The Annals of Statistics
2021, Vol. 49, No. 1, 49–75
https://doi.org/10.1214/19-AOS1908

CONCORDANCE AND VALUE INFORMATION CRITERIA FOR OPTIMAL TREATMENT DECISION

BY CHENGCHUN SHI[1], RUI SONG[2,*] AND WENBIN LU[2,†]

[1]*Department of Statistics, London School of Economics, C.Shi7@lse.ac.uk*
[2]*Department of Statistics, North Carolina State University, *rsong@ncsu.edu; †wlu4@ncsu.edu*

Personalized medicine is a medical procedure that receives considerable scientific and commercial attention. The goal of personalized medicine is to assign the optimal treatment regime for each individual patient, according to his/her personal prognostic information. When there are a large number of pretreatment variables, it is crucial to identify those important variables that are necessary for treatment decision making. In this paper, we study two information criteria: the concordance and value information criteria, for variable selection in optimal treatment decision making. We consider both fixed-p and high dimensional settings, and show our information criteria are consistent in model/tuning parameter selection. We further apply our information criteria to four estimation approaches, including robust learning, concordance-assisted learning, penalized A-learning and sparse concordance-assisted learning, and demonstrate the empirical performance of our methods by simulations.

1. Introduction. Personalized medicine is a medical procedure that receives considerable scientific and commercial attention. The goal of personalized medicine is to assign the optimal treatment regime for each individual patient, according to his/her personal information, such as a patient's genetic content, clinical response and demographic characteristics, etc. A treatment regime is a decision rule that assigns treatments to patients based on their observed covariates. Among the set of all possible treatment regimes, the one that optimize patients' expected outcomes of interest is referred to as the optimal treatment regime. Classical methods for estimating optimal treatment regime include Q-learning (Watkins and Dayan (1992), Chakraborty, Murphy and Strecher (2010)) and A-learning (Murphy (2003), Robins, Hernan and Brumback (2000)). Recently, many authors proposed to estimate the optimal treatment regime by directly maximizing the estimated expected outcome, that is, the value function. References include Zhang et al. (2012, 2013), Zhao et al. (2012, 2015). In addition, Fan et al. (2017) introduced a type of concordance function for prescribing treatment and proposed a concordance-assisted learning for estimating the optimal treatment regime.

When there are many pretreatment variables, how to organize and use these variables for treatment decision making becomes a big challenge. This makes it clinically important to implement the variable selection technique in personalized medicine. There are a large amount of works considering variable selection in linear and generalized linear models (GLMs) in the literature (see discussions in Fan and Lv (2010)). Parameters in the model can be consistently estimated even in the ultrahigh dimension where the number of covariates p grows exponentially fast with respect to the sample size n. However, the literature on estimating the optimal treatment regime in high dimension is scarce, especially when p is much larger than n. For a single stage study, Qian and Murphy (2011) proposed to construct the optimal

Received March 2017; revised August 2019.
MSC2020 subject classifications. 62E99.
Key words and phrases. Concordance and value information criteria, optimal treatment regime, tuning parameter selection, variable selection.

treatment regime by estimating the conditional mean of the response given treatment of predictors with l_1 penalty function. Lu, Zhang and Zeng (2013) proposed a convenient loss-based framework for variable selection. Liang et al. (2017) proposed a sparse concordance-assisted learning algorithm. For multiple treatment decision points, Shi et al. (2018) proposed a high-dimensional A-learning method which estimates the optimal treatment regime by solving penalized A-learning estimating equations. All these regularization methods require appropriate choices of the tuning parameters.

Akaike information criterion (AIC, Akaike (1973)) and Bayes information criterion (BIC, Schwarz (1978)) are widely applied to linear models and generalized linear models. In the ultrahigh dimension, Fan and Tang (2013) proposed a generalized information criterion (GIC) and showed its model selection consistency. These information criteria are all constructed based on the likelihood function. However, the optimal treatment regime is usually estimated by some semiparametric or nonparametric methods. These methods are typically not likelihood-based. An alternative approach is to consider information criterion constructed by an empirical objective function, such as the information criterion proposed by Zhang et al. (2016) for support vector machines (SVMIC$_H$). However, how to derive meaningful and suitable information criteria for selecting important covariates for optimal treatment decision remains challenging. Shi et al. (2018) used a BIC-type criterion to select tuning parameters for their estimation methods. However, there is no theoretical guarantees for the BIC procedure.

In this paper, we consider model selection and tuning parameter selection for estimating the optimal treatment regime. Specifically, we propose value information criterion (VIC) and concordance information criterion (CIC) for model selection. VIC and CIC are constructed based on the empirical estimators for the value (Zhang et al. (2012)) and the concordance function (Fan et al. (2017)), respectively. The concordance function stands for the average difference of the benefit in receiving a treatment for two patients, if one is more likely to be assigned to this treatment compared to another under a given regime.

There are several technical challenges for establishing the asymptotic properties of the proposed information criteria. Different from AIC, BIC, GIC and SVMIC$_H$ that rely heavily on the smoothness of the log-likelihood function and Lipschitz continuity of the loss function, the empirical value and concordance functions involve indicator functions that are neither continuous nor concave. In addition, the derivation of the asymptotic properties of the proposed information criteria is further complicated due to the curse of dimensionality. For example, the estimated concordance function is a U-process of order two. In the fixed-p scenario, applying the maximal inequality for degenerate U-process (cf. Nolan and Pollard (1987), Sherman (1994)), it can be uniformly approximated by a smooth function with the approximation error $O(1/n)$. Such results no longer hold when $p \gg n$.

The contributions of this paper are summarized as follows. First, a more general class of models is considered. More specifically, in this paper, we assume a monotonic linear index model for the contrast function. In contrast, previous work on variable selection for optimal treatment regime mainly assume a linear interaction for the contrast (cf. Lu, Zhang and Zeng (2013), Shi et al. (2018)). Other information criteria such as AIC, BIC and GIC focus on linear models or GLMs where the link function needs to be specified.

Second, we not only establish the consistency of our proposed information criteria, but also provide upper bounds for the probabilities that VIC or CIC chooses an underfitted model and an overfitted model. To the best of our knowledge, such type of nonasymptotic bounds are rarely established for other information criteria previously. Proofs of our major theorems (Theorems 3.3 and 3.4) rely on some newly developed empirical process and U-process techniques, which are important in their own rights. First, we provide a Bernstein-type concentration inequality (Theorem 7.1) for the unbounded degenerate U-process. Our theorem

generalizes existing results and relaxes classical assumptions that require the underlying class of functions to be bounded (Arcones (1995), Clémençon, Lugosi and Vayatis (2008), Li, Ren and Li (2014)). In addition, we develop the tail inequalities and uniform consistencies of empirical maximizers of the estimated value and concordance functions (Lemma 7.1) that are useful to show selection consistencies of VIC and CIC. This is a very challenging task due to the nonsmoothness of the objective function and curse of dimensionality.

Third, our proposed information criteria are generally applicable and are not specifically tailored to certain estimating procedures. For any estimation procedure, as long as the true model can be recovered and the estimator satisfies certain convergence rates, we show that VIC and CIC are consistent, in both fixed-p and ultrahigh dimension cases. Specifically, we apply our information criteria to four estimation approaches, including robust learning (Zhang et al. (2012)), concordance-assisted learning (Fan et al. (2017)), penalized A-learning (Shi et al. (2018)) and sparse concordance-assisted learning (Liang et al. (2017)), and demonstrate that our information criteria are able to achieve consistent model/tunning parameter selection in these examples.

We briefly summarize our key findings here. Comparatively speaking, CIC is more reliable than VIC in model selection, although both criteria are consistent. In our numerical experiments, CIC achieves smaller false negative and false positive when compared with VIC. In our theoretical results, conditions to ensure model selection consistency for VIC are more restrictive than those for CIC. Moreover, the probability that CIC chooses a wrong model decays much faster than that of VIC, under certain cases. This is because the estimated concordance function in CIC is a U-process of order two, and is more "smooth" than the estimated value function in VIC, which is an empirical process that involves summation of indicator functions.

The rest of the article is organized as follows. We introduce VIC and CIC in Section 2. Consistencies of these criteria in selecting variables for optimal treatment decision are presented in Section 3. In Section 4, we introduce doubly-robust versions of VIC and CIC and investigate their properties. In Section 5, we apply our information criteria to four approaches for estimating the optimal treatment regime. Simulation studies are conducted in Section 6. Some technical results are provided in Section 7, with the detailed derivations provided in Section 9 and a Supplementary Material. Finally, we conclude our paper by a discussion section.

2. Concordance and value-based information criteria.

2.1. *Model setup and notation*. We only consider a single stage study with two treatments to illustrate the idea. Let Y_0 be a patient's response of interest, $A_0 \in \{0, 1\}$ denote the treatment a patient receives, and $X_0 \in \mathbb{R}^p$ denote the patient's baseline covariates. By convention, a larger value of Y_0 indicates a better clinical outcome. The number of covariates p is allowed to increase with n and can be potentially much larger than n.

The optimal treatment regime is defined in the potential outcome framework. Denoted by $Y_0^*(0)$ and $Y_0^*(1)$ the potential outcomes which represent the response that a patient would get if treated by treatment 0 and 1, respectively. A treatment regime d is a function that maps the covariate space to $\{0, 1\}$. For such a function d, define the potential outcome

$$Y_0^*\{d(X_0)\} = Y_0^*(0)\{1 - d(X_0)\} + Y_0^*(1)\, d(X_0).$$

Let $\mathcal{D}$ denote the set of all possible treatment regimes. The optimal treatment regime (OTR) d^{opt} is the maximizer of the expected potential outcome $\mathrm{E}[Y_0^*\{d(X_0)\}]$ among the set $\mathcal{D}$, that is,

$$d^{\text{opt}} \in \arg\max_{d \in \mathcal{D}} \mathrm{E}\big[Y_0^*\{d(X_0)\}\big].$$

The OTR may not be unique. Denote by $\mathcal{D}^{\text{opt}}$ the set of all OTRs. Let $\tau(x) = \mathrm{E}\{Y_0^*(1) - Y_0^*(0)|X_0 = x\}$. Under the following two assumptions:

(A1.) SUTVA: $Y_0 = A_0 Y_0^*(1) + (1 - A_0)Y_0^*(0)$,
(A2.) No unmeasured confounders: $Y_0^*(0), Y_0^*(1) \perp\!\!\!\perp A_0|X_0$,

we can show that

$$d^{\text{opt},0}(x) \equiv \mathbb{I}\{\tau(x) > 0\} \in \mathcal{D}^{\text{opt}}, \tag{2.1}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function.

We assume $\tau(x) = Q(x^T\beta_0)$ for some $\beta_0 \in \mathbb{R}^p$ and some monotonically increasing function Q. Function Q can either be specified as linear or remain completely unspecified. Assume there exists some unique $c_0 \in \mathbb{R}$ such that $Q(c_0) = 0$. It follows from (2.1) that $d^{\text{opt}}(x) = \mathbb{I}(x^T\beta_0 + c_0 > 0)$. Hence, finding the optimal treatment regime is equivalent to estimating the high dimensional parameter $\theta_0 = (c_0, \beta_0^T)^T$. Assume β_0 is sparse. Let $\mathcal{M}_{\beta_0} = \text{supp}(\beta_0)$ be the support of β_0 consisting of indices of all nonzero elements. The aim of this paper is to identify the set $\mathcal{M}_{\beta_0}$.

2.2. *Value and concordance function.* For a given treatment regime d, the expected potential outcome $V(d) = \mathrm{E}[Y_0^*\{d(X_0)\}]$ is referred to as the value function of d. Recall that d^{opt} is the maximizer of $V(d)$.

Assume data can be summarized as $\{O_i = (Y_i, A_i, X_i), i = 1, \ldots, n\}$, which are i.i.d. copies of $O_0 = (Y_0, A_0, X_0)$. In the high dimensional case, the distribution of O_0 is allowed to vary with n and it is more proper to write $O_0 = O_0^{(n)} = (Y_0^{(n)}, A_0^{(n)}, X_0^{(n)})$. However, we will omit the superscript n for notational convenience. Let $\pi_0(x) = \Pr(A_0 = 1|X_0 = x)$ be the propensity score. In a randomized study, $\pi_{0,i} = \pi_0(X_i)$ is known for each patient. To estimate $V(d)$, Zhang et al. (2012) proposed an inverse propensity-score weighted estimator (IPWE),

$$\widehat{V}(d) = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i\, d(X_i) + (1 - A_i)\{1 - d(X_i)\}}{A_i\pi_{0,i} + (1 - A_i)(1 - \pi_{0,i})} Y_i.$$

In this paper, we focus on the class of linear decision rules. For any $\theta = (c, \beta^T)^T$, we write $V(d)$, $\widehat{V}(d)$ as $V(\theta)$, $\widehat{V}(\theta)$ if d takes the form $d(x) = \mathbb{I}(x^T\beta + c > 0)$. Hence, it follows from (2.1) that

$$\theta_0 \in \underset{\theta\in\mathbb{R}^{p+1}}{\arg\max}\, V(\theta).$$

Fan et al. (2017) proposed to obtain β_0 by maximizing the estimated concordance function. For any linear treatment regime $\mathbb{I}(x^T\beta + c > 0)$, the concordance function $C(\beta)$ is defined as

$$C(\beta) = \mathrm{E}\big[\{Y_i^*(1) - Y_i^*(0)\} - \{Y_j^*(1) - Y_j^*(0)\}\big]\mathbb{I}(X_i^T\beta > X_j^T\beta),$$

for two subjects $i \neq j$. The rationale behind their method is that if $Y_i^*(1) - Y_i^*(0) > Y_j^*(1) - Y_j^*(0)$, the optimal treatment regime should be more likely to assign treatment 1 to subject i compared with subject j. In our setting where $\tau(x) = Q(x^T\beta_0)$, we have by Conditions (A1) and (A2) that

$$C(\beta) = \mathrm{E}\{Q(X_i^T\beta_0) - Q(X_j^T\beta_0)\}\mathbb{I}(X_i^T\beta > X_j^T\beta).$$

It follows that

$$C(\beta_0) - C(\beta) = \mathrm{E}\{Q(X_i^T\beta_0) - Q(X_j^T\beta_0)\}\{\mathbb{I}(X_i^T\beta_0 > X_j^T\beta_0) - \mathbb{I}(X_i^T\beta > X_j^T\beta)\}.$$

When $X_i^T\beta_0 > X_j^T\beta_0$, it follows from the monotonicity of Q that $Q(X_i^T\beta_0) > Q(X_j^T\beta_0)$. Therefore, we have for any $\beta \in \mathbb{R}^p$,

$$\{Q(X_i^T\beta_0) - Q(X_j^T\beta_0)\}\{\mathbb{I}(X_i^T\beta_0 > X_j^T\beta_0) - \mathbb{I}(X_i^T\beta > X_j^T\beta)\}\mathbb{I}(X_i^T\beta_0 > X_j^T\beta_0) \geq 0.$$

One can similarly show

$$\{Q(X_i^T\beta_0) - Q(X_j^T\beta_0)\}\{\mathbb{I}(X_i^T\beta_0 > X_j^T\beta_0) - \mathbb{I}(X_i^T\beta > X_j^T\beta)\}\mathbb{I}(X_i^T\beta_0 \leq X_j^T\beta_0) \geq 0.$$

It follows that $C(\beta_0) \geq C(\beta)$, for any $\beta \in \mathbb{R}^p$. Hence, we have

$$\beta_0 \in \arg\max_{\beta\in\mathbb{R}^p} C(\beta).$$

When the propensity score is known, the estimated concordance function is given by

$$\widehat{C}(d) = \frac{1}{n(n-1)}\sum_{i\neq j}\left\{\frac{Y_i(A_i - \pi_{0,i})}{\pi_{0,i}(1-\pi_{0,i})} - \frac{Y_j(A_j - \pi_{0,j})}{\pi_{0,j}(1-\pi_{0,j})}\right\}\mathbb{I}\{d(X_i) > d(X_j)\}.$$

Analogous to the likelihood-based information criteria, we define the following value information criterion (VIC):

$$\mathrm{VIC}_{\kappa_n}(\theta) = n\widehat{V}(\theta) - \kappa_n\|\beta\|_0, \tag{2.2}$$

and concordance information criterion (CIC)

$$\mathrm{CIC}_{\kappa_n}(\beta) = n\widehat{C}(\beta) - \kappa_n\|\beta\|_0, \tag{2.3}$$

for some sequence κ_n, where $\|\beta\|_0$ denotes the number of nonzero elements in the p-dimensional vector β. To ease the presentation, we suppress the dependence of VIC and CIC on κ_n whenever there is no confusion. In the next section, we show selection consistencies of VIC and CIC.

3. Model selection consistency.

3.1. *VIC and CIC in fixed-p case.* For any q-dimensional vector $\nu \in \mathbb{R}^q$ and any sets $J \in \{1,\ldots,q\}$, we denote by ν^J the subvector of ν formed by elements in J. When J is a single-element set, that is, $J = \{j_0\}$ for some $1 \leq j_0 \leq q$, we write ν^J as ν^{j_0}. Let $\Omega = \{\mathcal{M} : \mathcal{M} \subseteq \{1,\ldots,p\}\}$ be the set of all possible candidate models. In the fixed-p scenario, total number of elements in Ω is also fixed. For each $\mathcal{M} \in \Omega$, let $(\hat{c}_{\mathcal{M}}, \bar{\beta}_{\mathcal{M}}^T)^T \in \mathbb{R}^{|\mathcal{M}|+1}$ be the estimator based on covariates $X_0^{\mathcal{M}}$. Denote by $\hat{\beta}_{\mathcal{M}}$ the vector in $\mathbb{R}^p$ that has the same coordinates as $\bar{\beta}_{\mathcal{M}}$ on $\mathcal{M}$ and zero components on the complement $\mathcal{M}^c$ of $\mathcal{M}$.

For any triple $o = (x, a, y)$, define function

$$\begin{aligned} &g(o,\beta)\\ &\quad= \frac{1}{2}\mathrm{E}\left\{\frac{\{A_0 - \pi_0(X_0)\}Y_0}{\pi_0(X_0)\{1-\pi_0(X_0)\}} - \frac{\{a-\pi_0(x)\}y}{\pi_0(x)\{1-\pi_0(x)\}}\right\}\mathbb{I}(X_0^T\beta > x^T\beta)\\ &\qquad+ \frac{1}{2}\mathrm{E}\left\{\frac{\{a-\pi_0(x)\}y}{\pi_0(x)\{1-\pi_0(x)\}} - \frac{\{A_0 - \pi_0(X_0)\}Y_0}{\pi_0(X_0)\{1-\pi_0(X_0)\}}\right\}\mathbb{I}(x^T\beta > X_0^T\beta). \end{aligned} \tag{3.1}$$

Write Δ_m for the mth partial derivative operator with respect to β, and define

$$\partial_i g(o,\beta) = \frac{\partial g(o,\beta)}{\partial\beta^i} \quad\text{and}\quad \partial_{ij} g(o,\beta) = \frac{\partial^2 g(o,\beta)}{\partial\beta^i\,\partial\beta^j}.$$

Let δ be some positive constant such that $\delta < \min_{j\in\mathcal{M}_{\beta_0}}|\beta_0^j|$. For any $\varepsilon > 0$, define the ε-neighborhood of θ_0,

$$\tilde{N}_\varepsilon = \{\theta \in \mathbb{R}^{p+1} : \|\theta_0 - \theta\|_2 \leq \varepsilon\},$$

and $\tilde{S}(\theta_0) = \{\theta \in \mathbb{R}^{p+1} : \|\theta\|_2 = \|\theta_0\|_2\}$. Similarly, define

$$N_\varepsilon = \{\beta \in \mathbb{R}^{p+1} : \|\beta_0 - \beta\|_2 \le \varepsilon\},$$

and $S(\beta_0) = \{\beta \in \mathbb{R}^p : \|\beta\|_2 = \|\beta_0\|_2\}$. We first introduce some assumptions.

(A3.) There exist some constants c_1, c_2 that satisfy $0 < c_1 \le \inf_x \pi_0(x) \le \sup_x \pi_0(x) \le c_2 < 1$.

(A4.) $\hat{\beta}_{\mathcal{M}_{\beta_0}} = \beta_0 + O_p(R_n^{(1)})$, $\hat{c}_{\mathcal{M}_{\beta_0}} = c_0 + O_p(R_n^{(2)})$ for some sequences $R_n^{(1)}$ and $R_n^{(2)}$ that satisfy $n^{-1/2} \le R_n^{(1)}, R_n^{(2)} \ll 1$.

(A5.) (i) $V(\theta_0) > V(0)$ and $V(\theta_0) > \sup_{\theta \in \tilde{N}_{\varepsilon_0}^c \cap \tilde{S}(\theta_0)} V(\theta)$ for some constants $0 < \varepsilon_0 \le \delta$.

(ii) The following holds for any sufficiently small $\varepsilon > 0$:

$$\mathrm{E} \sup_{\substack{\|\theta - \theta_0\|_2 \le \varepsilon \\ \theta = (c, \beta^T)^T}} \left| \mathbb{I}(X_0^T \beta > -c) - \mathbb{I}(X_0^T \beta_0 > -c_0) \right| = O(\varepsilon).$$

(iii) There exist some constants $\bar{c}_1, \bar{c}_2 > 0$ such that

$$\bar{c}_1 \|\theta_0 - \theta\|_2^2 \le V(\theta_0) - V(\theta) \le \bar{c}_2 \|\theta_0 - \theta\|_2^2 \quad \text{for all } \theta \in \tilde{N}_{\varepsilon_0} \cap \tilde{S}(\theta_0).$$

(A6.) (i) $C(\beta_0) > C(0)$ and $C(\beta_0) > \sup_{\beta \in N_{\varepsilon_0}^c \cap S(\beta_0)} C(\beta)$ for some constants $0 < \varepsilon_0 \le \delta$.

(ii) There exist some constants $\bar{c}_1, \bar{c}_2 > 0$ such that

$$\bar{c}_1 \|\beta_0 - \beta\|_2^2 \le C(\beta_0) - C(\beta) \le \bar{c}_2 \|\beta_0 - \beta\|_2^2 \quad \text{for all } \beta \in N_{\varepsilon_0} \cap S(\beta_0).$$

(iii) Function $g(o, \beta)$ is twice continuously differentiable for all $\beta \in N_{\varepsilon_0}$.

(iv) There is an integrable function $K(o)$ such that for all o and $\beta \in N_{\varepsilon_0}$,

$$\|\Delta_2 g(o, \beta) - \Delta_2 g(o, \beta_0)\|_2 \le K(o) \|\beta - \beta_0\|_2.$$

(v) $\mathrm{E}|\partial_i g(O_0, \beta_0)|^2 < \infty$, $\mathrm{E}|\partial_{ij} g(O_0, \beta_0)| < \infty$.

Assumption (A4) requires that $\hat{\beta}_{\mathcal{M}_{\beta_0}}$ converges to β_0. The sequences $R_n^{(1)}$ and $R_n^{(2)}$ depend on the estimating procedure and are known to us. When $\hat{\beta}_{\mathcal{M}}$ is estimated by solving Q-learning or A-learning estimating equations for any $\mathcal{M}$, we can show $R_n^{(1)} = n^{-1/2}$. This requires Q-function to be correctly specified. When Q-function remains unspecified, we can apply robust learning or concordance assisted-learning to estimate β_0. The convergence rates $R_n^{(1)}$ for these two estimators are $n^{-1/3}$ and $n^{-1/2}$, respectively.

Assumptions (A5)(i) and (A6)(i) require functions V and C have unique maximizers on the L_2 sphere. These conditions guarantee that with probability tending to 1, VIC and CIC will not pick underfitted models for $\kappa_n = o(n)$. Assumption (A5)(ii) holds when the angular component of X_0 has a bounded and continuous density with respect to the surface measure on the unit sphere (see Section 6.4 in Kim and Pollard (1990)). When the derivative $\frac{dQ(x-c_0)}{dx}|_{x=0} \neq 0$, it implies the margin assumption $\Pr(0 < |\tau(X_0)| < t) = O(t^\alpha)$ (see Luedtke and van der Laan (2016), Qian and Murphy (2011)) holds with $\alpha = 1$ (see Section B.1 in the Supplementary Material (Shi, Song and Lu (2020))).

Assumption (A5)(iii) is satisfied if V is twice continuously differentiable and possess a unique maximizer on $S(\beta_0)$. This condition holds when X_0 has a continuous density q which has a compact support. The explicit form of the first- and second-order derivatives of V can be derived by some standard arguments in classical differential geometry (see Sections 5 and 6.4 in Kim and Pollard (1990)). Assumptions (A6)(iii), (iv), (v) are standard to establish the limiting distribution of concordance and maximum rank correlation estimators (cf. Sherman (1993), Fan et al. (2017)). In Section B.2 of the Supplementary Material, we give detailed discussion on Assumption (A6)(iii).

Under the scenario of treatment effect homogeneity, that is, $\theta_0 = 0$, $V(\theta)$ and $C(\beta)$ are constants as functions of θ and β. (A5)(i) and (A6)(i) are violated under this scenario. As a result, VIC and CIC are not consistent.

Denote by $\widehat{\mathcal{M}}_V$ and $\widehat{\mathcal{M}}_C$ the models chosen by VIC and CIC, respectively,

$$\widehat{\mathcal{M}}_V = \arg\max_{\mathcal{M}\in\Omega} \mathrm{VIC}(\hat{\theta}_{\mathcal{M}}), \qquad \widehat{\mathcal{M}}_C = \arg\max_{\mathcal{M}\in\Omega} \mathrm{CIC}(\hat{\beta}_{\mathcal{M}}),$$

where $\hat{\theta}_{\mathcal{M}} = (\hat{c}_{\mathcal{M}}, \hat{\beta}_{\mathcal{M}}^T)^T$. Define $R_n = \max(R_n^{(1)}, R_n^{(2)})$. The following theorem states the model selection consistencies of these criteria.

THEOREM 3.1. *Suppose* $\sup_x \mathrm{E}(Y_0^2|X_0 = x) \le \bar{C}$ *for some constant* $\bar{C} > 0$. *Set* $\kappa_n = c_n \max(nR_n^2, \sqrt{nR_n}, n^{1/3})$ *for some* $c_n \to \infty$, *if* $\kappa_n = o(n)$, *under Assumptions* (A1)–(A5), *we have*

$$\Pr(\widehat{\mathcal{M}}_V = \mathcal{M}_{\beta_0}) \to 1.$$

Set $\kappa_n = n(R_n^{(1)})^2 \log(n)$, *if* $\kappa_n = o(n)$, *under Assumptions* (A1)–(A4) *and* (A6), *we have*

$$\Pr(\widehat{\mathcal{M}}_C = \mathcal{M}_{\beta_0}) \to 1.$$

REMARK 3.2. The choice of κ_n depends on $R_n^{(1)}$ and $R_n^{(2)}$. Suppose $R_n = n^{-1/2}$. Then we have $\kappa_n = \log(n)$ for CIC and $\kappa_n = c_n n^{1/3}$ for VIC. Unlike BIC, when setting $\kappa_n = \log(n)$, VIC fails to select the correct model if $R_n = n^{-1/2}$. The penalty term $c_n n^{1/3}$ accounts for the nonsmoothness of the indicator function in $\widehat{V}$. On the contrary, CIC directly follows the spirit of BIC. The estimated concordance function $\widehat{C}$ is a U-statistic of order two. Due to Hoeffding's decomposition theorem and the maximal inequality for degenerate U-process (Sherman (1994)), we have

$$\widehat{C}(\beta) = \frac{2}{n}\sum_{i=1}^{n} \mathrm{g}(O_i, \beta) - C(\beta) + O_p\Big(\frac{1}{n}\Big), \tag{3.2}$$

uniformly for all β.

We now sketch a few lines to see why VIC can fail when $\kappa_n = \log(n)$ and $R_n = n^{-1/2}$. Recall that V is maximized at θ_0. Under Assumptions (A4) and (A5)(iii), we have that

$$nV(\hat{\theta}_{\mathcal{M}_{\beta_0}}) = nV(\theta_0) + O\big(n\|\hat{\theta}_{\mathcal{M}_{\beta_0}} - \theta_0\|_2^2\big) = nV(\theta_0) + O(1). \tag{3.3}$$

For any overfitted model $\mathcal{M}$ that satisfies $\mathcal{M}_{\beta_0} \subseteq \mathcal{M}$, let $\tilde{\theta}_{\mathcal{M}}$ be the maximizer of $\widehat{V}$ with the constraint $\tilde{\theta}_{\mathcal{M}}^{\mathcal{M}^c} = 0$. Notice that $\widehat{V}$ is a nonsmooth function of θ. Under the given conditions, each $\tilde{\theta}_{\mathcal{M}}$ exhibits cube root asymptotics, and we have

$$\tilde{\theta}_{\mathcal{M}} = \theta_0 + O_p\big(n^{-1/3}\big). \tag{3.4}$$

Consider the stochastic process $\widehat{V}(\cdot) - V(\cdot) - \widehat{V}(\theta_0) + V(\theta_0)$ indexed by θ. Under Assumption (A5)(ii), using some standard arguments in the empirical process theory (cf. van der Vaart and Wellner (1996)), we can show

$$\sup_{\|\theta-\theta_0\|_2\le\epsilon} \big|\widehat{V}(\theta) - V(\theta) - \widehat{V}(\theta_0) + V(\theta_0)\big| = O_p\big(n^{-1/2}\epsilon^{1/2}\big), \tag{3.5}$$

for some sufficiently small $\epsilon > 0$. This together with (3.4) yields that

$$n\big[\widehat{V}(\tilde{\theta}_{\mathcal{M}}) - V(\tilde{\theta}_{\mathcal{M}}) - \{\widehat{V}(\theta_0) - V(\theta_0)\}\big] = O_p\big(n^{1/3}\big).$$

Notice that $V(\tilde{\theta}_{\mathcal{M}}) \le V(\theta_0)$. It follows that

$$n\{\widehat{V}(\tilde{\theta}_{\mathcal{M}}) - \widehat{V}(\theta_0)\} = O_p(n^{1/3}),$$

for any $\mathcal{M}$ that satisfies $\mathcal{M}_{\beta_0} \subseteq \mathcal{M}$. Similarly, we can show $n\{\widehat{V}(\tilde{\theta}_{\mathcal{M}_{\beta_0}}) - \widehat{V}(\theta_0)\} = O_p(n^{1/3})$. Hence,

$$\max_{\mathcal{M}:\mathcal{M}_{\beta_0}\subseteq\mathcal{M}} n\{\widehat{V}(\tilde{\theta}_{\mathcal{M}}) - \widehat{V}(\hat{\theta}_{\mathcal{M}_{\beta_0}})\} = O_p(n^{1/3}).$$

Since $\widehat{V}(\hat{\theta}_{\mathcal{M}}) \le \widehat{V}(\tilde{\theta}_{\mathcal{M}})$, we have

$$\max_{\mathcal{M}:\mathcal{M}_{\beta_0}\subseteq\mathcal{M}} n\{\widehat{V}(\hat{\theta}_{\mathcal{M}}) - \widehat{V}(\hat{\theta}_{\mathcal{M}_{\beta_0}})\} \le R_n, \tag{3.6}$$

for some random variable R_n that satisfies $R_n = O_p(n^{1/3})$.

On the other hand,

$$\begin{aligned}&\max_{\mathcal{M}:\mathcal{M}_{\beta_0}\subseteq\mathcal{M}} \mathrm{VIC}(\hat{\theta}_{\mathcal{M}}) - \mathrm{VIC}(\hat{\theta}_{\mathcal{M}_{\beta_0}})\\ &\quad= \max_{\mathcal{M}:\mathcal{M}_{\beta_0}\subseteq\mathcal{M}} n\{\widehat{V}(\hat{\theta}_{\mathcal{M}}) - \widehat{V}(\hat{\theta}_{\mathcal{M}_{\beta_0}})\} - \kappa_n(|\mathcal{M}| - |\mathcal{M}_{\beta_0}|).\end{aligned} \tag{3.7}$$

The difference $|\mathcal{M}| - |\mathcal{M}_{\beta_0}|$ is always positive. When $\kappa_n = \log(n)$, it follows from (3.6) that the sign of (3.8) can be positive in the limit. Equation (3.8) also implies VIC is not able to pick overfitted models if $\kappa_n = c_n n^{1/3}$ for some diverging sequence c_n.

3.2. *VIC and CIC in the ultrahigh dimension.* The problem becomes far more challenging in the ultrahigh dimension when p is allowed to grow exponentially fast with respect to n. Assume $\log(p) = O(n^{a_0})$ for some $0 < a_0 < 1$. For notational convenience, in this paper, we assume the nonzero indices $\mathcal{M}_{\beta_0}$ and $\beta_0^{\mathcal{M}_{\beta_0}}$ are fixed. In the ultrahigh dimension, it is computationally infeasible to estimate $\theta_{\mathcal{M}}$ for all $\mathcal{M} \in \Omega$. Instead, we use some penalization methods to simultaneously select and estimate θ_0, with some tuning parameter λ.

For each $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ where $\lambda_{\min}$ and $\lambda_{\max}$ are allowed to vary with n, denote $\widehat{\mathcal{M}}(\lambda)$ as the nonzero entries selected by our estimating procedure and $\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)} = (\hat{c}_{\widehat{\mathcal{M}}(\lambda)}, \hat{\beta}^T_{\widehat{\mathcal{M}}(\lambda)})^T \in \mathbb{R}^{p+1}$ the corresponding estimator. We define

$$\widehat{\mathcal{M}}_V = \mathop{\arg\max}_{\substack{|\widehat{\mathcal{M}}(\lambda)|\le s_n\\ \lambda\in[\lambda_{\min},\lambda_{\max}]}} \mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}) \quad \text{and} \quad \widehat{\mathcal{M}}_C = \mathop{\arg\max}_{\substack{|\widehat{\mathcal{M}}(\lambda)|\le s_n\\ \lambda\in[\lambda_{\min},\lambda_{\max}]}} \mathrm{CIC}(\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}).$$

The sequence s_n is allowed to vary with n in the order $s_n = O(n^{l_0})$ for some $0 \le l_0 < 1$. To show the model selection consistency, we need Conditions (A4′)–(A6′). (A5′) and (A6′) are high dimensional versions of (A5) and (A6), and are provided in Section A of the Supplementary Material to save space.

(A4′.) There exists some $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$ such that with probability tending to 1, we have $\widehat{\mathcal{M}}(\lambda_0) = \mathcal{M}_\beta$ and

$$\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)} - \beta_0\|_2 = O(R_n^{(1)}), \qquad \|\hat{c}_{\widehat{\mathcal{M}}(\lambda_0)} - c_0\|_2 = O(R_n^{(2)})$$

for some sequence $R_n^{(1)}$, $R_n^{(2)}$ that $R_n^{(j)} \to 0$ and $R_n^{(j)} \ge n^{-1/2}$ for $j = 1, 2$. The tuning parameter λ_0 is allowed to vary with n.

Assumption (A4′) requires the true model to be recovered by the regularization methods and assumes the convergence rate of parameters for the true model. In the following, we

establish consistencies of our information criteria. We use $\bar{c}$ to denote some generic constant. Let $\|Y\|_{\psi_p}$ be the Orlicz norm of any random variable Y,

$$\|Y\|_{\psi_p} \triangleq \inf_{C>0}\left\{\mathrm{E}\exp\left(\frac{|Y|^p}{C^p}\right) \le 2\right\}.$$

THEOREM 3.3. *Let* $R_n = \max(R_n^{(1)}, R_n^{(2)})$. *Assume* $s_n^2 \log(p)\log(n) = o(n)$, (A1)–(A3) *and* (A4′), (A5′) *hold*, $\|Y_0\|_{\psi_1} = O(1)$, *and* $\sup_x \mathrm{E}(Y_0^2 | X_0 = x) \le \bar{C}$ *for some constant* $\bar{C} > 0$. *If* κ_n *satisfies* $\kappa_n = o(n)$, *and*

$$\kappa_n \gg n R_n^2 + \sqrt{n R_n} + n^{1/3}\log^{2/3}(p), \tag{3.8}$$

then VIC is consistent. In addition, conditional on the events $\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)} - \beta_0\|_2 = O(R_n^{(1)})$ *and* $\|\hat{c}_{\widehat{\mathcal{M}}(\lambda_0)} - c_0\|_2 = O(R_n^{(2)})$, *we have*

$$\Pr(\widehat{\mathcal{M}}_V \ne \mathcal{M}_{\beta_0}) \le \exp\left(-\frac{\bar{c}\kappa_n^2}{n R_n}\right) + \exp\bigl(-\bar{c}\log(p)\bigr). \tag{3.9}$$

THEOREM 3.4. *Assume* $s_n^2 \log(p)\log(n) = o(n)$, (A1)–(A3) *and* (A4′), (A6′) *hold*, $\|Y_0\|_{\psi_1} = O(1)$ *and* $\sup_x \mathrm{E}(Y_0^2 | X_0 = x) \le \bar{C}$ *for some constant* $\bar{C} > 0$. *If* κ_n *satisfies* $\kappa_n = o(n)$, *and*

$$\kappa_n \gg n\bigl(R_n^{(1)}\bigr)^2 + \log(p)\log(n), \tag{3.10}$$

then CIC is consistent. In addition, conditional on the event $\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)} - \beta_0\|_2 = O(R_n^{(1)})$, *we have*

$$\Pr(\widehat{\mathcal{M}}_C \ne \mathcal{M}_{\beta_0}) \le \exp\left(-\frac{\bar{c}\kappa_n^2}{n(R_n^{(1)})^2}\right) + \exp\bigl(-\bar{c}\log(p)\bigr). \tag{3.11}$$

REMARK 3.5. Equations (3.9) and (3.11) provide nonasymptotic bounds on the probabilities that VIC and CIC do not select the correct model. Under the assumptions in (3.8) and (3.10), these upper bounds go to 0. Consistencies of these two criteria thus follow. The second term $\exp(-\bar{c}\log(p))$ on the RHS of (3.9) and (3.11) bounds the probability that VIC or CIC selects an underfitted model. The first term on the RHS bounds the probability that VIC or CIC picks an overfitted model. When $\kappa_n \ll \sqrt{n R_n \log(p)}$, the RHS in (3.9) is dominated by

$$\exp\left(-\frac{\bar{c}\kappa_n^2}{n R_n}\right),$$

which is much larger than those in the RHS of (3.11). This suggests that CIC is more likely to choose the correct model compared with VIC.

REMARK 3.6. Conditions on κ_n in Theorem 3.3 are more restrictive than those in Theorem 3.4. This means that the consistency of VIC is more sensitive to the choice of κ_n. Denote k_V and k_C as the RHS of (3.8) and (3.10), respectively. Since $R_n \ge R_n^{(1)}$ and $n \gg \log(p)$, it is immediate to see that $k_C = O(k_V)$. In addition, when $R_n = O(\sqrt{\log(p)/n})$, we have $k_V \gg k_C$. This is in line with results given in the fixed-p scenario (see Theorem 3.1), where VIC can fail for $R_n = n^{-1/2}$ if $\kappa_n = \log(n)$.

Proofs of Theorems 3.3 and 3.4 are more involved than those of the fixed-p scenario. Define

$$\Omega_+ = \{\lambda \in [\lambda_{\min}, \lambda_{\max}] : \mathcal{M}_{\beta_0} \subsetneq \widehat{\mathcal{M}}(\lambda), |\widehat{\mathcal{M}}(\lambda)| \le s_n\}.$$

The major technical challenge lies in bounding

$$\Pr\Big(\mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) \le \sup_{\lambda\in\Omega_+} \mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)})\Big),$$

and

$$\Pr\Big(\mathrm{CIC}(\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}) \le \sup_{\lambda\in\Omega_+} \mathrm{CIC}(\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)})\Big),$$

where λ_0 is the tuning parameter defined in (A4′). Unlike the fixed-p scenario, inequalities (3.2) and (3.5) no longer hold in the ultrahigh dimension.

4. Doubly-robust information criteria. In an observational study, the propensity score is unknown and needs to be estimated from data. Usually, a parametric model $\pi(x, \alpha)$ is assumed for the propensity score. To calculate our doubly-robust information criteria, we also fit a parametric model $h(x, \eta)$ for the baseline function $h_0(x) = \mathrm{E}(Y_0|A_0 = 0, X_0 = x)$. We assume estimators $\hat{\alpha}$ and $\hat{\eta}$ converge to some $\alpha^* \in \mathbb{R}^{q_1}$ and $\eta^* \in \mathbb{R}^{q_2}$. When the models are correct, α^* and η^* correspond to the true parameters in the model, that is, $\pi_0(x) = \pi(x, \alpha^*)$, $h_0(x) = h(x, \eta^*)$. Otherwise, these parameters stand for some population-level least false parameters. Let $\theta = (c, \beta^T)^T$. Define

$$\begin{aligned} V^{DR}(\theta) = \mathrm{E}\Big\{ &\frac{A_0\mathbb{I}(X_0^T\beta > -c)}{\pi(X_0, \alpha^*)} + \frac{(1 - A_0)\mathbb{I}(X_0^T\beta \le -c)}{1 - \pi(X_0, \alpha^*)}\Big\} Y_0 \\ &- \mathrm{E}\Big\{ \frac{A_0\mathbb{I}(X_0^T\beta > -c)}{\pi(X_0, \alpha^*)} + \frac{(1 - A_0)\mathbb{I}(X_0^T\beta \le -c)}{1 - \pi(X_0, \alpha^*)} - 1\Big\} h(X_0, \eta^*) \end{aligned}$$

and

$$\begin{aligned} C^{DR}(\beta) = \mathrm{E}\Big\{ &\frac{\{A_i - \pi(X_i, \alpha^*)\}\{Y_i - h(X_i, \eta^*)\}A_j}{\pi(X_i, \alpha^*)\{1 - \pi(X_i, \alpha^*)\}\pi(X_j, \alpha^*)} \\ &- \frac{\{A_j - \pi(X_j, \alpha^*)\}\{Y_j - h(X_j, \eta^*)\}A_i}{\pi(X_j, \alpha^*)\{1 - \pi(X_j, \alpha^*)\}\pi(X_i, \alpha^*)}\Big\} \mathbb{I}(X_i^T\beta > X_j^T\beta). \end{aligned}$$

Under Assumptions (A1) and (A2), when either the propensity score model or the baseline model is correct, we can show

$$\begin{aligned} V^{DR}(\theta) &= \mathrm{E}\Big\{h(X_0) + \frac{\pi(X_0)}{\pi(X_0, \alpha^*)} Q(X_0^T\beta_0)\mathbb{I}(X_0^T\beta > -c)\Big\}, \\ C^{DR}(\beta) &= \mathrm{E}\Big\{\frac{\pi(X_i)\pi(X_j)}{\pi(X_i, \alpha^*)\pi(X_j, \alpha^*)}\{Q(X_i^T\beta_0) - Q(X_j^T\beta_0)\}\mathbb{I}(X_i^T\beta > X_j^T\beta)\Big\}. \end{aligned}$$

Therefore, when the propensity score model is correct, we have $V^{DR} = V$ and $C^{DR} = C$. This result generally does not hold when the propensity score model is not correct. However, $\theta_0(\beta_0)$ still maximizes $V^{DR}(C^{DR})$ as long as either of the models is correct. This suggests V^{DR} and C^{DR} can be used to construct information criteria. Define

$$\mathrm{VIC}^{DR}(\theta) = n\widehat{V}^{DR}(\theta) - \kappa_n\|\beta\|_0, \mathrm{CIC}^{DR}(\beta) = n\widehat{V}^{DR}(\beta) - \kappa_n\|\beta\|_0,$$

where $\widehat{V}^{DR}$ and $\widehat{C}^{DR}$ are empirical estimators for V^{DR} and C^{DR}, namely,

$$\begin{aligned}\widehat{V}^{DR}(\theta) &= \frac{1}{n}\sum_i \left\{ \frac{A_i \mathbb{I}(X_i^T\beta > -c)}{\pi(X_i,\hat{\alpha})} + \frac{(1-A_i)\mathbb{I}(X_i^T\beta \le -c)}{1-\pi(X_i,\hat{\alpha})} \right\} Y_i \\ &\quad - \left\{ \frac{A_i \mathbb{I}(X_i^T\beta > -c)}{\pi(X_i,\hat{\alpha})} + \frac{(1-A_i)\mathbb{I}(X_i^T\beta \le -c)}{1-\pi(X_i,\hat{\alpha})} - 1 \right\} h(X_i,\hat{\eta}), \\ \widehat{C}^{DR}(\beta) &= \frac{1}{n}\sum_{i\neq j} \left\{ \frac{\{A_i - \pi(X_i,\hat{\alpha})\}\{Y_i - h(X_i,\hat{\eta})\}A_j}{\pi(X_i,\hat{\alpha})\{1-\pi(X_i,\hat{\alpha})\}\pi(X_j,\hat{\alpha})} \right. \\ &\quad \left. - \frac{\{A_j - \pi(X_j,\hat{\alpha})\}\{Y_j - h(X_j,\hat{\eta})\}A_i}{\pi(X_j,\hat{\alpha})\{1-\pi(X_j,\hat{\alpha})\}\pi(X_i,\hat{\alpha})} \right\} \mathbb{I}(X_i^T\beta > X_j^T\beta).\end{aligned}$$

In Section 10 of the Supplementary Material, we derive the consistencies of VIC^{DR} and CIC^{DR} under the fixed-p scenario. When p is comparable or much larger than n, we can fit the baseline or propensity score models via penalized regression with folded-concave penalty functions (Fan and Lv (2011)). In practice, we recommend a linear regression model for the baseline model and a logistic regression model for the propensity score model with SCAD penalty function (Fan and Li (2001)). Under certain conditions on these estimators, consistencies of VIC^{DR} and CIC^{DR} can be similarly proven. We omit the technical details to save space.

5. Applications. In this section, we apply our information criteria to four applications estimating the optimal treatment regime, including robust learning, concordance-assisted learning (CAL), penalized A-learning (PAL) and sparse concordance-assisted learning (SCAL). The first two consider a fixed-p setting while the last two can be applied in a diverging-p setting. For each application, we introduce its estimating procedure and discuss the choice of κ_n in our information criteria.

5.1. *Robust learning.*

5.1.1. *Estimating procedure.* Zhang et al. (2012) proposed a robust method for estimating the optimal treatment regime within the class of linear decision rules. For a given candidate model $\mathcal{M}$, when the propensity score is known, the estimator $\hat{\theta}_{\mathcal{M}} = (\hat{c}_{\mathcal{M}}, \hat{\beta}_{\mathcal{M}}^T)^T$ is obtained by solving

$$\arg\max_{\theta=(c,\beta^T)^T} \widehat{V}(\theta) \quad \text{subject to } \beta^{\mathcal{M}^c} = 0.$$

In an observational study, they first fit some parametric models $\pi(x,\alpha)$, $h(x,\eta)$, $\Phi(x,\zeta)$ for $\pi_0(x)$, $h_0(x)$ and $\mathrm{E}(Y_0|A_0=1, X_0=x)$, to obtain estimators $\hat{\alpha}$, $\hat{\eta}$ and $\hat{\zeta}$. Then they proposed to compute $\hat{\theta}_{\mathcal{M}}$ by maximizing the following augmented inverse propensity score weighted estimator,

$$\begin{aligned}&\arg\max_{\theta=(c,\beta^T)^T} \frac{1}{n}\sum_{i=1}^n \left[\left\{ \frac{A_i}{\hat{\pi}_i} Y_i - \left(\frac{A_i}{\hat{\pi}_i} - 1 \right) \hat{h}_i \right\} \mathbb{I}(X_i^T\beta > -c) \right. \\ &\quad \left. + \left\{ \frac{1-A_i}{1-\hat{\pi}_i} Y_i - \left(\frac{1-A_i}{1-\hat{\pi}_i} - 1 \right) \hat{\Phi}_i \right\} \mathbb{I}(X_i^T\beta \le -c) \right] \quad \text{subject to } \beta^{\mathcal{M}^c} = 0,\end{aligned} \tag{5.1}$$

where $\hat{\pi}_i$, $\hat{h}_i$ and $\hat{\Phi}_i$ are plug-in estimators $\pi(X_i,\hat{\alpha})$, $h(X_i,\hat{\eta})$ and $\Phi(X_i,\hat{\zeta})$, respectively.

5.1.2. *Choice of* κ_n. Using some standard arguments in the cube root asymptotics (cf. Example 6.4, Kim and Pollard (1990)), we can show $\hat{\theta}_{\mathcal{M}_{\beta_0}} = \theta_0 + O_p(n^{-1/3})$. Since the dimension of covariates p is fixed, in order to implement variable selection, we can apply robust learning (by solving (5.1)) to all 2^p models and choose the one that maximizes VIC^{DR} or CIC^{DR}. Hence, we have $R_n = n^{-1/3}$ in this application. Therefore, it follows from Theorem 10.1 that CIC^{DR} and VIC^{DR} are both consistent when $\kappa_n = c_n n^{1/3}$ for $c_n \to \infty$. Instead of choosing a single κ_n, one can alternatively select a set $\{\kappa_{n,j}\}_j$ that satisfy $\kappa_{n,j} \gg n^{1/3}$ and $\kappa_{n,j} = o(n)$ for each j, and apply cross-validation to determining which $\kappa_{n,j}$ to use. More details about the cross-validation procedure are given in Section J.

5.2. *Concordance-assisted learning.*

5.2.1. *Estimating procedure.* Fan et al. (2017) proposed concordance-assisted learning to estimate β_0 by maximizing the estimated concordance function. Specifically, for a given candidate model $\mathcal{M}$, $\hat{\beta}_{\mathcal{M}}$ is computed by solving

$$\arg\max_{\beta} \widehat{C}(\beta) (\text{or } \widehat{C}^{DR}(\beta)) \quad \text{subject to } \beta^{\mathcal{M}^c} = 0.$$

Assume the estimator $\hat{\beta}_{\mathcal{M}}$ is obtained, they proposed to compute $\hat{c}_{\mathcal{M}}$ by maximizing the estimated value function among the class of regimes $\mathbb{I}(c + x^T \hat{\beta}_{\mathcal{M}} > 0)$, indexed by c.

5.2.2. *Choice of* κ_n. To implement variable selection, we can apply CAL to all 2^p models. Similar to Theorem 1, 2 and 5 in Fan et al. (2017), we can show $\hat{\beta}_{\mathcal{M}_{\beta_0}} = \beta_0 + O_p(n^{-1/2})$ and $\hat{c}_{\mathcal{M}_{\beta_0}} = c_0 + O_p(n^{-1/3})$, under certain regularity conditions. Therefore, we have

$$R_n^{(1)} = n^{-1/2} \quad \text{and} \quad R_n^{(2)} = n^{-1/3}.$$

By Theorem 10.1, CIC^{DR} is consistent when $\kappa_n = \log(n)$, and VIC^{DR} is consistent if $n^{1/3} \ll \kappa_n \ll n$. In practice, we recommend to set $\kappa_n = n^{1/3} \log(\log(n))$. In our simulation studies, we find out that VIC^{DR} works well under such choices of κ_n.

5.3. *Penalized A-learning.*

5.3.1. *Estimating procedure.* When the contrast function is linear, that is, $\tau(x) = x^T \beta_0 + c_0$, Shi et al. (2018) proposed a penalized A-learning method for estimating the optimal treatment regime. Specifically, they proposed to first estimate $\pi_0(x)$ and $h_0(x)$ by penalized regression. Denoted by $\hat{\pi}_i$ and $\hat{h}_i$ the estimated propensity score and baseline function for the ith patient. For a given tuning parameter λ, they estimated θ_0 by

$$(\bar{c}_{\mathcal{M}(\lambda)}, \bar{\beta}_{\mathcal{M}(\lambda)}^T)^T = \underset{(c,\beta^T)^T \in \Lambda}{\arg\min} \|\beta\|_1, \tag{5.2}$$

where

$$\Lambda = \left\{ c \in \mathbb{R}, \beta \in \mathbb{R}^p : \left\| \frac{1}{n} \sum_i X_i (A_i - \hat{\pi}_i) \{ Y_i - \hat{h}_i - A_i (X_i^T \beta + c) \} \right\|_\infty \le \lambda \right\}.$$

The estimating procedure is similar in rationale to the Dantzig selector (Candès and Tao (2007)) in a linear regression setting. Let $\widehat{\mathcal{M}}(\lambda)$ be the support of $\bar{\beta}_{\mathcal{M}(\lambda)}$. We can compute $\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}, \hat{c}_{\widehat{\mathcal{M}}(\lambda)}$ by solving the following A-learning estimating equations:

$$\sum_i (A_i - \hat{\pi}_i)(Y_i - \hat{h}_i - A_i X_i^T \hat{\beta}_{\widehat{\mathcal{M}}(\lambda)} - A_i \hat{c}_{\widehat{\mathcal{M}}(\lambda)}) = 0,$$

$$\sum_i X_i^{\widehat{\mathcal{M}}(\lambda)} (A_i - \hat{\pi}_i)(Y_i - \hat{h}_i - A_i X_i^T \hat{\beta}_{\widehat{\mathcal{M}}(\lambda)} - A_i \hat{c}_{\widehat{\mathcal{M}}(\lambda)}) = 0,$$

with $\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}^{\widehat{\mathcal{M}}(\lambda)^c} = 0$.

5.3.2. *Choice of* κ_n. Gai, Zhu and Lin (2013) proved the model selection consistency of the Dantzig selector for linear regression, under the irrepresentable condition. Using their arguments, $\bar{\beta}_{\mathcal{M}(\lambda_0)}$ can achieve selection consistency with some tuning parameter λ_0, and we can show Assumption (A4′) holds with $R_n = n^{-1/2}$.

It follows from Theorem 3.3 that VIC is consistent when $\kappa_n = c_n n^{1/3} \log^{2/3}(p)$ for some $c_n \to \infty$. By Theorem 3.4, CIC is consistent when $\kappa_n = c_n' \log(p) \log_{10}(n)$ for some $c_n' \to \infty$. Similarly, we can show VIC^{DR} and CIC^{DR} are consistent under these choices of κ_n. In practice, we recommend to choose $c_n' = \log(\log_{10}(n))$, and $c_n = \log(\log(n))$. We demonstrate the performance of these information criteria via simulations.

5.4. *Sparse concordance-assisted learning.*

5.4.1. *Estimating procedure.* Liang et al. (2017) proposed a sparse concordance-assisted learning algorithm that extends CAL to the setting allowing p to be much larger than n. The concordance function $\widehat{C}^{DR}$ involves indicators, making it computationally difficult to optimize. Instead of directly maximizing $\widehat{C}^{DR}$, they considered a convex surrogate objective function with L_1 penalty term on the coefficients to facilitate the computation and ensure sparsity of the estimator.

Using SCAL, for any tuning parameter λ, we can estimate β_0 by

$$\bar{\beta}_{\widehat{\mathcal{M}}(\lambda)} = \arg\max_{\beta} \left\{ \frac{2}{n(n-1)} \sum_{\omega_{i,j} > \omega_{j,i}} (\omega_{j,i} - \omega_{i,j}) \{1 - \beta^T (X_i - X_j)\}_+ - \lambda \|\beta\|_1 \right\},$$

where

$$\omega_{i,j} = \frac{\{A_i - \pi(X_i, \hat{\alpha})\}\{Y_i - h(X_i, \hat{\eta})\} A_j}{\pi(X_i, \hat{\alpha})\{1 - \pi(X_i, \hat{\alpha})\}\pi(X_j, \hat{\alpha})},$$

where $\hat{\alpha}$ and $\hat{\eta}$ denote some penalized regression estimators in the propensity score and baseline model. Let $\widehat{\mathcal{M}}(\lambda)$ be the support of $\bar{\beta}_{\widehat{\mathcal{M}}(\lambda)}$. We can calculate $\hat{\beta}_{\mathcal{M}}$ by maximizing $\hat{C}^{DR}(\beta)$ subject to the constraint that $\beta^{\widehat{\mathcal{M}}(\lambda)^c} = 0$, and obtain $\hat{c}_{\mathcal{M}}$ by maximizing $\hat{V}^{DR}$ among the class of treatment regimes $\mathbb{I}(\hat{\beta}_{\mathcal{M}}^T x > -c)$.

5.4.2. *Choice of* κ_n. Assume there exists some λ_0 such that $\bar{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}$ is selection consistent, then Assumption (A4′) holds with $R_n^{(1)} = n^{-1/2}$, $R_n^{(2)} = n^{-1/3}$. By Theorem 3.3 and Theorem 3.4, we can show VIC is consistent when $\kappa_n = c_n n^{1/3} \log^{2/3} p$ for some $c_n \to \infty$, and CIC is consistent when $\kappa_n = c_n' \log(p) \log_{10}(n)$ for some $c_n' \to \infty$. Similarly, we can show VIC^{DR} and CIC^{DR} are consistent under these choices of κ_n.

6. Simulations. In this section, we conduct simulation studies to examine the numerical performance of our proposed information criteria. In Section 6.1, we consider a fixed-p scenario where the optimal treatment regime is estimated via CAL. In Section 6.2, we design a high dimensional setting and estimate the optimal treatment regime by PAL. Additional simulations results can be found in Section I of the Supplementary Material.

6.1. *Concordance-assisted learning.* Data are generated from the following model:

$$Y_i = h_0(X_i^1, X_i^3) + A_i Q(X_i^1 + X_i^2) + \varepsilon_i,$$

where $A_i \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(0.5)$, $X_i \overset{\text{i.i.d}}{\sim} N_p(0, I_p)$, $\varepsilon_i \overset{\text{i.i.d}}{\sim} N(0, 0.5^2)$, where $N_p(\mu, \Sigma)$ stands for the p-dimensional multivariate normal distribution with mean μ, covariance matrix Σ and I_p denotes the $p \times p$ identity matrix.

TABLE 1
Simulation settings in Section 6.1

	S1	S2	S3	S4
$h_0(x, y)$	$1 + x - y$	$1 + x - y$	$1 + xy$	$1 + xy$
$Q(x)$	x	$\exp(x) - 1$	x	$\exp(x) - 1$

We design four settings by considering two choices of h_0 and two choices of Q. The functional forms of h_0 and Q in each setting are listed in Table 1. It can be verified that in all four settings, the optimal treatment regime takes the form:

$$d^{\text{opt}}(x) = \mathbb{I}(x^1 + x^2 > 0).$$

We set $p = 8$, and consider two choices of the sample size, $n = 100$ and $n = 200$, respectively. This gives a total of 8 scenarios. For each scenario, we report the false positives (FP) rate (the percentage of unimportant variables that are selected),

$$\text{FP} = \frac{1}{L}\sum_{l=1}^{L}\frac{|\mathcal{M}_{\beta_0}^c \cap \widehat{\mathcal{M}}^{(l)}|}{|\mathcal{M}_{\beta_0}^c|},$$

the false negatives (FN) rate (the percentage of important variables that are missed),

$$\text{FN} = \frac{1}{L}\sum_{l=1}^{L}\frac{|\mathcal{M}_{\beta_0} \cap (\widehat{\mathcal{M}}^{(l)})^c|}{|\mathcal{M}_{\beta_0}|},$$

the percentage of selecting the true models (TP),

$$\text{TP} = \frac{1}{L}\sum_{l=1}^{L}\mathbb{I}(\mathcal{M}_{\beta_0} = \widehat{\mathcal{M}}^{(l)}),$$

the average error rate (ER) and average ratio of value (VR) of the estimated optimal treatment regime,

$$\text{ER} = \frac{1}{L}\sum_{l=1}^{L}\text{E}|\hat{d}^{(l)}(X_0) - d^{\text{opt}}(X_0)|, \qquad \text{VR} = \frac{1}{L}\sum_{l=1}^{L}\frac{\text{E}Y_0^*(\hat{d}^{(l)})}{\text{E}Y_0^*(d^{\text{opt}})}, \tag{6.1}$$

where $\hat{d}^{(l)}(x) = \mathbb{I}(\hat{c}_{\widehat{\mathcal{M}}^{(l)}} + x^T\hat{\beta}_{\widehat{\mathcal{M}}^{(l)}} > 0)$, $\widehat{\mathcal{M}}^{(l)}$ is the set of important variables selected by our information criteria in the lth simulation and L is the total number of simulations. In our implementation, we set $L = 100$ and approximate the expectations in (6.1) by the use of 1000 Monte Carlo samples.

We use CAL to estimate the parameters. Specifically, we first fit a logistic regression model with SCAD penalty function for the propensity score, and a linear model with SCAD penalty for the baseline function. Next, we obtain $\hat{\beta}_{\mathcal{M}}$ by maximizing $\widehat{C}^{DR}$ for all $2^8 = 256$ models. The threshold $\hat{c}_{\mathcal{M}}$ is obtained by maximizing the estimated value function $\widehat{V}^{DR}$ among the class of regimes $\mathbb{I}(c + x^T\hat{\beta}_{\mathcal{M}} > 0)$. We use the genetic algorithm implemented in the R package `rgenoud` (Mebane et al. (2011)) to compute the maximizers of the value and concordance functions. The package `rgenoud` combines evolutionary search algorithms with derivative-based methods to solve difficult optimization problems. In our experiments, we find the maximizers are very close to the true parameters. However, there is no guarantee that the searching algorithms will find the global maximizer in general, due to nonconvexity of the optimization problem. We use CIC^{DR} and VIC^{DR} for model selection. The model complexity penalty κ_n is chosen according to the discussion in Section 5.2. The propensity score model is always

TABLE 2
Simulation results (%, standard deviations in parenthesis)

		S1		S2	
	n	100	200	100	200
CIC^{DR}	TP	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)
	FN	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	FP	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	ER	7.58 (0.53)	5.71 (0.38)	7.54 (0.51)	5.42 (0.39)
	VR	98.95 (0.15)	99.38 (0.08)	99.49 (0.06)	99.73 (0.03)
VIC^{DR}	TP	77.00 (4.23)	99.00 (1.00)	71.00 (4.56)	97.00 (1.71)
	FN	11.5 (2.11)	0.50 (0.50)	14.50 (2.28)	1.50 (0.86)
	FP	0.17 (0.17)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	ER	11.49 (0.92)	5.78 (0.42)	12.84 (1.02)	5.93 (0.54)
	VR	96.65 (0.48)	99.33 (0.12)	97.94 (0.27)	99.56 (0.11)
		S3		S4	
	n	100	200	100	200
CIC^{DR}	TP	69.00 (4.65)	78.00 (4.16)	82.00 (3.86)	93.00 (2.56)
	FN	4.00 (1.36)	0.50 (0.50)	0.00 (0.00)	0.00 (0.00)
	FP	6.50 (1.16)	4.17 (0.87)	3.5 (0.83)	1.33 (0.51)
	ER	14.90 (0.89)	9.87 (0.53)	13.04 (0.87)	9.12 (0.71)
	VR	95.50 (0.51)	98.09 (0.2)	98.41 (0.18)	99.22 (0.11)
VIC^{DR}	TP	42.00 (4.96)	70.00 (4.61)	43.00 (4.98)	71.00 (4.56)
	FN	23.50 (2.61)	8.50 (1.89)	32.00 (3.37)	12.50 (2.18)
	FP	6.17 (1.02)	3.17 (0.77)	5.50 (0.95)	1.33 (0.51)
	ER	19.69 (1.1)	11.85 (0.81)	22.49 (1.44)	13.2 (1.06)
	VR	91.88 (0.74)	96.78 (0.41)	93.93 (0.79)	98.12 (0.24)

correct, hence our information criteria are consistent. We use 100 simulations replications. Results were given in Table 2.

We make the following observations. First, CIC^{DR} perform much better than VIC^{DR} in all scenarios. For example, in Settings 1 and 2, CIC^{DR} always chooses the correct model while TPs of VIC^{DR} are below 80% when $n = 100$. In Settings 3 and 4, TPs of CIC^{DR} are still much higher than those of VIC^{DR}. In addition, in all scenarios, CIC^{DR} achieves a smaller ER and a higher VR compared to VIC^{DR}. Moreover, the model selection results improve when sample size increases. This illustrates the selection consistencies of our information criteria.

6.2. *Penalized A-learning*. Consider the high dimensional setting where p is set to be 1000. We generate the response from the following model:

$$Y_i = h_0\big(X_i^1, X_i^3\big) + A_i\big(X_i^1 + X_i^2\big) + \varepsilon_i,$$

where $X_i \overset{\text{i.i.d}}{\sim} N_p(0, I_p)$, $A_i \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(\pi_0(X_i))$, $\varepsilon_i \overset{\text{i.i.d}}{\sim} N(0, 0.5^2)$. The contrast function takes the linear form, $\tau(x) = x^1 + x^2$ and the optimal treatment regime is

$$d^{\text{opt}}(x) = \mathbb{I}\big(x^1 + x^2 > 0\big).$$

We design four settings by considering two choices of the baseline function, and two choices of the propensity score function. Table 3 gives the propensity and baseline function in each setting. We fit a penalized linear regression model for the baseline and a penalized

TABLE 3
Simulation settings in Section 6.2

	S1	S2	S3	S4
$h_0(x, y)$	$1 + x - y$	$1 + xy$	$1 + x - y$	$1 + xy$
$\pi_0(x)$	0.5	0.5	$\Phi(x_{p-1} - x_p)$	$\Phi(x_{p-1} - x_p)$

$\Phi(\cdot)$ stands for the cumulative distribution function of a standard normal variable.

logistic regression model for the propensity score, and choose SCAD as the penalty function. Hence, both the propensity score and baseline models are correctly specified in Setting 1. One of them is misspecified in Settings 2 and 3. In Setting 4, both models are misspecified. In implementation, instead of directly optimizing (5.2), we solve its dual problem:

$$\bar{\theta} = \min_{(c,\beta^T)} \left\| \sum_i \tilde{X}_i (A_i - \hat{\pi}_i)\big(Y_i - \hat{h}_i - A_i c - A_i X_i^T \beta\big) \right\|_\infty,$$
$$\text{subject to } \|\beta\|_1 \le \lambda.$$

We compute $\bar{\beta}$ for a series of log-spaced values $\exp(-3) = \lambda_0, \lambda_1, \ldots, \lambda_{100} = \exp(2)$, and obtain θ by refitting the A-learning estimating equation. Tuning parameters are selected by CIC^{DR} and VIC^{DR}. In CIC^{DR}, we set

$$\kappa_n = \log(p) \log_{10}(n) \log\big(\log_{10}(n)\big),$$

as discussed in Section 5.3. In VIC^{DR}, we set

$$\kappa_n = n^{1/3} \log^{2/3}(p) \log\big(\log(n)\big)/\kappa,$$

where κ is a constant from a set $\{3, 4, 5\}$. For each κ, we denote the corresponding information criterion as VIC_κ^{DR}.

We further compare our information criteria with the BIC-type criterion (Shi et al. (2018)), which is used for tuning parameter selection for the PAL method. For any $\theta = (c, \beta^T)^T$, define

$$\text{BIC}(\theta) = n \log\big(\text{RSS}(\theta)/n\big) + \|\beta\|_0 \big\{\log(n) + \log(p+1)\big\},$$

where

$$\text{RSS}(\theta) = \sum_{i=1}^n (A_i - \hat{\pi}_i)^2 \big(Y_i - \hat{h}_i - A_i c - A_i X_i^T \beta\big)^2.$$

It remains unknown whether this information criterion is consistent.

Tables 4 and 5 report the results with sample size $n = 200/300$ and 100 simulation replications. CIC^{DR} outperform VIC^{DR} and BIC in all settings, in terms of TP. For example, in Setting 2 with $n = 200$, CIC^{DR} correctly recover 77% of the models, while TPs for other criteria are smaller than 70%. In addition, except for Setting 2, VIC^{DR} outperforms BIC in all other settings. Take Setting 3 with $n = 300$ as an example, TPs for VIC_3^{DR}, VIC_4^{DR}, VIC_5^{DR} are all very close to 1 while BIC only correctly recovers 61% of the models. False positives of BIC are much higher compared to our information criteria in Setting 3. Moreover, all the information criteria work extremely well in Setting 1 where both the propensity score and baseline models are correctly specified, and perform much worse in Setting 4 where both models are misspecified. Except for BIC and VIC_5^{DR}, all other criteria always select the true model in Setting 1. In Setting 4 with $n = 200$, however, TPs of all criteria are below 50%.

TABLE 4
Simulation results for Settings 1 *and* 2 (%, *standard deviations in parenthesis*)

		S1		S2	
	n	200	300	200	300
CIC^{DR}	TP	100.00 (0.00)	100.00 (0.00)	77.00 (4.23)	90.00 (3.02)
	FN	0.00 (0.00)	0.00 (0.00)	7.00 (1.88)	0.50 (0.50)
	FP	0.00 (0.00)	0.00 (0.00)	0.03 (0.01)	0.01 (0.00)
	ER	1.18 (0.08)	1.17 (0.09)	8.34 (1.09)	4.08 (0.43)
	VR	99.96 (0.00)	99.97 (0.00)	97.02 (0.66)	99.39 (0.15)
BIC	TP	95.00 (2.19)	94.00 (2.39)	69.00 (4.65)	89.00 (3.14)
	FN	0.00 (0.00)	0.00 (0.00)	4.50 (1.60)	0.00 (0.00)
	FP	0.01 (0.00)	0.01 (0.00)	0.03 (0.01)	0.02 (0.00)
	ER	1.44 (0.12)	1.34 (0.13)	7.98 (0.85)	4.00 (0.41)
	VR	99.94 (0.01)	99.94 (0.02)	97.77 (0.47)	99.44 (0.11)
VIC_3^{DR}	TP	100.00 (0.00)	100.00 (0.00)	61.00 (4.90)	86.00 (3.49)
	FN	0.00 (0.00)	0.00 (0.00)	13.50 (2.34)	2.00 (0.98)
	FP	0.00 (0.00)	0.00 (0.00)	0.02 (0.01)	0.01 (0.00)
	ER	1.16 (0.08)	1.08 (0.08)	10.64 (1.08)	4.67 (0.56)
	VR	99.97 (0.00)	99.97 (0.00)	96.13 (0.61)	99.12 (0.21)
VIC_4^{DR}	TP	100.00 (0.00)	100.00 (0.00)	56.00 (4.99)	80.00 (4.02)
	FN	0.00 (0.00)	0.00 (0.00)	11.00 (2.20)	2.00 (0.98)
	FP	0.00 (0.00)	0.00 (0.00)	0.05 (0.01)	0.02 (0.00)
	ER	1.08 (0.08)	1.11 (0.09)	11.05 (1.13)	5.09 (0.54)
	VR	99.97 (0.00)	99.97 (0.00)	95.98 (0.66)	99.02 (0.21)
VIC_5^{DR}	TP	99.00 (1.00)	100.00 (0.00)	54.00 (5.01)	72.00 (4.51)
	FN	0.00 (0.00)	0.00 (0.00)	9.00 (2.06)	2.00 (0.98)
	FP	0.00 (0.00)	0.00 (0.00)	0.07 (0.01)	0.03 (0.01)
	ER	1.18 (0.09)	1.1 (0.08)	11.20 (1.14)	5.80 (0.59)
	VR	99.96 (0.01)	99.97 (0.00)	95.92 (0.72)	98.81 (0.22)

7. Some technical results. In this section, we summarize some major technical results used in the proof of our theorems. They are generally applicable and self-important. In Section 7.1, we present a tail inequality for unbounded degenerate U-process that is useful to show model selection consistency of CIC and CIC_{DR}. In Section 7.2, we show uniform consistencies of empirical maximizers of $\widehat{V}$ and $\widehat{C}$, which enable us to bound the probability that VIC or CIC selects an overfitted model in the ultrahigh dimension.

7.1. *Tail inequality for unbounded degenerate U-process.* In this subsection, we provide a tail inequality for the supremum of order two U-process with finite ψ_1 Orlicz norm. We first introduce some notation. Let $X_1, \ldots, X_n$ be i.i.d. random variables taking values on $\mathcal{X}$, $\mathcal{F}$ a countable class of measurable and symmetric functions from $\mathcal{X} \times \mathcal{X}$ to $\mathbb{R}$.

THEOREM 7.1. *Assume* f *satisfies* $\mathrm{E}f(X_i, x) = \mathrm{E}f(x, X_i) = 0$, $f(x, x) = 0$ *for any* x, *and* $\omega_n = \|\max_{i \neq j} F(X_i, X_j)\|_{\psi_1} < \infty$, *where the function* F *satisfies* $F(x, y) \geq \sup_f |f(x, y)|$ *for any* x, y. *Define the following degenerate U-process*:

$$Z = \sup_{f \in \mathcal{F}} \left| \sum_{i,j} f(X_i, X_j) \right|.$$

TABLE 5
Simulation results for Settings 3 *and* 4 (%, *standard deviations in parenthesis*)

		S3		S4	
	n	200	300	200	300
CIC^{DR}	TP	91.00 (2.88)	99.00 (1.00)	48.00 (5.02)	66.00 (4.76)
	FN	3.00 (1.19)	0.00 (0.00)	27.00 (3.21)	14.00 (2.47)
	FP	0.00 (0.00)	0.00 (0.00)	0.03 (0.01)	0.03 (0.01)
	ER	3.15 (0.61)	1.42 (0.11)	16.50 (1.50)	10.90 (1.27)
	VR	99.23 (0.28)	99.95 (0.01)	92.09 (1.03)	95.63 (0.75)
BIC	TP	55.00 (5)	61.00 (4.9)	32.00 (4.69)	40.00 (4.92)
	FN	0.00 (0.00)	0.00 (0.00)	18 (3.06)	8.00 (2.10)
	FP	0.08 (0.01)	0.11 (0.02)	0.14 (0.02)	0.15 (0.02)
	ER	4.23 (0.33)	3.69 (0.35)	17.4 (1.35)	13.99 (1.2)
	VR	99.48 (0.06)	99.57 (0.06)	92.27 (0.95)	94.52 (0.74)
VIC_3^{DR}	TP	82.00 (3.86)	98.00 (1.41)	39.00 (4.90)	59.00 (4.94)
	FN	7.00 (1.74)	0.00 (0.00)	29.50 (3.26)	17.50 (2.60)
	FP	0.01 (0.00)	0.00 (0.00)	0.05 (0.01)	0.03 (0.01)
	ER	5.16 (0.85)	1.38 (0.12)	18.02 (1.46)	12.39 (1.29)
	VR	98.4 (0.38)	99.94 (0.01)	91.4 (1.01)	94.79 (0.77)
VIC_4^{DR}	TP	89.00 (3.14)	98.00 (1.41)	43.00 (4.98)	59.00 (4.94)
	FN	2.50 (1.10)	0.00 (0.00)	26.00 (3.29)	14.50 (2.49)
	FP	0.01 (0.00)	0.00 (0.00)	0.06 (0.01)	0.05 (0.01)
	ER	2.97 (0.54)	1.42 (0.13)	17.41 (1.53)	12.2 (1.28)
	VR	99.35 (0.24)	99.94 (0.01)	91.61 (1.06)	94.88 (0.78)
VIC_5^{DR}	TP	90.00 (3.02)	97.00 (1.71)	43.00 (4.98)	54.00 (5.01)
	FN	2.00 (0.98)	0.00 (0.00)	23.00 (3.21)	13.50 (2.45)
	FP	0.01 (0.00)	0.00 (0.00)	0.07 (0.01)	0.07 (0.01)
	ER	2.85 (0.52)	1.43 (0.13)	16.95 (1.48)	12.52 (1.28)
	VR	99.4 (0.25)	99.94 (0.01)	91.98 (1.02)	94.73 (0.78)

Let $\varepsilon_1, \ldots, \varepsilon_n$ *be i.i.d. Rademacher random variables independent of* $\{X_1, \ldots, X_n\}$, *and introduce the random variables*:

$$Z_\varepsilon = \sup_{f\in\mathcal{F}}\left|\sum_{i,j}\varepsilon_i\varepsilon_j f(X_i, X_j)\mathbb{I}\big(F(X_i, X_j) \le 8\omega_n\big)\right|,$$

$$U_\varepsilon = \sup_{f\in\mathcal{F}}\sup_{\alpha:\|\alpha\|_2\le 1}\sum_{i,j}\varepsilon_i\alpha_j f(X_i, X_j)\mathbb{I}\big(F(X_i, X_j) \le 8\omega_n\big),$$

$$M_\varepsilon = \sup_{f\in\mathcal{F}},\ \sup_{k=1,\ldots,n}\left|\sum_i \varepsilon_i f(X_i, X_k)\mathbb{I}\big(F(X_i, X_k) \le 8\omega_n\big)\right|.$$

Then there exists some constants $C > 0$ *such that for all* n *and* $t > 0$,

$$\begin{aligned}&\Pr(Z > C\mathrm{E}Z_\varepsilon + t)\\ &\le 3\exp\left(-\min\left(\frac{t^2}{(\mathrm{E}U_\varepsilon)^2}, \frac{t}{\mathrm{E}M_\varepsilon}, \frac{t}{n\omega_n}, \left(\frac{t}{\omega_n\sqrt{n}}\right)^{2/3}, \sqrt{\frac{t}{\omega_n}}\right)\right).\end{aligned} \tag{7.1}$$

REMARK 7.2. For bounded degenerate U-process, that is, $F \le F_0$ for some constant F_0, Clémençon, Lugosi and Vayatis (2008) showed LHS of (7.1) can be bounded by

$$\exp\left(-\min\left(\frac{t^2}{(\mathrm{E}U_\varepsilon)^2}, \frac{t}{\mathrm{E}M_\varepsilon}, \frac{t}{nF_0}, \left(\frac{t}{F_0\sqrt{n}}\right)^{2/3}, \sqrt{\frac{t}{F_0}}\right)\right). \tag{7.2}$$

For unbounded U-process whose envelope function has finite ψ_1 Orlicz norm, it is natural to replace the uniform bound F_0 in (7.2) by ω_n. Upper bounds for the Rademacher complexities $\mathrm{E}Z_\varepsilon$, $\mathrm{E}M_\varepsilon$ and $\mathrm{E}U_\varepsilon$ can be obtained as in Clémençon, Lugosi and Vayatis (2008).

7.2. *Uniform consistency of empirical maximizers.* Recall that

$$\Omega_+ = \{\lambda \in [\lambda_{\min}, \lambda_{\max}] : \mathcal{M}_{\beta_0} \subsetneq \widehat{\mathcal{M}}(\lambda), |\widehat{\mathcal{M}}(\lambda)| \le s_n\}.$$

For any $\lambda \in \Omega_+$, define

$$\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)} = \mathop{\arg\max}_{\substack{\theta=(c,\beta^T)^T \in \tilde{S}(\theta_0) \\ \beta^{\widehat{\mathcal{M}}(\lambda)^c}=0}} \widehat{V}(\theta), \qquad \tilde{\beta}_{\widehat{\mathcal{M}}(\lambda)} = \mathop{\arg\max}_{\substack{\beta \in S(\beta_0) \\ \beta^{\widehat{\mathcal{M}}(\lambda)^c}=0}} \widehat{C}(\beta).$$

By the definitions of CIC and VIC, the probabilities that VIC and CIC choose an overfitted model are upper bounded by

$$\Pr\Big(\mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) \le \sup_{\lambda\in\Omega_+} \{n\widehat{V}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}) - \kappa_n \|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\}\Big), \tag{7.3}$$

$$\Pr\Big(\mathrm{CIC}(\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}) \le \sup_{\lambda\in\Omega_+} \{n\widehat{C}(\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}) - \kappa_n \|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\}\Big). \tag{7.4}$$

Notice that $\widehat{V}(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)}) \ge \widehat{V}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)})$, $\widehat{C}(\tilde{\beta}_{\widehat{\mathcal{M}}(\lambda)}) \ge \widehat{C}(\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)})$. Therefore, (7.3) and (7.4) are upper bounded by

$$\Pr\Big(\mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) \le \sup_{\lambda\in\Omega_+} \{n\widehat{V}(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)}) - \kappa_n \|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\}\Big), \tag{7.5}$$

$$\Pr\Big(\mathrm{CIC}(\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}) \le \sup_{\lambda\in\Omega_+} \{n\widehat{C}(\tilde{\beta}_{\widehat{\mathcal{M}}(\lambda)}) - \kappa_n \|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\}\Big), \tag{7.6}$$

respectively. To bound (7.5) and (7.6), we need uniform convergence rates of $\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)}$ and $\tilde{\beta}_{\widehat{\mathcal{M}}(\lambda)}$ over all $\lambda \in \Omega_+$, summarized as follows.

LEMMA 7.1. *Under the conditions in Theorem 3.3, there exists some constant $t_0 > 0$ such that for all $t \ge t_0$,*

$$\begin{aligned}&\Pr\Big(\bigcap_{\lambda\in\Omega_+} \{\|\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)} - \theta_0\|_2 \ge t n^{-1/3} |\widehat{\mathcal{M}}(\lambda)|^{1/3} \log^{1/3} p\}\Big) \\ &\le \exp(-\bar{c}t^3 \log(p)) + \exp\Big(-\frac{\bar{c}t^2 n^{2/3} \log^{1/3} p}{\log(n)}\Big) + \exp\Big(-\frac{\bar{c}n}{\log(n)}\Big).\end{aligned} \tag{7.7}$$

Under the conditions in Theorem 3.4, there exists some constant $t_0 > 0$ such that for all $t \ge t_0$,

$$\begin{aligned}&\Pr\Big(\bigcap_{\lambda\in\Omega_+} \{\|\tilde{\beta}_{\widehat{\mathcal{M}}(\lambda)} - \beta_0\|_2 \ge t n^{-1/2} |\widehat{\mathcal{M}}(\lambda)|^{1/2} \log^{1/2} p \log^{1/2}(n)\}\Big) \\ &\le \exp(-\bar{c}t^2 \log(p)) + \exp(-\bar{c}t\sqrt{n\log(p)}) + \exp\Big(-\frac{\bar{c}n}{\log(n)}\Big).\end{aligned} \tag{7.8}$$

REMARK 7.3. In the fixed-p scenario, $\tilde{\theta}_{\mathcal{M}}$ converges at a rate of $O_p(n^{-1/3})$. In comparison, the uniform convergence rate in (7.7) is slower by a factor of $|\widehat{\mathcal{M}}(\lambda)|^{1/3} \log^{1/3} p$. This is the price we pay to search over the entire overfitted model space. By assumption, we have $\log(p) = O(n^{a_0})$, $s_n = O(n^{l_0})$, $\sup_{\lambda\in\Omega_+} |\widehat{\mathcal{M}}(\lambda)| \le s_n$. When $a_0 + l_0 < 1$, we have

$$n^{-1/2} |\widehat{\mathcal{M}}(\lambda)|^{1/2} \log^{1/2} p \log^{1/2}(n) \ll n^{-1/3} |\widehat{\mathcal{M}}(\lambda)|^{1/3} \log^{1/3} p, \quad \text{for all } \lambda \in \Omega_+.$$

Therefore, it follows from (7.7) and (7.8) that $\sup_{\lambda\in\Omega^+}\|\tilde{\beta}_{\widehat{\mathcal{M}}(\lambda)}-\beta_0\|_2$ converges faster than $\sup_{\lambda\in\Omega_+}\|\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}-\theta_0\|_2$.

8. Discussion. In this paper, we propose the concordance and value information criteria (CIC and VIC) to select important variables that are involved in the optimal treatment regime. We consider both fixed-p and high dimensional settings, and show that VIC and CIC are able to correctly identify those important variables in both scenarios when the contrast is a monotonic function of a linear combination of baseline covariates. In addition, we show CIC is more reliable than VIC both theoretically and empirically.

8.1. *Extensions to multiple stages.* The proposed concordance and value information criteria can be extended to multistage settings, where models are selected via backward induction. These results are provided in Section 11 of the Supplementary Material. We find out that if the contrast function on each stage is a monotonic function of a linear combination of available covariates and previous treatments up to that stage, our information criteria are consistent. Otherwise, estimators selected by our information criteria will converge to some least false parameters and it is likely that CIC and VIC choose different models. In addition, conditions on κ_n are strengthened in backward induction, due to the variability in the estimation of the contrast function of previous stages.

8.2. *Model misspecification.* In Section 12 of the Supplementary Material, we further investigate the performance of the proposed information criteria when the contrast function does not take the monotonic linear index form. Theorem 12.1 shows the model CIC and VIC choose will converge to the support of some least false parameters. We further conduct simulation studies in Section 12.2. We find CIC achieves better model selection results when compared to VIC in finite samples. In addition, all the numerical results improve when sample size increases, validating our theoretical findings.

8.3. *Nonregularity.* Our method requires assuming the uniqueness of the optimal treatment. In the nonregular cases where $\Pr(\tau(X_0)=0)>0$, Conditions (A5)(ii), (A5′)(ii), (A6)(iii) and (A6′)(iii) are likely to be violated. More detailed discussions can be found in Section B.1.2 and Section B.2.3 of the Supplementary Material. Thus, selection consistencies of our proposed information criteria are not guaranteed. We further investigate the numerical performance of our proposed information criteria in the nonregular cases. Results are provided in Section I.1 of the Supplementary Material. We find CIC still works better when compared to VIC. However, increasing the sample size does not improve the performance of CIC. This suggests that our information criteria might not be consistent in this case.

9. Proof of Theorem 3.3. Here, we only present the proof of Theorem 3.3. Proofs of other theorems and lemmas are given in the Supplementary Material. Let Ω_- be the underfitted model space,

$$\Omega_-=\{\lambda\in[\lambda_{\min},\lambda_{\max}]:\mathcal{M}_{\beta_0}\not\subset\widehat{\mathcal{M}}(\lambda),|\widehat{\mathcal{M}}(\lambda)|\le s_n\}.$$

Assumption (A4′) states that

$$\Pr(\{\|\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}-\theta_0\|_2=O(R_n)\}\cap\{\widehat{\mathcal{M}}(\lambda_0)=\mathcal{M}_{\beta_0}\})\to 1. \tag{9.1}$$

Under the events defined in (9.1), to prove Theorem 3.3, we provide tail inequalities for

$$\Pr\Big(\mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})\le\sup_{\lambda\in\Omega_-}\mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)})\Big). \tag{9.2}$$

Then we bound

$$\Pr\Big(\mathrm{VIC}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})\le\sup_{\lambda\in\Omega_+}\{n\widehat{V}(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})-\kappa_n\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\}\Big). \tag{9.3}$$

9.1. *Underfitted model space.* Since $V(\theta_0) > V(0)$, we have $\theta_0 \neq 0$. For any $\theta = (c, \beta^T)^T \neq 0$ and x, we have

$$\mathbb{I}(\beta^T x > -c) = I\Big(\frac{\beta^T}{\|\theta\|_2}\|\theta_0\|_2 x > -\frac{c}{\|\theta\|_2}\|\theta_0\|_2\Big).$$

This implies we have

$$V(\theta) = V\big(\|\theta_0\|_2\theta/\|\theta\|_2\big), \tag{9.4}$$

for any $\theta \neq 0$. The vector $\|\theta_0\|_2\theta/\|\theta\|_2$ lies on the L_2 surface $\tilde{S}(\theta_0)$.

For any $\lambda \in \Omega_-$, we have $\beta_0^j \neq 0$ and $\hat{\beta}^j_{\widehat{\mathcal{M}}(\lambda)} = 0$ for some j. By the definition of δ, this implies

$$\Big\|\theta_0 - \frac{\|\theta_0\|_2\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}}{\|\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}\|_2}\Big\|_2 \geq |\beta_0^j| > \delta,$$

or $\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)} \notin \tilde{N}_\delta$, if $\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)} \neq 0$. Since $\delta \leq \varepsilon_0$, it follows from Assumption (A5′)(i) that there exists some constant $\xi > 0$ such that

$$V(\theta_0) > V\Big(\frac{\|\theta_0\|_2\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}}{\|\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}\|_2}\Big) + 3\xi.$$

It follows from (9.4) that

$$V(\theta_0) > V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}) + 3\xi. \tag{9.5}$$

By assumption (A5′)(i), we have $V(\theta_0) > V(0)$. Without loss of generality, assume $3\xi < V(\theta_0) - V(0)$. Then (9.5) holds for any $\lambda \in \Omega_-$. Assumptions (A5′)(iii) and the event defined in (9.1) imply that

$$V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) \geq V(\theta_0) - O\big(R_n^2\big). \tag{9.6}$$

It follows from (9.5) and (9.6) that

$$V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) \geq \sup_{\lambda\in\Omega_-} V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}) + 3\xi - O\big(R_n^2\big).$$

Since the sequence $R_n \to 0$, for sufficiently large n, we have $\xi \geq O(R_n^2)$. Hence,

$$V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) - \sup_{\lambda\in\Omega_-} V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}) \geq 2\xi, \tag{9.7}$$

for sufficiently large n. Since the number of nonzero elements in $\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}$ is fixed, we have

$$\kappa_n\big(\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}\|_0 - \|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\big) \leq O(\kappa_n) \quad \text{for all } \lambda \in \Omega_-.$$

Together with (9.7) and the condition $\kappa_n = o(n)$, we obtain that for sufficiently large n and all $\lambda \in \Omega_-$,

$$\big\{V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) - V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)})\big\} - \frac{\kappa_n}{n}\big(\|\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}\|_0 - \|\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\big) \geq \xi. \tag{9.8}$$

By (9.8) and the definition of VIC, the event defined in (9.2) happens if

$$\sup_{\lambda\in\Omega_-} \big|\big\{\widehat{V}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) - V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)}) - \widehat{V}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)}) + V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda)})\big\}\big| \geq \xi,$$

or

$$\sup_{\|\beta\|_0\leq s_n, c\in\mathbb{R}} \big|\widehat{V}(\theta) - V(\theta)\big| \geq \frac{\xi}{2}.$$

Therefore, we can bound (9.2) by

$$\Pr\Big(\sup_{c\in\mathbb{R},\|\beta\|_0\le s_n}|\widehat{V}(\theta)-V(\theta)|\ge\frac{\xi}{2}\Big). \tag{9.9}$$

We now provide an upper bound for (9.9). Define $B_{\mathcal{M}}=\{\beta\in\mathbb{R}^p:\beta^{\mathcal{M}^c}=0\}$. We define $\Omega^*=\{\mathcal{M}\in\Omega:|\mathcal{M}|=s_n\}$. It follows from Bonferroni's inequality that (9.9) is bounded by

$$\sum_{\mathcal{M}\in\Omega^*}\Pr\Big(\sup_{c\in\mathbb{R},\beta\in B_{\mathcal{M}}}|\widehat{V}(\theta)-V(\theta)|\ge\frac{\xi}{2}\Big). \tag{9.10}$$

For any triple $o=(y,a,x)$, define

$$\psi_\theta^V(o)=\Big(\frac{a}{\pi_0(x)}-\frac{1-a}{1-\pi_0(x)}\Big)y\mathbb{I}\big(x^T\beta>-c\big)+\frac{1-a}{1-\pi_0(x)}y.$$

For each fixed $\mathcal{M}$, the class of functions $\mathcal{V}_{\mathcal{M}}=\{\psi_\theta^V:c\in\mathbb{R},\beta\in B_{\mathcal{M}}\}$ has finite VC index s_n+3 (see Lemma 2.6.15 and 2.6.18 in van der Vaart and Wellner (1996)). Therefore, we have

$$\begin{aligned}J(1,\mathcal{V})&\equiv\sup_Q\int_0^1\sqrt{1+\log N\big(\varepsilon\|V_{\mathcal{M}}\|_{Q,2},\mathcal{V}_{\mathcal{M}},L_2(Q)\big)}\,d\varepsilon\\&\le\int_0^1\sqrt{1+(s_n+3)\log(K/\varepsilon)}\,d\varepsilon=O(\sqrt{s_n}),\end{aligned} \tag{9.11}$$

for some constant K, where $V_{\mathcal{M}}$ stands for an envelope function of $\mathcal{V}_{\mathcal{M}}$, and the supremum is taken over all discrete measures Q with $\|V_{\mathcal{M}}\|_{Q,2}>0$. The definition of the entropy number $N(\cdot,\cdot,\cdot)$ can be found in van der Vaart and Wellner (1996). The above bound is uniform for all $\mathcal{M}\in\Omega^*$.

For any $\mathcal{M}$, $V_{\mathcal{M}}(O_i)$ is bounded by

$$\begin{aligned}&\sup_{\theta=(c,\beta^T)^T}\Big|\frac{A_i\mathbb{I}(X_i^T\beta>-c)+(1-A_i)\mathbb{I}(X_i^T\beta\le-c)}{A_i\pi_{0,i}+(1-A_i)(1-\pi_{0,i})}Y_i\Big|\\&\le\sup_{c,\beta}\Big|\frac{A_i\mathbb{I}(X_i^T\beta>-c)+(1-A_i)\mathbb{I}(X_i^T\beta\le-c)}{A_i\pi_{0,i}+(1-A_i)(1-\pi_{0,i})}Y_i\Big|\le\frac{1}{(1-c_2)c_1}|Y_i|,\end{aligned} \tag{9.12}$$

by Assumption (A3). In addition, it follows from Lemma H.1 and Cauchy–Schwarz inequality that

$$\big(\mathrm{E}|Y_i|\big)^2\le\mathrm{E}|Y_i|^2\le2\|Y_i\|_{\psi_1}^2=O(1). \tag{9.13}$$

Therefore, we have $\mathrm{E}V_{\mathcal{M}}^2(O_1)=O(1)$ where the big-O notation is uniform in $\mathcal{M}$.

It follows from (9.11) and Theorem 2.14.1 in van der Vaart and Wellner (1996) that

$$\mathrm{E}\sup_{\substack{c\in\mathbb{R}\\\beta\in B_{\mathcal{M}}}}|\widehat{V}(\theta)-V(\theta)|\le O(1)\frac{\sqrt{s_n}}{n}\sqrt{n\mathrm{E}V_{\mathcal{M}}^2(O_1)}. \tag{9.14}$$

Here, $O(1)$ denotes a universal constant that is independent of $\mathcal{M}$.

This together with (9.12) and (9.13) implies

$$\sup_{\mathcal{M}\in\Omega^*}\mathrm{E}\Big(\sup_{c\in\mathbb{R},\beta\in B_{\mathcal{M}}}|\widehat{V}(\theta)-V(\theta)|\Big)=O\Big(\sqrt{\frac{s_n}{n}}\Big). \tag{9.15}$$

For sufficiently large n, RHS of (9.15) goes to 0. It follows from (9.15) that (9.10) is bounded by

$$\sum_{\mathcal{M}\in\Omega^*}\Pr\Big(\sup_{c\in\mathbb{R},\beta\in B_{\mathcal{M}}}\big|\widehat{V}(\theta)-V(\theta)\big|-(1+\eta)\mathrm{E}\sup_{c\in\mathbb{R},\beta\in B_{\mathcal{M}}}\big|\widehat{V}(\theta)-V(\theta)\big|\geq\frac{\xi}{4}\Big),$$

for some fixed $\eta>0$.

For any β and c, it follows from (9.12) that $\sup_{\mathcal{M}\in\Omega^*}\|V_{\mathcal{M}}(O_i)\|_{\psi_1}=O(1)$. Similarly we have $\sup_{\mathcal{M}\in\Omega^*}\mathrm{E}V^2_{\mathcal{M}}(O_i)=O(1)$. Take $\eta=0.5$, it follows from Lemma H.4 that the above probability can be bounded by

$$|\Omega^*|\{\exp(-\bar{c}n)+3\exp\big(-\bar{c}n/\log(n)\big)\},\tag{9.16}$$

for some constant $\bar{c}>0$. Observe that $|\Omega^*|=O(p^{s_n})$. It follows from the condition $n\gg s_n\log(p)\log(n)$ that (9.16) is bounded by

$$\begin{aligned}&\exp\big(-\bar{c}n+k_1s_n\log(p)\big)+3\exp\big(-\bar{c}n/\log(n)+k_1s_n\log(p)\big)\\&\quad\leq 4\exp\big(-\bar{c}n/\log(n)\big)\leq\exp\big\{-k_2n/\big(2\log(n)\big)\big\}\leq\exp\big(-k_2\log(p)\big),\end{aligned}$$

for some constants k_1, $k_2>0$ and sufficiently large n. This provides the tail inequality that VIC chooses an underfitted model.

9.2. *Overfitted model space.* It follows from Lemma 7.1 that

$$\begin{aligned}&\Pr\Big(\bigcap_{\lambda\in\Omega_+}\{\|\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)}-\theta_0\|_2\leq t_0n^{-1/3}|\widehat{\mathcal{M}}(\lambda)|^{1/3}\log^{1/3}(p)\}\Big)\\&\geq 1-\exp\big(-\bar{c}t_0^3\log(p)\big)-\exp\Big(-\frac{\bar{c}t_0^2n^{2/3}\log^{1/3}p}{\log(n)}\Big)-\exp\Big(-\frac{\bar{c}n}{\log(n)}\Big),\\&\geq 1-3\exp\big(-\bar{c}t_0^3\log(p)\big)\geq 1-\exp\big(\log 3-\bar{c}t_0^3\log(p)\big)\geq 1-\exp\big(-\bar{c}^*\log(p)\big),\end{aligned}\tag{9.17}$$

for some $\bar{c},\bar{c}^*>0$, where the second inequality is due to the condition $\log(p)=O(n^{a_0})$ for some $0<a_0<1$, which further implies $n^{2/3}\log^{1/3}p\gg\log(n)\log(p)$ and $n\gg\log(p)\log(n)$.

On the event defined in (9.1), it follows from Assumption (A6′)(iii) that

$$\big|V(\theta_0)-V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})\big|=O\big(R_n^2\big).$$

This together with $\sup_{\lambda\in\Omega_+}V(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})\leq V(\theta_0)$ implies that

$$V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})\geq\sup_{\lambda\in\Omega_+}V(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})-O\big(R_n^2\big).\tag{9.18}$$

Denoted by s_β the number of nonzero elements in β_0. For any $\lambda\in\Omega_+$, we have $\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0>s_\beta$. Therefore, for any $\lambda\in\Omega_+$, we obtain

$$\frac{\kappa_n}{|\widehat{\mathcal{M}}(\lambda)|}\big(\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0-\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}\|_0\big)=\kappa_n\Big(1-\frac{s_\beta}{|\widehat{\mathcal{M}}(\lambda)|}\Big)\geq\frac{\kappa_n}{s_\beta+1}.\tag{9.19}$$

Since s_β is fixed, under the condition $\kappa_n\gg nR_n^2$, it follows from (9.18) and (9.19) that for any $\lambda\in\Omega_+$ and sufficiently large n,

$$\frac{1}{|\widehat{\mathcal{M}}(\lambda)|}\big\{nV(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})-nV(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})-\kappa_n\big(\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda_0)}\|_0-\|\hat{\beta}_{\widehat{\mathcal{M}}(\lambda)}\|_0\big)\big\}\geq\frac{\kappa_n}{2(s_\beta+1)}.$$

Hence, the event defined in (9.3) happens if

$$\sup_{\lambda\in\Omega_+}\frac{n}{|\widehat{\mathcal{M}}(\lambda)|}\big|\{\widehat{V}(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})-V(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})-\widehat{V}(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})+V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})\}\big|\geq\frac{\kappa_n}{2(s_\beta+1)},$$

or

$$\sup_{\lambda\in\Omega_+}\frac{n}{|\widehat{\mathcal{M}}(\lambda)|}\big|\widehat{m}_V(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})-m_V(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})\big|\geq\frac{\kappa_n}{4(s_\beta+1)},$$

and

$$\sup_{\lambda\in\Omega_+}\frac{n}{|\widehat{\mathcal{M}}(\lambda)|}\big|\widehat{m}_V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})-m_V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})\big|\geq\frac{\kappa_n}{4(s_\beta+1)},\tag{9.20}$$

where $\widehat{m}_V(\theta)=\widehat{V}(\theta)-\widehat{V}(\theta_0)$ and $m_V(\theta)=V(\theta)-V(\theta_0)$. Since $|\widehat{\mathcal{M}}(\lambda)|\geq 1$, for any $\lambda\in\Omega_+$, LHS of (9.20) is smaller than $n|\widehat{m}_V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})-m_V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})|$. In the following, we show that conditional on the event defined in the LHS of (9.17),

$$\begin{aligned}&\Pr\Big(\sup_{\lambda\in\Omega_+}\frac{n}{|\widehat{\mathcal{M}}(\lambda)|}\big|\widehat{m}_V(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})-m_V(\tilde{\theta}_{\widehat{\mathcal{M}}(\lambda)})\big|\geq\frac{\kappa_n}{4(s_\beta+1)}\Big)\\&\quad\leq\exp\big(-k_3\log(p)\big),\end{aligned}\tag{9.21}$$

for some constant $k_3>0$. Similarly, we can show

$$\Pr\Big(n\big|\widehat{m}_V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})-m_V(\hat{\theta}_{\widehat{\mathcal{M}}(\lambda_0)})\big|\geq\frac{\kappa_n}{4(s_\beta+1)}\Big)\leq\exp\Big(-\frac{k_4\kappa_n^2}{nR_n}\Big),$$

for some constant $k_4>0$. This together with (9.21) and (9.17) yields (3.9).

Let $R^V_{\mathcal{M}}=t_0n^{-1/3}|\mathcal{M}|^{1/3}\log^{1/3}p$, and $\Omega^*_+=\{\mathcal{M}\in\Omega:\mathcal{M}_{\beta_0}\subsetneq\mathcal{M},|\mathcal{M}|\leq s_n\}$, LHS of (9.21) is bounded by

$$\sum_{\mathcal{M}\in\Omega^*_+}\Pr\Big(\sup_{\substack{\theta=(c,\beta^T)^T\\c\in\mathbb{R},\beta\in B_{\mathcal{M}}\\\|\theta-\theta_0\|_2\leq R^V_{\mathcal{M}}}}\frac{n}{|\mathcal{M}|}\big|\widehat{m}_V(\theta)-m_V(\theta)\big|\geq\frac{\kappa_n}{4(s_\beta+1)}\Big),\tag{9.22}$$

using Bonferroni's inequality. Observe that

$$\widehat{m}_V(\theta)-m_V(\theta)=\frac{1}{n}\sum_i\{\psi^V_\theta(O_i)-\mathrm{E}\psi^V_\theta(O_i)\}.$$

Let $\Theta^V_{\mathcal{M}}=\{\theta=(c,\beta^T)^T:\beta\in B_{\mathcal{M}},c\in\mathbb{R},\|\theta-\theta_0\|_2\leq R^V_{\mathcal{M}}\}$. By Assumption (A3), the class of functions $\{|\psi^V_\theta(o)|:\theta\in\Theta^V_{\mathcal{M}}\}$ is bounded by

$$\Psi^V_{\mathcal{M}}(o)=\bar{c}\Big(\sup_{\theta\in\Theta^V_{\mathcal{M}}}|y|\{\big|\mathbb{I}(x^T\beta>-c)-\mathbb{I}(x^T\beta_0>-c_0)\big|\}\Big),$$

for some $\bar{c}>0$. Therefore, we have $\sup_{\mathcal{M}}\|\Psi^V_{\mathcal{M}}(O_i)\|_{\psi_1}=O(1)$. In addition, it follows from the Cauchy–Schwarz inequality that $\mathrm{E}|\Psi^V_{\mathcal{M}}(O_i)|^2$ is bounded by

$$\begin{aligned}&2\bar{c}^2\mathrm{E}Y_i^2\Big|\sup_{\theta\in\Theta^V_{\mathcal{M}}}\Big|\mathbb{I}(X_i^T\beta>-c)-\mathbb{I}(X_i^T\beta_0>-c_0)\Big|\Big|^2\\&\leq 2\bar{c}^2\mathrm{E}\Big(\mathrm{E}(Y_i^2|X_i)\Big|\sup_{\theta\in\Theta^V_{\mathcal{M}}}\Big|\mathbb{I}(X_i^T\beta>-c)-\mathbb{I}(X_i^T\beta_0>-c_0)\Big|\Big|\Big)\\&\leq 2\bar{c}^2\bar{C}\mathrm{E}\Big|\sup_{\theta\in\Theta^V_{\mathcal{M}}}\Big|\mathbb{I}(X_i^T\beta>-c)-\mathbb{I}(X_i^T\beta_0>-c_0)\Big|\Big|=O(R^V_{\mathcal{M}}),\end{aligned}\tag{9.23}$$

where the second inequality is due to the condition that $\sup_x\{\mathrm{E}Y_0^2|X_0=x\}\le\bar{C}$, and the second equality is due to Assumption (A5′)(ii). The big-O term on the right-hand side is uniform in $\mathcal{M}\in\Omega_+^*$. Similar to (9.11) and (9.14), it follows from (9.23) that

$$\mathrm{E}\Big(\sup_{\theta\in\Theta_{\mathcal{M}}^V}|\widehat{m}_V(\theta)-m_V(\theta)|\Big)\le O(1)\frac{\sqrt{|\mathcal{M}|R_{\mathcal{M}}^V}}{\sqrt{n}},$$

and hence

$$\frac{1}{|\mathcal{M}|}\mathrm{E}\Big(\sup_{\theta\in\Theta_{\mathcal{M}}^V}|\widehat{m}_V(\theta)-m_V(\theta)|\Big)\le O(1)n^{-2/3}\log^{1/6}p. \tag{9.24}$$

Since $\kappa_n\gg n^{1/3}\log^{2/3}p$, we have $\kappa_n\gg n^{1/3}\log^{1/6}p$. For sufficiently large n, (9.22) is bounded by

$$\begin{aligned}&\sum_{\mathcal{M}\in\Omega_+^*}\Pr\Big(\sup_{\theta\in\Theta_{\mathcal{M}}^V}\frac{n}{|\mathcal{M}|}|\widehat{m}_V(\theta)-m_V(\theta)|\\&\quad-\frac{3}{2}\mathrm{E}\sup_{\theta\in\Theta_{\mathcal{M}}^V}\frac{n}{|\mathcal{M}|}|\widehat{m}_V(\theta)-m_V(\theta)|\ge\frac{\kappa_n}{8(s_\beta+1)}\Big).\end{aligned} \tag{9.25}$$

It follows from (9.23) and Lemma H.4 that (9.25) is bounded by

$$\sum_{\mathcal{M}\in\Omega_+^*}\Big\{\exp\Big(-\frac{\bar{c}\kappa_n^2|\mathcal{M}|^2}{nR_{\mathcal{M}}^V}\Big)+3\exp\Big(-\frac{\bar{c}\kappa_n|\mathcal{M}|}{\log(n)}\Big)\Big\}, \tag{9.26}$$

for some constants $\bar{c}>0$.

Define $\Omega_s^*=\{\mathcal{M}\in\Omega_+^*:|\mathcal{M}|=s\}$, it is immediate to see that $\Omega_+^*\subseteq\bigcup_{s=1}^{s_n}\Omega_s^*$. Hence, (9.26) is bounded by

$$\sum_{s=1}^{s_n}|\Omega_s^*|\Big\{\exp\Big(-\frac{\bar{c}\kappa_n^2s^{5/3}}{n^{2/3}\log^{1/3}p}\Big)+3\exp\Big(-\frac{\bar{c}\kappa_ns}{\log(n)}\Big)\Big\}. \tag{9.27}$$

For each s, the number of elements in $|\Omega_s^*|$ is bounded by $O(p^s)$. By assumption, we have $\kappa_n\gg n^{1/3}\log^{2/3}p$ and hence $\kappa_n\gg\log(p)\log(n)$. This implies

$$\frac{\kappa_n^2s^{5/3}}{n^{2/3}\log^{1/3}p}\gg s\log(p)\text{and}\frac{\kappa_ns}{\log(n)}\gg s\log(p).$$

Hence, for sufficiently large n, (9.27) is bounded by

$$\begin{aligned}&|s_n|\min_{s\ge1}\Big\{\exp\Big(-\frac{\bar{c}\kappa_n^2s^{5/3}}{2n^{2/3}\log^{1/3}p}\Big)+3\exp\Big(-\frac{\bar{c}\kappa_ns}{2\log(n)}\Big)\Big\}\\&=O(n)\Big\{\exp\Big(-\frac{\bar{c}\kappa_n^2}{2n^{2/3}\log^{1/3}p}\Big)+3\exp\Big(-\frac{\bar{c}\kappa_n}{2\log(n)}\Big)\Big\}\\&\le\exp\Big(-\frac{\bar{c}\kappa_n^2}{3n^{2/3}\log^{1/3}p}\Big)+3\exp\Big(-\frac{\bar{c}\kappa_n}{3\log(n)}\Big)\\&\le\frac{1}{2}\exp\Big(-\frac{\bar{c}\kappa_n^2}{4n^{2/3}\log^{1/3}p}\Big)+\frac{1}{2}\exp\Big(-\frac{\bar{c}\kappa_n}{4\log(n)}\Big)\le\exp(-\bar{c}\log(p)),\end{aligned} \tag{9.28}$$

where the last inequality is due to $\kappa_n\gg n^{1/3}\log^{2/3}(p)$ and $\kappa_n\gg\log(p)\log(n)$. This proves (9.21). The proof is hence complete.

Acknowledgments. Chengchun Shi is Assistant Professor of Statistics at London School of Economics. Rui Song is Professor of Statistics at North Carolina State University. Wenbin Lu is Professor of Statistics at North Carolina State University. The bulk of the work was done when Chengchun Shi was a graduate student at North Carolina State University. Rui Song's research was partially supported by a NSF grant DMS 1555244 and a NIH grant P01 CA142538.

SUPPLEMENTARY MATERIAL

Supplement to "Concordance and value information criteria for optimal treatment decision" (DOI: 10.1214/19-AOS1908SUPP; .pdf). This supplement file includes some proofs and simulation results.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) 267–281. Akadémiai Kiadó, Budapest. MR0483125

ARCONES, M. A. (1995). A Bernstein-type inequality for U-statistics and U-processes. *Statist. Probab. Lett.* **22** 239–247. MR1323145 https://doi.org/10.1016/0167-7152(94)00072-G

CANDÈS, E. and TAO, T. (2007). Rejoinder: "The Dantzig selector: Statistical estimation when p is much larger than n" [Ann. Statist. **35** (2007), no. 6, 2313–2351; MR2382644]. *Ann. Statist.* **35** 2392–2404. MR2382651 https://doi.org/10.1214/009053607000000532

CHAKRABORTY, B., MURPHY, S. and STRECHER, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat. Methods Med. Res.* **19** 317–343. MR2757118 https://doi.org/10.1177/0962280209105013

CLÉMENÇON, S., LUGOSI, G. and VAYATIS, N. (2008). Ranking and empirical minimization of U-statistics. *Ann. Statist.* **36** 844–874. MR2396817 https://doi.org/10.1214/009052607000000910

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

FAN, J. and LV, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica* **20** 101–148. MR2640659

FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inf. Theory* **57** 5467–5484. MR2849368 https://doi.org/10.1109/TIT.2011.2158486

FAN, Y. and TANG, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 531–552. MR3065478 https://doi.org/10.1111/rssb.12001

FAN, C., LU, W., SONG, R. and ZHOU, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1565–1582. MR3731676 https://doi.org/10.1111/rssb.12216

GAI, Y., ZHU, L. and LIN, L. (2013). Model selection consistency of Dantzig selector. *Statist. Sinica* **23** 615–634. MR3086649

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

LI, H., REN, C. and LI, L. (2014). U-processes and preference learning. *Neural Comput.* **26** 2896–2924. MR3223194 https://doi.org/10.1162/NECO_a_00674

LIANG, S., LU, W., SONG, R. and WANG, L. (2017). Sparse concordance-assisted learning for optimal treatment decision. *J. Mach. Learn. Res.* **18** Paper No. 202, 26. MR3827090

LU, W., ZHANG, H. H. and ZENG, D. (2013). Variable selection for optimal treatment decision. *Stat. Methods Med. Res.* **22** 493–504. MR3190671 https://doi.org/10.1177/0962280211428383

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. MR3476615 https://doi.org/10.1214/15-AOS1384

MEBANE, W. R. JR., SEKHON, J. S. et al. (2011). Genetic optimization using derivatives: The rgenoud package for R. *J. Stat. Softw.* **42** 1–26.

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752 https://doi.org/10.1111/1467-9868.00389

NOLAN, D. and POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15** 780–799. MR0888439 https://doi.org/10.1214/aos/1176350374

QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351 https://doi.org/10.1214/10-AOS864

ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. MR1201705 https://doi.org/10.2307/2951780

SHERMAN, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22** 439–459. MR1272092 https://doi.org/10.1214/aos/1176325377

SHI, C., FAN, A., SONG, R. and LU, W. (2018). High-dimensional A-learning for optimal dynamic treatment regimes. *Ann. Statist.* **46** 925–957. MR3797992 https://doi.org/10.1214/17-AOS1570

SHI, C., SONG, R. and LU, W. (2021). Supplement to "Concordance and value information criteria for optimal treatment decision." https://doi.org/10.1214/19-AOS1908SUPP

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

WATKINS, C. J. C. H. and DAYAN, P. (1992). Q-learning. *Mach. Learn.* **8** 279–292.

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. MR3040007 https://doi.org/10.1111/j.1541-0420.2012.01763.x

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445 https://doi.org/10.1093/biomet/ast014

ZHANG, X., WU, Y., WANG, L. and LI, R. (2016). A consistent information criterion for support vector machines in diverging model spaces. *J. Mach. Learn. Res.* **17** Paper No. 16, 26. MR3491110 https://doi.org/10.1016/j.jocs.2016.07.001

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898 https://doi.org/10.1080/01621459.2012.695674

ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598. MR3367249 https://doi.org/10.1080/01621459.2014.937488