(v) = \log(v^T \widehat{M}_k v) + \log\{v^T(\widehat{M}_k + \widehat{U}_k)^{-1} v\}$. The resulting envelope estimator $\widehat{\theta}_u^{\text{1D}}$ is therefore $\sqrt{n}$-consistent [Cook and Zhang, 2015]. Now define,

$$\mathcal{I}_n^{\text{1D}}(k) = \sum_{j=1}^{k} \phi_{j,n}(\hat{v}_j) + \frac{Ck\log n}{n}, \qquad (k = 1, \ldots, p); \tag{11}$$

where $C > 0$ is a constant. When $C = 1$, the terms $n\mathcal{I}_n^{\text{1D}}(k)$ in (11) are BIC values corresponding to the asymptotic log likelihood the envelope model of dimension $k$. The envelope dimension selected is $\hat{u}_{\text{1D}} = \text{argmin}_{1\leq k\leq p} \mathcal{I}_n^{\text{1D}}(k)$. Theorem 3.2 in Zhang and Mai [2018] showed that $\mathbb{P}(\hat{u}_{\text{1D}} = u) \to 1$ as $n \to \infty$. We use $\mathcal{I}_n^{\text{1D}}(k)$ in (11) to construct $\widehat{\theta}_w^{\text{1D}}$ in (10). This construction yields consistent estimation of $\theta$ as seen in Section 3.3 and allows for reliable estimation of the variability of $\widehat{\theta}_u^{\text{1D}}$.

## 3.3 Consistency properties of weighted envelope estimators

Weighted envelope estimators exhibit desirable consistency properties. First of all, the weights in (5) and (10) can be constructed so that they both satisfy $w_u^{\text{FG}} \to 1$ and $w_u^{\text{1D}} \to 1$ as $n \to \infty$.

**Lemma 2.** *For any constant $C > 0$ and $\sqrt{n}$-consistent $\widehat{M}$ and $\widehat{U}$ in* (9) *and* (11)*, $w_u^{FG} \to 1$ and $w_u^{1D} \to 1$ as $n \to \infty$.*

The proof of Lemma 2 are included in the Supplementary Materials. This lemma facilitates consistent estimation of $\theta$ using $\widehat{\theta}_w^{\text{FG}}$ and $\widehat{\theta}_w^{\text{1D}}$.

**Lemma 3.** *For any constant $C > 0$ and $\sqrt{n}$-consistent $\widehat{M}$ and $\widehat{U}$ in* (9) *and* (11)*, both $\widehat{\theta}_w^{FG} \to \theta$ and $\widehat{\theta}_w^{1D} \to \theta$ as $n \to \infty$.*

The proof of Lemma 3 immediately follows from Lemmas 2 and Zhang and Mai [2018, Proposition 2.1]. While consistency is desirable, Lemma 3 does not provide knowledge about the asymptotic variability of $\widehat{\theta}_w^{\text{FG}}$ or $\widehat{\theta}_w^{\text{1D}}$. We expect that $\widehat{\theta}_w^{\text{FG}}$ and $\widehat{\theta}_w^{\text{1D}}$ will have lower asymptotic variance than $\widetilde{\theta}$ when $u < p$, but explicit computations of the asymptotic variance for both estimators are cumbersome. We will instead estimate the asymptotic variability of $\widehat{\theta}_w^{\text{FG}}$ and $\widehat{\theta}_w^{\text{1D}}$ with a nonparametric bootstrap that is developed in the next section.

# 4 Bootstrapping procedures

Let $\mathbf{X}_1$, ..., $\mathbf{X}_n$ be the original data. We will estimate the variability of $\widehat{\theta}_u^{\text{FG}}$ and $\widehat{\theta}_u^{\text{1D}}$ by bootstrapping with respect to the weighted estimators $\widehat{\theta}_w^{\text{FG}}$ and $\widehat{\theta}_w^{\text{1D}}$ via the nonparametric bootstrap. For each iteration of this nonparametric bootstrap procedure we denote the resampled data by $\mathbf{X}_1^*$, ..., $\mathbf{X}_n^*$ where each $\mathbf{X}_i^*$, $i = 1$, ..., $n$ is sampled, with replacement, from the original data with equal probability $1/n$. At candidate envelope dimension $k$ we define the bootstrapped envelope estimators ${\widehat{\theta}_k^{\text{FG}}}^* = \widehat{\theta}_k^{\text{FG}}\left(\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*\right)$, ${\widehat{\theta}_k^{\text{1D}}}^* = \widehat{\theta}_k^{\text{1D}}\left(\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*\right)$, and the bootstrapped version of the original estimator $\widetilde{\theta}$ as $\widetilde{\theta}^* = \widetilde{\theta}(\mathbf{X}_1^*, \ldots \mathbf{X}_n^*)$. Define $\widehat{M}^*$ and $\widehat{U}^*$ in the same manner as $\widehat{M}$ and $\widehat{U}$ with respect to the starred data. Following Remark 1 in Section 3 of Chang and Park [2003], we define $O_P^*$ as an analog to $O_P$ that is conditional on the original sample.

## 4.1 For quasi-likelihood optimization

In this section we provide justification for the nonparametric bootstrap as a method to estimate the variability of $\widehat{\theta}_w^{\text{FG}}$. Define

$$J_n^*(\Gamma) = \log \mid \Gamma^T \widehat{M}^* \Gamma \mid + \mid \Gamma^T \left(\widehat{M}^* + \widehat{U}^*\right)^{-1} \Gamma \mid, \tag{12}$$

as the starred analog to $J_n$ in (2) and define,

$$l_n^*(M, \theta) = \log \mid M \mid + \text{tr}\left[M^{-1}\left\{\widehat{M}^* + (\widetilde{\theta}^* - \theta)(\widetilde{\theta}^* - \theta)^T\right\}\right], \tag{13}$$

as the starred analog to $l_n$ in (6). Define $\widehat{\Gamma}^*$ as the minimizer to (12). When $\widehat{M}^* = n^{-1}\sum_{i=1}^n h(\mathbf{X}_i^*, \widetilde{\theta}^*)$ both Lemma 1 and our likelihood derivation in Section 3.1 give,

$$l_n^*(\Gamma) = \min_{\Omega,\Omega_o,\eta} n^{-1}\sum_{i=1}^n f(\mathbf{X}_i^*, \Gamma, \Omega, \Omega_o, \eta) = J_n^*(\Gamma) + \log | \widehat{M}^* + \widehat{U}^* | + p,$$

which is the starred analog to (8). Thus $\widehat{\Gamma}^*$ is an M-estimator, being the minimizer of the partially minimized objective function $l_n^*(\Gamma)$. We then let $\widehat{\theta}_u^{\mathrm{FG}*} = \widehat{\Gamma}^*\widehat{\Gamma}^{*T}\widetilde{\theta}^*$ where $\widetilde{\theta}^*$ is obtained an optimal solution to some objective function constructed from starred data, analogous to Stefanski and Boos [2002, pg. 29]. The envelope estimator $\widehat{\theta}_u^{\mathrm{FG}*}$ is then a product of M-estimators obtained from the same objective function. Therefore we can use the nonparametric bootstrap to consistently estimate the variability of $\widehat{\theta}_u^{\mathrm{FG}}$ [Andrews, 2002, Section 2]. We show that bootstrapping with respect to our weighted envelope estimator consistently estimates the variability of the envelope estimator $\widehat{\theta}_u^{\mathrm{FG}}$ at the true unknown dimension when $\widehat{M}^* = n^{-1}\sum_{i=1}^n g(\mathbf{X}_i^*, \widetilde{\theta}^*)$.

**Theorem 1.** *Let $\widetilde{\theta}$ be a $\sqrt{n}$-consistent and asymptotically normal estimator. Let $\widehat{\theta}_k^{FG}$ be the envelope estimator obtained from full Grassmannian optimization at dimension $k = 1$, ..., $p$ and let $\widehat{\theta}_w^{FG}$ be the weighted envelope estimator with weights $w^{FG}$. Let $\widehat{\theta}_k^{FG*}$ and $\widehat{\theta}_w^{FG*}$ denote the corresponding quantities obtained by resampled data. Then as $n$ tends to $\infty$,*

$$\sqrt{n}\left(\widehat{\theta}_w^{FG*} - \widehat{\theta}_w^{FG}\right) = \sqrt{n}\left(\widehat{\theta}_u^{FG*} - \widehat{\theta}_u^{FG}\right) + O_P^*\left\{n^{(1/2-C)}\right\} + O_P^*\left\{n^{(Cu+1/2)}\right\} e^{-n|O_P^*(1)|}. \tag{14}$$

**Remarks**:

1. Theorem 1 shows that our bootstrap procedure consistently estimates the asymptotic variability of $\widehat{\theta}_u^{\mathrm{FG}}$ when $u$ is unknown. We see that the second $O_P^*$ term in (14) vanishes quickly in $n$. These terms are associated with under selecting the true envelope dimension. Therefore it is more likely that our bootstrap procedures will conservatively estimate the variability of $\widehat{\theta}_u^{\mathrm{FG}}$ in finite samples.

2. We advocate for the case with $C = 1$ because of the close connection that $\mathcal{I}_n^{\mathrm{FG}}(k)$ has with BIC, similar reasoning was given in Zhang and Mai [2018]. The $Ck\log(n)/n$ penalty term in $\mathcal{I}_n^{\mathrm{FG}}(k)$ facilitates the decaying bias in $n$ represented by the $O_P^*$ terms in (14). Redefining $\mathcal{I}_n^{\mathrm{FG}}(k)$ to have a penalty term that is fixed in $n$, similar to that of AIC, fundamentally changes the $O_P^*$ terms in (14). Specifically, the $O_P(n^{-1/2})$ term

(when $C = 1$) disappears and the weights $w_k^{\text{FG}}$ fail to vanish for $k > u$. Therefore unknown non-zero asymptotic weight is given to candidate models with dimension $k > u$. Weighting in this manner is therefore suboptimal and is not advised.

3. The weights $w_k^{\text{FG}}$ have a similar form to the weights which appear in the model averaging literature [Buckland et al., 1997, Burnham and Anderson, 2004, Hjort and Claeskens, 2003, Claeskens and Hjort, 2008, Tsague, 2014]. These weights are of the form
$$w_k = \frac{\exp\left\{-n\mathcal{I}_n^{\text{FG}}(k)/2\right\}}{\sum_{j=0}^{p}\exp\left\{-n\mathcal{I}_n^{\text{FG}}(j)/2\right\}} \tag{15}$$
and they correspond to a posterior probability approximation for model $k$ under the prior that assigns equal weight to all candidate models, given the observed data. The weights (15) do not have the same asymptotic properties as our weights. The difference between the two is a rescaling of $C$. Weights (15) replace the constant $C$ in (14) with $C/2$. When $C = 1$, nonzero asymptotic weight would be placed on the envelope model with dimension $k = u + 1$. Therefore, weighting according to (15) leads to higher estimated variability than is necessary.

## 4.2 For the 1D algorithm

In this section we provide justification for the nonparametric bootstrap as a method to estimate the variability of $\widehat{\theta}^{\text{1D}}$. We cannot cast the 1D objective function $\phi_k(v)$ as an M-estimation problem. Instead of showing consistency, we verify that bootstrapping with respect to $\widehat{\theta}_w^{\text{1D}}$ is asymptotically the same as bootstrapping with respect to $\widehat{\theta}_u^{\text{1D}}$. We now define the quantities of the 1D algorithm applied to the starred data. Set $u_o \leq p-1$ to be the user inputted number of optimizations. For step $k = 0, \ldots, u_o$, let $\hat{g}_k^* \in \mathbb{R}^p$ denote the $k$-th direction to be obtained. Define $\widehat{G}_k^* = (\hat{g}_1^*, \ldots, \hat{g}_k^*)$, and $(\widehat{G}_k^*, \widehat{G}_{0k}^*)$ to be an orthogonal basis for $\mathbb{R}^p$ and set initial value $\hat{g}_o^* = \widehat{G}_0^* = 0$. Define $\widehat{M}_k^* = \widehat{G}_{0k}^{*T}\widehat{M}^*\widehat{G}_{0k}^*$, $\widehat{U}_k^* = \widehat{G}_{0k}^{*T}\widehat{U}^*\widehat{G}_{0k}^*$, and the objective function after $k$ steps as $\phi_{k,n}^*(v) = \log(v^T\widehat{M}_k^*v) + \log\{v^T(\widehat{M}_k^* + \widehat{U}_k^*)^{-1}v\}$. The $(k+1)$-th envelope direction is $\hat{g}_{k+1}^* = \widehat{G}_{0k}^* v_{k+1}^*$ where $v_{k+1}^* = \text{argmax}_{v^Tv}\phi_k^*(v)$. After $u_o$ steps we set $\widehat{\Gamma}^* = \widehat{G}_{u_o}^*$ and $\widehat{\theta}_u^{\text{1D}*} = \widehat{\Gamma}^*\widehat{\Gamma}^{*T}\widetilde{\theta}^*$. We show that bootstrapping $\widehat{\theta}_w^{\text{1D}}$ estimates the variability of the envelope estimator $\widehat{\theta}_u^{\text{1D}}$ at the true unknown dimension.

**Theorem 2.** *Let $\widetilde{\theta}$ be a $\sqrt{n}$-consistent and asymptotically normal estimator. Let $\widehat{\theta}_k^{1D}$ be the envelope estimator obtained from the 1D algorithm at dimension $k = 1, \ldots, p$ and let $\widehat{\theta}_w^{1D}$ be the weighted envelope estimator with weights $w^{1D}$. Let $\widehat{\theta}_k^{1D*}$ and $\widehat{\theta}_w^{1D*}$ denote the* *corresponding quantities obtained by resampled data. Then as $n$ tends to $\infty$,*

$$\sqrt{n}\left(\widehat{\theta}_w^{1D*} - \widehat{\theta}_w^{1D}\right) = \sqrt{n}\left(\widehat{\theta}_u^{1D*} - \widehat{\theta}_u^{1D}\right) + O_P^*\left\{n^{(1/2-C)}\right\} + O_P^*\left\{n^{(Cu+1/2)}\right\} e^{-n|O_P^*(1)|}. \tag{16}$$

The remarks to Theorem 2 are similar to those for Theorem 1. The second $O_P$ term in (16) vanishes quickly in $n$. These terms are associated with under selecting the true envelope dimension. Therefore it is more likely that our bootstrap procedures will conservatively estimate the variability of $\widehat{\theta}_u^{\mathrm{1D}}$ in finite samples. We advocate for the case with $C = 1$ because of the close connection that $\mathcal{I}_n^{\mathrm{1D}}(k)$ has with BIC. Manuals for available software recommend use of one-directional optimizations, such as the 1D algorithm or the ECD algorithm [Cook and Zhang, 2018], because they are faster, stable, and less sensitive to initial values [Wang et al., 2019, Zeng et al., 2020].

# 5 Examples

In this section we demonstrate the utility of weighted envelope estimation via the bootstrapping procedures developed in the previous section. We will consider several bootstrap estimators and evaluation metrics that are summarised below:

| Quantity | Description |
|---|---|
| $\widehat{\theta}_w^{\mathrm{FG}*}$ or $\widehat{\theta}_w^{\mathrm{1D}*}$ | The weighted envelope estimator $\widehat{\theta}_w^{\mathrm{FG}}$ or $\widehat{\theta}_w^{\mathrm{1D}}$ computed with respect to resampled data. |
| $\widehat{\theta}_k^{\mathrm{FG}*}$ or $\widehat{\theta}_k^{\mathrm{1D}*}$ | The envelope estimator $\widehat{\theta}_k^{\mathrm{FG}}$ or $\widehat{\theta}_k^{\mathrm{1D}}$ fit at dimension $k$ computed with respect to resampled data. |
| $\widehat{\theta}_{\hat{u}_{\mathrm{FG}}}^{\mathrm{FG}*}$ or $\widehat{\theta}_{\hat{u}_{\mathrm{1D}}}^{\mathrm{1D}*}$ | The envelope estimator $\widehat{\theta}_{\hat{u}_{\mathrm{FG}}}^{\mathrm{FG}}$ or $\widehat{\theta}_{\hat{u}_{\mathrm{1D}}}^{\mathrm{1D}}$ fit at the dimension estimated from the original sample $\hat{u}_{\mathrm{FG}}$ or $\hat{u}_{\mathrm{1D}}$ and is then computed with respect to resampled data. |
| $\widehat{\theta}_{\hat{u}_{\mathrm{FG}}^*}^{\mathrm{FG}*}$ or $\widehat{\theta}_{\hat{u}_{\mathrm{1D}}^*}^{\mathrm{1D}*}$ | The envelope estimator $\widehat{\theta}_{\hat{u}_{\mathrm{FG}}}^{\mathrm{FG}}$ or $\widehat{\theta}_{\hat{u}_{\mathrm{1D}}}^{\mathrm{1D}}$ fit at the dimension estimated from resampled data $\hat{u}_{\mathrm{FG}^*}$ or $\hat{u}_{\mathrm{1D}^*}$ and is then also computed with respect to resampled data. |
| $r(\widehat{\theta}_1, \widehat{\theta}_2)$ | A column vector containing ratios of bootstrapped standard errors $\mathrm{se}^*(\widehat{\theta_{1,j}})/\mathrm{se}^*(\widehat{\theta_{2,j}})$, $j = 1, \ldots, p$. |

## 5.1 Exponential family GLM simulations

We demonstrate our model free weighted envelope estimation techniques for exponential family glms using normal predictors as justified in Cook and Zhang [2015]. In this setting envelope estimators are not parameterized within the likelihood function as is typically the case in multivariate linear regression. Thus the only part of "model" that is used is the MLE and Fisher information matrix so that (1) holds. In these simulations we demonstrate that useful variance reduction is obtained while accounting for model selection variability. Estimation is performed using functionality in the `TRES` R package [Wang et al., 2019].

We simulate two different exponential family GLM settings where $p = 6$ and $u = 3$. One setting is for a logistic regression model and the other is for a Poisson regression model. Predictors are generated $\mathbf{X} \sim N(0, \Sigma_{\mathbf{X}})$, where $\Sigma_{\mathbf{X}} = \Gamma\Omega\Gamma^T + \Gamma_o\Omega_o\Gamma_o^T$. We construct the regression coefficient vector as $\theta = \Gamma\Gamma^T v$ where $\Gamma$ and $v$ are provided in the supplement materials. In the logistic regression simulations we generate $Y_i \sim \text{Bernoulli}(\text{logit}^{-1}(\theta^T\mathbf{X}_i))$, and in the Poisson regression simulations we generate $Y_i \sim \text{Poisson}(\exp(\theta^T\mathbf{X}_i))$. Our simulation settings are:

> Logistic regression: $\Omega$ has diagonal elements $0.5$, $1$, and $1.5$, $\Omega_o$ has diagonal elements $\exp(-3)$, $\exp(-2)$, and $\exp(1)$. We construct $\theta = -v_1/6 + 2v_2/3 + 2v_3/3$ where $v_j$ are the $j$th eigenvector of the envelope basis matrix $\Gamma$.
>
> Poisson regression: $\Omega$ has diagonal elements $\exp(0)$, $\exp(1)$ and $\exp(2)$, and $\Omega_o$ has diagonal elements $\exp(-4)$, $\exp(-3)$, and $\exp(-1)$. We construct $\theta = -v_1/2 - v_2/2 - v_3/10$ where $v_j$ are the $j$th eigenvector of the envelope basis matrix $\Gamma$.

Revised up to Here

Ratios of bootstrap standard deviations for estimators of the $j$th component of the canonical parameter vector across all simulation settings are depicted in Table 1. These ratios are of the form $r(\widetilde{\theta}, \widehat{\theta}_w^{\text{1D}}) = \widehat{\text{sd}}^*(\widetilde{\theta})/\widehat{\text{sd}}^*(\widehat{\theta}_w^{\text{1D}})$. The standard deviations $\widehat{\text{sd}}^*(\widetilde{\theta})$ and $\widehat{\text{sd}}^*(\widehat{\theta}_w^{\text{1D}})$ are the element in the first row and the first column of

$$\left(\frac{1}{B}\sum_{b=1}^{B}(\widetilde{\theta}_b^* - \widetilde{\theta})(\widetilde{\theta}_b^* - \widetilde{\theta})^T\right)^{1/2}, \qquad \left(\frac{1}{B}\sum_{b=1}^{B}(\widehat{\theta}_{w_b}^{\text{1D}*} - \widehat{\theta}_w^{\text{1D}})(\widehat{\theta}_{w_b}^{\text{1D}*} - \widehat{\theta}_w^{\text{1D}})^T\right)^{1/2};$$

respectively. When $r(\widetilde{\theta}, \widehat{\theta}_w^{\text{1D}}) > 1$ then variance reduction is obtained using $\widehat{\theta}_w^{\text{1D}}$ in place of the standard estimator.

In our simulations we find that envelope estimation provides variance reduction in all settings. However, the fixed $u$ regime exhibits erratic performance across $n$. This is due to the large variability in the estimated dimension across $n$, details of which are in the Supplementary Materials. Weighted envelope estimation exhibits a noticeable advantage to estimation under the variable $u$ regime. These simulations demonstrate the utility of weighted envelope estimation in the presence of model selection variability. Similar results are observed for other components of the regression coefficient vector as seen in the Supplementary Materials.

Following the recommendations in the TRES manuals, we use the 1D algorithm for estimation. We will therefore compare the performance of $\widehat{\theta}_w^{\mathrm{1D}}$ to $\widehat{\theta}_{\hat{u}_{\mathrm{1D}}}^{\mathrm{1D}}$, the envelope estimator of $\theta$ evaluated at estimated dimension $\hat{u}_{\mathrm{1D}}$. A nonparametric bootstrap with sample size $B = 5000$ is then used to estimate the variability of these estimators. Our bootstrap simulation will consider two model selection regimes for obtaining $\hat{u}_{\mathrm{1D}}$. In one regime, we estimate the envelope dimension at every iteration of the bootstrap (variable $u$ regime, estimated dimension denoted as $\hat{u}_{\mathrm{1D}}^*$). In the other regime, we estimate the dimension of the envelope space in the original data set and then condition on this estimated dimension as if it were the true dimension (fixed $u$ regime, estimated dimension denoted as $\hat{u}_{\mathrm{1D}}$). The fixed $u$ regime ignores the variability associated with model selection. Theorem 2 provides some guidance for the performance of the bootstrapping $\widehat{\theta}_w^{\mathrm{1D}}$. A formal analog does not exist for the other envelope estimators. Empirical evidence in Zhang and Mai [2018] and Eck et al. [2020] suggest that the variable $u$ regime will provide some robustness to variability in dimension selection.

## 5.2 Real data illustration

We examine the influence of several variables on a positive diagnosis of diabetes. We will let a positive diagnosis of diabetes be when an individual's hemoglobin percentage (also known as HbA1c) exceeds a value of 6.5% [World Health Organization, 2011]. We will consider an individual's height, weight, age, hip size, waist size, and gender, all of which are easy to measure, inexpensive, and do not require any laboratory testing, and a measure of their stabilized glucose as predictors for a positive diagnosis of diabetes. The data in this analysis come from a population-based sample of 403 rural African-Americans in Virginia [Willems et al., 1997], and is taken from the faraway R package [Faraway, 2016]. We considered a logistic regression model with response variable denoting a diagnosis of diabetes (1 when HbA1c $> 6.5\%$ and $0$ otherwise) that includes log transformed values for each continuous covariate and a main effect for gender. The log transformation was used to transform these variables to univariate normality while maintaining a scale that is interpretable.

Model free envelope estimation techniques and maximum likelihood estimation are then used to estimate the canonical parameter vector corresponding to this logistic regres-

Table 1: Ratios of standard deviations for envelope estimators relative to the MLE. The first and fourth ratio columns display the ratio of the bootstrap standard deviation of the MLE to that of the weighted envelope estimator. The second and fifth ratio columns display the ratio of the bootstrap standard deviation of the MLE to that of the envelope estimator under the variable $u$ dimension selection regime. The third and sixth ratio columns display the ratio of the bootstrap standard deviation of the MLE to that of the envelope estimator under the fixed $u$ dimension selection regime.

| | | Setting A | | | Setting B | | |
|---|---|---|---|---|---|---|---|
| model | $n$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{w})$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{\hat{u}^*_{1\mathrm{D}}})$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}})$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{w})$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{\hat{u}^*_{1\mathrm{D}}})$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}})$ |
| | 300 | 1.22 | 1.16 | 3.26 | 0.99 | 0.97 | 1.13 |
| | 500 | 1.88 | 1.76 | 2.88 | 1.04 | 1.02 | 1.14 |
| Logistic | 750 | 1.03 | 0.95 | 3.64 | 1.29 | 1.27 | 1.48 |
| | 1000 | 1.00 | 0.88 | 3.50 | 1.04 | 1.02 | 1.14 |
| | 300 | 1.15 | 1.11 | 2.09 | 1.15 | 1.13 | 1.41 |
| | 500 | 2.58 | 2.29 | 17.29 | 1.35 | 1.21 | 2.38 |
| Poisson | 750 | 3.81 | 3.48 | 61.88 | 1.12 | 1.10 | 1.17 |
| | 1000 | 4.09 | 3.62 | 93.11 | 1.58 | 1.49 | 2.85 |

sion model. Estimation is performed using functionality in the `TRES` R package [Wang et al., 2019]. We will compare the performance of $\widehat{\theta}^{1\mathrm{D}}_{w}$ to $\widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}}$. A nonparametric bootstrap with sample size $5000$ is then used to estimate the variability of these estimators. Our bootstrap simulation will consider both the variable $u$ and fixed $u$ model selection regimes. Performance results are displayed in Table 2. We see that $\widehat{\theta}^{1\mathrm{D}}_{w}$ and $\widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}}$ are very similar to each other and both are very different than the MLE $\widetilde{\theta}$. Similarity of $\widehat{\theta}^{1\mathrm{D}}_{w}$ and $\widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}}$ follows from empirical weights $w_1 = 0.982$, $w_2 = 0.0176$, and $w_k \approx 0$ for all $3 \leq k \leq 7$. Also observe that the bootstrap standard deviation estimates vary across the model selection procedures. Most notably, the fixed $u$ regime provides massive variance reduction while the weighted estimator and variable $u$ regime provide similar modest but appreciable variance reduction. The variance reduction discrepancy between the fixed $u$ regime and the the weighted estimator and variable $u$ regime is due to large model selection variability. Specifically, the selected dimension probabilities across our nonparametric bootstrap procedure are $p(\hat{u}_{1\mathrm{D}} = 1) = 0.568$, $p(\hat{u}_{1\mathrm{D}} = 2) = 0.358$, $p(\hat{u}_{1\mathrm{D}} = 3) = 0.067$, and $p(\hat{u}_{1\mathrm{D}} = 4) = 0.007$. It is clear that unaccounted model selection variability may lead users astray when they use the fixed $u$ regime in estimating standard deviations via bootstrapping. This example shows how difficult it can be to report reliable variance reduction in practice, and how tempting it can be to ignore model selection variability.

Table 2: Coefficient estimates for the logistic regression of diabetes diagnosis on seven predictors. Columns 1-3 display the weighted envelope estimator, the envelope estimator with $\hat{u}_{1\mathrm{D}} = 1$, and the MLE respectively. Columns 4-6 display the ratio of bootstrap standard deviations of all envelope estimators to the those of the MLE.

| | $\widehat{\theta}^{1\mathrm{D}}_w$ | $\widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}}$ | $\widetilde{\theta}$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_w)$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{\hat{u}^*_{1\mathrm{D}}})$ | $r(\widetilde{\theta}, \widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}})$ |
|---|---|---|---|---|---|---|
| $\log($Age$)$ | 1.78 | 1.78 | 2.03 | 1.24 | 1.21 | 1.19 |
| $\log($Weight$)$ | 0.70 | 0.70 | 1.26 | 1.58 | 1.48 | 7.07 |
| $\log($Height$)$ | 0.03 | 0.03 | -4.39 | 1.16 | 1.07 | 54.76 |
| $\log($Waist$)$ | 0.68 | 0.68 | 2.65 | 1.54 | 1.45 | 11.60 |
| $\log($Hip$)$ | 0.49 | 0.49 | -2.64 | 1.30 | 1.22 | 17.80 |
| Female | 0.40 | 0.40 | 0.17 | 1.19 | 1.17 | 1.20 |
| $\log($Stab. Gluc.$)$ | 5.11 | 5.11 | 5.03 | 1.15 | 1.15 | 1.29 |

## 5.3 Replicating Zhang and Mai [2018]

Here, we compare the performance of $\widehat{\theta}^{\mathrm{FG}}_w$ and $\widehat{\theta}^{1\mathrm{D}}_w$ to the consistent envelope estimators $\widehat{\theta}^{\mathrm{FG}}_{\hat{u}_{\mathrm{FG}}}$ and $\widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}}$ using the simulation settings in Zhang and Mai [2018]. For our first comparison we reproduce the Monte Carlo simulations in Section 4.2 of Zhang and Mai [2018] and add both $\widehat{\theta}^{\mathrm{FG}}_w$ and $\widehat{\theta}^{1\mathrm{D}}_w$ to the list of estimators under comparison. Performance of all estimators at a sample size of $n = 75$ is also assessed. The data generating models that are considered are a single predictor linear regression model with 10 responses, a logistic regression model with a 10 predictors, and a Cox proportional hazards model with 10 predictors. In all three modeling setups, the true dimension of the envelope space is set at $u = 2$. In-depth details about this simulation setup are presented in Zhang and Mai [2018]. It is important to note $\widehat{\theta}^{\mathrm{FG}}_{\hat{u}_{\mathrm{FG}}}$ and $\widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}}$ are estimated according to the variable $u$ regime. The Monte Carlo sample size is 200, as in Zhang and Mai [2018].

Table 3 displays the results. From Table 3 we see that the weighted envelope estimators perform very similarly to the consistent envelope estimators. This suggests that the variability in model selection is captured by all envelope estimators. This finding is expected in larger samples when the correct dimension selected percentage approaches 1, and it is a direct consequence of Lemma 3 and Theorem 3.2 in Zhang and Mai [2018]. On the other hand, this finding is illuminating for sample sizes where the correct dimension selected percentages are nowhere near 1. Some variability in selection of $u$ which was used to construct both $\widehat{\theta}^{\mathrm{FG}}_{\hat{u}_{\mathrm{FG}}}$ and $\widehat{\theta}^{1\mathrm{D}}_{\hat{u}_{1\mathrm{D}}}$ is incorporated into these simulations since $u$ is estimated at every iteration.

We now estimate the variability of envelope estimators under the simulation settings in

Table 3: Monte Carlo simulation results for different envelope estimators with respect to three different envelope models in the spirit of Table 3 from Zhang and Mai [2018]. Left panel includes percentages of correct selection for these envelope estimators. Right panel includes means and standard errors of $\|\hat{\theta} - \theta\|_F$ for the standard estimator and the envelope estimators with either true or estimated dimensions.

| | Correct Selection % | | | Estimation Error $\|\hat{\theta} - \theta\|_F$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Standard | Envelope | | | | |
| Model | $n$ | 1D | FG | | true $u$ | 1D | FG | W1D | WFG |
| | 75 | 74 | 63.5 | 0.69 | 0.50 | 0.55 | 0.55 | 0.54 | 0.55 |
| Linear | 150 | 93 | 81 | 0.49 | 0.31 | 0.33 | 0.33 | 0.34 | 0.33 |
| | 300 | 99 | 92 | 0.33 | 0.19 | 0.19 | 0.20 | 0.19 | 0.19 |
| | 600 | 99 | 92.5 | 0.23 | 0.13 | 0.14 | 0.14 | 0.14 | 0.14 |
| | 75 | 22.5 | 42 | 4.04 | 1.04 | 1.06 | 1.00 | 1.09 | 1.08 |
| Logistic | 150 | 72 | 77.5 | 2.16 | 0.56 | 0.67 | 0.60 | 0.67 | 0.64 |
| | 300 | 92 | 89.5 | 1.40 | 0.34 | 0.35 | 0.34 | 0.37 | 0.36 |
| | 600 | 98 | 94 | 0.98 | 0.22 | 0.22 | 0.24 | 0.24 | 0.24 |
| | 75 | 35 | 38 | 2.07 | 1.99 | 1.95 | 1.96 | 2.04 | 2.05 |
| Cox | 150 | 57.5 | 53.5 | 1.33 | 1.24 | 1.21 | 1.22 | 1.27 | 1.28 |
| | 300 | 83 | 75.5 | 0.98 | 0.90 | 0.89 | 0.90 | 0.93 | 0.93 |
| | 600 | 100 | 93 | 0.79 | 0.72 | 0.72 | 0.72 | 0.75 | 0.75 |

Zhang and Mai [2018] which were not designed to showcase weighted envelope estimation techniques. The Cox proportional hazards model is ignored in our simulation since appreciable envelope estimation was not observed in the original Monte Carlo simulations. For this simulation, we generated one data set corresponding to the linear and logistic regression models in the previous simulation at sample sizes $n = 75, 150, 300$. We then perform a nonparametric bootstrap to estimate the variability of each envelope estimator using a bootstrap sample size of 200 iterations. We repeat this process 25 times, and report the average ratios of standard deviations relative to the standard estimator across these 25 Monte Carlo samples. Note that estimates of $u$ are allowed to (and do) vary across the iterations of the 25 Monte Carlo samples.

Table 4 displays the results with respect to the first component of the parameter vector (other components behave similarly) in both regression settings. In Table 4 we see that weighted envelope estimation provides larger variance reduction then given by $\widehat{\theta}^{\mathrm{1D}}_{\hat{u}^*_{\mathrm{1D}}}$ and $\widehat{\theta}^{\mathrm{FG}}_{\hat{u}^*_{\mathrm{FG}}}$ and is comparable to oracle estimation in most settings. The estimators $\widehat{\theta}^{\mathrm{FG}}_{\hat{u}_{\mathrm{FG}}}$ and $\widehat{\theta}^{\mathrm{1D}}_{\hat{u}_{\mathrm{1D}}}$ outperform weighted envelope estimation. However, this variance reduction is due to

Table 4: Ratios of standard deviations for envelope estimators relative to the MLE.

| Model | $n$ | $r(\tilde{\theta}, \hat{\theta}_u)$ | $r(\tilde{\theta}, \hat{\theta}^{\mathrm{1D}}_{\hat{u}_{\mathrm{1D}}})$ | $r(\tilde{\theta}, \hat{\theta}^{\mathrm{FG}}_{\hat{u}_{\mathrm{FG}}})$ | $r(\tilde{\theta}, \hat{\theta}^{\mathrm{1D}}_{\hat{u}^*_{\mathrm{1D}}})$ | $r(\tilde{\theta}, \hat{\theta}^{\mathrm{FG}}_{\hat{u}^*_{\mathrm{FG}}})$ | $r(\tilde{\theta}, \hat{\theta}^{\mathrm{1D}}_{w})$ | $r(\tilde{\theta}, \hat{\theta}^{\mathrm{FG}}_{w})$ |
|---|---|---|---|---|---|---|---|---|
| | 75 | 0.992 | 2.024 | 1.768 | 0.991 | 0.947 | 1.094 | 1.024 |
| Linear | 150 | 1.076 | 1.592 | 1.524 | 1.033 | 1.008 | 1.105 | 1.046 |
| | 300 | 1.236 | 2.219 | 2.108 | 1.173 | 1.102 | 1.264 | 1.171 |
| | 75 | 1.013 | 1.054 | 1.022 | 0.978 | 0.966 | 1.079 | 1.033 |
| Logistic | 150 | 1.548 | 2.741 | 2.459 | 1.231 | 1.008 | 1.374 | 1.079 |
| | 300 | 4.525 | 7.338 | 5.738 | 1.331 | 1.003 | 1.450 | 1.042 |

underestimation of $u$ in many of the original samples. Thus, weighted envelope estimation provides a desirable balance between model variance reduction and robustness to model misspecification.

# 6 Discussion

One limitation of bootstrapping weighted envelope estimators is that it can be computationally expensive, especially when $p$ is large [Yau and Volaufova, 2019]. In such settings, we recommend investigating if the range of candidate dimensions can reasonably be reduced to a less computationally burdensome set of values or using the variable $u$ approach when estimating the envelope dimension at every iteration of the nonparametric bootstrap. Existing envelope software implements the former approach in the context of multivariate linear regression [Lee and Su, 2019]. Our simulations provide some empirical justification for the performance of the latter approach.

Yau and Volaufova [2019] developed a novel hypothesis testing procedure with respect to the multivariate linear envelope model. They showed that model averaging in Eck and Cook [2017] is successful and exhibits comparable performance to their proposed methodology. They dismissed the model averaging technique by saying, "the model average estimator is not that viable. We may recall that the original motivation for applying the envelope model is to achieve dimension reduction. When one obtains $\hat{\theta}_w$,. . . it becomes unclear which subspace is being projected to as a result." The motivation for envelope methodology is not to "achieve dimension reduction," rather the motivation for envelope methodology is to increase efficiency in multivariate analyses without altering traditional objectives [Cook, 2018, first sentence of page 1]. Dimension reduction is at the core of envelope methodology, but it is just a means to an end for achieving useful variance reduction. The reporting of a specific subspace is not of foundational importance to practitioners seeking variance reduction, especially when there is both uncertainty in the subspace selected and its dimension. When there is uncertainty about the correct envelope dimension, model averaging with our weighted envelope estimator provides a desirable balance between variance reduction and correct model specification.


# Acknowldegement

The author would like to thank R. Dennis Cook, Forrest W. Crawford, Karl Oskar Ekvall, Dootika Vats, and Xin Zhang for valuable feedback that improved the presentation of this paper. This work was partially supported by NIH grants NICHD DP2 HD091799-01.


# Supplementary Materials

Supplementary materials are available with this paper. This supplement includes proofs of all technical results and it doubles as a fully reproducible technical report that makes all R based analyses transparent. The simulations in Section 5.3 are not included in the supplementary materials. These simulations are adopted from Matlab code that accompanied Zhang and Mai [2018]. This code is readily available upon request.

# References


D. W. K. Andrews. Higher-order improvements of a computationally attractive $k$-step bootstrap for extremum estimators. *Econometrica*, 70:1:119–162, 2002.

S. T. Buckland, K. P. Burnham, and N. H. Augustin. Model selection: An integral part of inference. *Biometrics*, 53:603–618, 1997.

K. P. Burnham and D. R. Anderson. Multimodel inference. *Sociological and Methods Research*, 33:261–304, 2004.

Yoosoon Chang and Joon Y Park. A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, 24(4):379–400, 2003.

G. Claeskens and N. L. Hjort. *Model Selection and Model Averaging*. Cambridge University Press, Cambridge, 2008.

R. D. Cook. *An introduction to envelopes: dimension reduction for efficient estimation in multivariate statistics*, volume 401. John Wiley & Sons, 2018.

R. D. Cook and Z. Su. Scaled envelopes: scale-invariant and efficient estimation in multivariate linear regression. *Biometrika*, 100:939–954, 2013.

R. D. Cook and Z. Su. Scaled predictor envelopes and partial least squares regression. *Technometrics*, 58:155–165, 2016.

R. D. Cook and X. Zhang. Foundations for envelope models and methods. *Journal of the American Statistical Association*, 110(510):599–611, 2015.

R. D. Cook and X. Zhang. Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics*, 25(1):284–300, 2016. doi: 10.1080/10618600.2015.1029577.

R. D. Cook and X. Zhang. Fast envelope algorithms. *Statistica Sinica*, 28(3):1179–1197, 2018.

R. D. Cook, B. Li, and F. Chiaromonte. Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, 20:927–1010, 2010.

R. D. Cook, L. Forzani, and Z. Su. A note on fast envelope estimation. *Journal of Multivariate Analysis*, 150:42–54, 2016.

S. Ding and R. D. Cook. Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B*, 80(2):387–408, 2018a.

S. Ding and R. D. Cook. Supplement: Proofs, technical details and additional results for “matrix-variate regressions and envelope models”. *Journal of the Royal Statistical Society: Series B*, 80(2):387–408, 2018b.

D. J. Eck. Bootstrapping for multivariate linear regression models. *Statistics and Probability Letters*, 134:141–149, 2018.

D. J. Eck and R. D. Cook. Weighted envelope estimation to handle variability in model selection. *Biometrika*, 104(3):743–749, 2017.

D. J. Eck, C. J. Geyer, and R. D. Cook. Supporting data analysis for ”combining envelope methodology and aster models for variance reduction in life history analyses”. 2018. http://hdl.handle.net/11299/178384.1.

D. J. Eck, C. J. Geyer, and R. D. Cook. Combining envelope methodology and aster models for variance reduction in life history analyses. *Journal of Statistical Planning and Inference*, (205):283–292, 2020.

B. Efron. Estimation and accuracy after model selection. *Journal of the American Statistical Association*, 109:507:991–1007, 2014.

Julian Faraway. *faraway: Functions and Datasets for Books by Julian Faraway*, 2016. `https://cran.r-project.org/web/packages/faraway/index.html`.

C. J. Geyer, S. Wagenius, and R. G. Shaw. Aster models for life history analysis. *Biometrika*, 94:415–426, 2007.

N. L. Hjort and G. Claeskens. Frequentist model average estimators. *Journal of the American Statistical Association*, 98:464:879–899, 2003.

M. Lee and Z. Su. *Renvlp: Computing Envelope Estimators*, 2019. `http://cran.radicaldevelop.com/web/packages/Renvlp/Renvlp.pdf`.

Lexin Li and Xin Zhang. Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519):1131–1146, 2017.

R. G. Shaw, C. J. Geyer, S. Wagenius, H. Hangelbroek, and J. R. Etterson. Unifying life-history analyses for inference of fitness and population growth. *American Naturalist*, 172:E35–E47, 2008.

L. A. Stefanski and Dennis D. Boos. The calculus of m-estimation. *American Statistician*, 56:29–38, 2002.

Z. Su and R. D. Cook. Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika*, 98:133–146, 2011.

Z. Su and R. D. Cook. Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika*, 99:687–702, 2012.

Z. Su, G. Zhu, X. Chen, and Y. Yang. Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression. *Biometrika*, 103:579–593, 2016.

G. N. Tsague. On optimal weighting scheme in model averaging. *American Journal of Applied Mathematics and Statistics*, 2(3):150–156, 2014.

W. Wang, J. Zeng, and X. Zhang. *TRES: Tensor Regression with Envelope Structure and Three Generic Envelope Estimation Approaches*, 2019. https://cran.r-project.org/web/packages/TRES/index.html.

J. P. Willems, J. T. Saunders, D. E. Hunt, and J. B. Schorling. Prevalence of coronary heart disease risk factors among rural blacks: a community-based study. *Southern Medical Journal*, 90(8):814–820, 1997.

World Health Organization. Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. Technical report, Geneva: World Health Organization, 2011.

Tat S. Yau and J. Volaufova. A note on hypothesis testing in the envelope model. *Communications in Statistics*, 2019. DOI: 10.1080/03610918.2019.1672737.

J. Zeng, W. Wang, and Z. Zhang. Tres: An r package for tensor regression and envelope algorithms. *Journal of Statistical Software*, 2020.

X. Zhang and Q. Mai. Model-free envelope dimension selection. *Electronic Journal of Statistics*, 12:2193–2216, 2018.

X. Zhang, C. E. Lee, and X. Shao. Envelopes in multivariate regression models with nonlinearity and heteroscedasticity. *Biometrika*, 2020. In Press.