Submitted to the Annals of Applied Statistics
arXiv: arXiv:0000.0000

VARIABLE SELECTION FOR ESTIMATING THE OPTIMAL TREATMENT REGIMES IN THE PRESENCE OF A LARGE NUMBER OF COVARIATES

By Baqun Zhang[†], and Min Zhang[*,‡]

Shanghai University of Finance and Economics [†] *and University of Michigan*[‡]

Most of existing methods for optimal treatment regimes, with few exceptions, focus on estimation and are not designed for variable selection with the objective of optimizing treatment decisions. In clinical trials and observational studies, often numerous baseline variables are collected and variable selection is essential for deriving reliable optimal treatment regimes. Although many variable selection methods exist, they mostly focus on selecting variables that are important for prediction (predictive variables) instead of variables that have a qualitative interaction with treatment (prescriptive variables) and hence are important for making treatment decisions. We propose a variable selection method within a general classification framework to select prescriptive variables and estimate the optimal treatment regime simultaneously. In this framework, an optimal treatment regime is equivalently defined as the one that minimizes a weighted misclassification error rate and the proposed method forward sequentially select prescriptive variables by minimizing this weighted misclassification error. A main advantage of this method is that it specifically targets selection of prescriptive variables and in the meantime is able to exploit predictive variables to improve performance. The method can be applied to both single- and multiple- decision point setting. The performance of the proposed method is evaluated by simulation studies and application to an clinical trial.

1. Introduction. Personalized medicine that explicitly recognizes individual heterogeneity in response to treatments and focuses on making treatment decisions for a patient based on his/her own characteristics (eg., demographic, clinical, genetic information, etc.) has received much attention recently. The idea of personalized medicine can be formalized using the concept of treatment regimes, which are one or a sequence of decision rules that specify which treatment (among available options) a given subject receives based on a subject's characteristics at the time of the decision. In the last decade, there has been increasing interest and more vigorous research

[*] To whom correspondence should be addressed to

Keywords and phrases: C-learning, Classification, High-dimensional data, Misclassification error, Personalized medicine, Dynamic treatment regime, Variable selection

on developing methodologies for estimating the optimal treatment regimes (Murphy, 2003; Robins, 2004; Moodie, et al., 2007; Robins, et al., 2008; Brinkley, et al., 2009; Qian and Murphy, 2011; Chakraborty et al., 2010; Zhang et al., 2012ab, 2013; Zhao et al., 2012 and 2015; Geng et al., 2015; Barrett et al., 2014; Young et al., 2011; Tian, et al., 2014).

Most of existing methods, with few exceptions, focus on estimation and are not designed for selecting important variables from among a large number of covariates for optimizing treatment decisions. Clinical trials and observational studies (e.g., clinical registries), on which estimation of the optimal treatment regime is based, often collect a large amount of potentially useful patient information. Although it is likely that many of those variables are useful for predicting outcomes, realistically perhaps only a small number of patient characteristics are useful in making treatment decisions since only those variables with a qualitative interaction with treatment are useful in making treatment decisions. The importance of variables that have qualitative interactions with treatments in medical decision-making setting has been noted previously (Peto, 1982) and are referred to as prescriptive variables. In the presence of high-dimensional set of covariates, many existing methods developed for estimating the optimal treatment regime may lead to unnecessarily complicated decision rules that are of little practical use. Often times these methods may even fail to work due to the difficulty of handling high dimensional set of covariates. Therefore, variable selection from a high dimensional set of covariates targeted towards optimal decision making is an essential step in constructing a meaningful and practically useful treatment decision rule.

Variable selection has been an active research area in statistics; however, as pointed out by Gunter, Zhu and Murphy (2011), current variable selection work has been focused on prediction and their use in decision making has not been well developed and tested. As a matter of fact, variable selection approaches focused on prediction may neglect variables vital for decision making since the effect of interactions is often weaker than that of the main effect. Fairly recent literature started to see more research on variable selection methods for making treatment decisions (Gunter, Zhu and Murphy, 2011; Qian and Murphy, 2011; Lu, Zhang and Zeng, 2013; Fan, Lu and Song, 2015). Penalized least squares methods were proposed in the framework of Q-learning by Qian and Murphy (2011) and in the framework of A-learning by Lu, Zhang and Zeng (2013) to select important variables in the corresponding outcome regression, leading to estimated regimes with fewer variables. However, the variable selection in the two penalized methods are not directly targeted towards selecting prescriptive variables. Gunter, Zhu and

Murphy (2011) proposed a variable selection ranking method for variable selection, where the ranking is based a measure that specifically characterizes the qualitative interaction of a variable with treatment. As a result, the method of Gunter, Zhu and Murphy (2011) focuses specifically on selection of prescriptive variables. As noted by Biernot and Moodie (2010) and Fan, Lu and Song (2015), since the ranking method considers each variable separately and ignores correlations between covariates, they may identify too many covariates as potential prescriptive variables or miss some variables important for decision making. Building upon the work of Gunter, Zhu and Murphy (2011), Fan, Lu and Song (2015) proposed a sequential advantage selection method which takes into account variables already selected in previous steps and assesses the additional value of a new variable instead of considering variables individually. One advantage of this method is that it avoids selecting variables that are marginally important for decision making but are jointly unimportant. Although not directly focused on estimating optimal treatment regimes, one other recent relevant work is that of Tian, et al, (2014), which considers estimating interactions between treatment and a large number of covariates.

In this paper, we propose a new method to select important prescriptive variables for estimating the optimal treatment regimes in a classification framework. The proposed method is motivated by directly targeting and optimizing the objective function of the optimal treatment regime, ie., the expectation of potential outcomes under the optimal regime if it is followed by the entire population. This is in contrast with existing methods discussed above, which select/rank variables by focusing on studying models for outcomes or interaction terms. We show that optimizing the objective function is equivalent to minimizing a function that can be interpreted as weighted misclassification error rate for classifying patients to classes corresponding to their optimal treatment. The weighted misclassfication error is defined in terms of contrast between treatments given covariates and hence directly targets on selection of variables with qualitative interaction with treatments. Our proposed algorithm is then based on forward sequentially minimizing the weighted misclassification error rate and, as Fan, Lu and Song (2015), in each step it takes into account previously selected variables. The performance and merit of the proposed method relative to the sequential advantage selection method of Fan, Lu and Song (2015) are evaluated by various simulation studies.

The remainder of the paper is organized as follows. In Section 2, we describe the framework and objective function for variable selection for optimal treatment regimes and propose a forward minimal misclassification error selection algorithm that selects variables important for decision making. We evaluate the performance of the proposed method by simulations studies in Section 3 and illustrate the use of the proposed method using data from the Nefazodone CBASP trial in Section 4, followed by a discussion in Section 5.

2. Method.

2.1. *Notation and Assumptions.* We first focus on presenting the method in the simpler setting where only a single treatment decision point is involved; Section 2.4 considers the extension to multiple decision point setting. Consider a clinical trial or observational study involving n subjects, who receive either treatment $A = 0$ or $A = 1$. Let X be a p-dimensional vector of subject characteristics collected before the treatment. Let Y denote the observed outcome of interest and, without loss of generality, assume that larger values of Y are preferred. The observed data are then (X_i, A_i, Y_i), $i = 1, \ldots, n$, which are assumed to be independent and identically distributed (i.i.d.) across i. The goal is to use the data to find the optimal treatment decision rule which determines which treatment a patient should receive based on his/her baseline characteristics.

Formally, a treatment regime or rule, g, is a function which maps the values of X to the domain of A, eg, $\mathcal{A} = \{0, 1\}$. Let $Y^*(0)$ and $Y^*(1)$ denote the potential outcomes for a subject that would be observed had the subject received treatment 0 or 1, respectively. Then for each treatment regime g, there is a corresponding potential outcome, which is defined as $Y^*(g) = Y^*(1)g(X) + Y^*(0)\{1 - g(X)\}$. The expectation of potential outcomes if the entire population had followed regime g, $E\{Y^*(g)\}$, is referred to as the value of regime g and the optimal treatment regime g^{opt} is the one that leads to the optimal value, i.e., $g^{opt} = \arg\max_{g \in \mathcal{G}} E\{Y^*(g)\}$, where $\mathcal{G}$ is the class of all regimes under consideration. We make the commonly assumed stable unit treatment value assumption (SUTVA), which states that the observed outcome is the same as the potential outcome under the treatment actually received; ie., $Y = Y^*(1)A + Y^*(0)(1 - A)$. This assumption allows identification of the optimal treatment regime based on the observed data. We also assume the standard no unmeasured confounders assumption, ie., $\{Y^*(0), Y^*(1)\} \perp\!\!\!\perp A | X$, where $\perp\!\!\!\perp$ denotes statistical independence. This assumption holds automatically for clinical trials and has to be evaluated for observational studies. See a review paper by Schulte, et al. (2014) for more background on potential outcomes and optimal treatment regimes. As most existing methods on estimating optimal treatment regimes, in this article we focus on considering regimes that are of a linear form, i.e., for any $g \in \mathcal{G}$, $g(X) = I(\beta^T X > 0)$ for some β.

2.2. *Framework for Variable Selection for Optimal Treatment Regimes.* In terms of estimating the optimal treatment regimes, approaches can be categorized into two broad classes: outcome regression-based methods (Q- and A-learning) and direct optimization methods (e.g., Zhao, et al., 2012, 2015, Zhang, et al. 2012ab, Zhang, et al. 2013). In outcome regression-based methods, one aims to build good (parametric or semiparametric) regression models for outcomes given covariates and treatments and then optimal treatment regimes are estimated by inverting the relationship. Specifically, in Q-learning, one postulates models for $\mu(A, X) \equiv E(Y|A, X)$ by some parametric model and in A-learning one models $\mu(A, X)$ by some semiparametric model where only the treatment contrast, i.e., $C(X) = \mu(1, X) - \mu(0, X)$, is modeled parametrically leaving the other part unspecified (e.g., Watkins and Dayan, 1992; Murphy, 2003). These methods work well if the posited regression models are correctly specified and if they are misspecified the estimated regime may far from optimal. This is due to that, as firstly pointed out by Murphy (2005), there is a mismatch between the target of outcome regression-based methods and the goal of learning the optimal treatment regime. That is, outcome regression-based methods targets good models for the outcome instead of optimizing decision rules to yield the maximum expected potential outcomes. More recently direct optimization approaches are developed to mitigate the concern of outcome model misspecification and they aim to directly maximize estimates of $E\{Y^*(g)\}$ across a class of regimes, consistent with the definition. If $E\{Y^*(g)\}$ can be robustly and efficiently estimated then the estimated regimes from direct optimization methods will have good property. By viewing the problem as a missing data problem, that is, $Y^*(g)$ is observed if the observed treatment is the same as the one that is dictated by regime g, $E\{Y^*(g)\}$ can be more robustly estimated by the (augmented) inverse probability weighted estimator (AIPWE, or IPWE). In IPWE one models treatment probability and in AIPWE one additionally incorporates outcome regression models through augmentation terms to further improve efficiency and robustness. It is well-known in the missing data and causal inference literature that, AIPWE/IPWE are always consistent if treatments are randomized and, in observational studies, AIPWE is consistent if either treatment or outcome regression models, but not necessarily both, are correctly specified, referred to as the double-robustness property (e.g., Bang and Robins, 2005). Therefore, direct optimization methods that directly targeting optimizing $E\{Y^*(g)\}$ can be more robust. The advantages of direct optimization methods are discussed in detail in Zhao, et al. (2012, 2015), Zhang, et al. (2012ab, 2013) and Kang et al.(2014) and discussion papers.

Variable selection methods are proposed within the framework of outcome regression-based methods (Qian and Murphy, 2011; Lu, Zhang and Zeng,2013). As discussed in Section 1, these methods are not directly targeted towards selecting prescriptive variables. Moreover, as they are developed within the framework of outcome regression-based methods, they suffer from the mismatch problem as well.

Following the principle of direct optimization, naturally one should also aim to maximize $E\{Y^*(g)\}$ in selecting prescriptive variables for constructing the optimal treatment regime. However, to the best of our knowledge, none of existing prescriptive variable selection methods directly target this objective function. In this article, we propose a direct optimization method for selecting variables useful for making treatment decisions and as a result this method does not suffer from the mismatch problem of outcome-regression based methods and enjoys advantages of direct optimization methods for estimating the optimal treatment regimes discussed above. Specifically we adapt the classification framework for estimating optimal treatment regimes developed by Zhang, et al. (2012b); we refer to this general classification framework as C-learning with "C" stands for "classification." Within the C-learning framework we propose a method for simultaneously estimating the optimal treatment regime and selecting prescriptive variable by explicitly optimizing $E\{Y^*(g)\}$, where optimization of $E\{Y^*(g)\}$ is achieved by equivalently minimizing an objective function that can be intuitively interpreted as a weighted misclassification error rate. We provide a brief review of the C-learning framework below.

Intuitively, it is easy to see that we can view subjects as coming from two latent classes corresponding to $g^{opt}(X) = 0$ or 1; i.e., the class corresponding to $g^{opt}(X) = a$ includes all subjects whose expected potential outcome under treatment a is greater than that under treatment $1-a$. Equivalently, only the contrast between treatments is relevant in determining which class a subject belongs to as well as whether a variable is prescriptive or not. Denoting $\mu(a, X) = E(Y|A = a, X)$, it is easy to see that

$$E\{Y^*(g)\} = E_X\{C(X)g(X) + \mu(0, X)\}, \tag{1}$$

where $C(X) = \mu(1, X) - \mu(0, X)$, denoting the contrast between treatment 1 and 0. According to (1), it is clear that the optimal treatment regime corresponds to $I\{C(X) > 0\}$. Zhang, et al. (2012b) show that maximizing (1) is equivalent to

$$g^{opt}(X) = \arg\min_{g\in\mathcal{G}} E[|C(X)|I\{Z \neq g(X)\}]. \tag{2}$$

This alternative definition corresponds exactly to the intuition described above. That is, we can view each subject as belonging to one of two latent classes with the class label denoted by $Z = I\{C(X) > 0\}$. We can then view $E[|C(X)|I\{Z \neq g(X)\}]$ as a weighted misclassification error rate corresponding to treatment regime (or classifier) $g(X)$. That is, if $g(X) \neq Z$, then an error is made since the treatment decision according to $g(X)$ is not optimal and the loss corresponding to this error is $|C(X)|$, the difference in expected outcomes between $g(X)$ and $g^{opt}(X)$. In practice, the class label Z_i as well as the weight $|C(X_i)|$ is unknown and Zhang, et al. (2012b) discussed various ways to estimate $C(X_i)$ and then Z_i. Zhang, et al. (2012b) recommend estimating $C(X_i)$ by the more robust AIPWE estimator defined as

(3)

$$\widehat{C}_{AIPWE}(X_i) = \frac{A_i}{\widehat{\pi}_i}Y_i - \frac{A_i - \widehat{\pi}_i}{\widehat{\pi}_i}\widehat{\mu}(1, X_i) - \left\{\frac{1 - A_i}{1 - \widehat{\pi}_i}Y_i - \frac{\widehat{\pi}_i - A_i}{1 - \widehat{\pi}_i}\widehat{\mu}(0, X_i)\right\},$$

where $\widehat{\pi}_i$ estimates $\pi_i = P(A = 1|X_i)$ and is simply the sample proportion corresponding to $A = 1$ for a randomized clinical trial, and terms involving $\widehat{\mu}(a, X_i)$ are referred to as augmentation terms. One may also postulate parametric models respectively for $\mu(a, X)$, $a = 0, 1$, and estimate $C(X_i)$ by $\widehat{C}_{reg}(X_i) \equiv \widehat{\mu}(1, X_i) - \widehat{\mu}(0, X_i)$, where $\widehat{\mu}(a, X)$ estimates $\mu(a, X)$ based on the fitted model. Denoting an estimate of $C(X_i)$ by $\widehat{C}(X_i)$ and $\widehat{Z}_i = I\{\widehat{C}(X_i) > 0\}$, then $g^{opt}(X)$ can be estimated by $\arg\min_{g \in \mathcal{G}} \sum_{i=1}^{n} |\widehat{C}(X_i)|I\{\widehat{Z}_i \neq g(X_i)\}$.

It may seem that the definition (2) and the resulting C-learning are unnecessarily complicated since by (1) one can directly estimate $g^{opt}(X)$ by $I\{\widehat{C}(X) > 0\}$, whereas the C-learning involves an additional step of optimization after obtaining $\widehat{C}(X)$. As a matter of fact, decoupling the optimization step from the step for building good outcome models has several advantages and is the key to variable selection for prescriptive variables. Among then, one advantage is that, as opposed to outcome-regression based methods, it leads to a direct optimization method that optimizes estimate of $E\{Y^*(g)\}$ or $E[|C(X)|I\{Z \neq g(X)\}]$ explicitly, which can lead to more robust estimation of the optimal treatment regime by using robust estimator of $C(X)$. As explained in Zhang, et al. (2012b), some algebra can show that, as far as estimating the optimal treatment regime is concerned, estimating $C(X)$ by $\widehat{C}_{AIPWE}(X)$ is equivalent to estimating $E\{Y^*(g)\}$ by the well-known doubly-robust AIPWE estimator, which is consistent if either the model for A or the model for Y used in augmentation terms, but not necessarily both, is correctly specified. In a randomized study, as one can always model treatment correctly, regardless of whether the postulated model for $\mu(a, X)$ is correct or not, AIPWE always consistently estimates $E\{Y^*(g)\}$

and leads to robust estimation of the optimal treatment regime. Therefore, as opposed to outcome regression-based methods, the performance of the estimated treatment regimes is not completely dictated by outcome regression models used for estimating $\mu(a, X)$ and is more robust to model misspecification. See Zhang, et al., (2012a and b) and Zhang and Zhang (2015) for detailed discussions on AIPWE-based direct optimization method and the classification framework. We also comment that when $\widehat{C}_{AIPWE}(X_i)$ is used then the optimization step within $\mathcal{G}$ is necessary to obtain a valid regime because $\widehat{C}_{AIPWE}(X_i)$ still depends on Y_i and is not a function of covariates only. When $\widehat{C}_{reg}(X_i)$ is used and $\mu(a, X_i)$, $a = 0, 1$, are modeled using parametric models with implied regimes in the class of $\mathcal{G}$, then $I\{\widehat{C}_{reg}(X) > 0\}$ is a valid treatment regime in $\mathcal{G}$. However, postulating models for $\mu(a, X_i)$, $a = 0, 1$, separately is equivalent to including interaction of treatment with all available covariates, which leads to unnecessarily overcomplicated decision rules. In this case, an additional optimization step is still important for prescriptive variable selection as we describe below. Equally important, (2) provides a natural objective function for prescriptive variable selection for estimating the optimal treatment regime since it only depends on the contrast function, the part relevant for optimizing treatment decisions. See also discussion at the end of Section 2.3 for advantages of decoupling optimization step from the outcome model building step.

2.3. *Forward Minimal Misclassification Error Rate (ForMMER) Selection.* In C-learning, the estimated weighted misclassification error rate corresponding to regime $g(X)$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}[\widehat{W}_i I\{\widehat{Z}_i \neq g(X_i)\}], \tag{4}$$

where $\widehat{Z}_i = I\{\widehat{C}(X_i) > 0\}$, $\widehat{W}_i = |\widehat{C}(X_i)|$ and $\widehat{C}(X_i)$ is an estimate of $C(X_i)$. This weighted misclassification error rate is for a given regime $g(X) \equiv g(X_1, \ldots, X_p)$. Now we use this to define a measure that is helpful for quantifying the importance of a potential prescriptive variable given a set of already selected prescriptive variables. For that, we define the weighted misclassification error rate corresponding to a set of variables $\{X_{j^1}, ..., X_{j^m}\}$ as

$$\begin{aligned} & err(X_{j^1}, ..., X_{j^m}) \\ & = \min_{\beta=\{\beta_0,...,\beta_m\}} n^{-1}\sum_{i=1}^{n}\widehat{W}_i I\{\widehat{Z}_i \neq I(\beta_0 + \beta_1 X_{j^1} + ... + \beta_m X_{j^m} > 0)\}, \end{aligned}$$

which can be interpreted as the minimum weighted misclassification error rate among a subclass of regimes that are constructed by linear combinations of the set of variables $(X_{j^1}, ..., X_{j^m})$. Then naturally we can quantity the importance of a potential prescriptive variable, say X_j, given a set of selected prescriptive variables $\{X_{j^1}, ..., X_{j^m}\}$ by the difference in misclassification error, i.e., $err(X_{j^1}, ..., X_{j^m}) - err(X_{j^1}, ..., X_{j^m}, X_j)$.

Based on the idea described above, we propose the following forward minimal misclassification error rate (ForMMER) selection algorithm to sequentially select variables that are important for treatment decision making. Our algorithm starts with an empty set corresponding to the case where there is no prescriptive variable and hence the optimal treatment is a fixed treatment for everyone.

Step 1 (**Initial Step**). Let

$$err^{(0)} \equiv err(\text{null set}) = \min\{n^{-1}\sum_{i=1}^{n}\widehat{W}_i I(\widehat{Z}_i \neq 1), n^{-1}\sum_{i=1}^{n}\widehat{W}_i I(\widehat{Z}_i \neq 0)\},$$

where the equality is due to that there are only two treatment regimes ($a = 0, 1$) when the set of covariates under consideration is null. Here $err^{(0)}$ is the weighted misclassification error rate by assigning the treatment with better average treatment effect to all patients regardless of their characteristics. We term this as the baseline weighted misclassification error rate and use it as a reference in the criterion for the initial selection that selects the first important prescriptive variable.

When the number of candidate variables is huge, it is preferable to initially screen variables for consideration in subsequent steps. For each X_j, let $err(X_j) = \min_{\beta=\{\beta_0,\beta_1\}} n^{-1}\sum_{i=1}^{n}\widehat{W}_i I\{\widehat{Z}_i \neq I(\beta_0+\beta_1 X_j > 0)\}$ and calculate

$$err^{(0)} - err(X_j), j = 1, \ldots, p.$$

As explained above, this difference characterizes the degree of reduction in weighted misclassification error rate under the optimal treatment regime within a subclass of regimes based on variable X_j, relative to the optimal treatment regime based on a null set of covariates. Therefore, the ranking (from the largest to the lowest) of $err^{(0)} - err(X_j), j = 1, \ldots p$, quantifies the relative importance of variables in treatment decision making. We propose to use the ranking of $err^{(0)} - err(X_j)$ to initially select the set of covariates considered in subsequent steps when the dimension of covariates is high. For example, one may consider the first 30 or 40 variables based on the ranking of $err^{(0)} - err(X_j)$ and include them into a set $\mathcal{F}^0$, which include variables considered in subsequent steps. In addition, one may add into

$\mathcal{F}^0$ any variables that are thought to be potentially important in making treatment decisions based on clinical/scientific reasons or on evidence from empirical data. The purpose of this step is only to screen variables when the number of candidate variables is very large to make it computationally easier; for example, in our simulations we considered 1000 covariates. When the dimension is not super large, for example, in our real data analysis there are only 50 baseline covariates, then the screening step can be omitted, i.e., all candidate variables are included in $\mathcal{F}^0$.

We comment that, even with the screen, the proposed variable selection method is still selecting prescriptive variables from the entire p-dimensional covariates instead of only selecting among variables in $\mathcal{F}^0$. The reason is that in the first step in selecting X_{j^1} (as is clear from the formula below), it is selecting among all $(X_1, \dots, X_p)$ using a principled approach based on importance of variables. Our simulations demonstrate that it works well in practice even with a large number of candidate variables.

Step 2 (**Forward Selection**). Let

$$X_{j^1} = \arg \min_{X_j \in (X_1, \dots, X_p)} err(X_j), \text{ and } err^{(1)} \equiv \min_{X_j \in (X_1, \dots, X_p)} err(X_j).$$

Then X_{j^1} is the first selected variable and $\mathcal{S}^{(1)} = \{X_{j^1}\}$, denoting the set of selected variables from step (1). We note that the selected X_{j^1} is always included in $\mathcal{F}^0$ as $\mathcal{F}^0$ includes variables ranked high based on $err^{(0)} - err(X_j)$ and hence X_{j^1}.

In the m-th step ($m > 1$), we have $\mathcal{S}^{(m-1)} = \{X_{j^1}, ..., X_{j^{m-1}}\}$, which denotes the set of selected variables in steps prior to the m-th step. For every $X_j \in \mathcal{F}^0 \backslash \mathcal{S}^{(m-1)}$, we compute each $err(\mathcal{S}^{(m-1)}, X_j)$, which is the minimum weighted misclassification error rate for regimes constructed using variables in $\mathcal{S}^{(m-1)}$ and X_j. The m-th variable to be selected is the one with the smallest weighted misclassification error rate in this step, i.e.

$$X_{j^m} = \arg \min_{X_j \in \mathcal{F}^0 \backslash \mathcal{S}^{(m-1)}} err(\mathcal{S}^{(m-1)}, X_j).$$

We update the set of selected variables, i.e., $\mathcal{S}^{(m)} = \mathcal{S}^{(m-1)} \cup \{X_{j^m}\}$. weighted misclassification error rate corresponding to the optimal treatment regime among regimes that are based on the m variables in $\mathcal{S}^{(m)}$ is also updated accordingly as follows,

$$err^{(m)} \equiv \min_{X_j \in \mathcal{F}^0 \backslash \mathcal{S}^{(m-1)}} err(\mathcal{S}^{(m-1)}, X_j).$$

Step 3 (**Stopping Criterion**). Continue forward selection until $prop^{(m)} \leq \alpha$, where α is a cut-off point and

$$prop^{(m)} = \frac{err^{(m-1)} - err^{(m)}}{err^{(m-1)}}.$$

Several ways can be used to choose the tuning parameter α. The simplest way is to pre-specify α, say at 0.05, which says that the algorithm would stop when the reduction in error rate is less than 5%. Alternatively and more rigorously, as in many other statistical methods involving tuning parameters, we may choose the cut-off point using, for example, five- or ten-fold, cross-validation, where the final α is chosen as the one that leads to estimated regimes with the smallest value of (4) overall on validation data sets. Similar to scree plot used in selecting the number of principal components to be included in principal component analysis, one other alternative is to make a plot of err^m against m and identify the point where error rate starts to level off (the so-called "elbow"). It can be argued that, when values of decision rules are similar, rules with less number of variables and rules with variables that are easy to measure should be preferred from a practical point of view. Then when to stop the algorithm can even be determined subjectively by balancing consideration of the reduction in err^m, the number of additional variables to be selected, the cost of collecting variables and other factors based on clinical/subject matter knowledge.

In our simulations as well as the data analysis, the optimization is implemented using a genetic algorithm discussed by Goldberg (1989), implemented in the `rgenoud` package in R (Mebane and Sekhon, 2011). As well as the sequential advantage selection (SAS) method of Fan, Lu and Song (2015), the proposed ForMMER algorithm sequentially select potential prescriptive variables by assessing the added advantage of a new variable relative to existing ones, in contrast to the S-score based method of Gunter, Zhu and Murphy (2011) that considers each variable individually. Therefore, ForMMER enjoys the same advantage as SAS, namely, it tends not to select those unnecessary variables that are only marginally important but not important given other variables. One key difference between ForMMER and SAS lies in the function used in quantifying the advantage. In SAS, the sequential advantage of a variable, say X_j, given a set of selected prescriptive variables $S^{m-1} = \{X_{j^1}, ..., X_{j^{m-1}}\}$, is given by

$$\frac{1}{n}\sum_{i=1}^{n}\{\max_a \widehat{E}(Y_i|S_i^{m-1}, X_{ij}, A_i = a) - \widehat{E}(Y_i|S_i^{m-1}, X_{ij}, A_i = a_{opt}(S_i^{m-1}))\}, \tag{5}$$

where $a_{opt}(S_i^{m-1})$ is the optimal decision based on variables in S_i^{m-1}. This sequential advantage extends the S-score of Gunter, Zhu and Murphy (2011) in that the second term of the sequential advantage additionally conditions on S^{k-1}, whereas the second term of S-score conditions on X_j only. Subject i contributes to the sequential advantage only if the optimal decision based on (S_i^{m-1}, X_{ij}) is different from the optimal decision based on S_i^{m-1} and hence the sequential advantage quantifies the importance of X_j in addition to S^{m-1} for decision making. At each step in the sequential advantage selection, it fits a model conditional on potential prescriptive variables selected in previous steps and a new variable. As a result of the SAS algorithm, variables with only main effect but no qualitative interactive effect tend to be not selected, which of course is an intended property; however, since at each step it builds conditional models conditional on only variables selected in previous steps, those only predictive but not prescriptive variables will not be able to be used in the outcome-regression models in subsequent steps, which is clearly not desirable as it misses the chance of exploiting those predictive variables to improve performance. In the proposed ForMMER method, the forward selection algorithm for selecting prescriptive variables are separated from estimation of the contrast function. In principle any model selection methods developed for prediction can be used to best model the outcome given covariates including those predictive but not prescriptive variables in the estimation of the contrast function. Therefore, ForMMER is able to exploit predictive variables for improving efficiency in the outcome-regression step or equivalently the contrast function estimation step. In the meantime, the variable selection step focuses on selecting prescriptive variables, aiming towards minimizing a weighted misclassification error rate or equivalently maximizing the expected potential outcome of a regime. This difference explains the superior performance of ForMMER relative to SAS, especially when the number of predictive variables is large, as illustrated by our simulation studies.

2.4. *Extension to Multiple Decision Point Setting.* The ForMMER algorithm extends naturally to multi-stage treatment decision problems where decisions are made at K decision points and at each stage there are two treatment options (0 or 1). Suppose data are obtained from sequentially randomized clinical trials or observational studies where the no unmeasured confounders assumption holds. We denote the treatment received at stage k as A_k and the observed treatment history up to decision k as $\bar{A}_k = (A_1, \ldots, A_k)$. Let X_k be the covariate information observed between decision $k-1$ and k and $\bar{X}_k = (X_1, \ldots, X_k)$ be the observed covariate history up to k. The overall outcome of interest is still denoted by Y. A dynamic treatment regime is a set of sequential decision rules, $g = (g_1, \ldots, g_K)$, where g_k is a function of $\bar{x}_k$ and $\bar{a}_{k-1}$, denoted as $g_k(\bar{x}_k, \bar{a}_{k-1})$, that determines the treatment decision at stage k based on patient's covariate and treatment history available up to decision k. We denote $L_k = (\bar{X}_k, \bar{A}_k)$.

Similar to the single decision point setting, only the treatment contrast at each stage is relevant for treatment decision. Analogously we define a contrast function at each stage, ie., $C_k(L_k) = Q_k(L_k, a_k = 1) - Q_k(L_k, a_k = 0)$, where a_k is a treatment decision at stage k and $Q_k(L_k, a_k)$ is the so-called Q-functions with "Q" for "quality". At the last stage K, $Q_K(L_K, a_K) = E(Y|L_K, A_K = a_K)$. The Q-functions at stage $k < K$ are defined recursively and can be interpreted as the conditional expected outcomes given that the optimal decisions are made in the future. Therefore, the contrast function $C_k(L_k)$ represents the contrast in the quality between treatment 1 or 0 at stage k assuming the optimal decisions are made in the future. We refer readers to Schulte, et al. (2014) for more details.

Intuitively, at each stage, subjects can be viewed as coming from two latent classes for whom the optimal decision at stage k is 1 or 0 (or equivalently $Z_k \equiv I\{C_k(L_k) > 0\}$), assuming the optimal decisions are made in the future. Zhang and Zhang (2015) show that for multiple-stage decision problems, the optimal treatment regimes can be identified by backward sequentially minimizes $E[|C_k(L_k)|I\{Z_k \neq g_k(L_k)\}]$, which is the expected loss of misclassifying a patient at stage k by decision rule g_k. In practice, Z_k and $C_k(L_k)$ have to be estimated and, as in the single-decision point setting, various methods (eg, parametric regression method $\widehat{C}_{reg}$ and the AIPWE $\widehat{C}_{AIPWE}$) can be used to estimate the contrast functions as well as Z_k. Then the empirical analog of $E[|C_k(L_k)|I\{Z_k \neq g_k(L_k)\}]$ can be used as a objective function in estimating the optimal treatment regime at each stage. We refer readers to Zhang and Zhang (2015) for details on the C-learning framework for multiple-decision point setting and for discussions on the connection and distinction of C-learning with existing methods. The proposed ForMMER can naturally be embedded in the C-learning framework to select variables important for decision-making and estimate the optimal treatment regime at each stage. Specifically, one only needs to modify the objective function in (4) to $\frac{1}{n}\sum_{i=1}^{n}[\widehat{W}_{ki} I\{\widehat{Z}_{ki} \neq g_k(L_i)\}]$, where $\widehat{W}_{ki} = |C_k(L_{ki})|$, and modify $err^{(m)}$ accordingly. ForMMER can then be used to identify the linear decision rule that minimizes the weighted misclassification error rate at each stage.

3. Simulations. We conducted simulation studies to evaluate the performance of the proposed methods. Data generating scenarios are adopted from Fan, Lu and Song (2015) and additionally we considered two new scenarios. We consider both single-decision point and multiple-decision point settings. We compare our methods with SAS developed by Fan, Lu and Song (2015) since in their simulations they demonstrated that SAS has superior performance than the S-score method of Gunter, Zhu and Murphy (2011) and the method of Lu, Zhang and Zeng (2013) with LASSO selection.

For the single decision point setting, data were generated according to six scenarios, where scenarios I-IV are directly adopted from Fan, Lu and Song (2015). Specifically, Covariates $X = (X_1, \dots, X_p)^T$, $p = 1000$, are generated from multivariate normal distribution with mean zero, variance 1 and correlation $corr(X_j, X_k) = \rho^{|j-k|}$, where $\rho = 0.2$ or 0.8. Treatment A is generated from a Bernoulli distribution with probability 0.5 and the error term ϵ is normally distributed with mean 0 and variance 0.25. Defining $\tilde{X} = (1, X^T)^T$ and $\mathbf{0}_p$ as a p-dimensional vector with all zero elements, the outcomes are generated according to:

- Scenario I: $Y = 1 + \gamma_1^T X + A\beta^T \tilde{X} + \epsilon$ with $\gamma_1 = (1, -1, \mathbf{0}_{p-2})^T$, $\beta = (0.1, 1, \mathbf{0}_7, -0.9, 0.8, \mathbf{0}_{p-10})$;
- Scenario II: $Y = 1 + 0.5\sin(\pi\gamma_1^T X) + 0.25(1 + \gamma_2^T X)^2 + A\beta^T \tilde{X} + \epsilon$ with γ_1 and β the same as in scenario I and $\gamma_2 = (1, 0_2, -1, \mathbf{0}_5, 1, \mathbf{0}_{p-10})^T$;
- scenario III: $Y = 1 + \gamma_1^T X + A\beta^T \tilde{X} + \epsilon$ with γ_1 the same as in scenario I, and
  $\beta = (0.1, 1, \mathbf{0}_7, -0.9, 0.8, \mathbf{0}_{10}, 1, 0.8, -1, \mathbf{0}_5, 1, -0.8, \mathbf{0}_{p-30})$;
- Scenario IV: $Y = 1 + 0.5\sin(\pi\gamma_1^T X) + 0.25(1 + \gamma_2^T X)^2 + A\beta^T \tilde{X} + \epsilon$ with γ_1, and γ_2 the same as in scenario II, and β the same as in scenario III.
- Scenario V: $Y = 1 + \gamma_1^T X + A\beta^T \tilde{X} + \epsilon$ with $\gamma_1 = (1, -0.8, 1, 0.9, 0.8, 1, 0.9, 0.8, \mathbf{0}_{p-8})^T$, $\beta = (0.1, \mathbf{0}_8, 1, 0.8, \mathbf{0}_{p-10})$.
- Scenario VI: $Y = 2 + \gamma_1^T X - |\gamma_2^T \tilde{X}|\{A - I(\beta^T \tilde{X} > 0)\}^2 + \epsilon$ with $\gamma_1 = (1, -1, \mathbf{0}_{p-2})^T$, $\gamma_2 = (0.5, 1.5, -2, \mathbf{0}_{p-2})^T$ and $\beta = (0.1, \mathbf{0}_8, 1, 0.8, \mathbf{0}_{p-10})$.

Scenarios I and II have three prescriptive variables and scenarios III and IV have eight prescriptive variables. In scenarios I and III, the relationship between outcome and covariates are linear, whereas in scenarios II and IV, the relationship is nonlinear. In scenarios I-IV, the number of prescriptive variables is more than the number of predictive variables. However, in reality, it is perhaps more plausible or often believed that many covariates have a main effect but not a qualitative interaction effect with treatment. Considering this, we also generated data from scenario V, which is modified based on scenario I but more variables have a main effect and less variables have a qualitative interaction with treatment. Scenario VI also considers a scenario where treatment interacts with covariates nonlinearly but still the optimal treatment regime is of a linear form. Based on Scenario VI, the subgroup of subjects whose optimal treatment option is 1 is determined by $I(\beta^T\tilde{X} > 0)$ and the contrast between optimal treatment option and the other option is $|\gamma_2^T\tilde{X}|$. The sample size we considered in the single-decision point setting is $n = 200$ and 400.

As for the multiple-decision point setting, again we adopted the same data generating process as in Fan, Lu and Song (2015). Specifically, data are generated to mimic a two-stage decision problem, where the outcome Y is generated according to

$$Y = A_1A_2 + A_2(a + \beta_{12}^T X_1 + \beta_{21}^T X_2) + A_1(a + \beta_{11}^T X_1) + \epsilon,$$

where treatment A_1 and A_2 follow Bernoulli (0.5), baseline covariats $X_1 = (X_{1,1}, \ldots, X_{1,p_1})$ follow multivariate normal distribution with mean 0, variance 1 and $corr(X_{1,j}, X_{1,k}) = 0.2^{|j-k|}, j \neq k$, and ϵ follows normal distribution with mean 0 and variance 0.25. The intermediate covariate X_2 is generated according to $X_2 = c_0 + c_1X_{1,1} + c_2A_2 + C_3A_1X_{1,1} + e$ with e generated from normal with mean 0 and variance 0.25. The parameter values are chosen as: $\beta_{12} = (0, 0, 1, -1, \mathbf{0}_{p_1-4})$, $\beta_{11} = (\mathbf{0}_4, 1, -1, \mathbf{0}_{p_1-6})^T$, $a = 0$, and $c = (0, 1, 0, 0)^T$. Also we chose $p_1 = 500$.

We implemented the proposed methods using two ways. In ForMMER-reg, parametric regression models are used to model $\mu(a, X)$ and the contrast functions are estimated by $\widehat{C}_{reg}$. Specifically, forward selection based on AIC is used to build models for $\mu(0, X)$ and $\mu(1, X)$ respectively, where the maximum number of steps is set to be 10. In ForMMER-AIPWE, the contrast functions are estimated using $\widehat{C}_{AIPWE}$, where the same parametric models as in ForMMER-reg are used in the augmentation terms of the AIPWEs. We used the union of the variables selected in the model for $\mu(1, X)$ and the first 10 variables based on the ranking of $err^{(0)} - err(X_j)$ as the initially selected set to be considered in the ForMMER procedure. We implemented ForMMER by setting $\alpha = 0.02, 0.05$ and 0.08 and also we used ten-fold cross-validation to select α for scenario VI. In addition to SAS, we compare our method to the usual regression method where we model outcomes given treatment, covariates and treatment and covariates interaction and select variables to be included in the final model by forward selection based on AIC. In the usual regression method with forward selection, the final estimated regime is then $I\{\widehat{\mu}(1, X) - \widehat{\mu}(0, X) > 0\}$.

We evaluate the performance of each methods using three metrics. TP (true positive) is the number of correctly identified prescriptive variables. VR (value ratio) is the value ratio of the estimated regime relative to the true optimal regime, i.e., VR=$E\{Y^*(\widehat{g}^{opt})\}/E\{Y^*(g^{opt})\}$, where the value of a regime $E\{Y^*(g)\}$ is calculated by the average of outcomes generated from the true model with treatment determined by the regime using 100,000 Monte Carlo Replicates. ER (error rate) is the rate of incorrect treatment decision of the estimated regime, ie, an incorrect decision is made if the treatment decision determined by a regime is different from the correct optimal decision. Reported results are averages across 500 Monte Carlo simulations and the standard deviation are reported in parenthesis. Although all three metrics are useful in evaluating the performance of a method, from the perspective of optimizing expected potential outcomes, VR is in our view the most relevant one as it takes into account whether or not an correct decision is made and the magnitude of the consequence of an incorrect decision, whereas ER only accounts for whether the optimal decision is made ignoring the magnitude of loss associated with an incorrect decision. Although TP can provide us some useful information, we note that it cannot be used as a metric alone to evaluate methods since it does not account for the size of the selected variables and also it does not for the different importance of variables in terms of treatment decision making . In addition, the number (size) of prescriptive variables in the estimated regime is also reported, which is important for interpreting TP.

Main results are shown in Tables 1-3 and additional results are shown in supplementary materials. We report results on ForMMER with $\alpha = 0.05$ here and results on ForMMER with different α values are reported in supplementary materials, which show that for scenarios considered in this paper all three choices of α (0.02, 0.05, and 0.08) lead to comparable results. Results on ForMMER with α chosen by cross-validation is marked with a †.

Results on scenarios adopted from Fan, Lu and Song (2015), i.e., scenarios I-IV, for sample size $n = 200$ and $\rho = 0.2$ are summarized in Table 1; additional results on $n = 200$, $\rho = 0.8$ and other values of α are in Tables S1-S3 in the supplementary material. Table 1 shows that, under scenarios II, III and IV, ForMMER-AIPWE and ForMMER-reg are comparable and the proposed methods (both implementations) have better performances than SAS. Under scenario I the proposed methods and SAS have comparable performance with SAS being slightly better. We also note that the usual forward selection method works well in these scenarios and are even better than SAS.

Table 2 shows results under scenario V, which is similar to scenario I except for that there are more predictive variables and less prescriptive variables. Under this scenario, the proposed ForMMER (both implementations) has considerably better performances than SAS. This difference in performance is expected to be even larger under other scenarios (e.g., scenarios similar to II and IV) since scenario V is modified based on scenario I where SAS has relatively the best performance compared with other scenarios in Table 1. This result is consistent with and supports our conjecture that SAS may not perform well when many covariates are predictive but not prescriptive as explained at the end of Section 2.2. This is because in SAS predictive but not prescriptive variables will not be selected in previous steps and as a result cannot be used in the conditional models in subsequent steps to improve performance, even though predictive variables (regardless of being prescriptive or not) are useful for improving the performance of the models and estimation of optimal regimes. Our methods do not suffer from this issue and is able to take advantages of predictive variables to improve efficiency while still targeting selection of only prescriptive variables in the forward selection algorithm. This is achieved by decoupling the step for estimating the contrast function and the step for variable selection in the optimization step. This difference also explains the better performance of ForMMER in Table 1.

We also note under scenarios I-V, scenarios either directly adopted from or modified based upon Fan, Lu and Song (2015), not only the proposed methods have overall better performance than SAS, the usual forward selection method has also better performance than SAS and in addition ForMMER-reg has comparable but slightly better performance than the more robust ForMMER-AIPWE. We think this is due to the particular data generating process. For example, in these scenarios even though nonlinear models are considered but linear models can provide a good approximation and treatment only interacts linearly with covariates. Under scenarios VI, ForMMER-AIPWE considerably outperforms SAS, the usual forward selection method and in addition, ForMMER-AIPWE has better performance than ForMMER-reg. For example, when $n = 200$ and $\rho = 0.2$, relative to SAS, ForMMER-AIPWE with α set at 0.05 decreases the error rate of treatment decision from 32.7% to 7.2% and increases the value from 70.0 to 93.4. Yet the decision rules from ForMMER-AIPWE are much simpler, involving about 3 variables as opposed to about 13 variables from SAS. As expected, performance of ForMMER are further improved when the tuning parameter α is selected by cross-validation as shown in Table 2, scenarios VI.

Results on the two-stage setting are shown in Table 3. We report on the size of selected prescriptive varaibles, true positive (TP), and error rate (ER) of treatment decisions for each stage separately. The value ratio (VR) of the final estimated dynamic two-stage treatment regime is reported as overall VR. Table 3 shows that the proposed methods and SAS have comparable performance at the last stage (stage 2) with SAS selects a slightly larger number of variables and hence slightly larger number of true prescriptive variables. At the stage 1, SAS tends to select considerably larger number of variables with true positive only slightly different from the proposed methods. ForMMER (both implementations) have considerably better error rate than SAS at stage 1. Overall, for both sample size $n = 200$ and 400, the proposed methods have better value than SAS; for example, for n=200, the value ratio using SAS is 79.1% and 91.6% using the proposed ForMMER-reg. As the value of a regime takes into account the magnitude of incorrect decisions and the overall and long-term effect of decisions at multiple stages, it is the most relevant metric in evaluating the overall performance of an estimated dynamic treatment regime.

To summarize, overall forMMER-AIPWE has the most robust performance in terms of higher values and lower error rates of identifying the optimal treatment decision for an individual patient. In addition, the proposed method achieved high values with considerably simpler decision rules, relative to SAS and the usual forward selection based on AIC, across all scenarios. This is obviously an advantage of the proposed method as it can greatly improves practicality of the estimated treatment regime because simpler decision rules can increase interpretability and also reduce the burden of collecting patient-level information useful for decision making. Finally, we note that the proposed method is applicable in practice in terms of computational cost. Taking a simulated data set from scenario IV (p=1000) as an example, the proposed method takes less than 4 minutes to run using R on a PC with an Intel(R) Core(TM)i7-6700 CPU@3.4GHz and 8GB RAM, which we think is affordable in practice.

4. Application to Nefazodone CBASP trial.

We applied the proposed method to the Nefazodone CBASP trial, where 681 patients with nonpsychotic chronic major depressive disorder (MDD) were randomized to receive either Nefazodone, cognitive behavioral analysis system of psychotherapy (CBASP) or the combination of the two treatments (Keller, et al, 2000). Subjects were followed for 12 weeks with various assessments taken throughout the study. We consider the score on the 24-item Hamilton Rating Scale for Depression (HRSD) at 12 weeks after treatment as our outcome of interest and our analysis includes 577 subjects for whom the HRSD score at 12 weeks are available. We considered a total of 50 baseline variables in constructing optimal treatment regimes. Baseline variables considered in this analysis are listed in Table 5. Lower HRSD indicates low depression and better outcome. Previous analyses have found that the combination treatment leads to lower HRSD score than the other two single treatments, whereas there is no significance difference between the two single treatments. Based on this results, in our analysis, we firstly combined the two single treatment arms into one arm and the treatment decision to be made is either combination treatment or single treatment. Then, we limit our analysis to patients who were randomized to receive single treatment (Nefazodone or CBASP) and consider the treatment decision being either Nefazodone or CBASP. Also in our analysis, we consider −HRSD as our outcome such that larger value means better outcome.

We analyzed the data using the proposed ForMMER-reg and SAS. For ForMMER, we first build parametric linear regression model for each treatment group and the final models were chosen using forward selection based on BIC information number. Then the contrast for each subjects were used in the ForMMER algorithm to estimate the optimal treatment regime. Since the number of candidate variables in our data is only 50 and far less than the number of covariates (1000) in our simulations, we chose to use a less stringent cutoff value $\alpha = 0.02$. For combination ($a = 1$) versus single treatment ($a = 0$), the estimated optimal regimes by ForMMER and SAS are, respectively,

$$\begin{aligned} &\widehat{g}^{opt}_{ForMMER} = I(55 - X_6 > 0), \\ &\widehat{g}^{opt}_{SAS} = I(4.55 - 1.97X_{14} + 0.16X_{18} - 5.88X_{35} > 0). \end{aligned}$$

The forms of regimes from the two methods seem very different and completely different sets of covariates are selected. However, we note that, despite the form of the decision rules, the actual treatment decisions are actually very similar. Both rules suggest that for the majority of patients the optimal treatment option is the combination treatment ($a = 1$). For patients in our data set, ForMMER suggests that 555 patients out of 557 should receive the combination treatment and SAS suggests 545 patients should receive treatment 1. There is a lot of overlap in terms of the estimated optimal decisions and all the 545 patients identified by SAS that should receive the combination treatment are also identified by ForMMER. In Table S4 in the supplementary material, the estimated optimal treatment decisions from the two methods are given as a two-by-two table. The estimated values of the two regimes using the inverse probability weighted method are almost the same: -9.8 (95% CI: -10.9,-8.7) and -9.8 (95% CI: -10.9, -8.6). We note that, in constructing confidence intervals, regimes are taken to be given without taking into account that regimes are estimated from the data. The estimated values are very close to the value (-9.9) of a regime that assigns everyone to the combined treatment ($a = 1$) and much better than a regime that assigns everyone to single treatment (Table 4). These results are consistent with the previously published results, which indicate that the combined treatment is superior to single treatment. Although ForMMER and SAS lead to similar treatment decisions and estimated values, the form of decision rule from ForMMER is much simpler than SAS, which is consistent with our simulation studies. From a practical point of view, a simpler decision rule with less variables are more convenient to use in practice and should be preferred.

For Nefazodone ($a = 1$) versus CBASP ($a = 0$), the estimated optimal regimes by ForMMER and SAS are, respectively,

$$\widehat{g}^{opt}_{ForMMER} = I(-0.55 - 0.30X_1 - 0.17X_8 + 0.20X_{12} - 0.12X_{13} - 0.60X_{15}$$
$$+0.33X_{16} + 0.71X_{40} - 0.88X_{50} > 0),$$

$$\widehat{g}^{opt}_{SAS} = I(-15.03 - 5.15X_1 - 2.71X_8 + 1.01X_{12} + 6.60X_{14} + 4.17X_{16} +$$
$$0.09X_{22} - 2.49X_{28} + 6.16X_{31} - 14.69X_{33} - 7.69X_{36} - 12.53X_{37} - 4.56X_{38}$$
$$+5.37X_{40} - 17.87X_{42} - 5.87X_{46} - 7.59X_{48} + 7.98X_{49} - 6.71X_{50} > 0).$$

Again the estimated optimal treatment regime from ForMMER is much simpler than that from SAS. Six out of eight selected variables by ForMMER are also selected by SAS. There is also a fair amount of overlap in terms of treatment decisions. ForMMER recommends 152 patients receive Nefazodone and 121 of them are also suggested by SAS to receive Nefazodone; see Table S4 in the supplementary material for treatment decisions from the two methods. The estimated value of $\widehat{g}^{opt}_{ForMMER}$ and $\widehat{g}^{opt}_{SAS}$are -11.0 (95% CI: -12.6, -9.4) and -12.1 (95% CI: -13.8, -10.4) respectively, with $\widehat{g}^{opt}_{ForMMER}$ having slightly larger estimated value while selecting less variables. The value of both regimes perform much better than the two regimes that assign everyone to Nefazodone or CBASP. More results are reported in Table 4. In addition to Table S4, Tables S4 and S5 in the supplementary material also provide additional summary statistics on this data application.

5. Discussion. Within the classification framework (C-learning) for estimating the optimal treatment regimes, in this article we further developed a variable selection algorithm for selecting variables that have qualitative interactions with treatment and hence are important for making treatment decisions, namely, prescriptive variables. This variable selection algorithm directly targets prescriptive variables with the objective of optimizing treatment rules, in contrast to methods focusing on selecting predictive variables and prediction. Within the C-learning framework, the optimal treatment regime can be equivalently defined as the classifier that minimizes a weighted misclassification error, where the objective of the classifier is to, based on patient's characteristics, classify patients to the treatment option that leads to larger expected potential outcomes. A major advantage of this framework is that it naturally accommodates a strategy for variable selection targeting prescriptive variables, since only prescriptive variables are relevant in determining the contrast functions and the weighted misclassification error. In the proposed ForMMER algorithm, it forward sequentially selects important prescriptive variables and estimates the optimal treatment regimes simultaneously. The proposed prescriptive variable selection method is based on a direct optimization strategy by directly optimizing the value of treatment decision rules and as a result it enjoys the advantages of direct optimization methods for estimating the optimal treatment regimes discussed in detail in Section 2.2. This also explains the superior performance of the proposed method as demonstrated in our simulations.

The ForMMER algorithm selects prescriptive variables sequentially and at each step it assesses the additional merit of a new variable given variables that have already been selected. As a result, similar to SAS, it tends not to select those variables that are only marginally important for decision making but are not important jointly. Therefore, as SAS, it tends to select fewer variables overall but more true prescriptive variables than methods that consider each variables individually. Furthermore, the proposed ForMMER algorithm decouples the step for estimating the contrast functions from the step for optimization and prescriptive variable selection and, as a result, it is able to target directly on prescriptive variables while still taking advantage of predictive variables in the outcome-regression step to improve performance. This is one of the main differences between ForMMER and SAS and in SAS variables that are only predictive but not prescriptive tend not to be selected and hence will not be able to be exploited in subsequent steps to improve performance. This point is discussed in detail at the end of Section 2.3 and illustrated in simulations, especially in Table 2. To summarize, the flexibility of modeling the contrast functions using various ways, the sequential selection strategy, and the separation of the optimization step for variable selection and optimizing decision rules from the estimation of the contrast functions together contribute to the superior performance of the proposed ForMMER method. As demonstrated by our simulations and real data application, ForMMER selects considerably less variables yet with better value and lower error rate than SAS and the same statement can be made for its performance relative to other methods evaluated in Fan, Lu and

Song (2015) (i.e., the S-score method of Gunter, Zhu and Murphy, 2011, and the method of Lu, Zhang and Zeng, 2013, with LASSO selection) as our simulation settings are adopted from Fan, Lu and Song (2015). Finally, we note the measure (weighted misclassification error) used in the forward sequential variable selection in our method is directly related to the definition of an optimal treatment regime and has a very intuitive interpretation, making it easier to communicate with clinicians. As argued at the end of Section 3, overall we think the proposed method offers a reasonable and practical solution to a clinically very important issue.

Acknowledgements. This work was supported by National Natural Science Foundation of China (71701120 and 61472205), Program for Innovative Research Team of Shanghai University of Finance and Economics.

References.

[1] Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61,** 962-972.

[2] Barrett, J.K., Henderson, R.& Rosthøj, S. (2014). Doubly robust estimation of optimal dynamic treatment regimes. *Statistics in Biosciences* **6,** 244–260.

[3] Biernot,P. & Moodie, E.E.M. (2010). A Comparison of Variable Selection Approaches for Dynamic Treatment Regimes. *The International Journal of Biostatistics* **6**,1557–4679.

[4] Brinkley, J., Tsiatis, A.A. & Anstrom, K.J. (2009). A generalized estimator of the attributable benefit of an optimal treatment regime. *Biometrics* **21,** 512–522.

[5] Chakraborty, B., Murphy, S.A. & Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* **19** 317–343.

[6] Fan, A., Lu, W. & Song, R. (2015). Sequential Advantage Selection for Optimal Treatment Regimens. *Annals of Applied Statistics*, in press.

[7] Geng, Y., Lu, W. & Zhang, H.H. (2015). On Optimal Treatment Regimes Selection for Mean Survival Time *Statistics in Medicine* **34**, 1169–1184.

[8] Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley.

[9] Gunter, L., Zhu, J. & Murphy, S. (2011). Variable selection for qualitative interactions. *Statistical methodology* **8**, 42-55.

[10] Keller, M.B., McCullough, J.P., Klein, D.N., Arnow, B., Dunner, D.L., Gelenberg, A.J., Marekowitz, J.C., Nemeroff, C.B., Russell, J.M., Thase, M.E., Trivedi, M.H. & Zajecka, J. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for treatment of chronic depression. *New England Journal of Medicine* **342**, 331–336.

[11] Kang, C., Janes, H. & Huang, Y. (2014). Combining biomarkers to optimize patient treatment recommendations. *Biometrics* **70**, 695–720.

[12] Lu, W., Zhang, H.H. & Zeng, D. (2013). Variable selection for optimal treatment decision. *Statistical methods in medical research* **22**, 493–504.

[13] Mebane, W.R. & Sekhon, J.S. (2011). Genetic optimization using derivatives: the rgenoud package for R. *Journal of Statistical Software* **42**, 1–26.

[14] Moodie, E.E.M., Richardson, T.S. & Stephens, D.A. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics* **63**, 447–455.

[15] Murphy, S.A. (2003). Optimal dynamic treatment regimes (with discussion). *Journal of Royal Statistical Society Series B* **58**, 331–366.

[16] Peto, R. (1982). Statistical aspects of cancer trials. In *Treatment of Cancer* (K. Halnan, ed.) 867-871. Chapman, London, UK.

[17] Qian, M. & Murphy, S.A. (2011). Performance Guarantees for Individualized Treatment Rules. *Annals of Statistics* **39,** 1180–1210.

[18] Robins, J.M. (2004). Optimal structured nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium on Biostatistics*, D. Y. Lin and P. J. Heagerty (eds), 189–326. New York: Springer.

[19] Robins, J., Orellana, L. & Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine* **27**, 4678-4721.

[20] Schulte, P.J., Tsiatis, A.A., Laber, E.B. & Davidian, M. (2014). Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Statistical Science* **29** 640–661.

[21] Tian, L., Alizadh, A.A., Gentles, A.J. & Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association* **109,** 1517-1532.

[22] Watkins, C. J. C. H. & Dayan, P. (1992). Q-learning. *Mach. Learn.* **8**, 279–292.

[23] Young, J.G., Cain, L.E., Robins, J.M., O'Reilly, E.J., Hernán, M.A. (2011). Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences* **3,** 119–143.

[27] Zhang, B., Tsiatis, A.A., Laber, E.B. & Davidian, M. (2012a). A robust method for estimating optimal treatment regimes. *Biometrics* **68**, 1010–1018.

[27] Zhang, B., Tsiatis, A.A., Laber, E.B. Davidian, M., Zhang, M. & Laber, E.B.(2012b). Estimating optimal treatment regimes from a classification perspective *Stat* **1**, 103–114.

[27] Zhang, B., Tsiatis, A.A., Laber, E.B., & Davidian, M.(2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions *Biometrika* **100**, 681–694.

[27] Zhang, B., Zhang, M.(2015). C-learning: a New Classification Framework to Estimate Optimal Dynamic Treatment Regimes. The University of Michigan Department of Biostatistics Working Paper Series, paper 116.

[29] Zhao, Y., Zeng, D., Rush, A.J. & Kosorok, M.R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* **107**, 1106–1118.

[29] Zhao, Y., Zeng, D., Laber, E.B & Kosorok, M.R. (2015). New Statistical Learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Journal of the American Statistical Association* **510**, 583–598.

School of Statistics and Management,
Shanghai University of Finance and Economics,
Shanghai, P.R.China., P.R.China.
E-mail: zhang.baqun@mail.shufe.edu.cn

Department of Biostatistics,
University of Michigan,
Ann Arbor, MI, U.S.A.
E-mail: mzhangst@umich.edu

Table 1

Simulation results for single-decision point setting based on 500 replications based on sample size n=200 (Scenarios adopted from Fan, Lu and Song, 2015). Size: number of selected prescriptive variables; TP: number of true positive (important) prescriptive variables; ER: error rate of the treatment decision; VR: ratio of the value of the estimated regime relative to that of the true optimal regime. Numbers in parenthesis are Monte Carlo standard deviation.

Method	ρ	Size	TP	ER	VR
Scenario I					
SAS	0.2	6.81 (1.65)	2.98 (0.25)	6.2 (4.8)	99.0 (3.4)
Forward		14.50 (2.63)	2.93 (0.25)	12.7 (5.2)	96.8 (3.4)
ForMMER-AIPWE		4.93 (1.05)	2.93 (0.26)	10.7 (5.1)	97.6 (2.6)
ForMMER-reg		5.78 (1.47)	2.93 (0.26)	10.3 (6.3)	97.6 (3.3)
Scenario II					
SAS	0.2	11.90 (2.87)	2.09 (1.12)	32.7 (10.3)	88.7 (6.2)
Forward		12.16 (3.03)	2.72 (0.54)	24.7 (6.9)	93.3 (3.6)
ForMMER-AIPWE		4.82 (1.23)	2.49 (0.69)	22.6 (9.2)	94.0 (4.4)
ForMMER-reg		6.37 (1.58)	2.76(0.48)	20.7 (7.5)	95.0 (3.4)
Scenario III					
SAS	0.2	11.06 (2.92)	5.08 (2.56)	23.0 (15.0)	84.2 (15.5)
Forward		20.03 (3.48)	7.42 (0.74)	12.9 (7.1)	95.0 (5.1)
ForMMER-AIPWE		7.50 (1.28)	6.79 (1.39)	14.2 (7.2)	94.3 (5.6)
ForMMER-reg		8.58 (1.13)	7.17 (1.14)	12.0 (7.5)	95.5(5.4)
Scenario IV					
SAS	0.2	12.02 (2.28)	3.36 (2.02)	35.0 (10.0)	81.5 (8.2)
Forward		16.07 (2.64)	6.89 (1.34)	20.3 (6.9)	93.0 (4.6)
ForMMER-AIPWE		6.08 (1.48)	5.11 (1.62)	23.4 (6.9)	91.1 (4.8)
ForMMER-reg		7.84 (1.47)	6.33 (1.52)	18.6 (7.3)	94.0(4.5)

Table 2

Simulation results for single-decision point setting based on 500 replications. Size: number of selected prescriptive variables; TP: number of true positive (important) prescriptive variables; ER: error rate of the treatment decision; VR (value ratio): ratio of the value of the estimated regime relative to that of the true optimal regime. Numbers in parenthesis are Monte Carlo standard deviations).

Method	ρ	Size	TP	ER	VR
		Scenario V			
		n=200			
SAS	0.2	14.49 (2.07)	0.72 (0.60)	43.9 (5.3)	71.8 (5.6)
Forward		10.46 (2.21)	1.98 (0.13)	12.9 (4.5)	96.9 (2.3)
ForMMER-AIPWE		3.99 (0.95)	1.97 (0.16)	11.8 (5.1)	97.3 (2.5)
ForMMER-reg		4.39 (1.04)	1.98 (0.15)	12.5 (6.3)	96.7 (2.9)
SAS	0.8	8.55 (2.03)	0.96 (0.35)	21.3 (6.3)	90.9 (5.1)
Forward		11.59 (2.06)	2.00 (0.00)	8.7 (1.4)	98.5 (0.5)
ForMMER-AIPWE		3.57 (1.00)	1.81 (0.39)	7.7 (2.9)	98.7 (0.9)
ForMMER-reg		3.84 (1.00)	1.99 (0.11)	4.4 (1.8)	99.6 (0.4)
		Scenario VI			
		n=200			
SAS	0.2	12.98 (2.24)	1.67 (0.47)	32.7 (4.9)	70.0 (4.5)
Forward		8.84 (2.44)	1.47 (0.53)	32.1 (7.2)	70.1 (6.6)
ForMMER-AIPWE		3.39 (1.23)	1.96 (0.23)	7.2 (7.6)	93.4 (6.9)
ForMMER-reg		5.75 (1.59)	1.85 (0.38)	19.5 (8.2)	81.9 (7.6)
ForMMER-AIPWE†		2.82(0.92)	1.96 (0.25)	6.0(7.2)	94.5 (6.6)
ForMMER-reg†		5.14(1.61)	1.78 (0.44)	18.6 (8.6)	82.8 (8.0)
SAS	0.8	11.33 (2.23)	1.01 (0.27)	31.6 (3.8)	83.7 (2.0)
Forward		10.42 (2.45)	0.99 (0.43)	31.6 (6.7)	83.6 (3.5)
ForMMER-AIPWE		3.66 (1.28)	1.61 (0.59)	13.7 (10.8)	92.9 (5.6)
ForMMER-reg		5.44 (1.64)	1.33 (0.57))	22.2 (8.7)	88.5 (4.5)

Table 3
Simulation results for two-stage setting based on 500 replications (scenario adopted from Fan, Lu and Song, 2015). Size: number of selected prescriptive variables; TP: number of true positive (important) prescriptive variables; ER: error rate of the treatment decision at the stage; overall VR (value ratio): ratio of the value of the estimated regime relative to that of the true optimal regime. Numbers in parenthesis are Monte Carlo standard deviations.

		stage 2			stage 1		
Method	Size	TP	ER	Size	TP	ER	Overall VR
			n=200				
SAS	6.59	3.53	14.0	11.70	2.19	30.2	79.1
	(2.16)	(0.59)	(5.5)	(2.68)	(0.93)	(7.0)	(8.7)
ForMMER-AIPWE	4.59	3.14	14.0	4.07	2.19	13.4	87.3
	(1.05)	(0.61)	(4.1)	(1.23)	(0.82)	(4.4)	(6.4)
ForMMER-reg	5.12	3.34	11.8	4.75	2.61	11.3	91.6
	(1.21)	(0.54)	(3.6)	(1.35)	(0.82)	(3.8)	(4.2)
			n=400				
SAS	5.77	3.93	7.6	13.00	3.95	16.2	93.3
	(1.80)	(0.26)	(3.6)	(3.75)	(1.13)	(5.3)	(3.9)
ForMMER-AIPWE	3.88	3.27	10.8	3.00	2.20	10.5	94.1
	(0.81)	(0.48)	(3.0)	(0.78)	(0.51)	(3.3)	(2.7)
ForMMER-reg	4.36	3.37	9.1	3.55	2.45	8.7	96.0
	(0.96)	(0.49)	(3.4)	(1.01)	(0.62)	(3.5)	(1.6)

Table 4
Data analysis results: estimated value of the estimated optimal treatment. The values are estimated using inverse probability weighted method. $g = 1$ is a regime that assigns everyone to treatment 1 and $g = 0$ is a regime that assign everyone to treatment 0. Numbers in parenthesis are 95% confidence intervals.

	Combination ($a = 1$)	Nefazodone ($a = 1$)
	vs. Single ($a = 0$)	vs. CBASP ($a = 0$)
ForMMER	-9.8 (-10.9,-8.7)	-11.0 (-12.6,-9.4)
SAS	-9.8 (10.9,-8.6)	-12.1 (-13.8,-10.4)
$g = 1$	-9.9 (-11.0,-8.8)	-14.9 (-16.4,-14.9)
$g = 0$	-14.9 (-15.8,-13.9)	-14.8 (-16.2,-13.4)

Table 5

List of covariates used in the analysis of Nefazodone CBASP trial

1 Female
2 White
3-4 Marital Status (single, widowed/divorced/separated)
5 Body mass index
6 Age of MDD onset
7 Age at screening
8 Live alone
9 IDSSR Anxiety/Arousal Score
10 IDSSR General/Mood Cognition Score
11 IDSSR total score
12 IDSSR sleep score 1
13 IDSSR sleep score 2
14 HAMD Anxiety/Somatic Symptoms
15 HAMD Cognitive Disturbance
16 HAMD Retardation Score
17 HAMD Sleep Disturbance factor score
18 Total HAMD-17 score
19 Total HAMD-24 score
20 MOS36 Cognitive Functioning Factor Score
21 MOS36 General Health Factor Score
22 MOS36 Mental Health Factor Score
23 MOS36 Social Functioning
24 Total HAMA score
25 HAMA Psychic Anxiety Score
26 HAMA Somatic Anxiety Score
27-28 MDD type (neither melancholic or atypical, melancholic)
29-30 Main study diagnosis (no antecedent, continuous)
31-32 MDD current severity (mild, moderate)
33 Anxiety disorder NOS
34-35 Alcohol (abuse, dependence)
36-37 Anxiety(sub-threshold, threshold)
38 Other psychological problems
39 Body dysmorphic current
40 Drug abuse
41 Anorexia or bulimia nervosa
42 Obsessive compulsive
43-44 Specific phobia (sub-threshold, threshold)
45-46 Social phobia (sub-threshold, threshold)
47-48 Post traumatic stress(sub-threshold, threshold)
49-50 Panic (sub-threshold, threshold)