

# General model-free weighted envelope estimation\*

Daniel J. Eck

*Department of Statistics, University of Illinois,  
605 E. Springfield Ave.,  
Champaign, IL 61820 USA  
e-mail: [dje13@illinois.edu](mailto:dje13@illinois.edu)*

**Abstract:** Envelope methodology is succinctly pitched as a class of procedures for increasing efficiency in multivariate analyses without altering traditional objectives [5, 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 and phrases:** Bootstrap smoothing, dimension reduction, model averaging, model selection, nonparametric bootstrap.

Received July 2021.

## 1. Introduction

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

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N(0, \Sigma), \quad (1.1)$$

as  $n \rightarrow \infty$ . 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  $\theta$  [7, 10, 5]. Envelope methodology originated as a method to reduce the variability

---

\*This work was partially supported by NIH grants NICHD DP2 HD091799-01.

of a regression coefficient matrix  $\beta$  in the multivariate linear regression model [7, 27, 28, 8, 5]. 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. [10] 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 (1.1), and a  $\sqrt{n}$  consistent estimator  $\widehat{\Sigma}$  of  $\Sigma$ .

In both the multivariate linear regression model and the general estimation framework (1.1), variance reduction obtained through envelope methodology arises from exploiting a subspace of the spectral structure of  $\Sigma$  with dimension  $u < p$  that contains  $\text{span}(\theta)$ . 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. [36] 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. [16] 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 [15]. A double bootstrap procedure that incorporates model selection variability for envelope estimation of expected Darwinian fitness from an aster model [21, 26] has also been developed [18]. This procedure 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 [16] are constructed from the Bayesian Information Criterion (BIC) values of the multivariate linear regression model evaluated at all envelope estimators fit at dimension  $u = 1, \dots, p$ . Model selection volatility is taken into account in [18] 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 (1.1). We will not require the existence of a likelihood function as in [16] 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. Formal asymptotic justification for these procedures is provided. 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  $n \rightarrow \infty$  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. One concern with bootstrapping a weighted envelope estimator is that the computational cost of

the procedure can be burdensome [33]. This concern is alleviated through the parallel nature of computations. Both estimation for different envelope dimensions and the bootstrap procedure can be done in parallel. 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 [16] and [36].

Others have studied bootstrapping procedures that incorporate variability in model selection and have demonstrated applicability for envelope models. A double bootstrap procedure that incorporate variability in model selection was proposed by [19]. This procedure is applicable for exponential families, and it has been applied to envelope methodology within this context [18, 17]. [18] did not provide any asymptotic justification for the bootstrap procedures that are implemented. [36] also considered bootstrap schemes that incorporate model selection variability in their numerical studies, but they do not provide theoretical properties. In addition to providing asymptotic justification for our bootstrap procedures, the weighted envelope estimators that we develop are appropriate for a more general class of envelope models than either [19] or [18] can claim. The weighted envelope estimation methodology that we develop in this paper extends to partial envelopes [27], inner envelopes [28], scaled envelopes [8], predictor envelopes [9], sparse response envelopes [29], tensor response regression [24], matrix-variate response regression [13, 14], and envelope models with non-linearity and heteroscedasticity [35]. 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 2.1** (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  $M\mathcal{R}^\perp \subset \mathcal{R}^\perp$  where  $\mathcal{R}^\perp$  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 2.2** (Envelope). *The  $M$  envelope of  $\text{span}(U)$  is defined as the intersection of all reducing subspaces  $\mathcal{R}$  of  $M$  which satisfies  $\text{span}(U) \subseteq \mathcal{R}$ . 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  $\text{span}(U)$ . 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, [36] 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:

$$J_n(\Gamma) = \log |\Gamma^T \widehat{M} \Gamma| + \log |\Gamma^T (\widehat{M} + \widehat{U})^{-1} \Gamma|, \quad (2.1)$$

where  $\widehat{M}$  and  $\widehat{U}$  are  $\sqrt{n}$ -consistent estimators of  $M$  and  $U$ . In the general setting,  $\widehat{M} = \widehat{\Sigma}$  and  $\widehat{U} = \widehat{\theta}\widehat{\theta}^T$ . We denote  $\widehat{\Gamma}$  as the minimizer of (2.1). The motivations for (2.1) comes from its population counterpart,  $J(\Gamma) = \log |\Gamma^T M \Gamma| + \log |\Gamma^T (M + U)^{-1} \Gamma|$ , where [11] showed that any  $\Gamma$  which minimizes  $J(\Gamma)$  must satisfy the envelope condition that  $\text{span}(U) \subseteq \text{span}(\Gamma)$ . 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 [11]. We will therefore refer to the minimization of (2.1) 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 [11, 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 (1.1). We will also consider envelope estimators constructed with a sequential one-dimensional (1D) coordinate optimization algorithm [11, 10], 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  $\widehat{\theta}$  computed using FG or 1D optimizations, or any other valid envelope optimization routine [6, 23, 31]. 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  $\widehat{\theta}$  with the true dimension specified. Therefore, we will often write  $\widehat{\theta}_u$  to emphasize this point. We will also write  $\widehat{\theta}_k$  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

$$\widehat{\theta}_w = \sum_{k=1}^p w_k \widehat{\theta}_k, \quad w_k = f_k(\mathcal{I}_n(1), \dots, \mathcal{I}_n(p)), \quad (3.1)$$

where  $\widehat{\theta}_k$  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 (3.1), 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,  $\widehat{\theta}_k$  is consistent for all weight choices that satisfy  $\sum_{k=u}^p w_k \rightarrow 1$  as  $n \rightarrow \infty$  where  $w_k \geq 0$  for  $k \geq u$ . Such weight choices induce a consistent estimator since the envelope estimator  $\widehat{\theta}_k$  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  $w_u \rightarrow 1$ .

We specifically study two specific choices of  $\mathcal{I}_n(\cdot)$  and one choice of  $f_k$  so that  $w_u \rightarrow 1$  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

$$w_k = \frac{\exp\{-n\mathcal{I}_n(k)\}}{\sum_{j=1}^p \exp\{-n\mathcal{I}_n(j)\}}. \quad (3.2)$$

The choice of  $f_k$  that yields the weights (3.2) is motivated by [16], 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 (1.1). 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. 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. Our simulations, and the simulations presented in [36] find that the choice of  $\mathcal{I}_n^{\text{1D}}(k)$  exhibits greater empirical variance reduction than the choice of  $\mathcal{I}_n^{\text{FG}}(k)$ . However, we find that FG optimization offers more stability than the 1D algorithm which conflicts with advice in envelope software packages [31, 34].

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

In this section we construct  $\hat{\theta}_w$  in (3.1) using  $\mathcal{I}_n^{\text{FG}}(k)$  as the chosen the information criteria. At candidate dimension  $k$ , define  $\hat{\Gamma}_k \in \mathbb{R}^{p \times k}$  as the minimizer of the objective function (2.1). Thus  $\hat{\Gamma}_k$  is the estimated basis of the envelope subspace with dimension  $k$ . After obtaining  $\hat{\Gamma}_k$ , the envelope estimator at candidate dimension  $k$  is  $\hat{\theta}_k^{\text{FG}} = \hat{\Gamma}_k \hat{\Gamma}_k^T \tilde{\theta}$ . This envelope estimator is the original estimator projected into the estimated envelope subspace at dimension  $k$ . When the true dimension  $u$  is known, then  $\hat{\theta}_u^{\text{FG}}$  is  $\sqrt{n}$ -consistent and has been shown to have lower variability than  $\tilde{\theta}$  in finite samples [7, 10, 5]. The weighted envelope estimator corresponding to  $\mathcal{I}_n^{\text{FG}}(k)$  is

$$\hat{\theta}_w^{\text{FG}} = \sum_{k=1}^p w_k^{\text{FG}} \hat{\theta}_k^{\text{FG}}, \quad w_k^{\text{FG}} = \frac{\exp\{-n\mathcal{I}_n^{\text{FG}}(k)\}}{\sum_{j=1}^p \exp\{-n\mathcal{I}_n^{\text{FG}}(j)\}}, \quad (3.3)$$

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

[36] showed that optimization of  $J_n$  in (2.1) is the same as optimization of a partially minimized quasi-likelihood function. Define this quasi-likelihood function as,

$$l_n(M, \theta) = \log |M| + \text{tr} \left[ M^{-1} \left\{ \widehat{M} + (\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T \right\} \right], \quad (3.4)$$

and, for some candidate dimension  $k = 1, \dots, p$ , define the constraint set for the minimization of (3.4) to be,

$$\begin{aligned} \mathcal{A}_k = \{ (M, \theta) : M = \Gamma \Omega \Gamma^T + \Gamma_o \Omega_o \Gamma_o^T > 0, \\ \theta = \Gamma \eta, \eta \in \mathbb{R}^k, (\Gamma, \Gamma_o)^T (\Gamma, \Gamma_o) = I_p \}. \end{aligned} \quad (3.5)$$

Minimization of (3.4) over the constraint set (3.5) is the same as minimizing  $J_n$  in (2.1).

**Lemma 3.1.** [36, Lemma 3.1]. *The minimizer of  $l_n(M, \theta)$  in (3.4) under the envelope parameterization (3.5) 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  $\widehat{\theta} = \widehat{\Gamma} \widehat{\Gamma}^T \widehat{\theta}$  where  $\widehat{\Gamma}$  is the minimizer of the partially optimized objective function  $l_n(\Gamma) = \min_{\Omega, \Omega_o, \eta} l_n(\Gamma, \Omega, \Omega_o, \eta) = J_n(\Gamma) + \log |\widehat{M} + \widehat{U}| + p$  where  $\widehat{U} = \widehat{\theta} \widehat{\theta}^T$ .*

Now define

$$\mathcal{I}_n^{\text{FG}}(k) = J_n(\widehat{\Gamma}_k) + \frac{Ck \log(n)}{n}, \quad (k = 1, \dots, p), \quad (3.6)$$

as in [36] where  $C > 0$  is a constant. The envelope dimension is selected as  $\hat{u}_{\text{FG}} = \operatorname{argmin}_{1 \leq k \leq p} \mathcal{I}_n(k)$ . Theorem 3.1 in [36] showed that  $\mathbb{P}(\hat{u}_{\text{FG}} = u) \rightarrow 1$  as  $n \rightarrow \infty$ , provided that  $C > 0$  and  $\widehat{M}$  and  $\widehat{U}$  are  $\sqrt{n}$ -consistent estimators of  $M$  and  $U$  respectively. We use  $\mathcal{I}_n^{\text{FG}}(k)$  in (3.6) to construct  $\hat{\theta}_w^{\text{FG}}$  in (3.3). This construction yields consistent estimation of  $\theta$  as seen in Section 3.3 and estimation of the variability of  $\hat{\theta}_u^{\text{FG}}$  through Theorem 4.1.

### 3.2. Weighted envelope estimation via the 1D algorithm

In this section we construct  $\hat{\theta}_w$  in (3.1) where the information criteria  $\mathcal{I}_n^{\text{1D}}(k)$  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  $\hat{\theta}_u^{\text{1D}} = \widehat{\Gamma} \widehat{\Gamma}^T \hat{\theta}$  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

$$\hat{\theta}_w^{\text{1D}} = \sum_{k=1}^p w_k^{\text{1D}} \hat{\theta}_k^{\text{1D}}, \quad w_k^{\text{1D}} = \frac{\exp \{ -n \mathcal{I}_n^{\text{1D}}(k) \}}{\sum_{j=1}^p \exp \{ -n \mathcal{I}_n^{\text{1D}}(j) \}}. \quad (3.7)$$

We briefly state the 1D algorithm: Set  $u_o \leq p - 1$  to be the user inputted number of optimizations. For step  $k = 0, \dots, u_o$ , let  $g_k \in \mathbb{R}^p$  denote the  $k$ -th direction to be obtained by the 1D algorithm. Define  $G_k = (g_1, \dots, g_k)$ , 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  $v_{k+1} = \operatorname{argmax}_v \phi_k(v)$  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  $\widehat{G}_k = (\widehat{g}_1, \dots, \widehat{g}_k) \in \mathbb{R}^{p \times k}$ ,  $k = 1, \dots, p$  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  $\widehat{\theta}_u^{1D}$  is therefore  $\sqrt{n}$ -consistent [10]. Now define,

$$\mathcal{I}_n^{1D}(k) = \sum_{j=1}^k \phi_{j,n}(\widehat{v}_j) + \frac{Ck \log n}{n}, \quad (k = 1, \dots, p); \quad (3.8)$$

where  $C > 0$  is a constant. When  $C = 1$ , the terms  $n\mathcal{I}_n^{1D}(k)$  in (3.8) are BIC values corresponding to the asymptotic log likelihood the envelope model of dimension  $k$ . The envelope dimension selected is  $\hat{u}_{1D} = \operatorname{argmin}_{1 \leq k \leq p} \mathcal{I}_n^{1D}(k)$ . Theorem 3.2 in [36] showed that  $\mathbb{P}(\hat{u}_{1D} = u) \rightarrow 1$  as  $n \rightarrow \infty$ . We use  $\mathcal{I}_n^{1D}(k)$  in (3.8) to construct  $\widehat{\theta}_w^{1D}$  in (3.7). This construction yields consistent estimation of  $\theta$  as seen in Section 3.3 and estimation of the variability of  $\widehat{\theta}_u^{1D}$  through Theorem 4.2.

### 3.3. Consistency properties of weighted envelope estimators

Weighted envelope estimators exhibit desirable consistency properties. First of all, the weights in (3.3) and (3.7) can be constructed so that they both satisfy  $w_u^{FG} \xrightarrow{P} 1$  and  $w_u^{1D} \xrightarrow{P} 1$  as  $n \rightarrow \infty$ , where  $\xrightarrow{P}$  means convergence in probability.

**Lemma 3.2.** *For any constant  $C > 0$  and  $\sqrt{n}$ -consistent  $\widehat{M}$  and  $\widehat{U}$  in (3.6) and (3.8),  $w_u^{FG} \xrightarrow{P} 1$  and  $w_u^{1D} \xrightarrow{P} 1$  as  $n \rightarrow \infty$ .*

The proof of technical results are included in the Mathematical Appendix. This lemma facilitates consistent estimation of  $\theta$  using  $\widehat{\theta}_w^{FG}$  and  $\widehat{\theta}_w^{1D}$ .

**Lemma 3.3.** *For any constant  $C > 0$  and  $\sqrt{n}$ -consistent  $\widehat{M}$  and  $\widehat{U}$  in (3.6) and (3.8), both  $\widehat{\theta}_w^{FG} \xrightarrow{P} \theta$  and  $\widehat{\theta}_w^{1D} \xrightarrow{P} \theta$  as  $n \rightarrow \infty$ .*

The proof of Lemma 3.3 immediately follows from Lemma 3.2 and [36, Proposition 2.1]. While consistency is desirable, Lemma 3.3 does not provide knowledge about the asymptotic variability of  $\widehat{\theta}_w^{FG}$  or  $\widehat{\theta}_w^{1D}$ . We expect that  $\widehat{\theta}_w^{FG}$  and  $\widehat{\theta}_w^{1D}$  will have lower asymptotic variance than  $\widehat{\theta}$  when  $u < p$ , but explicit computations of the asymptotic variance for both estimators are cumbersome. We will instead estimate the asymptotic variability of  $\widehat{\theta}_w^{FG}$  and  $\widehat{\theta}_w^{1D}$  with a nonparametric bootstrap that is developed in the next section.

## 4. Bootstrapping procedures

Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be the original data. We will estimate the variability of  $\widehat{\theta}_u^{FG}$  and  $\widehat{\theta}_u^{1D}$  by bootstrapping with respect to the weighted estimators  $\widehat{\theta}_w^{FG}$  and  $\widehat{\theta}_w^{1D}$  via the nonparametric bootstrap. For each iteration of this nonparametric

bootstrap procedure we denote the resampled data by  $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$  where each  $\mathbf{X}_i^*$ ,  $i = 1, \dots, n$  is sampled, with replacement, from the original data with equal probability  $1/n$ . At candidate envelope dimension  $k$  we define the bootstrapped envelope estimators  $\hat{\theta}_k^{FG*} = \hat{\theta}_k^{FG}(\mathbf{X}_1^*, \dots, \mathbf{X}_n^*)$ ,  $\hat{\theta}_k^{1D*} = \hat{\theta}_k^{1D}(\mathbf{X}_1^*, \dots, \mathbf{X}_n^*)$ , and the bootstrapped version of the original estimator  $\hat{\theta}$  as  $\hat{\theta}^* = \hat{\theta}(\mathbf{X}_1^*, \dots, \mathbf{X}_n^*)$ . Define  $\hat{M}^*$  and  $\hat{U}^*$  in the same manner as  $\hat{M}$  and  $\hat{U}$  with respect to the starred data. Following Remark 1 in Section 3 of [3], 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  $\hat{\theta}_w^{FG}$ . Define

$$J_n^*(\Gamma) = \log |\Gamma^T \hat{M}^* \Gamma| + |\Gamma^T (\hat{M}^* + \hat{U}^*)^{-1} \Gamma|, \quad (4.1)$$

as the starred analog to  $J_n$  in (2.1) and define,

$$l_n^*(M, \theta) = \log |M| + \text{tr} \left[ M^{-1} \left\{ \hat{M}^* + (\hat{\theta}^* - \theta)(\hat{\theta}^* - \theta)^T \right\} \right], \quad (4.2)$$

as the starred analog to  $l_n$  in (3.4). Define  $\hat{\Gamma}^*$  as the minimizer to (4.1). We then let  $\hat{\theta}_u^{FG*} = \hat{\Gamma}^* \hat{\Gamma}^{*T} \hat{\theta}^*$ . We show that bootstrapping with respect to our weighted envelope estimator estimates the variability of the envelope estimator  $\hat{\theta}_u^{FG}$  at the true unknown dimension.

**Theorem 4.1.** *Let  $\hat{\theta}$  be a  $\sqrt{n}$ -consistent and asymptotically normal estimator. Let  $\hat{\theta}_k^{FG}$  be the envelope estimator obtained from full Grassmannian optimization at dimension  $k = 1, \dots, p$  and let  $\hat{\theta}_w^{FG}$  be the weighted envelope estimator with weights  $w^{FG}$ . Let  $\hat{\theta}_k^{FG*}$  and  $\hat{\theta}_w^{FG*}$  denote the corresponding quantities obtained by resampled data. Then as  $n$  tends to  $\infty$ ,*

$$\begin{aligned} \sqrt{n} \left( \hat{\theta}_w^{FG*} - \hat{\theta}_w^{FG} \right) &= \sqrt{n} \left( \hat{\theta}_u^{FG*} - \hat{\theta}_u^{FG} \right) + O_P^* \left\{ n^{(1/2-C)} \right\} \\ &\quad + O_P^* \left\{ n^{(Cu+1/2)} \right\} e^{-n|O_P^*(1)|}. \end{aligned} \quad (4.3)$$

#### Remarks:

1. Theorem 4.1 shows that our bootstrap procedure can be used to estimate the asymptotic variability of  $\hat{\theta}_u^{FG}$  when  $u$  is unknown. We see that the second  $O_P^*$  term in (4.3) 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  $\hat{\theta}_u^{FG}$  in finite samples.



2. We advocate for the case with  $C = 1$  because of the close connection that  $\mathcal{I}_n^{\text{FG}}(k)$  has with BIC, similar reasoning was given in [36]. The  $Ck \log(n)/n$  penalty term in  $\mathcal{I}_n^{\text{FG}}(k)$  facilitates the decaying bias in  $n$  represented by the  $O_P^*$  terms in (4.3). Redefining  $\mathcal{I}_n^{\text{FG}}(k)$  to have a penalty term that is fixed in  $n$ , similar to that of AIC, fundamentally changes the  $O_P^*$  terms in (4.3). 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 [1, 2, 22, 4, 30]. These weights are of the form

$$w_k = \frac{\exp\{-n\mathcal{I}_n^{\text{FG}}(k)/2\}}{\sum_{j=0}^p \exp\{-n\mathcal{I}_n^{\text{FG}}(j)/2\}} \quad (4.4)$$

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 (4.4) do not have the same asymptotic properties as our weights. The difference between the two is a rescaling of  $C$ . Weights (4.4) replace the constant  $C$  in (4.3) 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 (4.4) 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  $\hat{\theta}^{1\text{D}}$ . We verify that bootstrapping with respect to  $\hat{\theta}_w^{1\text{D}}$  is asymptotically the same as bootstrapping with respect to  $\hat{\theta}_u^{1\text{D}}$ . 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  $k = 0, \dots, u_o$ , let  $\hat{g}_k^* \in \mathbb{R}^p$  denote the  $k$ -th direction to be obtained. Define  $\hat{G}_k^* = (\hat{g}_1^*, \dots, \hat{g}_k^*)$ , and  $(\hat{G}_k^*, \hat{G}_{0k}^*)$  to be an orthogonal basis for  $\mathbb{R}^p$  and set initial value  $\hat{g}_o^* = \hat{G}_0^* = 0$ . Define  $\hat{M}_k^* = \hat{G}_{0k}^{*T} \hat{M}^* \hat{G}_{0k}^*$ ,  $\hat{U}_k^* = \hat{G}_{0k}^{*T} \hat{U}^* \hat{G}_{0k}^*$ , and the objective function after  $k$  steps as  $\phi_{k,n}^*(v) = \log(v^T \hat{M}_k^* v) + \log\{v^T (\hat{M}_k^* + \hat{U}_k^*)^{-1} v\}$ . The  $(k + 1)$ -th envelope direction is  $\hat{g}_{k+1}^* = \hat{G}_{0k}^* v_{k+1}^*$  where  $v_{k+1}^* = \arg\max_{v^T v = 1} \phi_{k,n}^*(v)$ . After  $u_o$  steps we set  $\hat{\Gamma}^* = \hat{G}_{u_o}^*$  and  $\hat{\theta}_u^{1\text{D}*} = \hat{\Gamma}^* \hat{\Gamma}^{*T} \hat{\theta}^*$ . We show that bootstrapping  $\hat{\theta}_w^{1\text{D}}$  estimates the variability of the envelope estimator  $\hat{\theta}_u^{1\text{D}}$  at the true unknown dimension.

**Theorem 4.2.** *Let  $\tilde{\theta}$  be a  $\sqrt{n}$ -consistent and asymptotically normal estimator. Let  $\hat{\theta}_k^{1\text{D}}$  be the envelope estimator obtained from the 1D algorithm at dimension  $k = 1, \dots, p$  and let  $\hat{\theta}_w^{1\text{D}}$  be the weighted envelope estimator with weights  $w^{1\text{D}}$ . Let  $\hat{\theta}_k^{1\text{D}*}$  and  $\hat{\theta}_w^{1\text{D}*}$  denote the corresponding quantities obtained by resampled*

TABLE 1  
Description of bootstrap estimators.

Quantity	Description
$\hat{\theta}_w^{\text{FG}*}$ or $\hat{\theta}_w^{\text{1D}*}$	The weighted envelope estimator $\hat{\theta}_w^{\text{FG}}$ or $\hat{\theta}_w^{\text{1D}}$ computed with respect to resampled data. We will refer to these estimators as “weighted” in simulation results.
$\hat{\theta}_k^{\text{FG}*}$ or $\hat{\theta}_k^{\text{1D}*}$	The envelope estimator $\hat{\theta}_k^{\text{FG}}$ or $\hat{\theta}_k^{\text{1D}}$ fit at dimension $k$ computed with respect to resampled data. This estimator will be referred to as “true” in simulation results when $k = u$ .
$\hat{\theta}_{\hat{u}_{\text{FG}}}^{\text{FG}*}$ or $\hat{\theta}_{\hat{u}_{\text{1D}}}^{\text{1D}*}$	The envelope estimator $\hat{\theta}_{\hat{u}_{\text{FG}}}^{\text{FG}}$ or $\hat{\theta}_{\hat{u}_{\text{1D}}}^{\text{1D}}$ fit at the dimension estimated from the original sample $\hat{u}_{\text{FG}}$ or $\hat{u}_{\text{1D}}$ and is then computed with respect to resampled data. We call this the fixed $u$ regime, and these estimators will be referred to as “fixedu” in simulation results.
$\hat{\theta}_{\hat{u}_{\text{FG}}}^{\text{FG}*}$ or $\hat{\theta}_{\hat{u}_{\text{1D}}}^{\text{1D}*}$	The envelope estimator $\hat{\theta}_{\hat{u}_{\text{FG}}}^{\text{FG}}$ or $\hat{\theta}_{\hat{u}_{\text{1D}}}^{\text{1D}}$ fit at the dimension estimated from resampled data $\hat{u}_{\text{FG}*}$ or $\hat{u}_{\text{1D}*}$ and is then also computed with respect to resampled data. We call this the variable $u$ regime, and these estimators will be referred to as “varu” in simulation results. These estimators are similar to the consistent model-free estimators developed by [36].
$r(\tilde{\theta}, \hat{\theta}_{\text{env}})$	A column vector containing ratios of bootstrapped standard errors $\text{se}^*(\tilde{\theta}_j)/\text{se}^*(\hat{\theta}_{\text{env},j})$ , $j = 1, \dots, p$ where $\tilde{\theta}$ is the estimator (1.1) and $\hat{\theta}_{\text{env}}$ refers to one of the above envelope estimators.

data. Then as  $n$  tends to  $\infty$ ,

$$\begin{aligned} \sqrt{n} \left( \hat{\theta}_w^{\text{1D}*} - \hat{\theta}_w^{\text{1D}} \right) &= \sqrt{n} \left( \hat{\theta}_u^{\text{1D}*} - \hat{\theta}_u^{\text{1D}} \right) + O_P^* \left\{ n^{(1/2-C)} \right\} \\ &\quad + O_P^* \left\{ n^{(Cu+1/2)} \right\} e^{-n|O_P^*(1)|}. \end{aligned} \quad (4.5)$$

The remarks to Theorem 4.2 are similar to those for Theorem 4.1. The second  $O_P$  term in (4.5) 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  $\hat{\theta}_u^{\text{1D}}$  in finite samples. We advocate for the case with  $C = 1$  because of the close connection that  $\mathcal{I}_n^{\text{1D}}(k)$  has with BIC. Manuals for available software recommend use of one-directional optimizations, such as the 1D algorithm or the ECD algorithm [12], because they are faster, stable, and less sensitive to initial values than FG optimization [31, 34]. The simulations in the next section paint a more nuanced picture.

## 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. The specific bootstrap estimators under consideration are summarized in Table 1.

We demonstrate our model-free weighted envelope estimation techniques for exponential family generalized linear models (GLM) using normal predictors as justified in [10]. 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 (1.1) 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 [31].

### 5.1. Exponential family GLM simulations with $p = 6$

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 by  $\mathbf{X} \sim N(0, \Sigma_{\mathbf{X}})$ , 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 supplementary 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 0.25, 0.5, and 1,  $\Omega_o$  has diagonal elements  $\exp(-4)$ ,  $\exp(-2)$ , and  $\exp(1)$ . We construct  $\theta = -v_1/4 - v_2/4 - 3v_3/4$  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  $\Gamma$ .

We perform our simulations for four sample sizes,  $n = 200, 400, 600, 800$  and a Monte Carlo sample size of 400. A nonparametric bootstrap routine with bootstrap sample size of 1000 is performed for each Monte Carlo iterate at each considered sample size. Both FG and 1D envelope estimators are considered, see Table 1 for a description of these estimators.

We consider envelope estimation performance metrics for each simulation setting. First and foremost we consider componentwise ratios of the bootstrapped standard deviations of each envelope estimator to the bootstrapped standard deviation corresponding to the maximum likelihood estimator. These ratios  $r(\tilde{\theta}, \hat{\theta}_{\text{env}})$  are column vectors with elements  $\text{se}^*(\tilde{\theta}_j)/\text{se}^*(\hat{\theta}_{\text{env},j})$ ,  $j = 1, \dots, p$ . The standard deviation  $\text{se}^*(\hat{\theta}_j)$  is computed as

$$\left( \frac{1}{B} \sum_{b=1}^B (\tilde{\theta}_j^{*b} - \tilde{\theta}_j)(\tilde{\theta}_j^{*b} - \tilde{\theta}_j)^T \right)^{1/2},$$

where  $\tilde{\theta}_j^{*b}$  is  $\tilde{\theta}_j$  computed from the resampled data in the  $b$ th iteration of a nonparametric bootstrap routine. Any element of  $r(\tilde{\theta}, \hat{\theta}_{\text{env}})$  that is greater than

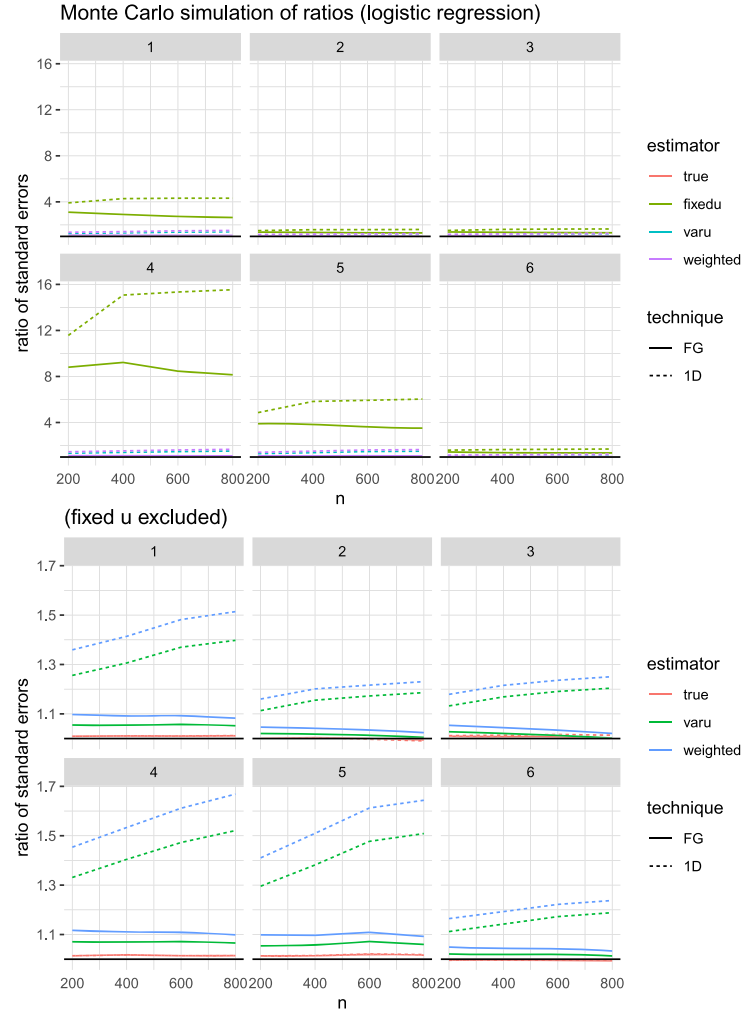


FIG 1. Averages of Monte Carlo simulations for ratios of envelope estimators to the MLE across all components of the parameter vector in our logistic regression example in Section 5.1. Ratios greater than 1 indicate superior performance for envelope estimation.

1 indicates that variance reduction is obtained using  $\hat{\theta}_{\text{env},j}$  in place of  $\tilde{\theta}_j$ . Componentwise ratios for both simulations are presented in Figures 1 and 4.

We also compute componentwise coverage probabilities averaged over Monte Carlo iterates for both simulations. For each component we compute coverage probabilities as the proportion of bootstrap iterations in which a 95% Wald-type confidence interval covers the true parameter, where the standard errors used to form these confidence intervals are estimated from the bootstrap samples. Coverage probabilities are presented in Figures 2 and 5.

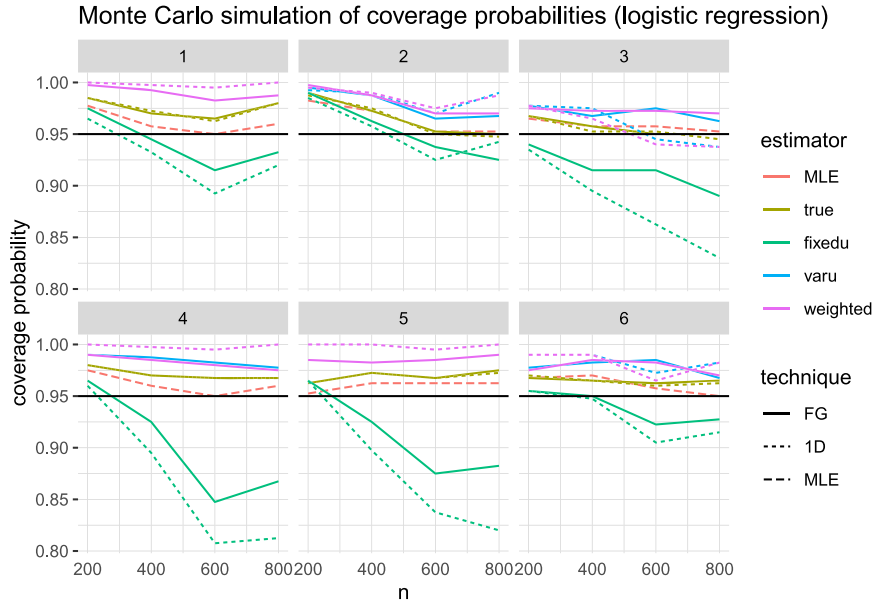


FIG 2. Coverage probabilities of all estimators averaged across Monte Carlo replicates for all components of the parameter vector in our logistic regression example in Section 5.1.

We also depict the empirical distribution of the estimated envelope dimension using FG optimization and the 1D algorithm in Figures 3 and 6. This distribution is averaged over Monte Carlo iterates. In both simulations the simulation truth is  $u = 3$ .

In each of these simulations we find that envelope estimation provides useful variance reduction. In the logistic regression setting we find that the fixed  $u$  regime (see Table 1) that is estimated with the 1D algorithm provides the largest variance reduction. However, these gains are offset by problematic under coverage that does not disappear as the sample size increases. See Figures 1 and 2 for the particulars. Moreover, we see that  $u$  is frequently underestimated by the 1D algorithm (Figure 3). These findings indicate that not accounting for model selection variability can lead to promising variance reduction which comes with the cost of inconsistent envelope estimation. These findings are in conflict with the recommendations in the TRES package manual which state that “the FG optimization is often associated with likelihood-based estimation but requires heavy computation and good initialization; the one-directional optimization approaches (ECD [12] and 1D algorithms) are faster, stable and does not require carefully chosen initial values” [31].

The weighted envelope estimator and the variable  $u$  regimes (see Table 1) estimated with 1D optimization provide the largest variance reduction after the fixed  $u$  regime in our logistic regression simulations. These envelope estimators exhibit far better coverage than the fixed  $u$  envelope estimation regime men-

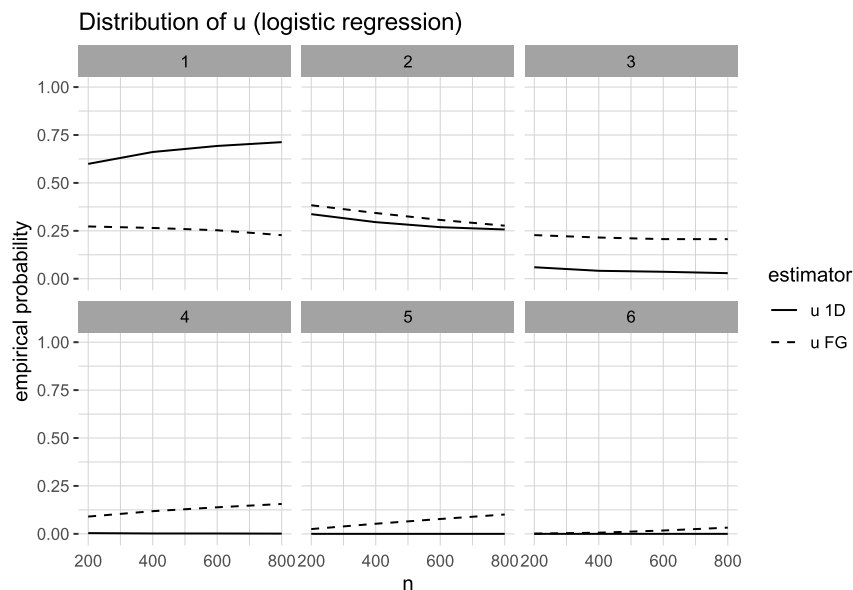


FIG 3. Empirical distribution of  $u$  averaged across Monte Carlo replicates for our logistic regression example in Section 5.1.

tioned in the previous paragraph. We find that the FG optimization routine yields envelope estimators with much more modest gains than those computed using the 1D algorithm. These modest gains come with less weight placed on an envelope dimension smaller than the truth.

Our Poisson regression simulations show massive variance reduction using envelope methods. In this simulation the fixed  $u$  regime now gives proper coverage while the weighted and variable  $u$  envelope estimators fit with 1D estimation exhibit problematic under coverage for small sample sizes. However, the under coverage improves quickly as the sample size increases. See Figures 4 and 5 for the particulars. We also see that envelope methods fit with FG optimization exhibit higher variance reduction and better coverage properties than the 1D algorithm. These findings are again in conflict with the recommendations in the TRES package manual which promote the 1D algorithm as more reliable. We find that both FG optimization and the 1D algorithm better estimate the correct envelope dimension as the sample size increases (Figure 6).

We now summarize results from these simulations. First and foremost, we find that envelope estimation provides useful and in some case massive variance reduction. We find that FG optimization is more reliable than the 1D algorithm, a finding that conflicts with the recommendations in the TRES package manuals. We find that the weighted envelope estimator exhibits a slight advantage but otherwise similar performance to the variable  $u$  regime. This finding was also confirmed in simulations conducted by Xin Zhang and his students (personal communication). We find that the fixed  $u$  regime can lead to misleading mas-

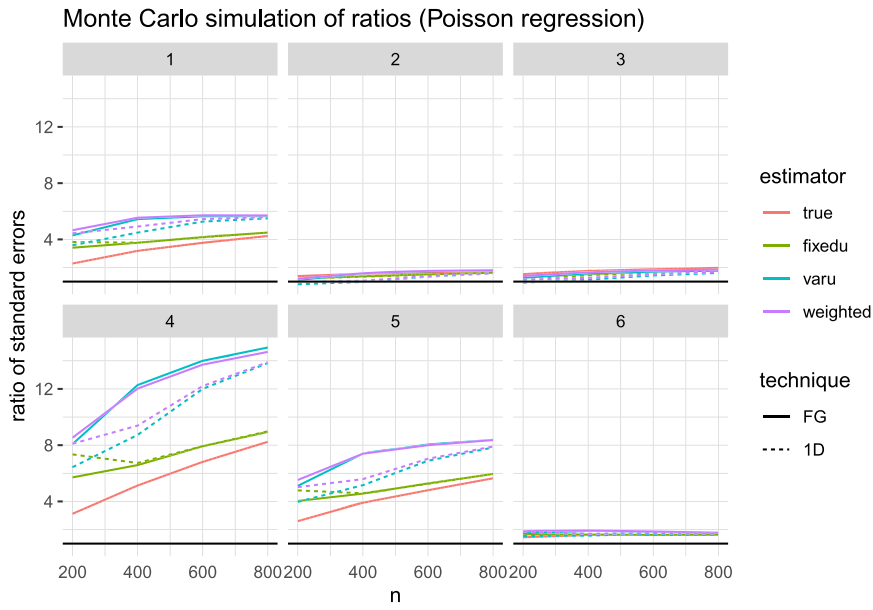


FIG 4. Averages of Monte Carlo simulations for ratios of envelope estimators to the MLE across all components in our Poisson regression example in Section 5.1. Ratios greater than 1 indicate superior performance for envelope estimation.

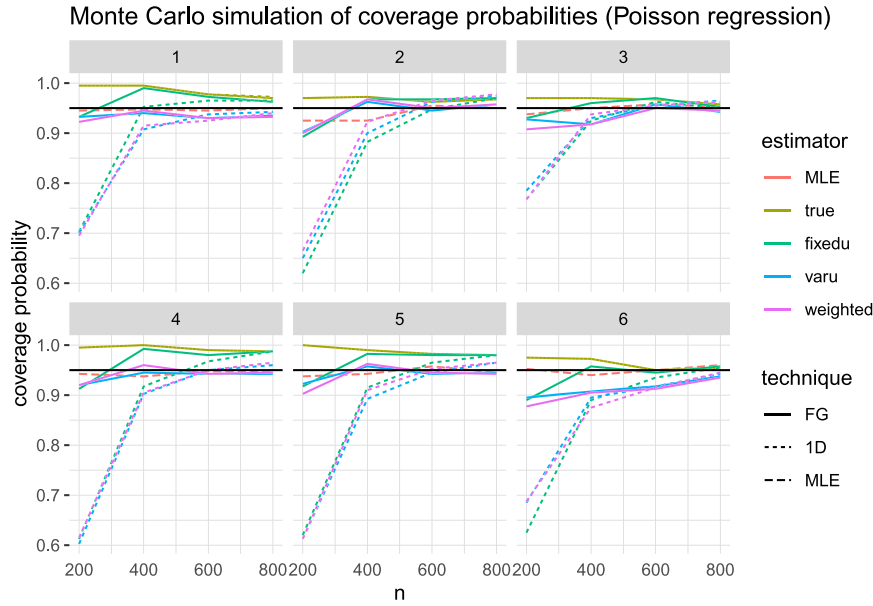


FIG 5. Coverage probabilities of all estimators averaged across Monte Carlo replicates.

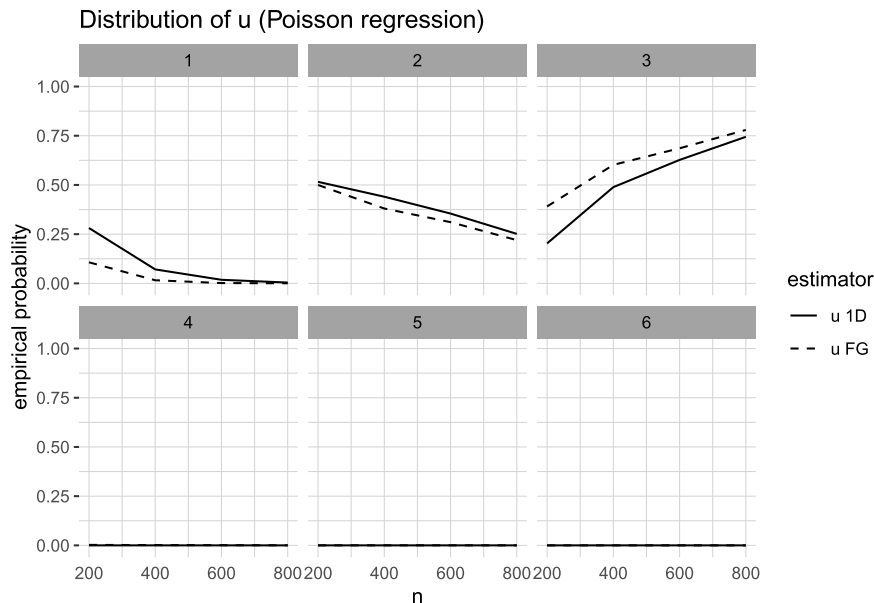


FIG 6. Empirical distribution of  $u$  averaged across Monte Carlo replicates for our Poisson regression example in Section 5.1.

sive variance reduction which comes at the expense of inconsistent estimation. This conflicts with the philosophy of envelope methodology which is that this methodology is a class of procedures for increasing efficiency in multivariate analyses without altering traditional objectives [5, first sentence of page 1]. We see that weighting and the variable  $u$  regimes fit with the 1D algorithms are not perfect, they can lead to under coverage in small sample sizes although that under coverage largely disappears as the sample size increases. The overall finding is that our weighted envelope estimation technique provides useful variance reduction that accounts for possible model selection variability.

## 5.2. Logistic regression simulation with $p = 15$

We also perform a logistic regression simulation when  $p = 15$  and  $u = 5$ . This simulation follows a similar setup of the previous simulations where predictors are generated by  $\mathbf{X} \sim N(0, \Sigma_{\mathbf{X}})$ ,  $\Sigma_{\mathbf{X}} = \Gamma\Omega\Gamma^T + \Gamma_o\Omega_o\Gamma_o^T$ , and we construct the regression coefficient vector as  $\theta = \Gamma\Gamma^T v$  where  $\Gamma$ ,  $\theta$ , and  $v$  are provided in the supplementary materials.

The results of increasing  $p$  in this simulation are interesting and more encouraging for the fixed  $u$  regime. First of all, we see that the fixed  $u$  estimator provides extremely large variance reduction, and variance reduction obtained from the weighted and variable  $u$  regimes is much more modest. See Figure 7 for the particulars and note that we only display results for the first 12 compo-



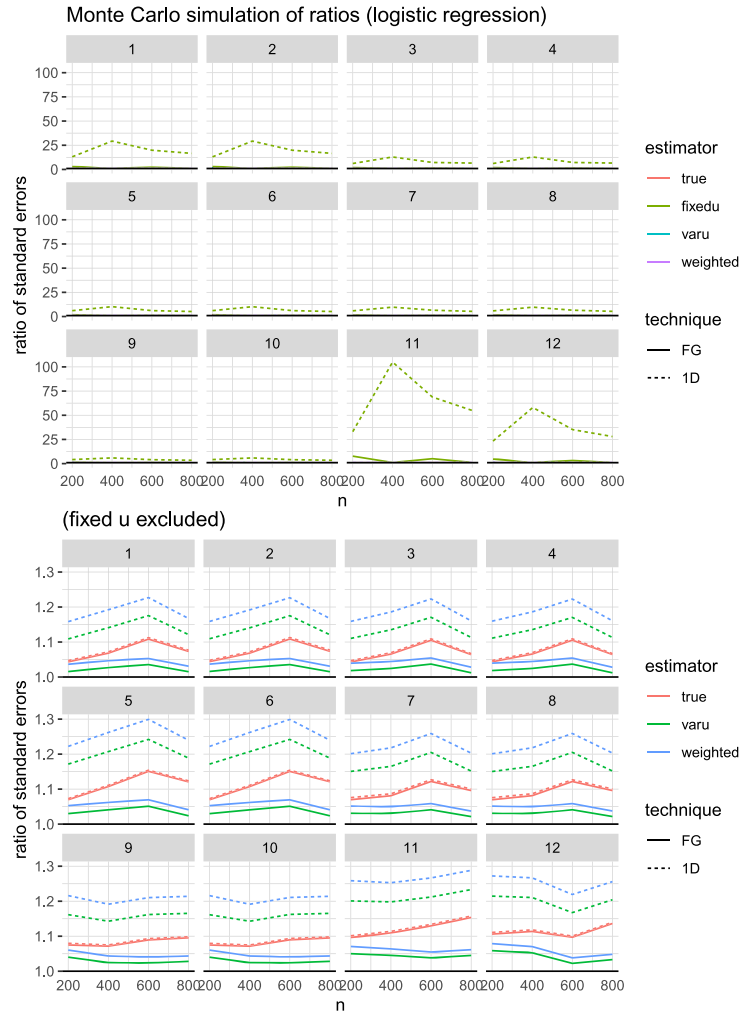


FIG 7. Averages of Monte Carlo simulations for ratios of envelope estimators to the MLE across all components for our logistic regression example in Section 5.2. Ratios greater than 1 indicate superior performance for envelope estimation.

nents for purely visual reasons. The remaining components are displayed in the supplementary materials.

As before, variance reduction provided by the fixed  $u$  regime yields subpar coverage for some of the components of the target parameter vector. Figure 8 displays each estimators coverage for the first component only. We see that the fixed  $u$  regime with estimation conducted via the 1D algorithm provides problematic under coverage. Coverage probability results for all components are included in the supplementary materials, and these materials reveal that

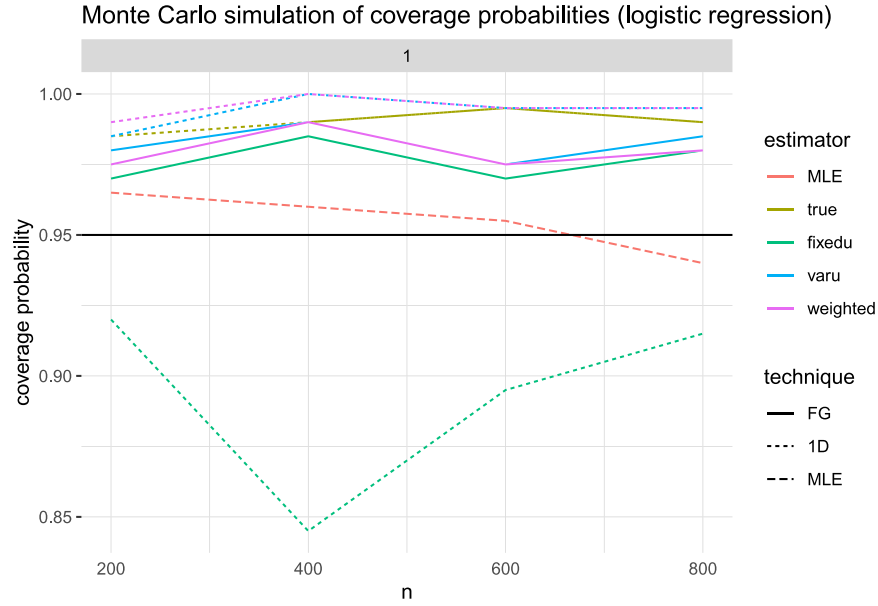


FIG 8. Coverage probabilities of all estimators averaged across Monte Carlo replicates for the first component of  $\theta$  for the logistic regression example in Section 5.2.

problematic under coverage is only observed for the first four components of the target parameter vector. Unlike before, the fixed  $u$  regime with estimation conducted via FG optimization provides large variance reduction while also yielding desirable coverage for all of the components of the target parameter vector. See the supplementary materials for these details. The tradeoffs in performances among the 1D algorithm and FG optimization for the fixed  $u$  regime likely resulted from dramatic under estimation of  $u$  when dimension estimation was conducted via the 1D algorithm.

### 5.3. 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% [25]. 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 [32], and is taken from the `faraway` R package [20]. We considered a logistic regression model with response variable denoting a diagnosis of diabetes (1 when  $\text{HbA1c} > 6.5\%$  and 0 otherwise) that includes log transformed values for each continuous covariate and a main effect

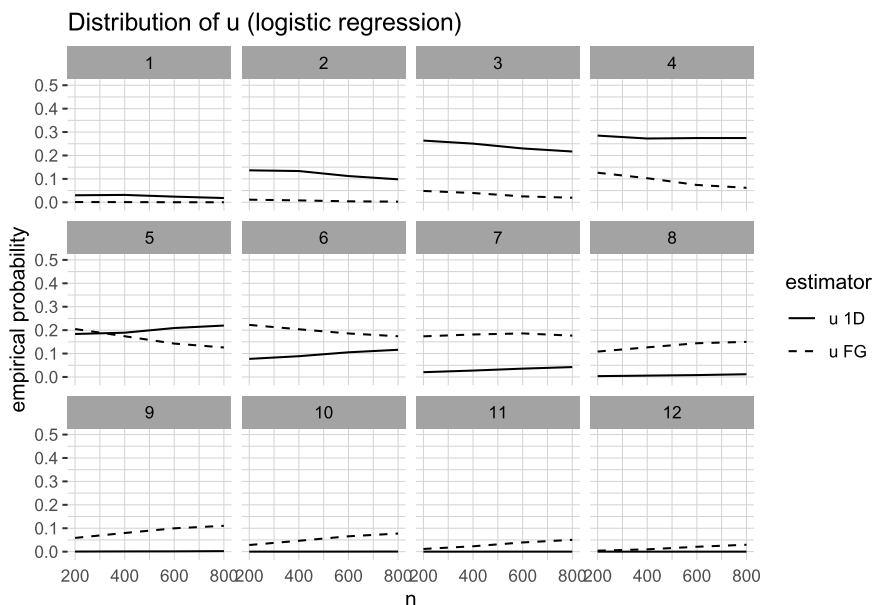


FIG 9. Empirical distribution of  $u$  averaged across Monte Carlo replicates for the logistic regression example in Section 5.2.

for gender. The log transformation was used to transform these variables to univariate normality while maintaining a scale that is interpretable.

We will compare the performance of weighted envelope estimation and both the fixed  $u$  and variable  $u$  regimes (see Table 1) using both FG optimization and the 1D algorithm. A nonparametric bootstrap with sample size 5000 is used to estimate the variability of these estimators. Performance will be assessed via variance reduction caveated with the empirical distribution of the envelope dimension. Ratios of standard deviations are displayed in Table 2. These ratios compare the bootstrapped standard deviation of each envelope estimator to the bootstrapped standard deviation of the MLE.

From Table 2 we see that 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 when estimation is performed using the 1D algorithm are  $p(\hat{u}_{1D} = 1) = 0.567$ ,  $p(\hat{u}_{1D} = 2) = 0.362$ ,  $p(\hat{u}_{1D} = 3) = 0.065$ , and  $p(\hat{u}_{1D} = 4) = 0.006$ . The empirical dimension of  $\hat{u}$  when estimation is performed using FG optimization is  $p(\hat{u}_{FG} = 1) = 0.238$ ,  $p(\hat{u}_{FG} = 2) = 0.354$ ,  $p(\hat{u}_{FG} = 3) = 0.216$ ,  $p(\hat{u}_{FG} = 4) = 0.113$ ,  $p(\hat{u}_{FG} = 5) = 0.054$ ,  $p(\hat{u}_{FG} = 6) = 0.019$ , and  $p(\hat{u}_{FG} = 7) = 0.005$ . It is clear that unaccounted model selection variability may lead users astray when they use the

TABLE 2  
The ratios of bootstrap standard deviations of all envelope estimators to the those of the MLE.

	$r(\tilde{\theta}, \hat{\theta}_{\hat{u}_{1D}}^{1D})$	$r(\tilde{\theta}, \hat{\theta}_{\hat{u}_{1D}^*}^{1D})$	$r(\tilde{\theta}, \hat{\theta}_w^{1D})$	$r(\tilde{\theta}, \hat{\theta}_{\hat{u}_{FG}}^{FG})$	$r(\tilde{\theta}, \hat{\theta}_{\hat{u}_{FG}^*}^{FG})$	$r(\tilde{\theta}, \hat{\theta}_w^{FG})$
log(Age)	1.20	1.23	1.26	1.20	1.08	1.10
log(Weight)	6.99	1.50	1.61	6.99	1.14	1.18
log(Height)	53.82	1.17	1.28	53.82	0.99	1.03
log(Waist)	11.43	1.51	1.64	11.43	1.12	1.17
log(Hip)	17.34	1.31	1.41	17.34	1.07	1.10
log(Stab. Gluc.)	1.25	1.13	1.13	1.25	1.05	1.06
Female	1.20	1.17	1.20	1.20	1.06	1.08

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.

#### 5.4. Replicating and expanding upon simulations in [36]

We now compare the performance of our weighted envelope estimators to the variable  $u$  regime using the simulation settings in [36]. For our first comparison we reproduce the Monte Carlo simulations in Section 4.2 of [36] and add both weighted estimators ( $\hat{\theta}_w^{FG}$  and  $\hat{\theta}_w^{1D}$ ) to the list of estimators under comparison in [36]. 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 and a logistic regression model with a 10 predictors, and a Cox proportional hazards model with 10 predictors. In these modeling setups, the true dimension of the envelope space is set at  $u = 2$ . In-depth details about this simulation setup are presented in [36]. The Monte Carlo sample size is 200, as in [36].

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 3.3 and Theorem 3.2 in [36]. 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 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 [36] 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

TABLE 3

Monte Carlo simulation results for different envelope estimators with respect to three different envelope models in the spirit of Table 3 from [36]. Left panel includes percentages of correct selection for these envelope estimators. Right panel includes means of  $\|\hat{\theta} - \theta\|_F$  for the standard estimator and the envelope estimators with either true or estimated dimensions.

Model	$n$	Correct Selection %		Estimation Error $\ \hat{\theta} - \theta\ _F$					
		1D	FG	Standard	Envelope true $u$	1D	FG	W1D	WFG
Linear	75	74	63.5	0.69	0.50	0.55	0.55	0.54	0.55
	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
Logistic	75	22.5	42	4.04	1.04	1.06	1.00	1.09	1.08
	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
Cox	75	35	38	2.07	1.99	1.95	1.96	2.04	2.05
	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

TABLE 4

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

Model	$n$	$r(\tilde{\theta}, \hat{\theta}_u)$	$r(\tilde{\theta}, \hat{\theta}_{u1D}^{1D})$	$r(\tilde{\theta}, \hat{\theta}_{uFG}^{FG})$	$r(\tilde{\theta}, \hat{\theta}_{u1D}^{1D})$	$r(\tilde{\theta}, \hat{\theta}_{uFG}^{FG})$	$r(\tilde{\theta}, \hat{\theta}_w^{1D})$	$r(\tilde{\theta}, \hat{\theta}_w^{FG})$
Linear	75	0.992	2.024	1.768	0.991	0.947	1.094	1.024
	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
Logistic	75	1.013	1.054	1.022	0.978	0.966	1.079	1.033
	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

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 slight but important efficiency gains over the variable  $u$  regime (see Table 1) and is comparable to oracle estimation in most settings. The fixed  $u$  regime (see Table 1) outperforms weighted envelope estimation. However, this variance reduction is due to 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

As previously mentioned, a limitation of bootstrapping weighted envelope estimators is that it can be computationally expensive, especially when  $p$  is large [33]. Parallel computing with several cores can alleviate these computations since both estimation for different envelope dimensions and the bootstrap procedure can be done in parallel. As an example, our largest simulation configuration, which bootstraps all candidate envelope estimators and the MLE with  $p = 15$ ,  $n = 800$ , and 500 bootstrap iterations, can be executed in less than 3 hours on a 20 core machine. In such settings where computational costs are burdensome even with parallel computing, 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  $u$  regime as in [36] 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 [23]. Our simulations provide some empirical justification for the performance of these approaches.

[33] developed a novel hypothesis testing procedure with respect to the multivariate linear envelope model. They showed that model averaging in [16] 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  $\hat{\theta}_w, \dots$  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 [5, 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.

## Appendix

### Mathematical results

*Proof of Lemma 2.* We first show that  $w_u^{\text{FG}} \xrightarrow{P} 1$ . Note that

$$w_k^{\text{FG}} = \frac{\exp\{-n\mathcal{I}_n^{\text{FG}}(k)\}}{\sum_{j=0}^p \exp\{-n\mathcal{I}_n^{\text{FG}}(j)\}} = \frac{\exp[n\{\mathcal{I}_n^{\text{FG}}(u) - \mathcal{I}_n^{\text{FG}}(k)\}]}{\sum_{j=0}^p \exp[n\{\mathcal{I}_n^{\text{FG}}(u) - \mathcal{I}_n^{\text{FG}}(j)\}]}.$$

By definition of  $\mathcal{I}_n^{\text{FG}}(k)$ , we have that

$$n\{\mathcal{I}_n^{\text{FG}}(k) - \mathcal{I}_n^{\text{FG}}(u)\} = n\{J_n(\hat{\Gamma}_k) - J_n(\hat{\Gamma}_u)\} + C(k - u)\log(n). \quad (\text{A.1})$$

We show that  $w_k^{\text{FG}} \xrightarrow{P} 0$  as  $n \rightarrow \infty$  for all  $k \neq u$  by following a similar argument as the proof of Theorems 3.1 and 3.2 in [36]. Lemma 3.2 in [36] states that  $J(\Gamma_u) < J(\Gamma_k) < 0$  for all  $k = 0, \dots, u-1$ , and  $J(\Gamma_k) = J(\Gamma_u)$  for all  $k = u+1, \dots, p$ . First suppose that  $k = 0, \dots, u-1$ . In this setting, we have that (A.1) tends to  $\infty$  as  $n \rightarrow \infty$ . Now suppose that  $k = u, \dots, p$ . In this setting, we have that  $n \{J_n(\hat{\Gamma}_k) - J_n(\hat{\Gamma}_u)\} = O_P(1)$  in (A.1). Therefore (A.1) tends to  $\infty$  as  $n \rightarrow \infty$  when  $k = u+1, \dots, p$ . Putting this together implies that  $w_k^{\text{FG}} \xrightarrow{P} 0$  for all  $k \neq u$  and  $w_u^{\text{FG}} \xrightarrow{P} 1$  as  $n \rightarrow \infty$ .

We now show that  $w_u^{\text{ID}} \xrightarrow{P} 1$ . Note that

$$w_k^{\text{ID}} = \frac{\exp\{-n\mathcal{I}_n^{\text{ID}}(k)\}}{\sum_{j=0}^p \exp\{-n\mathcal{I}_n^{\text{ID}}(j)\}} = \frac{\exp[n\{\mathcal{I}_n^{\text{ID}}(u) - \mathcal{I}_n^{\text{ID}}(k)\}]}{\sum_{j=0}^p \exp[n\{\mathcal{I}_n^{\text{ID}}(u) - \mathcal{I}_n^{\text{ID}}(j)\}]} \quad (\text{A.2})$$

We show that  $w_k^{\text{ID}} \xrightarrow{P} 0$  as  $n \rightarrow \infty$  for all  $k \neq u$  by following a similar argument as the proof of Theorems 3.1 and 3.2 in [36]. First suppose that  $k > u$  and observe that

$$n\{\mathcal{I}_n^{\text{ID}}(u) - \mathcal{I}_n^{\text{ID}}(k)\} = n\left\{\sum_{j=u+1}^k \phi_{j,n}(\hat{v}_j) + \frac{C(u-k)\log(n)}{n}\right\}.$$

We have that  $\phi_{j,n}(\hat{v}_j) = O_P(n^{-1})$ . Therefore

$$n\{\mathcal{I}_n^{\text{ID}}(u) - \mathcal{I}_n^{\text{ID}}(k)\} \rightarrow -\infty$$

as  $n \rightarrow \infty$ . From (A.2) we can conclude that  $w_k^{\text{ID}} \xrightarrow{P} 0$  as  $n \rightarrow \infty$  for all  $k > u$ .

Now suppose that  $k < u$ . Let

$$n\{\mathcal{I}_n^{\text{ID}}(u) - \mathcal{I}_n^{\text{ID}}(k)\} = n\left\{\sum_{j=k+1}^u \phi_{j,n}(\hat{v}_j) + \frac{C(u-k)\log(n)}{n}\right\}.$$

The function  $\phi_{j,n}(\hat{v}_j) \rightarrow \phi_j(v_j) < 0$  in probability as shown in the proof of Propositions 5 and 6 in [11]. Therefore  $n\{\mathcal{I}_n^{\text{ID}}(u) - \mathcal{I}_n^{\text{ID}}(k)\} \rightarrow -\infty$  as  $n \rightarrow \infty$ .

From (A.2) we can conclude that  $w_k^{\text{ID}} \xrightarrow{P} 0$  as  $n \rightarrow \infty$  for all  $k < u$ . Therefore  $w_k^{\text{ID}} \xrightarrow{P} 0$  as  $n \rightarrow \infty$  for all  $k \neq u$  which implies that  $w_u^{\text{ID}} \xrightarrow{P} 1$  as  $n \rightarrow \infty$ .  $\square$

We now proceed with the bootstrap Theorems in Section 4 of the main manuscript. Following Remark 1 in Section 3 of [3], we define  $O_P^*$  as an analog to  $O_P$  that is conditional on the original sample.

*Proof of Theorem 1.* Notice that

$$\sqrt{n}\left(\hat{\theta}_w^{\text{FG}*} - \hat{\theta}_w^{\text{FG}}\right) = \sqrt{n}\left(w_u^{\text{FG}*}\hat{\theta}_u^{\text{FG}*} - w_u^{\text{FG}}\hat{\theta}_u^{\text{FG}}\right)$$

$$\begin{aligned}
& + \sqrt{n} \left( \sum_{k \neq u}^p w_k^{\text{FG}*} \widehat{\theta}_k^{\text{FG}*} - \sum_{k \neq u}^p w_k^{\text{FG}} \widehat{\theta}_k^{\text{FG}} \right) \\
& = \sqrt{n} \left( \widehat{\theta}_u^{\text{FG}*} - \widehat{\theta}_u^{\text{FG}} \right) \\
& + \sqrt{n} \left\{ \sum_{k \neq u}^p w_k^{\text{FG}*} \left( \widehat{\theta}_k^{\text{FG}*} - \widehat{\theta}_u^{\text{FG}*} \right) - \sum_{k \neq u}^p w_k^{\text{FG}} \left( \widehat{\theta}_k^{\text{FG}} - \widehat{\theta}_u^{\text{FG}} \right) \right\}
\end{aligned}$$

We show that  $w_k^{\text{FG}}, w_k^{\text{FG}*} \rightarrow 0$  for all  $k \neq u$  such that

$$\begin{aligned}
& \sqrt{n} \left\| \sum_{k \neq u}^p w_k^{\text{FG}*} \left( \widehat{\theta}_k^{\text{FG}*} - \widehat{\theta}_u^{\text{FG}*} \right) - \sum_{k \neq u}^p w_k^{\text{FG}} \left( \widehat{\theta}_k^{\text{FG}} - \widehat{\theta}_u^{\text{FG}} \right) \right\| \\
& \leq \sum_{k \neq u}^p \left( \sqrt{n} w_k^{\text{FG}*} \left\| \widehat{\theta}_k^{\text{FG}*} - \widehat{\theta}_u^{\text{FG}*} \right\| + \sqrt{n} w_k^{\text{FG}} \left\| \widehat{\theta}_k^{\text{FG}} - \widehat{\theta}_u^{\text{FG}} \right\| \right) \rightarrow 0
\end{aligned}$$

as  $n \rightarrow \infty$  where the rates of the bound are given by the displayed equation in Theorem 1 of the main text. We have that

$$\begin{aligned}
\sqrt{n} w_k^{\text{FG}} \left\| \widehat{\theta}_k^{\text{FG}} - \widehat{\theta}_u^{\text{FG}} \right\| & = \frac{\sqrt{n} \exp \{ -n \mathcal{I}_n^{\text{FG}}(k) \}}{\sum_{j=0}^p \exp \{ -n \mathcal{I}_n^{\text{FG}}(j) \}} \left\| \widehat{\theta}_j^{\text{FG}} - \widehat{\theta}_u^{\text{FG}} \right\| \\
& \leq \sqrt{n} \exp \{ n \mathcal{I}_n^{\text{FG}}(u) - n \mathcal{I}_n^{\text{FG}}(k) \} \left\| \widehat{\theta}_k^{\text{FG}} - \widehat{\theta}_u^{\text{FG}} \right\| \\
& = O_P^* (\sqrt{n}) \exp [n \{ \mathcal{I}_n^{\text{FG}}(u) - \mathcal{I}_n^{\text{FG}}(k) \}] \\
& = O_P^* (\sqrt{n}) \exp [n \{ J_n(\widehat{\Gamma}_u) - J_n(\widehat{\Gamma}_k) \} + C(u-k) \log(n)] \\
& = O_P^* (n^{C(u-k)+1/2}) \exp [n \{ J_n(\widehat{\Gamma}_u) - J_n(\widehat{\Gamma}_k) \}].
\end{aligned} \tag{A.3}$$

The same steps as (A.3) yield

$$\begin{aligned}
& \sqrt{n} w_k^{\text{FG}*} \left\| \widehat{\theta}_k^{\text{FG}*} - \widehat{\theta}_u^{\text{FG}*} \right\| \\
& \leq O_P^* (n^{C(u-k)+1/2}) \exp [n \{ J_n^*(\widehat{\Gamma}_u^*) - J_n^*(\widehat{\Gamma}_k^*) \}].
\end{aligned} \tag{A.4}$$

For  $0 \leq k < u$ , we have that  $J_n(\widehat{\Gamma}_u) - J_n(\widehat{\Gamma}_k) = J(\Gamma_u) - J(\Gamma_k) + o_p(1)$  where  $J(\Gamma_u) < J(\Gamma_k)$  as in the proof of [36, Theorem 3.1]. Similarly we have that

$$J_n^*(\widehat{\Gamma}_u^*) - J_n^*(\widehat{\Gamma}_k^*) = J(\Gamma_u) - J(\Gamma_k) + o_P^*(1).$$

Therefore the rates for the exponent in the last line of (A.3) and the right hand side of (A.4) are  $-n|O_P^*(1)|$ . Notice that the rates in the last line of (A.3) and the right hand side of (A.4) are upper bounded when  $k = 0$ . Putting this together yields

$$\sqrt{n} w_k^{\text{FG}} \left\| \widehat{\theta}_k^{\text{FG}} - \widehat{\theta}_u^{\text{FG}} \right\| = O_P^* (n^{C_u+1/2}) \exp \{ -n|O_P(1)| \},$$



$$\sqrt{n}w_k^{\text{FG}*} \|\hat{\theta}_k^{\text{FG}*} - \hat{\theta}_u^{\text{FG}*}\| = O_P^* \left( n^{Cu+1/2} \right) \exp \left\{ -n|O_P^*(1)| \right\};$$

for all  $0 \leq k < u$ .

Now consider  $u < k \leq p$ . From the proof of [36, Theorem 3.1] we have that  $J_n(\hat{\Gamma}_u) - J_n(\hat{\Gamma}_k) = O_P^*(n^{-1})$ . Combining this result with the steps in (A.3) yields

$$\sqrt{n}w_k^{\text{FG}} \|\hat{\theta}_k^{\text{FG}} - \hat{\theta}_u^{\text{FG}}\| \leq O_P^* \left( n^{C(u-k)+1/2} \right). \quad (\text{A.5})$$

A similar argument applied to the starred data gives

$$\sqrt{n}w_k^{\text{FG}*} \|\hat{\theta}_k^{\text{FG}*} - \hat{\theta}_u^{\text{FG}*}\| \leq O_P^* \left( n^{C(u-k)+1/2} \right). \quad (\text{A.6})$$

The rates in both (A.5) and (A.6) are upper bounded when  $k = u - 1$ . Putting this together yields

$$\begin{aligned} \sqrt{n}w_k^{\text{FG}} \|\hat{\theta}_k^{\text{FG}} - \hat{\theta}_u^{\text{FG}}\| &= O_P^* \left( n^{1/2-C} \right), \\ \sqrt{n}w_k^{\text{FG}*} \|\hat{\theta}_k^{\text{FG}*} - \hat{\theta}_u^{\text{FG}*}\| &= O_P^* \left( n^{1/2-C} \right); \end{aligned}$$

for all  $u < k \leq p$ . Therefore

$$\begin{aligned} \sqrt{n} \left\{ \sum_{k \neq u}^p w_k^{\text{FG}*} \left( \hat{\theta}_k^{\text{FG}*} - \hat{\theta}_u^{\text{FG}} \right) - \sum_{k \neq u}^p w_k^{\text{FG}} \left( \hat{\theta}_k^{\text{FG}} - \hat{\theta}_u^{\text{FG}} \right) \right\} \\ = O_P^* \left( n^{1/2-C} \right) + O_P^* \left( n^{Cu+1/2} \right) \exp \left\{ -n|O_P^*(1)| \right\}, \end{aligned}$$

as desired and the conclusion follows.  $\square$

*Proof of Theorem 2.* Notice that

$$\begin{aligned} \sqrt{n} \left( \hat{\theta}_w^{\text{1D}*} - \hat{\theta}_w^{\text{1D}} \right) &= \sqrt{n} \left( w_u^* \hat{\theta}_u^{\text{1D}*} - w_u \hat{\theta}_u^{\text{1D}} \right) + \sqrt{n} \left( \sum_{k \neq u}^p w_k^* \hat{\theta}_k^* - \sum_{k \neq u}^p w_k \hat{\theta}_k \right) \\ &= \sqrt{n} \left( \hat{\theta}_u^* - \hat{\theta}_u \right) + \sqrt{n} \left\{ \sum_{k \neq u}^p w_k^* \left( \hat{\theta}_k^* - \hat{\theta}_u^* \right) - \sum_{k \neq u}^p w_k \left( \hat{\theta}_k - \hat{\theta}_u \right) \right\}. \end{aligned}$$

We show that  $w_k, w_k^* \rightarrow 0$  such that

$$\begin{aligned} \sqrt{n} \left\| \sum_{k \neq u}^p w_k^* \left( \hat{\theta}_k^* - \hat{\theta}_u^* \right) - \sum_{k \neq u}^p w_k \left( \hat{\theta}_k - \hat{\theta}_u \right) \right\| \\ \leq \sum_{k=1}^p \left( \sqrt{n}w_k^* \|\hat{\theta}_k^* - \hat{\theta}_u^*\| + \sqrt{n}w_k \|\hat{\theta}_k - \hat{\theta}_u\| \right) \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$  for all  $k \neq u$  and find the rates at which they vanish. We have that

$$\begin{aligned}
\sqrt{nw_k} \|\hat{\theta}_k - \hat{\theta}_u\| &= \frac{\sqrt{n} \exp \{-n\mathcal{I}_n^{1D}(k)\}}{\sum_{j=0}^p \exp \{-n\mathcal{I}_n^{1D}(j)\}} \|\hat{\theta}_k - \hat{\theta}_u\| \\
&\leq \sqrt{n} \exp \{n\mathcal{I}_n^{1D}(u) - n\mathcal{I}_n^{1D}(k)\} \|\hat{\theta}_k - \hat{\theta}_u\| \\
&= \sqrt{n} \exp \left\{ n \sum_{j=1}^u \phi_{j,n}(\hat{v}_j) - n \sum_{j=1}^k \phi_{j,n}(\hat{v}_j) + (u-k) \frac{C \log n}{n} \right\} \|\hat{\theta}_k - \hat{\theta}_u\| \\
&= n^{\{C(u-k)+1/2\}} \exp \left\{ n \sum_{j=1}^u \phi_{j,n}(\hat{v}_j) - n \sum_{j=1}^k \phi_{j,n}(\hat{v}_j) \right\} \|\hat{\theta}_k - \hat{\theta}_u\| \\
&= O_P^* \left[ n^{\{C(u-k)+1/2\}} \right] \exp \left\{ n \sum_{j=1}^u \phi_{j,n}(\hat{v}_j) - n \sum_{j=1}^k \phi_{j,n}(\hat{v}_j) \right\}
\end{aligned} \tag{A.7}$$

where the last equality follows from the fact that  $\|\hat{\theta}_k - \hat{\theta}_u\| = |O_P^*(1)|$  for all  $k = 1, \dots, p$ . This is because  $\hat{\theta}_k \rightarrow \theta$  for all  $k = u, \dots, p$  and  $\|\hat{\theta}_k\| \rightarrow a \leq \|\theta\|$  for all  $k = 1, \dots, u-1$  since the envelope estimator exhibits shrinkage when  $k = 1, \dots, u-1$ . First suppose that  $k = 1, \dots, u-1$ . In this setting  $\phi_{k,n}(\hat{v}_k) \rightarrow \phi_k(v_k) < 0$  as  $n \rightarrow \infty$  [11, proof of Theorems 5 and 6]. From (A.7) we have

$$\begin{aligned}
\sqrt{nw_k} \|\hat{\theta}_k - \hat{\theta}_u\| &\leq O_P^* \left[ n^{\{C(u-k)+1/2\}} \right] \exp \left\{ n \sum_{j=k-1}^u \phi_{j,n}(\hat{v}_j) \right\} \\
&= O_P^* \left[ n^{\{C(u-k)+1/2\}} \right] \exp \{ -n|O_P^*(1)| \}.
\end{aligned} \tag{A.8}$$

Now suppose that  $k = u+1, \dots, p$ . In this setting,  $\phi_{k,n}(\hat{v}_k) = O_P^*(n^{-1})$  [36, proof of Theorem 3.1]. From (A.7) we have

$$\begin{aligned}
\sqrt{nw_k} \|\hat{\theta}_k - \hat{\theta}_u\| &\leq O_P^* \left[ n^{\{C(u-k)+1/2\}} \right] \exp \left\{ -n \sum_{j=u+1}^k \phi_{j,n}(\hat{v}_j) \right\} \\
&= O_P^* \left[ n^{\{C(u-k)+1/2\}} \right].
\end{aligned} \tag{A.9}$$

The same steps in (A.8) and (A.9) apply to the starred data so that

$$\sqrt{nw_k^*} \|\hat{\theta}_k^* - \hat{\theta}_u^*\| \leq O_P^* \left[ n^{\{C(u-k)+1/2\}} \right] \exp \{ -n|O_P^*(1)| \}, \quad (k = 1, \dots, u-1), \tag{A.10}$$

and

$$\sqrt{nw_k^*} \|\hat{\theta}_k^* - \hat{\theta}_u^*\| = O_P^* \left[ n^{\{C(u-k)+1/2\}} \right], \quad (k = u+1, \dots, p). \tag{A.11}$$

Our conclusion follows by noting that (A.8), (A.9), (A.10), and (A.11) implies that

$$\begin{aligned}\sqrt{nw_k}\|\hat{\theta}_k - \hat{\theta}_u\| &\leq O_P^* \left[ n^{\{Cu+1/2\}} \right] \exp \{ -n|O_P^*(1)| \}, & (k = 1, \dots, u-1); \\ \sqrt{nw_k}\|\hat{\theta}_k - \hat{\theta}_u\| &\leq O_P^* \left\{ n^{(1/2-C)} \right\}, & (k = u+1, \dots, p); \\ \sqrt{nw_k^*}\|\hat{\theta}_k^* - \hat{\theta}_u^*\| &\leq O_P^* \left[ n^{\{Cu+1/2\}} \right] \exp \{ -n|O_P^*(1)| \}, & (k = 1, \dots, u-1); \\ \sqrt{nw_k^*}\|\hat{\theta}_k^* - \hat{\theta}_u^*\| &\leq O_P^* \left\{ n^{(1/2-C)} \right\}, & (k = u+1, \dots, p);\end{aligned}$$

respectively.  $\square$

## Acknowledgements

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. The author greatly appreciated three anonymous referees and an anonymous associate editor whose insightful comments led to a much improved paper.

## Supplementary materials

Supplementary materials are available at

<https://github.com/DEck13/general-weighted-envelope-supplement>

This GitHub repository contains a reproducible technical report that makes R based analyses transparent. The simulations in Section 5.4 are not included in the supplementary materials. These simulations are adopted from Matlab code that accompanied [36]. This code is available upon request.

## References

- [1] BUCKLAND, S. T., BURNHAM, K. P. and AUGUSTIN, N. H. (1997). Model Selection: An Integral Part of Inference. *Biometrics* **53** 603–618.
- [2] BURNHAM, K. P. and ANDERSON, D. R. (2004). Multimodel Inference. *Sociological and Methods Research* **33** 261–304. [MR2086350](#)
- [3] CHANG, Y. and PARK, J. Y. (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis* **24** 379–400. [MR1997120](#)
- [4] CLAESKENS, G. and HJORT, N. L. (2008). *Model Selection and Model Averaging*. Cambridge University Press, Cambridge. [MR2431297](#)
- [5] COOK, R. D. (2018). *An introduction to envelopes: dimension reduction for efficient estimation in multivariate statistics* **401**. John Wiley & Sons. [MR3774758](#)
- [6] COOK, R. D., FORZANI, L. and SU, Z. (2016). A Note on Fast Envelope Estimation. *Journal of Multivariate Analysis* **150** 42–54. [MR3534901](#)

- [7] COOK, R. D., LI, B. and CHIAROMONTE, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica* **20** 927–1010. [MR2729841](#)
- [8] COOK, R. D. and SU, Z. (2013). Scaled envelopes: scale-invariant and efficient estimation in multivariate linear regression. *Biometrika* **100** 939–954. [MR3142342](#)
- [9] COOK, R. D. and SU, Z. (2016). Scaled Predictor Envelopes and Partial Least Squares Regression. *Technometrics* **58** 155–165. [MR3488295](#)
- [10] COOK, R. D. and ZHANG, X. (2015). Foundations for Envelope Models and Methods. *Journal of the American Statistical Association* **110** 599–611. [MR3367250](#)
- [11] COOK, R. D. and ZHANG, X. (2016). Algorithms for Envelope Estimation. *Journal of Computational and Graphical Statistics* **25** 284–300. [MR3474048](#)
- [12] COOK, R. D. and ZHANG, X. (2018). Fast envelope algorithms. *Statistica Sinica* **28** 1179–1197. [MR3821000](#)
- [13] DING, S. and COOK, R. D. (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B* **80** 387–408. [MR3763697](#)
- [14] DING, S. and COOK, R. D. (2018). Supplement: Proofs, Technical Details and additional results for “Matrix-variate regressions and envelope models”. *Journal of the Royal Statistical Society: Series B* **80** 387–408. [MR3763697](#)
- [15] ECK, D. J. (2018). Bootstrapping for multivariate linear regression models. *Statistics and Probability Letters* **134** 141–149. [MR3758593](#)
- [16] ECK, D. J. and COOK, R. D. (2017). Weighted envelope estimation to handle variability in model selection. *Biometrika* **104** 743–749. [MR3694595](#)
- [17] ECK, D. J., GEYER, C. J. and COOK, R. D. (2018). Supporting Data Analysis for “Combining Envelope Methodology and Aster Models for Variance Reduction in Life History Analyses”. <http://hdl.handle.net/11299/178384.1>. [MR4011637](#)
- [18] ECK, D. J., GEYER, C. J. and COOK, R. D. (2020). Combining Envelope Methodology and Aster Models for Variance Reduction in Life History Analyses. *Journal of Statistical Planning and Inference* **205** 283–292. [MR4011637](#)
- [19] EFRON, B. (2014). Estimation and Accuracy After Model Selection. *Journal of the American Statistical Association* **109**:507 991–1007. [MR3265671](#)
- [20] FARAWAY, J. (2016). faraway: Functions and Datasets for Books by Julian Faraway <https://cran.r-project.org/web/packages/faraway/index.html>.
- [21] GEYER, C. J., WAGENIUS, S. and SHAW, R. G. (2007). Aster models for life history analysis. *Biometrika* **94** 415–426. [MR2380569](#)
- [22] HJORT, N. L. and CLAESKENS, G. (2003). Frequentist Model Average Estimators. *Journal of the American Statistical Association* **98**:464 879–899. [MR2041481](#)
- [23] LEE, M. and SU, Z. (2019). Renvlp: Computing Envelope Estimators <http://cran.radicaldevelop.com/web/packages/Renvlp/Renvlp.pdf>.

- [24] LI, L. and ZHANG, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association* **112** 1131–1146. [MR3735365](#)
- [25] WORLD HEALTH ORGANIZATION (2011). Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation Technical Report, Geneva: World Health Organization.
- [26] SHAW, R. G., GEYER, C. J., WAGENIUS, S., HANGELBROEK, H. and ETTERSON, J. R. (2008). Unifying life-history analyses for inference of fitness and population growth. *American Naturalist* **172** E35–E47.
- [27] SU, Z. and COOK, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98** 133–146. [MR2804215](#)
- [28] SU, Z. and COOK, R. D. (2012). Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika* **99** 687–702. [MR2966778](#)
- [29] SU, Z., ZHU, G., CHEN, X. and YANG, Y. (2016). Sparse Envelope Model: Efficient Estimation and Response Variable Selection in Multivariate Linear Regression. *Biometrika* **103** 579–593. [MR3551785](#)
- [30] TSAGUE, G. N. (2014). On Optimal Weighting Scheme in Model Averaging. *American Journal of Applied Mathematics and Statistics* **2(3)** 150–156.
- [31] WANG, W., ZENG, J. and ZHANG, X. (2021). TRES: Tensor Regression with Envelope Structure, version 1.1.4 <https://cran.r-project.org/web/packages/TRES/index.html>.
- [32] WILLEMS, J. P., SAUNDERS, J. T., HUNT, D. E. and SCHORLING, J. B. (1997). Prevalence of coronary heart disease risk factors among rural blacks: a community-based study. *Southern Medical Journal* **90** 814–820.
- [33] YAU, T. S. and VOLAUFOVA, J. (2019). A note on hypothesis testing in the envelope model. *Communications in Statistics*. DOI: 10.1080/03610918.2019.1672737. [MR4068477](#)
- [34] ZENG, J., WANG, W. and ZHANG, Z. (2020). TRES: An R Package for Tensor Regression and Envelope Algorithms. *Journal of Statistical Software*.
- [35] ZHANG, X., LEE, C. E. and SHAO, X. (2020). Envelopes in Multivariate Regression Models with Nonlinearity and Heteroscedasticity. *Biometrika*. In Press. [MR4186499](#)
- [36] ZHANG, X. and MAI, Q. (2018). Model-free Envelope Dimension Selection. *Electronic Journal of Statistics* **12** 2193–2216. [MR3829139](#)