HHS Public Access

Author manuscript

Electron J Stat. Author manuscript; available in PMC 2017 August 04.

Published in final edited form as:

Electron J Stat. 2016 ; 10: 2894–2921. doi:10.1214/16-EJS1178.

Robust learning for optimal treatment decision with NP-dimensionality

Chengchun Shi, **Rui Song**, and **Wenbin Lu***

Department of Statistics, North Carolina State University, Raleigh NC, U.S.A

Abstract

In order to identify important variables that are involved in making optimal treatment decision, Lu, Zhang and Zeng (2013) proposed a penalized least squared regression framework for a fixed number of predictors, which is robust against the misspecification of the conditional mean model. Two problems arise: (i) in a world of explosively big data, effective methods are needed to handle ultra-high dimensional data set, for example, with the dimension of predictors is of the non-polynomial (NP) order of the sample size; (ii) both the propensity score and conditional mean models need to be estimated from data under NP dimensionality.

In this paper, we propose a robust procedure for estimating the optimal treatment regime under NP dimensionality. In both steps, penalized regressions are employed with the non-concave penalty function, where the conditional mean model of the response given predictors may be misspecified. The asymptotic properties, such as weak oracle properties, selection consistency and oracle distributions, of the proposed estimators are investigated. In addition, we study the limiting distribution of the estimated value function for the obtained optimal treatment regime. The empirical performance of the proposed estimation method is evaluated by simulations and an application to a depression dataset from the STAR*D study.

Keywords and phrases

Non-concave penalized likelihood; optimal treatment strategy; oracle property; variable selection

1. Introduction

Personalized medicine, which has gained much attentions over the past few years, is a medical paradigm that emphasizes systematic use of individual patient information to optimize that patient's health care. In this paradigm, the primary interest lies in identifying the optimal treatment strategy that assigns the best treatment to a patient based on his/her observed covariates. Formally speaking, a treatment regime is a function that maps the sample space of patient's covariates to the treatments.

*The research of Chengchun Shi and Rui Song is supported in part by Grant NSF-DMS 1309465, 1555244 and Grant NCI P01 CA142538. The research of Wenbin Lu is supported in part by Grant NCI P01 CA142538.

Supplementary Material

Supplement to "Robust Learning for Optimal Treatment Decision with NP-Dimensionality"
(doi: 10.1214/16-EJS1178SUPP; .pdf).

There is a growing literature for estimating the optimal individualized treatment regimes. Existing literature can be casted into as model based methods and direct search methods. Popular model based methods include Q-learning (Watkins and Dayan, 1992; Chakraborty, Murphy and Strecher, 2010) and A-learning (Robins, Hernan and Brumback, 2000; Murphy, 2003), where Q-learning models the conditional mean of the response given predictors and treatment while A-learning models the interaction between treatment and predictors, better known as the contrast function. The advantage of A-learning is robustness against the misspecification of the baseline mean function, provided that the propensity score model is correctly specified. Recently, Zhang et al. (2012) proposed inverse propensity score weighted (IPSW) and augmented-IPSW estimators to directly maximize the mean potential outcome under a given treatment regime, i.e. the value function. Moreover, Zhao et al. (2012) recast the estimation of the value function from a classification perspective and use machine learning tools, to directly search for the optimal treatment regimes.

The rapid advances and breakthrough in technology and communication systems make it possible to gather an extraordinary large number of prognostic factors for each individual. For example, in the Sequenced Treatment Alternative to Relieve Depression (STAR*D) study, over 305 covariates are collected from each patient. With such data gathered at hand, it is of significant importance to organize and integrate information that is relevant to make optimal individualized treatment decisions, which makes variable selection as an emerging need for implementing personalized medicine. There have been extensive developments of variable selection methods for prediction, for example, LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001), MCP (Zhang, 2010) and many others in the context of penalized regression. Their associated inferential properties have been studied when the number of predictors is fixed, diverging with the sample size and of the non-polynomial order of the sample size.

In contrast to the large amount of work on developing variable selection methods for prediction, the variable selection tools for deriving optimal individualized treatment regimes have been less studied, especially when the number of predictors is much larger than the sample size. Among those available, Gunter, Zhu and Murphy (2011) proposed variable ranking methods for the marginal qualitative interaction of predictors with treatment. Fan, Lu and Song (2015) developed a sequential advantage selection method that extends the marginal ranking methods by selecting important variables with qualitative interaction in a sequential fashion. However, no theoretical justifications are provided for these methods. Qian and Murphy (2011) proposed to estimate the conditional mean response using a L_1-penalized regression and studied the error bound of the value function for the estimated treatment regime. However, the associated variable selection properties, such as selection consistency, convergence rate and oracle distribution, are not studied. Lu, Zhang and Zeng (2013) introduced a new penalized least squared regression framework, which is robust against the misspecification of the conditional mean function. However, they only studied the case when the number of covariates is fixed and the propensity score model is known as in randomized clinical trials. Song et al. (2015) proposed penalized outcome weighted learning for the case with the fixed number of predictors.

In this paper, we study the penalized least squared regression framework considered in Lu, Zhang and Zeng (2013) when the number of predictors is of the non-polynomial (NP) order of the sample size. In addition, we consider a more general situation where the propensity score model may depend on predictors and needs to be estimated from data, as common in observational studies. A two-step estimation procedure is developed. In the first step, penalized regression models are fitted for the propensity score and the conditional mean of the response given predictors. In the second step, the optimal treatment regime is estimated using the penalized least squared regression with the estimated propensity score and conditional mean models obtained in the first step. There are several challenges in both numerical implementation and derivation of theoretical properties, such as weak oracle and oracle properties, for the proposed estimation procedure. First, since the posited model for the conditional mean of the response given predictors may be misspecified, the associated estimation and variable selection properties under model misspecification with NP dimensionality is not standard. Second, it is unknown how the asymptotic properties of the estimators for the optimal treatment regime obtained in the second step will depend on the estimated propensity score and conditional mean models obtained in the first step under NP dimensionality. To our knowledge, these two challenges have never been studied in the literature. Moreover, we estimate the value function of the estimated optimal regime and study the estimator's theoretical properties.

The remainder of the paper is organized as follows. The proposed method for estimating the optimal treatment regime is introduced in Section 2. Simulation results are presented in Section 3. An application to a dataset from the STAR*D study is illustrated in Section 4. Section 5 and 6 demonstrate the weak oracle and oracle properties of the resulting estimators, respectively. The estimator for the value function of the estimated optimal treatment regime is given in Section 7, followed by a Conclusion Section. All the technical proofs are given in the Appendix.

2. Method

Let Y denote the response, $A \in \mathscr{A}$ denote the treatment received, where $\mathscr{A}$ is the set of available treatment options, and X denote the baseline covariates including constant one. For demonstration purpose, we focus on a binary treatment regime, i.e., $\mathscr{A} = \{0, 1\}$, with 0 for the standard treatment and 1 for the new treatment. We consider the following semiparametric model:

$$Y = h_0(X) + A(\beta_0^T X) + e, \quad (2.1)$$

where $h_0(X)$ is the unspecified baseline function, β_0 is the p-dimensional regression coefficients and e is an independent error with mean 0 and variance σ^2. Under the assumptions of stable unit treatment value (SUTVA) and no unmeasured confounders (Rubin, 1974), it can be shown that the optimal treatment regime $d^{\text{opt}}(x)$ for patients with baseline covariates $X = x$ takes the form

$$I(\mathrm{E}(Y|X=x, A=1) - \mathrm{E}(Y|X=x, A=0)>0)=I(\beta_0^T x>0),$$

where $I(\cdot)$ is the indicator function.

Our primary interest is in estimating the regression coefficients β_0 defining the optimal treatment regime. Let $\pi(x) = P(A = 1|X = x)$ be the propensity score. We assume a logistic regression model for $\pi(x)$:

$$\pi(x, \alpha_0)=\exp(x^T \alpha_0)/[1+\exp(x^T \alpha_0)], \quad (2.2)$$

with p-dimensional parameter α_0. Here, we allow the propensity score to depend on covariates, which is common in observational studies and the parameters α_0 can be estimated from the data. For randomized clinical trials, $\pi(x, \alpha_0)$ is a constant. We assume the majority of elements in β_0 and α_0 are zero and refer to the support supp(β_0), supp(α_0) as the true underlying sparse model of the indices.

Consider a study with n subjects. Assume $X = (x_1, \ldots, x_n)^T$ is deterministic. The observed data consist of $\{(Y_i, A_i, x_i) : i = 1, \cdots, n\}$. Define $\mu(x) = h_0(x) + \pi(x, \alpha_0)x^T\beta_0$, the conditional mean of the response given covariates $X = x$. We propose the following two-step estimation procedure to estimate the optimal treatment regime. In the first step, we posit a model $\Phi(x, \theta)$ for the conditional mean function $\mu(x)$, and consider the penalized estimation for the propensity score and conditional mean models as follows.

Define

$$\hat{\alpha}= \arg\min_{\alpha} \frac{1}{n}\sum_{i=1}^{n}[\log\{1+\exp(x_i^T \alpha)\} - A_i x_i^T \alpha]+\sum_{j=1}^{p}\lambda_{1n}\rho_1(|\alpha^j|, \lambda_{1n}), \quad (2.3)$$

and

$$\hat{\theta}= \arg\min_{\theta} \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \Phi(x_i, \theta)\}^2+\sum_{j=1}^{q}\lambda_{2n}\rho_2(|\theta^j|, \lambda_{2n}), \quad (2.4)$$

where α^j and θ^j refer to the jth element in α and θ, q is the dimension of θ, and ρ_1 and ρ_2 are folded concave penalty functions with the tuning parameters λ_{1n} and λ_{2n}, respectively. We allow p, q to be of NP order of n and assume $\log p = O(n^{1-2d_\beta})$ and $\log q = O(n^{1-2d_\theta})$ for some d_β and $d_\theta \in (0, \frac{1}{2})$, respectively. The posited model $\Phi(x, \theta)$ may be misspecified.

Define $\hat{\Phi}_i = \Phi(x_i, \hat{\theta})$ and $\hat{\pi}_i = \pi(x_i, \hat{\alpha})$. In the second step, we consider the following penalized least square estimation:

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{\Phi}_i - (A_i - \hat{\pi}_i)\beta^T x_i\}^2 + \sum_{j=1}^{p}\lambda_{3n}\rho_3(|\beta^j|, \lambda_{3n}), \tag{2.5}$$

where ρ_3 is a folded-concave penalty function with the tuning parameter λ_{3n}. Here the folded-concave penalty functions ρ_1, ρ_2 and ρ_3 are assumed to satisfy the following condition:

Condition 2.1. $\rho(t, \lambda)$ is increasing and concave in $t \in [0, \infty)$, and has a continuous derivative $\rho'(t, \lambda)$ with $\rho'(0+, \lambda) > 0$. In addition, $\rho'(t, \lambda)$ is increasing in $\lambda \in [0, \infty)$ and $\rho'(0+, \lambda)$ is independent of λ.

Popular penalties, such as LASSO, SCAD and MCP, satisfy Condition (2.1). In our implementation, we use SCAD penalty. Here, we adopt a two-step estimation procedure due to its computational simplicity. Alternatively, we can jointly estimate the parameters θ in the conditional mean model and β in the contrast function in a single penalized regression. However, this joint approach will require more computational effort since the tuning parameters for θ and β need to be selected simultaneously. In contrast, our two-step method only requires a single tuning parameter at each step and thus can be easily implemented by existing softwares, for example, the `R` package `ncvreg`.

3. Numerical studies

In this section, we evaluate the numerical performance of the proposed estimators in various settings. We generated the propensity score from the logistic regression model (2.2), with only one important covariate with the coefficient of 1.5. We chose three forms for the baseline function $h_0(x)$, including a simple linear form, a quadratic form and a complex non-linear form,

- Model I: $Y = \mathbf{1} + \theta_0^T X + A(\beta_0^T \tilde{X}) + \varepsilon$,
- Model II: $Y = \mathbf{1} + 0.5(1 + \theta_0^T X)^2 + A(\beta_0^T \tilde{X}) + \varepsilon$,
- Model III: $Y = \mathbf{1} + 1.5\sin(\pi\theta_0^T X) + X_1^2 + A(\beta_0^T \tilde{X}) + \varepsilon$,

where X is a p-dimensional vector of covariates and $\tilde{X} = (1, X^T)^T$. We set $p = 1000$. Covariates were generated independently from two distributions: standard normal or s shifted exponential distribution with mean 0 and variance 1.

For each model, the first two covariates were chosen as important variables both in the baseline mean function and the contrast function with $\theta_0 = (-2, -1, 0, \ldots, 0)^T$ and $\beta_0 = (0, -1.5, 1.5, 0, \ldots, 0)^T$. We considered two different sample sizes, $n = 300$ and $n = 500$. For each scenario, we conducted 1000 replications. In our method, we fitted a linear model for

$\Phi(X, \theta)$ and used the SCAD penalty for variable selection. The tuning parameter was chosen using 10-fold cross-validation.

To evaluate the performance of the proposed estimator, we also compared our method with the penalized Q-learning using the SCAD penalty. Specifically, we fitted a linear model with baseline covariate effects and treatment-covariates interaction. Note that it is correctly specified under model I but misspecified under models II and III.

Let $\hat{\beta}$ and $\tilde{\beta}$ denote our estimator and the penalized Q-learning estimator, respectively. We report the L_2 loss of $\hat{\beta}$ and $\tilde{\beta}$, the number of missed important variables (denoted as FN), the number of selected noisy variables (denoted as FP) and the average percentage of making correct decisions (denoted as PCD), which is defined as $1 - \sum_{i=1}^{n} |d(x_i) - I(\beta_0^T x_i > 0)|/n$ for treatment rules $\hat{d}(x) = I(x^T\hat{\beta} > 0)$ and $\tilde{d}(x) = I(x^T\tilde{\beta} > 0)$. In addition, we estimated $E\{Y^*(\hat{d})\}$, $E\{Y^*(\tilde{d})\}$ and $E\{Y^*(d^{opt})\}$, the value functions of the estimated optimal treatment regimes by our method and the penalized Q-learning method, and of the true optimal regime, respectively, using Monte Carlo simulations. For a given treatment rule $d(x)$, we compute $E\{Y^*(d)\}$ by averaging the responses for 20000 subjects generated from the true model with A being determined by $d(x)$. We report the averages of mean responses over 1000 replications as well as their standard deviations.

Table 1 summarizes the results. The penalized Q-learning method performs pretty well under Model I where the fitted linear model is correctly specified and is more efficient than the proposed method as expected. For example, when covariates are i.i.d normal and n = 300, the PCD is around 99.3% and the estimated value function is very close to the true optimal, $E\{Y^*(d^{opt})\}$. In contrast, under this setting, the PCD of our proposed method is 97.5%, and the estimated value function is slightly lower.

However, for Models II and III, the penalized Q-learning method could lead to substantial bias and works much worse than the proposed method. Taking the second model as an example, when covariates are normal and n = 300, $\|\tilde{\beta} - \beta_0\|_2$ = 4.86, approximately third times as large as $\|\hat{\beta} - \beta_0\|_2$. The PCD of the estimated treatment regime obtained by the penalized Q-learning is 55.0%, only a little better than a random guess. In contrast, for this scenario, the PCD of our proposed method is 73.4%. Moreover, when sample size increases, the performance of the penalized Q-learning method is even worse. This is due to the misspecification of the baseline mean function. For our method, there's a big increase in the PCD as the sample size gets larger. The L_2 loss and average number of missed important variables are also greatly reduced. This demonstrates the robustness of the proposed method to the misspecification of the baseline mean function.

4. Real data example

We applied our method to the data set from the STAR*D study for 4041 patients with nonpsychotic major depressive disorder (MDD). The aim of the study was to determine the effectiveness of different treatments for those people who have not responded to initial medication treatment. At Level 1, all patients received citalopram (CIT), an selective serotonin reuptake inhibit (SSRI) medication. After 8-12 weeks, three more levels of

treatments were offered to participants whose previous treatment didn't give an acceptable response. Available treatments at Level 2 included sertraline (SER), venlafaxine (VEN), bupropion (BUP) and cognitive therapy (CT) and augmenting CIT which combines CIT with one more treatment. At Level 2A, switch options to VEN or BUP treatment were provided for patients receiving CT but without sufficient improvement. Four treatments were available at Level 3 for participants without anticipated response, including medication switch to mirtazapine (MIRT), nortriptyline (NTP), and medication augmentation with either lithium (Li) and thyroid hormone (THY). Finally, treatment with tranylcypromine (TCP) or a combination of mirtazapine and venlafaxine (MIRT+VEN) were provided at Level 4 for those without sufficient improvement at Level 3.

Here, we only focused on a subset of data for those patients receiving treatment BUP (coded as 1) or SER (0) at Level 2. The outcome of interest was the 16-item Quick Inventory of Depressive Symptomatology-Clinician-Ratings (QIDS-C16), which indicated the severity of patient's depressive symptom. The maximum vale of QIDS-C16 was 24 and its distribution was highly skewed. Hence, we considered the transformation $Y_i = \log(25 - \text{QIDS-C16})$ as our response. Larger value of Y_i indicates better response. All baseline variables at Level 1 and intermediate outcomes at Level 2 were included as covariates in our study, yielding 305 covariates in total for each patient. There are 383 patients receiving treatment BUP or SER at Level 2, however, only 319 patients have complete records of all 305 covariates and the response. Among them, 153 were treated with BUP and 166 with SER. Our proposed method selected 14 variables that are important for treatment decision. We reestimate the coefficients of these variable by solving A-learning estimating equations (Robins, 2004) and obtained the resulting estimated optimal treatment regime.

To examine the performance of the estimated optimal treatment regime, we compared it with the fixed treatment regimes by assigning all patients to either BUP or SER, in terms of the estimated value functions obtained by the IPSW method (Zhang et al., 2012). The results for the estimated value functions were given in Table 2. In addition, we reported the 95% confidence intervals for the difference between the estimated values of the obtained optimal regime and the fixed regime based on 500 bootstrap samples. Our estimated optimal treatment regime gave larger estimated values than those of the fixed regimes, BUP and SER. The difference is significant when comparing to the BUP treatment at 5% level, but is less significant when comparing to the SER treatment. One reason is that our estimated optimal regime assigns the majority of patients (about two-thirds) to the SER treatment. Please refer to Table 3 for the numbers of patients receiving BUP or SER according to the estimated optimal regime.

In addition, as suggested by a referee, we examined the effects of missing data. Specifically, we deleted one patient whose response was missing, and imputed all the missing values in covariates using the `R` package `missForrest` available in `CRAN`. This package uses a random forest trained based on the observe entries in the design matrix to predict those missing values. The optimal treatment regime obtained based on the imputed data was similar to the one based on the complete-case analysis as shown above. It selected 14 variables among which 11 variables were also included in the estimated optimal treatment regime without imputation. In addition, the bootstrap results suggested that the estimated

value of the estimated optimal treatment regime is significantly larger than those of the fixed treatment regimes, under 0.05 significance level. Since results are similar, we omitted them here.

5. Non-asymptotic weak oracle properties

In this section we show that the proposed estimator enjoys the weak oracle property, that is, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$ defined in (2.3)-(2.5) are sign consistent with probability tending to 1, and are consistent with respect to the L_∞ norm. Weak oracle properties of $\hat{\theta}$ are established in the sense that it converges to some least false parameter $\theta^\star$ when the main effect model is misspecified.

Theorem 5.1 provides the main results. Some regularity conditions are discussed in subsections 5.1 and 5.2. A major technical challenge in deriving weak oracle properties of $\hat{\beta}$ is to analyze the deviation in (5.18), for which we develop a general empirical process result in the supplementary article (Shi et al., 2016). This result is important in its own right and can be used in analyzing many other high-dimensional semiparametric models where the index parameter of an empirical process is a plug-in estimator. The following notation is introduced to simplify our presentation.

Let **1** denote a vector of ones, E denote the identity matrix, O denote the zero matrix consisting of all zeros. For any matrix Ψ, let $P(\Psi)$ denote the projection matrix $\Psi(\Psi^T\Psi)^{-1}\Psi^T$. Ψ_M the submatrix of Ψ formed by columns in the subset M. For any vector a, b, let "$\bigcirc$" denote the Hadamard product: $a \bigcirc b = (a^1b^1, \ldots, a^nb^n)^T$, $|a| = (|a^1|, \ldots, |a^n|)^T$, diag($a$) as the diagonal matrix with elements of vector a and a_M the subvector of a formed by elements in M. The jth element in a is denoted as a^j. Let $\|\cdot\|_p$ be the L_p norm of vectors or matrices. Let $\|Y\|_{\psi m}$ be the Orlicz norm of a random variable Y,

$$\inf_u \left\{ u>0 : \mathrm{E}\exp\left(\frac{|Y|}{u}\right)^m \le 2 \right\},$$

for any $m \ge 1$.

Let $M_\alpha = \mathrm{supp}(\alpha_0)$, $M_\beta = \mathrm{supp}(\beta_0)$, $M_{\theta^\star} = \mathrm{supp}(\theta^\star)$, and M_α^c, M_β^c, $M_{\theta^\star}^c$ be their complements. Assume each x^j, is standardized such that $\|x^j\|_2 = \sqrt{n}$. Let $\Phi(\theta) = [\Phi(x_1, \theta), \ldots, \Phi(x_n, \theta)]^T$, $\phi(\theta) = [\phi^1(\theta), \ldots, \phi^q(\theta)]$ denote its Jacobian matrix. The derivatives are taken componentwise, i.e.,

$$\phi^l(\theta) = (\phi^l(x_1, \theta), \ldots, \phi^l(x_n, \theta)),$$

for all $l = 1, \ldots, q$. We denote $\Phi(\theta^\star)$ and $\phi(\theta^\star)$ as Φ and ϕ when there's no confusion. We use a short-hand $\hat{\Phi}$, $\hat{\phi}$ for $\Phi(\hat{\theta})$, $\phi(\hat{\theta})$.

5.1. The misspecified function

We first define the least false parameter under the misspecification due to the posited mean function $\Phi(x, \theta)$. For regression models with fixed number of predictors, the definition of the least false parameter under model misspecification has been widely studied in the literature (e.g, White, 1982; Li and Duan, 1989). However, for regression models with NP dimensionality, its definition is more tricky. Here, we define our least false parameter as follows.

For each $\theta \in \mathbb{R}^q$, let $d_{n\theta} = 1/2 \min_j\{|\theta^j| : \theta^j \neq 0\}$, M_θ be the support of θ, $\mu = (\mu(x_1), \ldots, \mu(x_n))^T$ and

$$H_\theta = \{\delta \in {}^d : \delta_{M_\theta^c} = 0, \|\delta_{M_\theta} - \theta_{M_\theta}\|_\infty \leq d_{n\theta}\}.$$

Consider the set

$$\Theta = \left\{\theta : \sup_{\delta \in H_\theta} \|\phi_{M_\theta^c}(\delta)^T[E - P\{\phi_{M_\theta}(\delta)\}]\{\mu - \Phi(\theta)\}\|_\infty \leq C_0 n^{1-d_\theta}\sqrt{\log n}, |M_\theta| \leq s_0\right\},$$

for some constant C_0, and $s_0 \ll n$. We assume the set Θ to be nonempty and define the least false parameter as

$$\theta^\star = \arg\min_{\theta \in \Theta} \sup_{\delta \in H_\theta} \|\{\phi_{M_\theta}(\delta)^T \phi_{M_\theta}(\delta)\}^{-1} \phi_{M_\theta}^T(\delta)\{\mu - \Phi(\theta)\}\|_\infty.$$

In addition, we assume

$$\sup_{\delta \in H_{\theta^\star}} \|\{\phi_{M_{\theta^\star}}(\delta)^T \phi_{M_{\theta^\star}}(\delta)\}^{-1} \phi_{M_{\theta^\star}}^T(\delta)(\mu - \Phi)\|_\infty = O(n^{-\gamma_0}\log n), \quad (5.1)$$

for some $\gamma_0 \geq 0$. By its definition, $\theta^\star$ satisfies

$$\sup_{\delta \in H_{\theta^\star}} \|\phi_{M_{\theta^\star}^c}(\delta)^T[E - P\{\phi_{M_{\theta^\star}}(\delta)\}](\mu - \Phi)\|_\infty = O(n^{1-d_\theta}\sqrt{\log n}), \quad (5.2)$$

and $|M_{\theta^\star}| \leq s_0$.

Remark 5.1. Conditions (5.1) and (5.2) are key assumptions determining the degree of model misspecification. Condition (5.1) requires that the posited working model Φ can provide a good approximation for μ. In that case, the residual $\mu - \Phi$ will be orthogonal to the jacobian matrix $\phi_{M_{\theta^\star}}$ and the left-hand side of (5.1) will be small. In general, our

assumptions are weaker than the weak sparsity assumption imposed for Lasso (Bunea, Tsybakov and Wegkamp, 2007), which assumes the L_2 approximation error $\|\mu - \Phi\|_2$ converges to 0 at some certain rate.

Condition 5.1. We assume the following conditions:

$$\sup_{\delta \in H_{\theta^\star}} \left\| \{\phi_{M_{\theta^\star}}(\delta)^T \phi_{M_{\theta^\star}}(\delta)\}^{-1} \right\|_\infty = O\left(\frac{b_{\theta^\star}}{n}\right), \tag{5.3}$$

$$\sup_{\delta \in H_{\theta^\star}} \left\| \phi_{M^c_{\theta^\star}}(\delta)^T \phi_{M_{\theta^\star}}(\delta) \{\phi_{M_{\theta^\star}}(\delta)^T \phi_{M_{\theta^\star}}(\delta)\}^{-1} \right\|_\infty \le \min\left\{ C\frac{\rho_3'(0+)}{\rho_3'(d_{n\theta})}, O(n^{a_3}) \right\}, \tag{5.4}$$

$$\max_{l=1}^{q} \|\phi^l \circ (\mathbf{1} + |X\beta_0|)\|_2 = O(\sqrt{n}), \tag{5.5}$$

$$\max_{l=1}^{q} \sum_{k \in M_{\theta^\star}} \sup_{\delta \in H_{\theta^\star}} \left\| \frac{\partial \phi^l(\delta)}{\partial \theta^k} \circ (\mathbf{1} + |X\beta_0|) \right\|_2 = O\left(\frac{n^{\frac{1}{2}+\gamma_{\theta^\star}}}{\sqrt{s_{\theta^\star}} \log n}\right), \tag{5.6}$$

$$\sup_{\delta_1 \in H_{\theta^\star}} \sup_{\delta_2 \in H_{\theta^\star}} \max_{l=1}^{q} \lambda_{\max}\left(\frac{\partial (|\phi^l(\delta_1)|)^T \phi_{M_{\theta^\star}}(\delta_2)}{\partial \theta_{M_{\theta^\star}}} \right) = O(n), \tag{5.7}$$

for some constants $0 \le a_3 \le 1/2$, $0 \le \gamma_{\theta^\star} \le \gamma_0$, $s_{\theta^\star} = |M_{\theta^\star}|$. If the response is unbounded, we require

$$\max_{l=1}^{q} (\|\phi^l\|_\infty) = o(n^{d_\theta} / \sqrt{\log n}), \tag{5.8}$$

and the right-hand side of (5.6) shall be modified to $O(n^{\frac{1}{2}+\gamma_{\theta^\star}} / \sqrt{s_{\theta^\star}} \log^2 n)$.

Remark 5.2. Conditions (5.6) and (5.7) put constraints on the derivatives of ϕ, requiring the misspecified function to be smooth. The right-hand side order in (5.6) is not too restrictive when $n^{\gamma_{\theta^\star}} \gg s_{\theta^\star} \log n$.

Two common examples of the main-effect function Φ are provided below to examine the validity of Condition 5.1.

Example 1. Set $\Phi = 0$. Then, no model is needed for Φ. It is easy to check that Condition 5.1 is satisfied.

Example 2. When a linear model is specified, i.e., $\Phi(x, \theta) = x^T\theta$, conditions (5.6) and (5.7) are automatically satisfied since the second-order derivative of Φ vanishes. In this example, $\theta^\star$ takes the form

$$\theta^\star_{M_{\theta^\star}} = (X^T_{M_{\theta^\star}} X_{M_{\theta^\star}})^{-1} X^T_{M_{\theta^\star}} \mu,$$

and $\theta^\star_{M^c_{\theta^\star}} = 0$. Note that $\theta^\star_{M_{\theta^\star}}$ is the regression coefficients between $X_{M_{\theta^\star}}$ and μ. Condition (5.1) holds automatically since

$$(X^T_{M_{\theta^\star}} X_{M_{\theta^\star}})^{-1} X_{M_{\theta^\star}} (\mu - X\theta^\star) = 0 >$$

Condition (5.2) becomes

$$\| X^T_{M^c_{\theta^\star}} \{ I - P(X_{M_{\theta^\star}}) \} \mu \|_\infty = O(n^{1-d_\theta} \sqrt{\log n}). \quad (5.9)$$

Each element in the left-hand side vector in (5.9) can be viewed as the inner product of the residuals obtained by fitting $X_{M_{\theta^\star}}$ on each noise variable in $X_{M^c_{\theta^\star}}$ and those fitted by regressing $X_{M_{\theta^\star}}$ on μ. When μ depends only on $X_{M_{\theta^\star}}$, (5.9) holds for Gaussian linear model.

5.2. The covariates

Condition 5.2. Assume that

$$\sup_{\delta \in H_{\theta^\star}} \| B^{-1}_{n\beta} X^T_{M_\beta} W(\delta) \Delta X_{M_\alpha} B^{-1}_{n\alpha} \|_\infty = O(\frac{b_{\alpha\beta}}{n}), \quad (5.10)$$

$$\sup_{\delta \in H_{\theta^\star}} \| X^T_{M^c_\beta} W_\beta W(\delta) X_{M_\alpha} B^{-1}_{n\alpha} \|_\infty = \min \left\{ o\left(\frac{\lambda_{2n} \rho_2'(0+)}{\lambda_{1n} \rho_1'(d_{n\beta})} \right), O(n^{a_2}) \right\}, \quad (5.11)$$

$$\max_{j=1}^{p} \| W(\theta^\star) x^j \|_2 = O(\sqrt{n}), \quad (5.12)$$

$$\max_{j=1}^{p} \sum_{k\in M_\alpha} \|x^k \circ x^j \circ (X\beta_0)\|_2 = O(\frac{n^{1/2+\gamma_\alpha}}{\log n}), \tag{5.13}$$

$$\max_{j=1}^{p} \sum_{k\in M_\beta} \|x^j \circ x^k\|_2 = O(\frac{n^{1/2+\gamma_\beta}}{\log n}), \tag{5.14}$$

$$\max_{j=1}^{p} \sum_{l\in M_{\theta^\star}} \sup_{\delta\in H_{\theta^\star}} \|x^j \circ \phi^l(\delta)\|_2 = O(\frac{n^{1/2+\gamma_{\theta^\star}}}{\sqrt{s_{\theta^\star}}\log^3 n}), \tag{5.15}$$

$$\sup_{\delta\in H_{\theta^\star}} \max_{j=1}^{p} \lambda_{\max}[X_{M_\alpha}^T \operatorname{diag}(|W(\delta)x^j|)X_{M_\alpha}] = O(n), \tag{5.16}$$

$$\max_{j=1}^{p} \lambda_{\max}[X_{M_\alpha}^T \operatorname{diag}|x^j \circ (X\beta_0)|X_{M_\alpha}] = O(n), \tag{5.17}$$

for some constants $0 \le \gamma_\alpha, \gamma_\beta, a_2 \le 1/2$, where

$$W(\delta)=\operatorname{diag}[\mu - \Phi(\delta)], \quad B_{n\alpha}=X_{M_\alpha}^T \Delta X_{M_\alpha}, \quad B_{n\beta}=X_{M_\beta}^T \Delta X_{M_\beta},$$
$$W_\beta=\Delta - \Delta^{\frac{1}{2}} P(\Delta^{\frac{1}{2}} X_{M_\beta})\Delta^{\frac{1}{2}}, \quad \Delta=\operatorname{diag}(\pi(x_1), \ldots, \pi(x_n)).$$

The sequence $b_{\alpha\beta}$ in (5.10) shall satisfy

$$b_{\alpha\beta} = \min\left\{o(n^{\frac{1}{2}-\gamma_\beta}\sqrt{\log n}), o(n^{2\gamma_\alpha-\gamma_\beta}/s_\alpha \log n)\right\}.$$

Remark 5.3. Conditions (5.10) and (5.11) control the impact of the deviation of the estimated propensity score from its true value on $\hat{\beta}$, thus are not needed when the propensity scores are known. By the definition of $W(\delta)$, magnitudes of the left-hand side in these two conditions depend on how accurate Φ models μ. The sequence $b_{\alpha\beta}$ in (5.10) can converge to 0 when X_{M_β} and X_{M_α} are weakly correlated. Each element in the left-hand side of (5.11) is the multiple regression coefficient of the corresponding variable in $X_{M_\beta^c}$ on $W(\delta)X_{M_\alpha}$, using weighted least squares with weights $\pi \circ (1 - \pi)$, after adjusted by X_{M_β}, which characterize their weak dependence given X_{M_β}. These two conditions are generally weaker than those imposed by Fan and Lv (2011) (Condition 2), and are therefore more likely to hold.

Remark 5.4. The right-hand side in (5.15) can be relaxed to $O(n^{1/2+\gamma_{\theta^\star}}/\log n)$ when using the linear model. The additional term $\sqrt{s_{\theta^\star}}$ is due to the penalty on the complexity of the main effect model. This condition typically controls the deviation

$$\|Z^T\{\Phi - \Phi(\hat{\theta})\}\|_\infty = O_p(\sqrt{\log p \log n}), \quad (5.18)$$

where $Z = \text{diag}(A - \pi)X$. A common approach to bound the deviation is to utilize the classical Bernstein's inequality. However this approach does not work here, because the indexing parameter in the process $\Phi(\cdot)$ in (5.18) is an estimator. To handle this challenge, we bound the left-hand side in (5.18) by

$$\sup_{\delta_1,\delta_2 \in H_{\theta^\star}} \|Z^T\{\Phi(\delta_1) - \Phi(\delta_2)\}\|_\infty.$$

A general theory that covers the above result is provided in Proposition C.1 in the supplementary article.

Remark 5.5. Conditions (5.16) and (5.17) aim to control the L_∞ norm of the quadratic term of the Taylor series as a function of $\hat{\alpha}$, expanded at α_0. Similar to (5.10) and (5.11), the two conditions are not needed when α_0 is known to us.

5.3. Weak oracle properties

Theorem 5.1 (Weak oracle property). Assume that conditions B.1 and B.3 in the supplementary Appendix and conditions 5.1, 5.2 hold, and $\max_i \|e_i\|_{\psi 1} < \infty$, where e_i is the residual for the ith patient in (2.1). Then there exist local minimizers $\hat{\alpha}$, $\hat{\theta}$ and $\hat{\beta}$ of the loss functions (2.3), (2.4), and (2.5) respectively, such that with probability at least $1 - \bar{c}/(n + p + q)$:

a. $\hat{\alpha}_{M_\alpha^c} = 0, \hat{\beta}_{M_\beta^c} = 0, \hat{\theta}_{M_{\theta^\star}^c} = 0,$

b. $\|\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}\|_\infty = O(n^{-\gamma_\alpha}\log n), \|\hat{\beta}_{M_\beta} - \beta_{0M_\beta}\|_\infty = O(n^{-\gamma_\beta}\log n),$
$\|\hat{\theta}_{M_{\theta^\star}} - \theta^\star_{M_{\theta^\star}}\|_\infty = O(n^{-\gamma_{\theta^\star}}\log n),$

for $\bar{c}$ is some positive constant.

Remark 5.6. In Theorem 5.1, part (a) corresponds to the sparse recovery while (b) gives the estimators' convergence rates. Weak oracle property of $\hat{\alpha}$ directly follows from Theorem 2 in Fan and Lv (2011). However, to prove this property of $\hat{\beta}$ requires further efforts, to account for the variability due to plugging in $\hat{\theta}$ and $\hat{\alpha}$. L_∞ convergence rate of $\hat{\alpha}_{M_\alpha}$ as well as the nonsparsity size s_α, play an important role in determining how fast $\hat{\beta}_{M_\beta}$ converges.

Remark 5.7. The convergence rate of $\hat{\theta}$ will not affect that of $\hat{\beta}$. This is because we require the posed propensity score model to be correct, the estimation of β is robust with respect to

the model misspecification of the main effect parameters θ. Simulation results also validate our theoretical findings.

6. Oracle properties

In this section we study the oracle property of the estimator $\hat{\beta}$. We assume that $\max(s_\alpha, s_\beta) \ll \sqrt{n}$ and $n^{\gamma_{\theta^\star}} \gg s_{\theta^\star} \log n$. The convergence rates of the estimators are established in Section 6.1 and their asymptotic distributions are provided in Section 6.2.

6.1. Rates of convergence

Condition 6.1. In addition to (5.16) and (5.17) in Condition 5.2, assume that the right-hand side of (5.15) is strengthened to $O(n^{\frac{1}{2}+\gamma_{\theta^\star}} / \sqrt{s_{\theta^\star} \log^3 n})$, and the following conditions hold,

$$\sup_{\delta \in H_{\theta^\star}} \| B_{n\beta}^{-1/2} X_{M_\beta}^T W(\delta) \Delta X_{M_\alpha} B_{n\alpha}^{-1/2} \|_\infty = O(1), \tag{6.1}$$

$$\sup_{\delta \in H_{\theta^\star}} \| X_{M_\beta^c}^T W_\beta W(\delta) X_{M_\alpha} \|_{2,\infty} = O(n), \tag{6.2}$$

$$\max_{j=1}^{p} \max_{k \in M_\beta} \| x^j \circ x^k \|_2 = O(\sqrt{n}), \tag{6.3}$$

$$\max_{j=1}^{p} \max_{k \in M_\alpha} \| x^j \circ x^k \circ (X\beta_0) \|_2 = O(\sqrt{n}), \tag{6.4}$$

$$\mathrm{tr}[X_{M_\beta}^T W(\theta^\star) \Delta W(\theta^\star) X_{M_\beta}] = O(s_\beta n). \tag{6.5}$$

Remark 6.1. Similar to the interpretation of (5.10) and (5.11), (6.1) corresponds to a notion of weak dependence between variables in X_{M_α} and X_{M_β} while (6.2) require $X_{M_\beta^c}$ and X_{M_α} are weakly correlated after adjusted by X_{M_β}. Besides, it can be verified that (6.3)-(6.5) hold with large probability when the baseline covariates possesses subgaussian tail.

Theorem 6.1. Assume that conditions 2.1, 5.1 and 6.1 and conditions B.2 and B.4 in the supplementary Appendix hold, and $\max_i \| e_i \|_{\psi 1} < \infty$. Constraints on $b_{\theta^\star}$, d_θ, $d_{n\theta}$ and λ_{3n} are same as in Theorem 5.1. Further assume $\max(l_1, l_2) < \frac{1}{2}$ with $s_\alpha = O(n^{l_1})$, $s_\beta = O(n^{l_2})$, and $n^{\gamma_{\theta^\star}} \gg s_{\theta^\star} \log n$. Then there exists a strict local minimizer $\hat{\beta}$ of the loss function (2.5), $\hat{\alpha}$ of

(2.3), such that $\hat{\alpha}_{M_\alpha^c}=0$, $\hat{\beta}_{M_\beta^c}=0$ with probability tending to 1 as $n\to\infty$, and $\|\hat{\alpha}_{M_\alpha}-\alpha_{0M_\alpha}\|_2=O(\sqrt{s_\alpha}n^{-1/2})$, $\|\hat{\beta}_{M_\beta}-\beta_{0M_\beta}\|_2=O(\sqrt{s_\alpha+s_\beta}n^{-1/2})$.

Remark 6.2. We note that when establishing the oracle property of $\hat{\beta}$, only the weak oracle property of $\hat{\theta}$ is required. This is due to the robustness of the A-learning methods and the fact that the propensity score is correctly specified.

Remark 6.3. Precision of $\hat{\beta}_{M_\beta}$ is affected by that of $\hat{\alpha}_{M_\alpha}$, since $\|\hat{\beta}_{M_\beta}-\beta_{0M_\beta}\|_2$ is at least the same order of magnitude as $\|\hat{\alpha}_{M_\alpha}-\alpha_{0M_\alpha}\|_2$. When the propensity score is known, convergence rate of $\hat{\beta}_{M_\beta}$ is improved to $\sqrt{s_\beta/n}$.

6.2. Asymptotic distributions

We define Σ_{12} and Σ_{22} as

$$\sum\nolimits_{12}=2B_{n\alpha}^{-1/2}X_{M_\alpha}^T\Delta W X_{M_\beta}B_{n\beta}^{-1/2},$$

$$\sum\nolimits_{22}=B_{n\alpha}^{-1/2}X_{M_\alpha}^T W\Delta^{1/2}(E-P_{\Delta^{\frac{1}{2}}X_{M_\alpha}})\Delta^{1/2}WX_{M_\beta}B_{n\beta}^{-1/2},$$

where W is a shorthand for $W(\theta^*)$.

To establish the weak convergence of the estimators, we introduce the following conditions.

Condition 6.2. Assume that

$$\lambda_{1n}\overline{\rho}_1(d_{n\alpha})=o(s_\alpha^{-1/2}n^{-1/2}),\quad \lambda_{2n}\overline{\rho}_2(d_{n\beta})=o(s_\beta^{-1/2}n^{-1/2}), \tag{6.6}$$

$$\sum_{i=1}^n(x_{M_\alpha i}^T B_{n\alpha}^{-1}x_{M_\alpha i})^{3/2}\to 0,\quad \sum_{i=1}^n(x_{M_\beta i}^T B_{n\beta}^{-1}x_{M_\beta i})^{3/2}\to 0, \tag{6.7}$$

$$\sum_{i=1}^n(x_{M_\beta i}^T B_{n\beta}^{-1}x_{M_\beta i})^{3/2}|\mu^i-\Phi^i|^3\to 0, \tag{6.8}$$

$$\lambda_{\max}\left(B_{n\beta}^{-1/2}X_{M_\beta}^T W^2 X_{M_\beta}B_{n\beta}^{-1/2}\right)=O(1), \tag{6.9}$$

$$\sup_{\delta \in H_{\theta^\star}} \| B_{n\beta}^{-1/2} X_{M_\beta}^T \mathrm{diag}[\Phi - \Phi(\delta)] \Delta X_{M_\alpha} B_{n\alpha}^{-1/2} \|_2 = o(1). \quad (6.10)$$

where $x_{M\alpha i}$ and $x_{M\beta i}$ stand for the ith row of the matrix X_{M_α} and X_{M_β} respectively.

Remark 6.4. Conditions (6.7) and (6.8) are the Lyapunov conditions which guarantee the normality of $\hat{\alpha}_{M_\alpha}$ and $\hat{\beta}_{M_\beta}$. Condition (6.9) puts constraints on the maximum eigenvalue of the variance-covariance matrix of $X_{M_\beta}^T \mathrm{diag}(A - \pi)(\mu\Phi)$ by requiring it to be finite. Condition (6.10) holds when $\Phi(\delta)$ converges to Φ uniformly in terms of L_∞ norm with δ in the region $H_{\theta^\star}$. When $\|\mu - \Phi\|_\infty$ is bounded, (6.8) and (6.9) are simultaneously satisfied.

Theorem 6.2 (Oracle property). Under conditions in Theorem 6.1 and Condition 6.2, assume $\max(s_\alpha, s_\beta) = o(n^{1/3})$, the right-hand side of (5.15) is strengthened to $O(n^{\frac{1}{2}+\gamma_{\theta^\star}} / \sqrt{s_\beta s_{\theta^\star} \log^3 n})$, as $n \to \infty$. Then with probability tending to 1, $\hat{\alpha} = (\hat{\alpha}_{M_\alpha}^T, \hat{\alpha}_2^T)^T$, $\hat{\beta} = (\hat{\beta}_{M_\beta}^T, \hat{\beta}_2^T)^T$ in Theorem 6.1 must satisfy

- **a.** $\hat{\alpha}_2 = 0$, $\hat{\beta}_2 = 0$,
- **b.** $[A_{1n} B_{n\alpha}^{1/2} (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}), A_{2n} B_{n\beta}^{1/2} (\hat{\beta}_{M_\beta} - \beta_{0M_\beta})]$ is asymptotically normally distributed with mean 0, covariance matrix Ω, which is the limit of

$$\begin{pmatrix} A_{1n} A_{1n}^T & A_{1n} \sum_{12} A_{2n}^T \\ A_{2n} \sum_{21} A_{1n}^T & \sigma^2 A_{2n} A_{2n}^T + A_{2n} \sum_{22} A_{2n}^T \end{pmatrix},$$

where A_{1n} is a $q_1 \times s_\alpha$ matrix and A_{2n} is a $q_2 \times s_\beta$ matrix such that

$$\lambda_{\max}(A_{1n} A_{1n}^T) = O(1), \quad \lambda_{\max}(A_{2n} A_{2n}^T) = O(1).$$

We note that conditions on the smoothness of the misspecified function (5.15) is strengthened. To better understand the above theorem, we provide the following two corollaries. The first corollary gives the limiting distribution when we specify both the propensity score and main-effect model while the second one corresponds to case when the propensity score is known in advance.

Corollary 6.1. Under conditions of Theorem 6.2, when we correctly specify the main-effect model, i.e., $\mu = \Phi$, $A_{1n} B_{n\alpha}^{1/2} (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha})$ and $A_{2n} B_{n\beta}^{1/2} (\hat{\beta}_{M_\beta} - \beta_{0M_\beta})$ are jointly asymptotically normally distributed, with the covariance matrix Ω', which is the limit of the following matrix,

$$\begin{pmatrix} A_{1n}^T A_{1n} & O \\ O & \sigma^2 A_{2n}^T A_{2n} \end{pmatrix}.$$

Remark 6.5. Comparing the results in Corollary 6.1 and in Theorem 6.2, the term $A_{2n}^T \sum_{22} A_{2n}$ accounts for the partial specification of model (2.1). In the most extreme case where we correctly specify Φ, $\hat{\beta}_{M\beta}$ will achieve its minimum variance and is independent of $\hat{\alpha}_{M\alpha}$. In general, we can gain efficiency by posing a good working model for Φ. Numerical studies also suggest that a linear model such as $\Phi = X\theta$ is preferred compared to the constant model. This is in line to our theoretical justification since W is a diagonal matrix with the ith diagonal element $\mu^i - \Phi^i$.

Corollary 6.2. When the propensity score is known, under conditions of Theorem 6.2 with all $\hat{\alpha}$'s replaced by α_0, then with probability tending to 1 as $n \to \infty$, $A_{2n} B_{n\beta}^{1/2} (\hat{\beta}_{M_\beta} - \beta_{0M_\beta})$ is asymptotically normally distributed with mean 0, co-variance matrix Ω'' which is the limit of

$$\sigma^2 A_{2n}^T A_{2n} + A_{2n}^T \sum\nolimits_{22}^{'} A_{2n},$$

where

$$\sum\nolimits_{22}^{'} = B_{n2}^{-1/2} X_{M_\beta}^T W \Delta W X_{M_\beta} B_{n2}^{-1/2}.$$

Remark 6.6. An interesting fact implied by Corollary 6.2 is that the asymptotic variance of $\hat{\beta}_{M\beta}$ will be smaller than that of the same estimator had we known the propensity score in advance. A similar result is given in the asymptotic distribution of the mean response for the value function in the next section. This is in line with the semiparametric theory in fixed p case where the variance of augmented-IPWS estimator would be smaller when we estimate the parameter in the coarsening probability model, even if we know what the true value is (see Chapter 9 in Tsiatis, 2006). By doing so, we can actually borrow information from the linear association between covariates in $WX_{M\beta}$ and those in $X_{M\alpha}$.

7. Evaluation of value function

In this section, we derive a non-parametric estimate for the mean response under the optimal treatment regime. By (2.1), define our average population-level response under a specific regime as

$$V_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[Y_i | A_i = I(x_i^T \beta > 0), X_i = x_i] = \frac{1}{n} \sum_{i=1}^{n} [h_0(x_i) + x_i^T \beta_0 I(x_i^T \beta > 0)],$$

where the treatment decision for the ith patient is given as $I(x_i^T\beta>0)$. The mean response under the true optimal regime is denoted as $V_n(\beta_0)$ and it is easy to verify that β_0 is the maximizer of the function V_n.

Similarly as in Murphy (2003), we propose to estimate $V_n(\beta_0)$ using

$$\hat{V}_n=\frac{1}{n}\sum_{i=1}^{n}[Y_i+x_i^T\hat{\beta}\{I(x_i^T\hat{\beta}>0)-A_i\}]. \tag{7.1}$$

This estimator is not doubly robust but offers protection against misspecification of the baseline function and improved efficiency It's not doubly robust because we require the propensity score model to be correctly specified to ensure the oracle property of $\hat{\beta}$. A key condition which guarantees asymptotic normality of (7.1) is given as follows.

Condition 7.1. Assume there exists some constant C', such that for all $\varepsilon > 0$,

$$\frac{1}{n}\sum_i I(|x_i^T\beta_0|<\varepsilon)\le C'\varepsilon.$$

Remark 7.1. The above condition has similar interpretation as Condition (3.3) in Qian and Murphy (2011), where random design were utilized. Condition 7.1 requires that the absolute value of the average contrast function can not be too small, which together with the condition $s_\beta = o(n^{1/4})$ ensures the following stochastic approximation condition:

$$\sqrt{n}\sum_i x_i^T\hat{\beta}\{I(x_i^T\hat{\beta}>0)-I(x_i^T\beta_0>0)\}=o_p(1). \tag{7.2}$$

Theorem 7.1. Assume that conditions in Theorem 6.2 hold. If Condition 7.1 holds and the nonsparsity size s_β satisfies $s_\beta = o(n^{1/4})$, then with probability going to 1, $\sqrt{n}\{\hat{V}_n - V_n(\beta_0)\}$ is asymptotically normally distributed with variance ν_0^2, which is limit of

$$\sigma^2+\sigma^2\upsilon_n^T X_{M_\beta}B_{n\beta}^{-1}X_{M_\beta}^T\upsilon_n+\upsilon_n^T X_{M_\beta}B_{n\beta}^{-1/2}\Sigma_{22}B_{n\beta}^{-1/2}X_{M_\beta}^T\upsilon_n, \tag{7.3}$$

where υ_n stands for the vector $[I(x_1^T\beta_0>0)-\pi(x_1),\ldots,I(x_n^T\beta_0>0)-\pi(x_n)]^T/\sqrt{n}$, and Σ_{22} is defined in Theorem 6.2.

Remark 7.2. Note that we only need $s_\beta = o(n^{1/2})$ to guarantee the weak oracle property of $\hat{\beta}$ or $O(\sqrt{s_\beta}/\sqrt{n})$ convergence rate of $\|\hat{\beta}_{M_\beta}-\beta_{0M_\beta}\|_2$. This condition is strengthened to $s_\beta =$

$o(n^{1/3})$ to show the asymptotic normality of $\hat{\beta}_{M_\beta}$. Theorem 7.1 further requires $s_\beta = o(n^{1/4})$ as to ensure the approximation condition (7.2).

Remark 7.3. When (7.2) is satisfied, the asymptotic normality of $\hat{V}_n$ follows immediately from the oracle property of the estimator $\hat{\beta}_{M_\beta}$. The first term σ^2 in (7.3) is due to variation of the error term e_i while the last two terms correspond to the asymptotic variance of $\hat{\beta}_{M_\beta}$.

We provide a corollary here which corresponds to the case where the main-effect model is correctly specified.

Corollary 7.1. In addition to the conditions in Theorem 7.1, if the main-effect model is correct, $\sqrt{n}\{\hat{V}_n - V_n(\beta_0)\}$ is asymptotically normally distributed with variance ν_1^2, which is defined as the limit of

$$\sigma^2+\sigma^2 \upsilon_n^T X_{M_\beta} B_{n\beta}^{-1} X_{M_\beta}^T \upsilon_n,$$

where υ_n is defined in Theorem 7.1.

Similar to the asymptotic distribution of $\hat{\beta}_{M_\beta}$, the following corollary suggests that the proposed estimator is more efficient in the case when we estimate the propensity score by fitting a penalized logistic regression.

Corollary 7.2. Assume the propensity score is known, and conditions in Theorem 7.1 hold with all $\hat{\alpha}$'s replaced by α_0, then with probability going to 1, $\sqrt{n}\{\hat{V}_n - V_n(\beta_0)\}$ is asymptotically normally distributed with variance ν_2^2, which is the limit of

$$\sigma^2+\sigma^2 \upsilon_n^T X_{M_\beta} B_{n\beta}^{-1} X_{M_\beta}^T \upsilon_n+\upsilon_n^T X_{M_\beta} B_{n\beta}^{-1/2} \sum\nolimits_{22}^{'} B_{n\beta}^{-1/2} X_{M_\beta}^T \upsilon_n,$$

with υ_n defined in Theorem 7.1, and $\sum_{22}^{'}$ defined in Corollary 6.2.

By the definition of υ_n and the condition that $\lambda_{\max}(X_{M_\beta}^T X_{M_\beta})=O(n)$, the asymptotic variance will reach its minimum when $I(x_i^T \beta_0>0)$ is close to the propensity score. We characterize this result in the following Corollary.

Corollary 7.3. Under the conditions in Theorem 7.1, if we further assume that

$$\frac{1}{n}\sum_{i=1}^{n}\{I(X_i^T \beta_0>0) - \pi(x_i)\}^2=o(1),$$

then with probability going to 1, $\sqrt{n}\{\hat{V}_n - V_n(\beta_0)\}$ is asymptotically normally distributed with the variance σ^2.

Remark 7.4. Such a result is expected with the following intuition: in an observational study, if the clinician or the decision maker has a high chance to assign the optimal treatment to an individual patient, i.e., the propensity score is close to $I(x_i^T\beta_0>0)$, the variation in estimating the value function will be decreased. In other words, the more skillful the clinician or the decision maker is, the closer the observed individual response Y_i approaches the potential outcome under the optimal treatment regime.

8. Conclusion

In this article, we propose a two-step estimator for estimating the optimal treatment strategy which selects variables and estimates parameters simultaneously in both propensity score and outcome regression models using penalized regression. Our methodology can handle data set whose dimensionality is allowed to grow exponentially fast compared to the sample size. Oracle properties of the estimators are given. Variable selection is also involved in the misspecified model and new mathematical techniques are developed to study the estimator's properties in a general form of optimization. The estimator is shown to be more efficient when the misspecified working model is “closer” to the conditional mean of the response, although our approach does not require correct specification of the baseline function. Numerical results demonstrate that the proposed estimator enjoys model selection consistency and has overall satisfactory performance.

In the case when there are multiple local solutions of our objective functions (2.5), (2.3) or (2.4), although our asymptotic theory only suggests the existence of a local minimum possessing the oracle property, it is worth mentioning that we can actually identify the desired oracle estimator using existing algorithms (see Fan, Xue and Zou, 2014; Wang, Kim and Li, 2013). Theoretical properties can be established in a similar fashion.

The proposed method requires to specify the propensity score model correctly. In randomized studies, the propensity score is known in advance and thus the assumption is automatically satisfied. However, for observational studies, there's no guarantee. In practice, some prior information on treatment decision mechanism used by physicians may be helpful for building a reasonable propensity score. In addition, model diagnostic tests can be used to check the goodness-of-fit of the posited propensity score model, such as a logistic regression model. In general, this might be easier than checking the goodness-of-fit of the regression model for the response. In addition, in our current work, we assume the design matrix X to be deterministic mainly for technical convenience. To the best of our knowledge, the penalized regression with the folded-concave penalties has never been studied in random design settings with NP dimensionality. To consider random design settings, we need to impose some tail conditions on X, and the derivation of some technical results needs to be modified. This is beyond the scope of our current paper and will be investigated elsewhere.

The current framework is focused on point treatment study. It will be interesting and practically useful to extend our results to dynamic treatment regimes. Significant efforts are needed to handle model misspecification in multiple stages. This is an interesting research topic that needs further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix

Here, we only give the proof of Theorem 5.1. More technical conditions and proofs for Theorems 6.1, 6.2 and 7.1 are given in the supplementary Appendix. To establish Theorem 5.1, we need the following lemmas. The proofs of these lemmas are also given in the supplementary Appendix.

Lemma 1. Let $z = (z^1, \ldots, z^n)^T$ be an n-dimensional independent random response vector with mean 0 and $a \in \mathbb{R}^n$.

a. If $z^1, \ldots, z^n$ are bounded in $[c, d]$, then for any $\varepsilon \in (0, \infty)$,

$$\Pr(|a^T z| > \varepsilon) \leq 2\exp\left(-\frac{\varepsilon^2}{2\|a\|_2^2 (d-c)^2}\right).$$

b. If $z^1, \ldots, z^n$ satisfy $\max_i \|z^i\|_{\psi 1} \leq \omega$, then for any $\varepsilon \in (0, \infty)$,

$$\Pr(|a^T z| > \varepsilon) \leq 2\exp\left(-\frac{1}{2}\frac{\varepsilon^2}{2\|a\|_2^2 \omega^2 + \|a\|\varepsilon\omega}\right).$$

Lemma 2. Define $\varepsilon = \cup_{k=1}^{16} \varepsilon_k$, where ε_k is defined in Appendix G, under conditions in Theorem 5.1, we have $P_I(\varepsilon) \geq 1 - \bar{c}'(n + p + q)$ for some $\bar{c} > 0$.

Notation. Let $Z = diag(A - \pi)X$, $\hat{Z} = diag(A - \hat{\pi})X$, and

$$\xi_1 = \hat{Z}^T e, \quad \xi_2 = Z^T(\mu - \Phi), \quad \xi_3 = \phi^T(e - Z\beta_0),$$
$$\xi_4 = Z^T \mathrm{diag}(X\beta_0)\Delta X_{M_\alpha}, \quad \xi_5 = X^T[\mathrm{diag}\{(A-\pi)\circ(A-\pi)\} - \Delta]X_{M_\beta},$$
$$\xi_6(\delta) = Z^T\{\Phi - \Phi(\delta)\}, \quad \xi_7(\delta) = \{\phi(\delta) - \phi\}^T(e - Z\beta_0),$$

and $\pi = (\pi(x_1), \ldots, \pi(x_n))$. For a given matrix Ψ, the superscript Ψ^j is used to refer to the vector which is the jth column of matrix Ψ while the subscript Ψ_i stands for the ith row of Ψ. We will write $\Phi(\theta)$, $\phi(\theta)$ with $\theta = (\theta_{M_{\theta^\star}}^T, 0^T)^T$ as $\Phi(\theta_{M_{\theta^\star}})$, $\phi(\theta_{M_{\theta^\star}})$ for convenience.

Proof of Theorem 5.1

We break the proof into three steps. Based on Theorem 1 in Fan and Lv (2011), it suffices to prove the existence of $\hat{\beta}_{M_\beta}$, $\hat{\theta}_{M_{\theta^\star}}$ inside the hypercube

$$\aleph=\{(\delta_\beta^T,\delta_\theta^T)^T:\|\delta_\beta-\beta_{0M_\beta}\|_\infty=n^{-\gamma_\beta}\log n,\|\delta_\theta-\theta^\star_{M_{\theta^\star}}\|_\infty=Kn^{-\gamma_{\theta^\star}}\log n\}$$

with K a large constant, conditional on the event ε, satisfying

$$\hat{Z}^T_{M_\beta}\{Y-\Phi(\hat{\theta})-\hat{Z}\hat{\beta}\}=n\lambda_{2n}\overline{\rho}_2(\hat{\beta}_{M_\beta}), \quad \text{(A.1)}$$

$$\hat{\phi}^T_{M_{\theta^\star}}\{Y-\Phi(\hat{\theta})\}=n\lambda_{3n}\overline{\rho}_3(\hat{\theta}_1), \quad \text{(A.2)}$$

$$\|\hat{Z}^T_{M^c_\beta}\{Y-\Phi(\hat{\theta})-\hat{Z}\hat{\beta}\}\|_\infty<n\lambda_{2n}\rho_2'(0+), \quad \text{(A.3)}$$

$$\|\hat{\phi}^T_{M^c_{\theta^\star}}\{Y-\Phi(\hat{\theta})\}\|_\infty<n\lambda_{3n}\rho_3'(0+), \quad \text{(A.4)}$$

$$\lambda_{\min}(\hat{Z}^T_{M_\beta}\hat{Z}_{M_\beta})>n\lambda_{2n}\kappa(\rho_2,\hat{\beta}_{M_\beta}), \quad \text{(A.5)}$$

$$\lambda_{\min}(\hat{\phi}^T_{M_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}})>n\lambda_{3n}\kappa(\rho_3,\hat{\theta}_{M_{\theta^\star}}). \quad \text{(A.6)}$$

Step 1. We first show the existence of a solution to equations (A.1) and (A.2) inside $\aleph$ for sufficiently large n. For any $\delta=(\delta^1,\ldots,\delta^{s_\beta+s_{\theta^\star}})^T\in\aleph$, since $d_{n\beta}\geq n^{-\gamma_\beta}\log n$, $d_{n\theta}\gg n^{-\gamma_{\theta^\star}}\log n$, we have

$$\min_{j=1}^{s_\beta}|\delta^j|\geq\min|\beta_0^j|-d_{n\beta}=d_{n\beta},\quad \min_{j=1}^{s_{\theta^\star}}|\delta^{j+s_\beta}|\geq\min|\theta^{\star j}|-d_{n\theta}=d_{n\theta}$$

and $\operatorname{sgn}(\delta_\beta)=\operatorname{sgn}(\beta_{0M_\beta})$, $\operatorname{sgn}(\delta_\theta)=\operatorname{sgn}(\theta^\star_{M_{\theta^\star}})$. The monotonicity condition of $\rho_2'(t)$, $\rho_3'(t)$ gives

$$\|n\lambda_{2n}\overline{\rho}_2(\delta)\|_\infty\leq n\lambda_{2n}\rho_2'(d_{n\beta}),\quad \|n\lambda_{3n}\overline{\rho}_3(\delta)\|_\infty\leq n\lambda_{3n}\rho_3'(d_{n\theta}). \quad \text{(A.7)}$$

We write the left hand side of (A.1) as

$$\begin{aligned}\hat{Z}_{M_\beta}^T\{&Y\\&-\Phi(\delta_\theta)\\&-\hat{Z}_{M_\beta}\delta_\beta\}\\=&\xi_{1M_\beta}+\xi_{2M_\beta}+(\hat{Z}_{M_\beta}-Z_{M_\beta})^T\{\mu-\Phi(\delta_\theta)\}\\&+\hat{Z}_{M_\beta}^T\hat{Z}_{M_\beta}(\beta_{0M_\beta}-\delta_\beta)+\hat{Z}_{M_\beta}^T(\hat{Z}_{M_\beta}\\&-Z_{M_\beta})\beta_{0M_\beta}\\&-Z_{M_\beta}^T\{\Phi(\delta_\theta)\\&-\Phi\}.\triangleq I_1\\&+I_2+I_3+I_4+I_5+I_6,\end{aligned}\tag{A.8}$$

on the set $\varepsilon_3 \cup \varepsilon_5 \cup \varepsilon_{13}$, we have

$$\|I_1\|_\infty+\|I_2\|_\infty+\|I_3\|_\infty=O(\sqrt{n\log n}).\tag{A.9}$$

Define

$$\eta_1=(\hat{Z}-Z)^T\{\mu-\Phi(\delta_\theta)\},\eta_2=(\hat{Z}-Z)^T(\hat{Z}_{M_\beta}-Z_{M_\beta})\beta_{0M_\beta}.$$

Note that $\eta_{1M_\beta} = I_3$ in (A.8), which we represent here using a second order Taylor expansion around α_{0M_α},

$$I_3=X_{M_\beta}^T W(\delta_\theta)\Delta X_{M_\alpha}(\alpha_{0M_\alpha}-\hat{\alpha}_{M_\alpha})+\frac{1}{2}r_{I_3},\tag{A.10}$$

where r_{I_3} in (A.10) corresponds to second order remainder, whose jth component is given as

$$(\hat{\alpha}_{M_\alpha}-\alpha_{0M_\alpha})^T X_{M_\alpha}^T W(\delta_\theta)\sum(\tilde{\alpha})\mathrm{diag}(x^j)X_{M_\alpha}(\hat{\alpha}_{M_\alpha}-\alpha_{0M_\alpha}),$$

where $\Sigma(\tilde{\alpha})$ is a diagonal matrix with the ith diagonal element $\pi''(x_{1\alpha i}^T\tilde{\alpha})$ with $\tilde{\alpha}$ lying in the line segment between $\hat{\alpha}_{M_\alpha}$ and α_{0M_α}. Since $\pi''(\cdot)$ is a bounded function, we can bound $\|r_{I_3}\|_\infty$ by

$$\max_j (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha})^T X_{M_\alpha}^T \mathrm{diag}(|W(\delta_\theta)x^j|) X_{M_\alpha} (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}), \quad \text{(A.11)}$$

whose order of magnitude is $O(s_\alpha n^{1-2\gamma_\alpha} \log^2 n)$ by (5.16).

We decompose I_4 in (A.8) as $\eta_{2M_\beta} + Z_{M_\beta}^T (\hat{Z}_{M_\beta} - Z_{M_\beta}) \beta_{0M_\beta}$. Using similar arguments, on the set $\mathcal{E}_9$, it follows from (5.17) that

$$\|Z_{M_\beta}^T (\hat{Z}_{M_\beta} - Z_{M_\beta}) \beta_{0M_\beta}\|_\infty \le \max_j (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha})^T X_{M_\alpha}^T \mathrm{diag}(|x^j \circ X\beta_0|) X_{M_\alpha} (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}) + \|\xi_{4M_\beta}\|_\infty = O(\sqrt{n \log n} + s_\alpha n^{1-2\gamma_\alpha} \log^2 n).$$

(A.12)

Using Taylor expansion, it is immediate to see that

$$\|\eta_{2M_\beta}\|_\infty \le \max_j (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha})^T X_{M_\alpha}^T \mathrm{diag}(|x^j \circ X\beta_0|) X_{M_\alpha} (\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}) = O(s_\alpha n^{1-2\gamma_\alpha} \log^2 n),$$

(A.13)

by (5.17). Combining (A.12) and (A.13) gives

$$\|\hat{Z}_{M_\beta}^T (\hat{Z}_{M_\beta} - Z_{M_\beta}) \beta_{0M_\beta}\|_\infty = O(\sqrt{n \log n}) + O(s_\alpha n^{1-2\gamma_\alpha} \log^2 n). \quad \text{(A.14)}$$

So far, we have

$$\|I_1 + I_2 + I_3 + I_4 + I_5 + I_6 - X_{M_\beta}^T W(\delta_\theta) \Delta X_{M_\alpha} (\alpha_{0M_\alpha} - \hat{\alpha}_{M_\alpha})\|_\infty = O(\sqrt{n \log n}) + O(s_\alpha n^{1-2\gamma_\beta} \log^2 n) + O(s_\beta n^{1-2\gamma_\beta} \log^2 n),$$

(A.15)

by (A.9), (A.10), (A.11) and (A.14). Now we approximate I_4 by $X_{M_\beta}^T \Delta X_{M_\beta} (\delta_\beta - \beta_{0M_\beta})$ and bound the magnitude of error $\|\omega_{M_\beta}\|_\infty$ where $\omega = (\hat{Z}^T \hat{Z}_{M_\beta} - X^T \Delta X_{M_\beta})(\delta_\beta - \beta_{0M_\beta})$. We present it as

$$\begin{aligned}\omega_{M_\beta}=&(\hat{Z}_{M_\beta}^T\hat{Z}_{M_\beta}\\&-X_{M_\beta}^T\Delta X_{M_\beta})(\delta_\beta\\&-\beta_{0M_\beta})=\hat{Z}_{M_\beta}^T(\hat{Z}_{M_\beta}\\&-Z_{M_\beta})(\delta_\beta\\&-\beta_{0M_\beta})+(\hat{Z}_{M_\beta}-Z_{M_\beta})^TZ_{M_\beta}(\delta_\beta\\&-\beta_{0M_\beta})+(Z_{M_\beta}^TZ_{M_\beta}\\&-X_{M_\beta}^T\Delta X_{M_\beta})(\delta_\beta\\&-\beta_{0M_\beta})\triangleq\omega_{1M_\beta}\\&+\omega_{2M_\beta}+\xi_{5M_\beta}(\delta_\beta-\beta_{0M_\beta}).\end{aligned}\tag{A.16}$$

It follows from first-order Taylor expansion that the jth element in ω_{1M_β} can be presented as

$$[(A-\hat{\pi})\circ x^j\circ\{\Delta(\tilde{\alpha}_{M_\alpha})X_{M_\alpha}(\hat{\alpha}_{M_\alpha}-\alpha_{0M_\alpha})\}]^TX_{M_\beta}(\delta_\beta-\beta_{0M_\beta}),\tag{A.17}$$

where $\Delta(\tilde{\alpha}_{M_\alpha})$ is a diagonal matrix with the ith diagonal component $\pi(x_i, \tilde{\alpha}_{M_\alpha})(1-\pi(x_i, \tilde{\alpha}_{M_\alpha}))$, where $\tilde{\alpha}_{M_\alpha}$ lies between the line segment of $\hat{\alpha}_{M_\alpha}$ and α_{0M_α}. We decompose x^j as the Hadamard product of two vectors, denoted by $\bar{x}^j\circ\tilde{x}^j$, where

$$\begin{aligned}\bar{x}^j&=\left(\sqrt{|x_1^j|},\ldots,\sqrt{|x_n^j|}\right),\\\tilde{x}^j&=\left(\mathrm{sgn}(x_1^j)\sqrt{|x_1^j|},\ldots,\mathrm{sgn}(x_n^j)\sqrt{|x_n^j|}\right).\end{aligned}$$

Let $\varphi=(A-\hat{\pi})\circ\bar{x}^j\circ\{\Delta(\tilde{\alpha}_{M_\alpha})X_{M_\alpha}(\hat{\alpha}_{M_\alpha}-\alpha_{0M_\alpha})\}$, we have

$$\begin{aligned}&\|[(A-\hat{\pi})\circ\tilde{x}^j\circ\{\Delta(\tilde{\alpha}_{M_\alpha})X_{M_\alpha}(\hat{\alpha}_{M_\alpha}-\alpha_{0M_\alpha})\}]^TX_{M_\beta}\|_2\|\delta_\beta-\beta_{0M_\beta}\|_2\\&=\sqrt{\varphi^T\mathrm{diag}(\bar{x}^j)X_{M_\beta}X_{M_\beta}^T\mathrm{diag}(\bar{x}^j)\varphi}\|\delta_\beta-\beta_{0M_\beta}\|_2\le\sqrt{\lambda_{\max}(X_{M_\beta}^T\mathrm{diag}(|x^j|)X_{M_\beta})}\|\delta_\beta-\beta_{0M_\beta}\|_2\|\varphi\|_2.\end{aligned}$$

(A.18)

Since $\|A-\hat{\pi}\|_\infty\le 1$, elements in $\Delta(\tilde{\alpha}_{M_\alpha})$ are bounded, we have

$$\|\varphi\|_2 \le \|\mathrm{diag}(\tilde{x}^j)X_{M_\alpha}(\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha})\|_2 \le \sqrt{\lambda_{\max}\{X_{M_\alpha}^T \mathrm{diag}(|x^j|)X_{M_\alpha}\}}\|\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}\|_2.$$

(A.19)

Combining (A.18) with (A.19) gives

$$\|\omega_{1M_\beta}\|_\infty \le \max_{j=1}^{p} \sqrt{\lambda_{\max}\{X_{M_\alpha}^T \mathrm{diag}(|x^j|)X_{M_\alpha}\}\|\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}\|_2^2} \max_{j=1}^{p} \sqrt{\lambda_{\max}\{X_{M_\beta}^T \mathrm{diag}(|x^j|)X_{M_\beta}\}\|\hat{\beta}_{M_\beta} - \beta_{0M_\beta}\|_2^2},$$

(A.20)

which is $O(\sqrt{s_\alpha s_\beta} n^{1-\gamma_a-\gamma_\beta}\log^2 n)$ by (B.4) and (B.5).

By the same argument, we can verify that $\|\omega_{2M_\beta}\|_\infty$ is of the same order. Note that on the set ε_{11},

$$\|\xi_{5M_\beta}(\delta - \beta_{0M_\beta})\|_\infty \le \|\xi_{5M_\beta}\|_\infty \|\delta - \beta_{0M_\beta}\|_\infty = O(s_\beta n^{1-2\gamma_\beta}\log^2 n),$$

these together with (A.20), yields

$$\|\omega_{M_\beta}\|_\infty = O(s_\alpha n^{1-2\gamma_\alpha}\log^2 n) + O(s_\beta n^{1-2\gamma_\beta}\log^2 n). \quad \text{(A.21)}$$

Define vector-valued function

$$\begin{aligned}
&\Psi_1(\delta_\beta, \delta_\theta) \\
&\quad = B_{n\beta}^{-1}[\hat{Z}_{M_\beta}^T\{y - \Phi(\delta_\theta) - \hat{Z}_{M_\beta}\delta_\beta\} - n\lambda_{2n}\overline{\rho}_2(\delta_\beta)] \\
&\quad = B_{n\beta}^{-1}\{I_1 + I_2 + I_3 + I_4 + I_5 + I_6 - n\lambda_{2n}\overline{\rho}_2(\delta_\beta)\} \\
&\quad = \delta_\beta - \beta_{0M_\beta} + B_{n\beta}^{-1}\{I_1 \\
&\qquad + I_2 + I_3 + \omega_{M_\beta} + I_5 + I_6 - n\lambda_{2n}\overline{\rho}_2(\delta_\beta)\} \triangleq \delta_\beta \\
&\quad - \beta_{0M_\beta} + u_\beta,
\end{aligned} \quad \text{(A.22)}$$

then equation (A.1) is equivalent to $\Psi_1(\delta_\beta, \delta_\theta) = 0$. It follows from (A.7), (A.15) and (A.21) that

$$\|u_\beta\|_\infty \le \sup_{\delta\in H_{\theta^\star}} \|B_{n\beta}^{-1}X_{M_\beta}^T W(\delta)\Delta X_{M_\alpha}(\hat\alpha_{M_\alpha}-\alpha_{0M_\alpha})\|_\infty + \|B_{n\beta}^{-1}\|_\infty\{O(s_\alpha n^{1-2\gamma_\alpha}\log^2 n)+O(s_\beta n^{1-2\gamma_\beta}\log^2 n)+O(\sqrt{n\log n})+n\lambda_{2n}\rho'(d_{n\beta})\}.$$

By similar arguments in the proof of Theorem 2 in Fan and Lv (2011), we have

$$\|B_{n\alpha}(\hat\alpha_{M_\alpha}-\alpha_{0M_\alpha})\|_\infty = O(s_\alpha n^{1-2\gamma_\alpha}\log^2 n)+O(\sqrt{n\log n})+n\lambda_{1n}\rho_1'(d_{n\alpha}), \quad \text{(A.23)}$$

on the set $\varepsilon_1 \cup \varepsilon_2$. Thus by (5.10), (B.1), (B.14) and (B.15), we have

$$\|u_\beta\|_\infty \le O[b_{\alpha\beta}\{s_\alpha n^{-2\gamma_\alpha}\log^2 n + \sqrt{\log n/n}+\lambda_{1n}\rho_1'(d_{n\alpha})\}]+O[b_\beta\{s_\alpha n^{-2\gamma_\alpha}\log^2 n+s_\beta n^{-2\gamma_\beta}\log^2 n+\sqrt{\log n/n}+\lambda_{2n}\rho_2'(d_{n\beta})\}].$$

Therefore by (A.20), for sufficiently large n, if $(\delta_\beta-\beta_{0M_\beta})^j = n^{-\gamma_\beta}\log n$,

$$\Psi_1^j(\delta_\beta,\delta_\theta)>0, \quad \text{(A.24)}$$

and if $(\delta-\beta_{0M_\beta})^j = -n^{-\gamma_\beta}\log n$,

$$\Psi_1^j(\delta_\beta,\delta_\theta)<0. \quad \text{(A.25)}$$

Similarly we write the left-hand side of (A.2) as

$$(\hat\phi_{M_{\theta^\star}}-\phi_{M_{\theta^\star}})^T(e-Z\beta_0)+\xi_{3M'_\theta}+\hat\phi_{M_{\theta^\star}}^T(\mu-\Phi)-\hat\phi_{M_{\theta^\star}}^T(\hat\Phi-\Phi). \quad \text{(A.26)}$$

It is immediately to see that

$$\|\xi_{3M_{\theta^\star}}\|_\infty = O(\sqrt{n\log n}), \quad \text{(A.27)}$$

on the set ε_5. The L_∞ norm of the first term in (A.26) is bounded by

$$\sup_{\delta\in H_{\theta^\star}} \|\xi_{7M_\beta}(\delta)\|_\infty = O(\sqrt{n\log n}), \quad \text{(A.28)}$$

on the set ε_{15}.

Using second-order Taylor expansion, we approximate the last term in (A.26) by its first-order term $\hat{\phi}^T_{M_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}}(\delta_\theta - \theta^\star_{M_{\theta^\star}})$. It follows from (5.7) that the L_∞ norm of the remainder term is bounded from above by

$$\max_{l=-1}^{s_{\theta^\star}} \lambda_{\max}\left\{\frac{\partial(|\phi^l(\delta_\theta)|)^T\phi_{M_{\theta^\star}}(\tilde{\delta}_\theta)}{\partial\theta_{M_{\theta^\star}}}\right\}\|\delta_\theta - \theta^\star_{M_{\theta^\star}}\|_2^2 = O(s_{\theta^\star}n^{1-2\gamma_{\theta^\star}}\log^2 n), \quad \text{(A.29)}$$

where $\tilde{\delta}_\theta$ lies between the line segment of $\theta^\star_{M_{\theta^\star}}$ and δ_θ.

Define $\Psi_2(\delta_\beta, \delta_\theta) = \{\phi_{M_{\theta^\star}}(\delta_\theta)^T\phi_{M_{\theta^\star}}(\delta_\theta)\}^{-1}[\phi_{M_{\theta^\star}}(\delta_\theta)^T\{Y - \Phi(\delta_\theta)\} - n\lambda_{3n}\bar{\rho}_3(\delta_\theta)]$, equation (A.2) is equivalent to $\Psi_2(\delta_\beta, \delta_\theta) = 0$. Similarly to $\Psi_1(\delta_\beta, \delta_\theta)$, we now show $\Psi_2(\delta_\beta, \delta_\theta)$ is mainly dominated by $\delta_\theta - \theta^\star_{M_{\theta^\star}}$. Define $u_\theta = \Psi_2(\delta_\beta, \delta_\theta) - \delta_\theta + \theta^\star_{M_{\theta^\star}}$, it follows from (5.1), (5.3), (B.13), (A.26), (A.27), (A.28) and (A.29) that

$$\begin{aligned}\|u_\theta\|_\infty &\le \|\Psi_2(\delta_\beta,\delta_\theta) - \delta_\theta + \theta^\star_{M_{\theta^\star}}\|_\infty \le \|\{\phi_{M_{\theta^\star}}(\delta_\theta)^T\phi_{M_{\theta^\star}}(\delta_\theta)\}^{-1}\|_\infty\{\|\xi_{3M'_\theta}\|_\infty + \|\xi_{7M'_\theta}(\delta_\theta)\|_\infty + \|\Phi(\delta_\theta) - \Phi - \phi(\delta_\theta)^T(\delta_\theta - \theta^\star_M\\ &+\|\{\phi_{M_{\theta^\star}}(\delta_\theta)^T\phi_{M_{\theta^\star}}(\delta_\theta)\}^{-1}\phi_{M_{\theta^\star}}(\delta_\theta)^T(\mu - \Phi)\|_\infty\\ &= o(n^{-\gamma_{\theta^\star}}\log n)\\ &+ O(n^{-\gamma_{\theta^\star}}\log n).\end{aligned} \quad \text{(A.30)}$$

Therefore, we can find a large constant $K < \infty$, for n large enough such that if $(\delta_\theta - \theta^\star_{M_{\theta^\star}})^j = Kn^{-\gamma_{\theta^\star}}\log n$,

$$\Psi_2^j(\delta_\beta, \delta_\theta) > 0, \quad \text{(A.31)}$$

and if $(\delta_\theta - \theta^\star_{M_{\theta^\star}})^j = -Kn^{-\gamma_{\theta^\star}}\log n$,

$$\Psi_2^j(\delta_\beta, \delta_\theta) < 0. \quad \text{(A.32)}$$

Combining (A.24), (A.25) with (A.31) and (A.32), an application of Miranda's existence theorem shows equations (A.1), (A.2) have a solution $(\hat{\beta}_{M_\beta}, \hat{\theta}_{M_{\theta^*}})$ in $\aleph$.

Step 2. Let $(\hat{\beta}^T, \hat{\theta}^T)^T$ be a solution to equations (A.1) and (A.2) with $\hat{\beta}_{M_\beta^c} = 0$ and $\hat{\theta}_{M_\beta^c} = 0$. We show that $(\hat{\beta}^T, \hat{\theta}^T)^T$ satisfies inequalities (A.3) and (A.4). Decompose (A.3) as the sum of the following terms,

$$\begin{aligned}
&\hat{Z}^T_{M_\beta^c}(Y \\
&\quad - \hat{\Phi} - \hat{Z}^T_{M_\beta}\hat{\beta}_{M_\beta}) \\
&= \xi_{1M_\beta^c} + \xi_{2M_\beta^c} + Z^T_{M_\beta}(\hat{Z}_{M_\beta} \\
&\quad - Z_{M_\beta})\beta_{0M_\beta} \\
&+ \xi_{5M_\beta^c}(\hat{\beta}_{M_\beta} - \beta_{0M_\beta}) + \omega_{1M_\beta^c} + \omega_{2M_\beta^c} + \eta_{1M_\beta^c} + X^T_{M_\beta^c}\Delta X_{M_\beta}(\hat{\beta}_{M_\beta} \\
&\quad - \beta_{0M_\beta}) + \eta_{2M_\beta^c} - Z_{M_\beta}(\hat{\Phi} - \Phi).
\end{aligned} \quad \text{(A.33)}$$

On the set $\varepsilon_4 \cup \varepsilon_6 \cup \varepsilon_{10} \cup \varepsilon_{12}$, it is immediately to see that

$$\|\xi_{1M_\beta^c}\|_\infty + \|\xi_{2M_\beta^c}\|_\infty + \|\xi_{5M_\beta^c}(\hat{\beta}_{M_\beta} - \beta_{0M_\beta})\|_\infty + \|Z^T_{M_\beta}(\hat{\Phi} - \Phi)\|_\infty = O(n^{1-d_\beta}\sqrt{\log n}). \quad \text{(A.34)}$$

By (B.4), (B.5) and (A.20), a first-order Taylor expansion gives

$$\|\omega_{1M_\beta^c}\|_\infty + \|\omega_{2M_\beta^c}\|_\infty = O(s_\alpha n^{1-2\gamma_\alpha}\log^2 n) + O(s_\beta n^{1-2\gamma_\beta}\log^2 n). \quad \text{(A.35)}$$

Similarly it follows from (5.17) and (A.13) that

$$\|\eta_{2M_\beta}\|_\infty = O(s_\alpha n^{1-2\gamma_\alpha}\log^2 n). \quad \text{(A.36)}$$

On the set ε_{10}, by (5.17) and (A.12), we have

$$\|Z^T_{M_\beta}(\hat{Z}_{M_\beta} - Z_{M_\beta})\beta\|_\infty = O(n^{1-d_\beta}\sqrt{\log n}) + O(s_\alpha n^{1-2\gamma_\alpha}\log^2 n). \quad \text{(A.37)}$$

Approximating $\eta_{1M_\beta^c}$ by $X^T_{M_\beta^c} W(\delta_\theta)\Delta X_{M_\alpha}(\alpha_{0M_\alpha} - \hat{\alpha}_{M_\alpha})$, the L_∞ norm of remainder error term is bounded from above by

$$(\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha})^T X^T_{M_\alpha} \mathrm{diag}(|W(\delta_\theta)x^j|) X_{M_\alpha}(\hat{\alpha}_{M_\alpha} - \alpha_{0M_\alpha}) = O(s_\alpha n^{1-\gamma_{\theta^\star}} \log^2 n), \quad \text{(A.38)}$$

by (5.16). Let $u'_\beta = \hat{Z}^T_{M_\beta^c}(Y - \hat{\Phi} - \hat{Z}_1^T \hat{\beta}_{M_\beta}) - X^T_{M_\beta^c}\Delta X_{M_\beta}(\hat{\beta}_{M_\beta} - \beta_{0M_\beta}) - X^T_{M_\beta^c} W(\hat{\theta}_1)\Delta X_{M_\alpha}(\alpha_{0M_\alpha} - \hat{\alpha}_{M_\alpha})$, it follows from (A.33)–(A.38) that

$$\|u'_\beta\|_\infty = O(n^{1-d_\beta}\sqrt{\log n} + s_\alpha n^{1-2\gamma_\alpha}\log^2 n + s_\beta n^{1-2\gamma_\beta}\log^2 n). \quad \text{(A.39)}$$

Since $\hat{\beta}_{M_\beta}$ solves (A.1), we have

$$\hat{\beta}_{M_\beta} - \beta_{0M_\beta} = -u_\beta, \quad \text{(A.40)}$$

where u_β is defined as $\Psi_1(\hat{\beta}_{M_\beta}, \hat{\theta}_{M_{\theta^\star}}) + \beta_{0M_\beta} - \hat{\beta}_{M_\beta}$. Combining (A.40) with (A.23) and (A.39) gives

$$\left\|\frac{1}{n\lambda_{2n}}\hat{Z}^T_{M_\beta^c}(Y - \hat{\Phi} - \hat{Z}^T_{M_\beta}\hat{\beta}_{M_\beta})\right\|_\infty \le \frac{1}{n\lambda_{2n}}\left[\|u'_\beta\|_\infty + \left\|X^T_{M_\beta^c}\Delta X_{M_\beta}(X^T_{M_\beta}\Delta X_{M_\beta})^{-1}\right\|_\infty \{u_\beta - X^T_{M_\beta} W(\hat{\theta}_{M_{\theta^\star}})\Delta X_{M_\alpha}(\alpha_{0M_\alpha} - \hat{\alpha}_{M_\alpha}\right.$$
$$\left. + C\rho'_2(0+),\right.$$

by (5.11), (B.3), (B.16) and (B.19). Since $C < 1$, for sufficiently large n, (A.3) is satisfied.

Now we verify (A.4), decomposing $\hat{\phi}^T_{M^c_{\theta^\star}}(Y - \hat{\Phi})$ as the sums of

$$(\hat{\phi}_{M^c_{\theta^\star}} - \phi_{M^c_{\theta^\star}})^T(e - Z\beta_0) + \xi_{3M^c_\theta} + \hat{\phi}^T_{M^c_{\theta^\star}}(\mu - \Phi) + \hat{\phi}^T_{M^c_{\theta^\star}}(\Phi - \hat{\Phi}), \quad \text{(A.41)}$$

on the set $\varepsilon_8 \cup \varepsilon_{16}$, we have

$$\|\xi_{3M'^c_\theta}\|_\infty + \|(\hat{\phi}_{M^c_{\theta^\star}} - \phi_{M^c_{\theta^\star}})^T(e - Z\beta_0)\|_\infty = O(n^{1-d_\theta}\sqrt{\log n}). \quad \text{(A.42)}$$

Similar to (A.29), a second-order Taylor expansion gives

$$\|\hat{\phi}^T_{M^c_{\theta^\star}}(\hat{\Phi}-\Phi)-\hat{\phi}^T_{M^c_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}}(\hat{\theta}_{M_{\theta^\star}}-\theta^\star_{M_{\theta^\star}})\|_\infty = O(s_{\theta^\star}n^{1-2\gamma_{\theta^\star}}\log^2 n), \tag{A.43}$$

by (5.7). Since $(\hat{\beta}_{M_\beta}, \hat{\theta}_{M_{\theta^\star}})$ is the solution to $\Psi_2(\delta_\beta, \delta_\theta) = 0$, it follows from (A.30) that

$$\begin{aligned}&\|\hat{\phi}^T_{M^c_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}}(\hat{\theta}_{M_{\theta^\star}}-\theta^\star_{M_{\theta^\star}})-\hat{\phi}^T_{M^c_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}}(\hat{\phi}^T_{M_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}})^{-1}(\mu-\Phi)\|_\infty\\ &=\|\hat{\phi}^T_{M^c_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}}(\hat{\phi}^T_{M_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}})^{-1}\|_\infty\{O(\sqrt{n\log n}+s_{\theta^\star}n^{1-2\gamma_{\theta^\star}}\log^2 n)+n\lambda_3\rho_3'(d_{n\theta})\}.\end{aligned} \tag{A.44}$$

By (A.41)–(A.44) and conditions in (5.2), (5.4), (B.15) and (B.20), the left-hand side of (A. 4) can be bounded by

$$\begin{aligned}&\frac{1}{n\lambda_{3n}}\{O(n^{1-d_\theta}\sqrt{\log n})\\ &\quad+O(s_{\theta^\star}n^{1-2\gamma_{\theta^\star}}\log^2 n)\}+\frac{1}{n\lambda_{3n}}\|\hat{\phi}^T_{M^c_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}}(\hat{\phi}^T_{M_{\theta^\star}}\hat{\phi}_{M_{\theta^\star}})^{-1}\|_\infty\{O(\sqrt{n\log n})\\ &\quad+O(s_{\theta^\star}n^{1-2\gamma_{\theta^\star}}\log^2 n)\\ &\qquad+n\lambda_{3n}\rho_3'(d_{n\theta})\}\\ &\quad+\frac{1}{n\lambda_{3n}}\|\hat{\phi}^T_{M^c_{\theta^\star}}\{I-P_{\phi M_{\theta^\star}}(\hat{\theta}_1)\}(\mu-\Phi)\|_\infty\\ &\quad=o(1)\\ &\quad+C\rho_3'(0+),\end{aligned}$$

for $C < 1$. Therefore (A.4) is satisfied.

Step 3. Now we show the second order conditions (A.5) and (A.6) hold. Because (A.6) is directly implied by (B.17), it suffices to show that $\lambda_{\min}(\hat{Z}^T_{M_\beta}\hat{Z}_{M_\beta}) \geq \lambda_{\min}(X^T_{M_\beta}\Delta X_{M_\beta})$ for sufficiently large n. Since $(\hat{Z}_{M_\beta} - Z_{M_\beta})^T(\hat{Z}_{M_\beta} - Z_{M_\beta})$ is positive semi-definite, we have

$$\lambda_{\min}(\hat{Z}^T_{M_\beta}\hat{Z}_{M_\beta}) \geq \lambda_{\min}(X^T_{M_\beta}\Delta X_{M_\beta})+\lambda_{\min}\{(\hat{Z}_{M_\beta}-Z_{M_\beta})^T Z_{M_\beta}+Z^T_{M_\beta}(\hat{Z}_{M_\beta}-Z_{M_\beta})+\xi_{5M_\beta}\}. \tag{A.45}$$

Since any symmetric matrix Ψ, the absolute value of minimum eigenvalue can be bounded by

$$|\lambda_{\min}(\Psi)| \leq \sqrt{\lambda_{\max}(\Psi^2)} \leq \sqrt{\|\Psi\|_\infty \|\Psi\|_1} = \|\Psi\|_\infty,$$

(A.5) follows if we can show $\|\xi_{5M_\beta} + (\hat{Z}_{M_\beta} - Z_{M_\beta})^T Z_{M_\beta} + Z_{M_\beta}^T (\hat{Z}_{M_\beta} - Z_{M_\beta})\|_\infty = o(n)$. But this is immediate to see because

$$\|\xi_{5M_\beta}\|_\infty = O(n^{1/2+\gamma_\beta} / \sqrt{\log n}) = o(n),$$

on the set ε_{11}. Similar to (A.20), $\|(\hat{Z}_{M_\beta} - Z_{M_\beta})^T Z_{M_\beta} + Z_{M_\beta}^T (\hat{Z}_{M_\beta} - Z_{M_\beta})\|_\infty$ can be bounded from above by

$$2 \max_j \sqrt{s_\beta \lambda_{\max}\{X_{M_\beta}^T \operatorname{diag}(|x^j|) X_{M_\beta}\} \lambda_{\max}\{X_{M_\alpha}^T \operatorname{diag}(|x^j|) X_{M_\alpha}\} \|\hat{\alpha}_{M_\beta} - \alpha_{0M_\alpha}\|_2^2}, \quad \text{(A.46)}$$

which is $O(\sqrt{s_\alpha s_\beta} n^{1-\gamma_\alpha} \log n) = o(n)$ implied by the constrain $\max(l_1, l_2) < \gamma_a$. This completes the proof.

References

Bunea F, Tsybakov A, Wegkamp M. Sparsity oracle inequalities for the Lasso. Electron J Stat. 2007; 1:169–194. MR2312149.

Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. Stat Methods Med Res. 2010; 19:317–343. MR2757118. [PubMed: 19608604]

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Amer Statist Assoc. 2001; 96:1348–1360. MR1946581 (2003k:62160).

Fan A, Lu W, Song R. Sequetial advantage selection for optimal treatment regime. Ann Appl Stat To appear. 2015 MR3480486.

Fan J, Lv J. Nonconcave penalized likelihood with NP-dimensionality. IEEE Trans Inform Theory. 2011; 57:5467–5484. MR2849368 (2012k:62211).

Fan J, Xue L, Zou H. Strong oracle optimality of folded concave penalized estimation. Ann Statist. 2014; 42:819–849. MR3210988.

Gunter L, Zhu J, Murphy SA. Variable selection for qualitative interactions. Stat Methodol. 2011; 8:42–55. MR2741508.

Li KC, Duan N. Regression analysis under link violation. Ann Statist. 1989; 17:1009–1052. MR1015136.

Lu W, Zhang HH, Zeng D. Variable selection for optimal treatment decision. Stat Methods Med Res. 2013; 22:493–504. MR3190671. [PubMed: 22116341]

Murphy SA. Optimal dynamic treatment regimes. J R Stat Soc Ser B Stat Methodol. 2003; 65:331–366. MR1983752 (2005b:62167).

Qian M, Murphy SA. Performance guarantees for individualized treatment rules. Ann Statist. 2011; 39:1180–1210. MR2816351 (2012e:62227).

Robins, JM. Proceedings of the Second Seattle Symposium in Biostatistics Lecture Notes in Statist. Vol. 179. Springer; New York: 2004. Optimal structural nested models for optimal sequential decisions; p. 189-326.MR2129402 (2006g:62007)

Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiol. 2000; 11:550–560.

Rubin DB. Estimating causal effects of treatments in randomized and non-randomized studies. J Edu Psychol. 1974; 66:688–701.

Shi C, Song R, Lu W. Supplement to “Robust Learning for Optimal Treatment Decision with NP-Dimensionality”. 2016; doi: 10.1214/16-EJS1178SUPP

Song R, Kosorok M, Zeng D, Zhao Y, Laber E, Yuan M. On sparse representation for optimal individualized treatment selection with penalized outcome weighted learning. Stat. 2015; 4:59–68. MR3405390. [PubMed: 25883393]

Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc Ser B Stat Methodol. 1996; 73:273–282. MR2815776 (2012e:62246).

Tsiatis, AA. Semiparametric theory and missing data Springer Series in Statistics. Springer; New York: 2006. MR2233926 (2007g:62009)

Wang L, Kim Y, Li R. Calibrating nonconvex penalized regression in ultra-high dimension. Ann Statist. 2013; 41:2505–2536. MR3127873.

Watkins CJCH, Dayan P. Q-learning. Mach Learn. 1992; 8:279–292.

White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982; 50:1–25. MR0640163.

Zhang CH. Nearly unbiased variable selection under minimax concave penalty. Ann Statist. 2010; 38:894–942. MR2604701 (2011d:62211).

Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. Biometrics. 2012; 68:1010–1018. MR3040007. [PubMed: 22550953]

Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. J Amer Statist Assoc. 2012; 107:1106–1118. MR3010898.

Table 1

Simulation results for L_2 loss, FN, FP, PCD and values

	Measures	n	Model I	Model II	Model III
Robust learning with covariates i.i.d normal	L_2 loss	300	0.276	1.743	1.171
		500	0.189	1.453	0.700
	FP	300	5.104	9.148	12.481
		500	4.143	9.742	12.616
	FN	300	0.000	0.893	0.125
		500	0.000	0.471	0.002
	PCD	300	0.975	0.734	0.834
		500	0.983	0.789	0.904
	$E\,Y^*(\hat{d})$	300	1.842(0.021)	4.544(0.157)	2.716(0.089)
		500	1.845(0.019)	4.643(0.116)	2.797(0.048)
	$E\,Y^*(d^{opt})$		1.847	4.846	2.847
Penalized Q-learning with covariates i.i.d normal	L_2 loss	300	0.080	4.861	1.729
		500	0.061	4.928	1.833
	FP	300	0.001	8.191	7.745
		500	0.000	4.438	7.972
	FN	300	0.000	0.050	0.757
		500	0.000	0.006	0.553
	PCD	300	0.993	0.550	0.714
		500	0.994	0.538	0.690
	$E\,Y^*(\hat{d})$	300	1.846(0.021)	4.117(0.165)	2.508(0.192)
		500	1.846(0.020)	4.091(0.093)	2.457(0.204)
	$E\,Y^*(d^{opt})$		1.847	4.846	2.847
Robust learning with covariates i.i.d exponential	L_2 loss	300	0.290	1.768	1.186
		500	0.199	1.495	0.730
	FP	300	6.596	9.700	13.240
		500	4.972	10.512	13.932
	FN	300	0.000	0.793	0.142

Electron J Stat. Author manuscript; available in PMC 2017 August 04.

	Measures	n	Model I	Model II	Model III
		500	0.000	0.466	0.003
	PCD	300	0.958	0.724	0.809
		500	0.971	0.761	0.871
	E $Y^*(\hat{d})$	300	1.744(0.018)	4.500(0.179)	2.670(0.095)
		500	1.747(0.018)	4.562(0.161)	2.736(0.041)
	E Y^* (d^{pn})		1.751	4.751	2.783
Penalized Q-learning with covariates i.i.d exponential	L_2 loss	300	0.264	2.580	2.225
		500	0.121	3.236	2.408
	FP	300	0.003	12.257	13.234
		500	0.000	21.383	15.479
	FN	300	0.045	0.824	0.288
		500	0.005	0.377	0.072
	PCD	300	0.954	0.610	0.609
		500	0.978	0.595	0.584
	E $Y^*(\tilde{d})$	300	1.744(0.018)	4.500(0.179)	2.670(0.095)
		500	1.743(0.037)	4.197(0.289)	2.201(0.224)
	E $Y^*(d^{pn})$		1.751	4.751	2.783

Table 2

Estimated value functions and confidence intervals for the difference of the estimated values

Treatment regime	Estimated value function	Diff	95% CI on Diff
Estimated optimal regime	3.10		
BUP	2.55	0.55	[0.07, 1.13]
SER	2.80	0.30	[−0.08, 0.64]

Table 3

Number of patients receiving BUP or SER, according to the estimated optimal treatment regime

	receives BUP	receives SER	total
assigns BUP	66	50	116
assigns SER	93	110	203
total	153	160	319