A Smoothed Q-Learning Algorithm for Estimating Optimal Dynamic Treatment Regimes

Yanqin Fan
Department of Economics, University of Washington

Ming He
Economics Discipline Group, University of Technology Sydney

Liangjun Su
School of Economics, Singapore Management University

Xiao-Hua Zhou
International Center for Mathematical Research, Peking University

July 5, 2018

Abstract

In this paper we propose a smoothed Q-learning algorithm for estimating optimal dynamic treatment regimes. In contrast to the Q-learning algorithm in which non-regular inference is involved, we show that under assumptions adopted in this paper, the proposed smoothed Q-learning estimator is asymptotically normally distributed even when the Q-learning estimator is not and its asymptotic variance can be consistently estimated. As a result, inference based on the smoothed Q-learning estimator is standard. We derive the optimal smoothing parameter and propose a data-driven method for estimating it. The finite sample properties of the smoothed Q-learning estimator are studied and compared with several existing estimators including the Q-learning estimator via an extensive simulation study. We illustrate the new method by analyzing data from the Clinical Antipsychotic Trials of Intervention Effectiveness-Alzheimer’s Disease (CATIE-AD) study.
Keywords: asymptotic normality; exceptional law; optimal smoothing parameter; sequential randomization; Wald-type inference

Running Headline: Y. Fan et al., Smoothed Q-learning to DTR
1 Introduction

A dynamic treatment regime (DTR) is a sequence of decision rules, one per stage or time period. The decision rule at each stage prescribes a recommended treatment on the basis of the subject’s treatment and covariate history. A DTR is optimal if it optimizes the mean outcome at the end of the final stage of the treatment. The optimal DTR must be estimated from data. Chakraborty and Moodie (2013) present a comprehensive discussion of existing statistical methods for the estimation and inference for DTRs based on sequential randomized trials. We refer interested readers to Chakraborty and Moodie (2013) for more discussion of existing methods and their applications.

One simple yet powerful method for estimating DTRs is the \( Q \)-learning algorithm. Consider the simplest \( Q \)-learning algorithm for a two-stage dynamic binary treatment based on linear models. The estimation using the \( Q \)-learning algorithm proceeds in two steps with each step involving a linear regression. However, statistical inference for the first stage parameter in the optimal DTR based on the \( Q \)-learning estimator is often complicated. The reason is that the dependent variable in the linear regression in the second step of the \( Q \)-learning algorithm is a generated variable which depends on the first stage parameter estimator through a max operation, see Section 2.1 below or Laber et al. (2014) for the exact expression. As shown in Robins (2004) for \( g \)-estimation and Laber et al. (2014) for the \( Q \)-learning algorithm, the max operation involved in the \( g \)-estimation or in the second step regression in the \( Q \)-learning algorithm leads to non-standard asymptotics for estimators of parameters in the first stage \( Q \)-function or the optimal DTR. First, these estimators are asymptotically biased for some data generating processes (DGPs) known as exceptional laws (see Moodie and Richardson, 2010). Second, their asymptotic distributions are normal for some DGPs but non-normal for other DGPs (namely, exceptional
laws), rendering statistical inference based on the $Q$-learning estimators difficult.

Various remedies have been proposed in the literature to alleviate both problems. To reduce asymptotic bias, Moodie and Richardson (2010) propose hard-thresholding for $g$-estimation, while Chakraborty et al. (2010) and Laber et al. (2014) study both hard- and soft-thresholding for the $Q$-learning algorithm. Although they may reduce the bias to some extent, statistical inference is still difficult due to the fact that they still involve non-smooth operations. Chakraborty et al. (2010) propose and study the finite sample performances of several bootstrap confidence intervals (CIs) based on the $Q$-learning algorithm and the hard- and soft-thresholding estimators via simulation. It is worth mentioning that none of these bootstrap CIs is asymptotically valid under exceptional laws. Using local sequences, Laber et al. (2014) propose and evaluate a locally consistent Fixed Adaptive Confidence Interval (FACI) for the parameters of the optimal DTR. Existing simulation results seem to suggest that the FACI is often conservative and none of these methods dominates the rest; see Chakraborty and Moodie (2013). In addition, Goldberg et al. (2014) propose a smoothing procedure using the SoftMax function. Luedtke and van der Laan (2016) develop an inference procedure for the value under the optimal treatment regime.

This paper develops a new algorithm, a smoothed $Q$-learning (SQ) algorithm, for estimating optimal DTRs. It involves the same two estimation steps as the $Q$-learning algorithm except that in the SQ algorithm, the non-smooth step function (in the max operation) in the second step of the $Q$-learning algorithm is replaced by a sequence of smooth distribution functions indexed by a smoothing parameter such that the sequence of distribution functions approaches the step function when the sample size approaches infinity. For any finite sample size, however, the SQ estimators of parameters in the first stage $Q$-function are smooth functions of the first step estimators. We show that
under assumptions adopted in this paper, the SQ estimators possess standard asymptotic properties, namely, they are asymptotically unbiased and normally distributed even under exceptional laws. As a result, statistical inference based on the SQ estimator is standard. Indeed we construct a consistent estimator of the asymptotic variance of the SQ estimator based on which Wald-type inference can be carried out. Moreover, by deriving higher order mean square error (MSE) expansions for the SQ estimator, we derive the optimal smoothing parameter and propose a data-driven method for estimating it.

To study the finite sample performance of our SQ estimator and compare it with some existing estimators, we conduct an extensive simulation study. We consider six designs ranging from fully regular to close to non-regular to fully non-regular, seven estimators, and eight methods of constructing the CIs. We find that our SQ estimator performs the best for all designs under consideration in terms of MSE, coverage probability (CP), and interval length (IL). Moreover, it is the least computationally expensive.

To illustrate the new method, we estimate the optimal dynamic treatment regime for the Alzheimer’s Disease, using data from the Clinical Antipsychotic Trials of Intervention Effectiveness-Alzheimer’s Disease (CATIE-AD) study. The study aims to assess the effectiveness of the second generation antipsychotic drugs (olanzapine, quetiapine, and risperidone) in treating delusions, hallucinations, aggression, and agitation for patients with Alzheimer’s Disease. The estimated optimal dynamic treatment regime allows the clinician to recommend treatment based on the Neuropsychiatric Inventory (NPI) score, age, and gender. We find that olanzapine is in general more effective than other medications in reducing symptoms associated with the Alzheimer’s Disease. In particular, olanzapine tends to be more effective for a patient with a higher NPI score, and more effective for female patients than for male patients. We do not find any statistically significant age effect.
We note that the same smoothing technique was first used in Horowitz (1992) to construct the smoothed maximum score estimator of parameters in a binary response model. Although the maximum score estimator of Manski (1975, 1985) converges at cubic rate and has a non-normal asymptotic distribution, the smoothed maximum score estimator has a faster rate of convergence and an asymptotic normal distribution. In our context, the SQ estimator retains the root-$n$ convergence rate of the $Q$-learning estimator, but unlike the $Q$-learning estimator, it is asymptotically normally distributed for all DGPs satisfying our assumptions.

The rest of this paper is organized as follows. Section 2 first reviews the basic set-up and the $Q$-learning algorithm, then presents the SQ algorithm. Section 3 establishes the asymptotic theory for the SQ estimator. Section 4 derives the optimal smoothing parameter and develops a data-driven method for estimating it. Section 5 presents numerical results from a simulation study. Section 6 uses the new method to study the optimal DTR for the CATIE-AD data. Some concluding remarks are provided in Section 7. All the technical proofs are relegated to Appendix A, and Appendix B presents the model used in the simulation study.

We close this section by introducing some notation. For a real matrix $A$, we denote its transpose as $A'$, its Frobenius norm as $\|A\| (\equiv \text{tr}(AA'))^{1/2})$. We will use $0_{a \times b}$ to denote the $a \times b$ matrix of zeros. Let $\Rightarrow$ and $\overset{p}{\Rightarrow}$ denote convergence in distribution and in probability, respectively.

2 The Basic Set-up and Two $Q$-learning Algorithms

We consider the basic set-up of a two-stage dynamic binary treatment. The sample information contains i.i.d. observations $\{X_{1i}, A_{1i}, Y_{1i}, X_{2i}, A_{2i}, Y_{2i}\}_{i=1}^n$, where for $t = 1, 2$,
$X_{ti}$ denotes individual $i$'s characteristics collected prior to the $t$-th treatment assignment, 
$A_{ti} \in \{0,1\}$ denotes individual $i$'s binary treatment assigned at stage $t$, with $A_{ti} = 1$ indicating treatment, and $A_{ti} = 0$ indicating no treatment, and $Y_{ti}$ is individual $i$'s outcome variable at stage $t$. We assume that $Y_{ti}$ has been coded such that a higher value corresponds to a better outcome. Let $H_{ti} = (X_{1i}, A_{1i}, \cdots, X_{ti})'$ be individual $i$'s history up to stage $t$, that is, the information available to the decision maker (who could be the individual) before the $t$-th treatment assignment $A_{ti}$. For a two-stage treatment, $H_{1i} = X_{1i}$ and $H_{2i} = (X_{1i}', A_{1i}, X_{2i}')'$.

A Dynamic Treatment Regime (DTR), denoted as $\pi = (\pi_1, \pi_2)$, is an ordered pair of functions $\pi_t : \mathcal{H}_t \to \{0,1\}$, where $\mathcal{H}_t$ is the domain of $H_t$, $t = 1, 2$. The objective is to learn a DTR $\pi$ which maximizes the expected outcome $E(Y_1^\pi + Y_2^\pi)$, where $Y_t^\pi$ denotes the potential outcome in period $t$ associated with treatment regime $\pi$. Chakraborty and Moodie (2013) offer a detailed discussion on various methods for estimating optimal DTRs.

We review one simple yet powerful approach, the $Q$-learning algorithm in the next subsection and propose a SQ algorithm in the second subsection. Throughout the paper, we adopt the following assumptions.

- **Stable unit treatment value assumption (SUTVA)**: A subject’s outcome is not influenced by other subjects’ treatment allocation (Rubin, 1980).

- **No unmeasured confounders (NUC)**: For any regime $(a_1, a_2) \in \{0,1\}^2$,

$$A_1 \perp (Y_1 (a_1), X_2 (a_1), Y_2 (a_1, a_2)) | H_1 \text{ and } A_2 \perp Y_2 (a_1, a_2) | H_2,$$

where $X_2 (a_1)$ is the potential covariate at stage 2 if the treatment in stage 1 is $a_1$, $Y_1 (a_1)$ is the potential outcome corresponding to treatment $a_1$ at stage 1, and $Y_2 (a_1, a_2)$ is the stage 2 potential outcome corresponding to treatment $(a_1, a_2)$ for
both stages.

As discussed in Chakraborty and Moodie (2013), the NUC assumption always holds under complete or sequential randomization and is sometimes called the sequential randomization assumption (SRA) or sequential ignorability.

2.1 The Q-learning Algorithm

To describe the Q-learning Algorithm, we define the Q-functions for the two stages as:

\[
Q_2(h_2, a_2) = \mathbb{E}[Y_2|H_2 = h_2, A_2 = a_2] \quad \text{and} \\
Q_1(h_1, a_1) = \mathbb{E}[Y_1 + \max_{a_2 \in \{0, 1\}} Q_2(H_2, a_2)|H_1 = h_1, A_1 = a_1],
\]

(2.1)

where \(\mathbb{E}[\cdot|\cdot]\) denotes the conditional expectation. Under Assumptions SUTVA and NUC, if the conditional expectations in the preceding display were known, then dynamic programming provides an optimal DTR given by

\[
\pi_{dp}^t(h_t) \in \arg \max_{a_t \in \{0, 1\}} Q_t(h_t, a_t), \ t = 1, 2,
\]

(2.2)

and studying the functions \(\pi_{dp}^t(\cdot)\) amounts to studying the Q functions. In most practical settings, these Q functions are unknown and need to be estimated from data.

To reduce dimensionality of the problem, most works in the existing literature adopt linear specifications for both stage Q functions:

\[
Q_t(H_{ti}, A_{ti}; \beta_t) = H_{t0,i}' \beta_{t0} + (H_{t1,i}' \beta_{t1}) A_{ti}, \ t = 1, 2,
\]

(2.3)

where \(\beta_t = (\beta_{t0}', \beta_{t1}')', H_{t0,i}, \text{and } H_{t1,i}\) are column vectors of features depending on indi-
vidual \(i\)'s history at stage \(t\). They can be subvectors or functions of \(H_{t,i}\) and there can be overlap between \(H_{t0,i}\) and \(H_{t1,i}\). Note the asymmetric roles of \(H_{t0,i}\) and \(H_{t1,i}\): \(H_{t0,i}\), which includes an intercept term, denotes the main effect of history; \(H_{t1,i}\), which also includes an intercept term, denotes the treatment effects of history. Under the above linear specification, an optimal DTR is given by

\[
\pi_{t}^{dp}(h_{t}) \in \arg \max_{a_{t} \in \{0, 1\}} \left[(h'_{t1} \beta^*_t) a_t\right], \ t = 1, 2, \tag{2.4}
\]

where \(h'_{t} = (h'_{t0}, h'_{t1})\), and \(\beta^*_t\) denotes the true value of \(\beta_t\). Two points are worth emphasizing here. First, any optimal policy only depends on \(h_{t1}\); Second, there is no unique optimal policy for stage \(t\) when \(h'_{t1} \beta^*_t = 0\).

Estimating the optimal DTR is equivalent to estimating the unknown parameters in the \(Q\) functions. Let \(\beta^*_t = (\beta'^*_{t0}, \beta'^*_{t1})'\) denote the true value of \(\beta_t = (\beta'_{t0}, \beta'_{t1})'\) for \(t = 1, 2\). The \(Q\)-learning algorithm proceeds to estimate \(\beta^*_1\) and \(\beta^*_2\) in two steps.

**Step 1.** Regress \(Y_{2i}\) on \(H_{2i}\) and \(A_{2i}\), \(i = 1, \cdots, n\), to obtain

\[
\hat{\beta}_2 = \arg \min_{\beta_2} \sum_{i=1}^{n} [Y_{2i} - Q_2 (H_{2i}, A_{2i}; \beta_2)]^2. \tag{2.5}
\]

**Step 2.** (a) Define the predicted future reward following the optimal policy as

\[
\bar{Y}_{1q,i} = Y_{1i} + \max_{A_{2i} \in \{0, 1\}} Q_2 \left(H_{2i}, A_{2i}; \hat{\beta}_2\right)
= Y_{1i} + H'_{20,i} \hat{\beta}_{20} + \left(H'_{21,i} \hat{\beta}_{21}\right) 1\{H'_{21,i} \hat{\beta}_{21} \geq 0\},
\]

where \(1\{\cdot\}\) is the indicator function.
(b) Regress $\tilde{Y}_{1q,i}$ on $H_{1i}$ and $A_{1i}$, $i = 1, \ldots, n$, to obtain

$$
\hat{\beta}_{1q} = \arg \min_{\beta_1} \sum_{i=1}^{n} \left( \tilde{Y}_{1q,i} - Q_1 (H_{1i}, A_{1i}; \beta_1) \right)^2.
$$

(2.6)

The $Q$-learning algorithm is popular in applied work because of its computational simplicity. However statistical inference for the first stage parameter $\beta^*_1$ using the $Q$-learning estimator $\hat{\beta}_{1q}$ is difficult and the standard Wald-type inference may be misleading. The reason is that the asymptotic distribution of $\hat{\beta}_{1q}$ changes discontinuously with the underlying DGP from normal when the underlying DGP is such that $\Pr \left( H'_{21,i} \beta^*_{21} = 0 \right) = 0$ (regular setting) to non-normal when $\Pr \left( H'_{21,i} \beta^*_{21} = 0 \right) > 0$ (exceptional law, or non-regular setting). This was first observed by Robins (2004) for $g$-estimation. Laber et al. (2014) establish precise expressions for the asymptotic distribution of $\hat{\beta}_{1q}$ in both cases. Interestingly this is related to the first stage treatment effect or equivalently whether there is a unique first stage optimal treatment. When the DGP is such that $\Pr \left( H'_{21,i} \beta^*_{21} = 0 \right) = 0$, the optimal treatment for the first stage is unique; otherwise, it is not.

### 2.2 A Smoothed $Q$-learning Algorithm

This section proposes a SQ algorithm. In sharp contrast to $\hat{\beta}_{1q}$, we show that under conditions adopted in Section 3 below, the asymptotic distribution of the SQ estimator is normal no matter $\Pr \left( H'_{21,i} \beta^*_{21} = 0 \right) = 0$ or $\Pr \left( H'_{21,i} \beta^*_{21} = 0 \right) > 0$.

The SQ algorithm also follows two steps. The first step is the same as that in the $Q$-learning algorithm and the second step replaces Step 2 in the $Q$-learning algorithm with Step $2^*$ below.

**Step $2^*$**. (a) Define the predicted future reward following the smoothed optimal policy
as
\[
\tilde{Y}_{1sq,i} = Y_{1i} + H'_{20,i} \hat{\beta}_{20} + \left(H'_{21,i} \hat{\beta}_{21}\right) K_\alpha(H'_{21,i} \hat{\beta}_{21}),
\] (2.7)
where \(K_\alpha(x) = K(x/\alpha)\), \(K(\cdot)\) is a cumulative distribution function (CDF) that admits a probability density function (PDF) \(k(\cdot)\), which is symmetric around 0. \(\alpha = \alpha_n \to 0\) is a smoothing parameter (bandwidth).

(b) Obtain the smoothed estimator of \(\beta_1\) as
\[
\hat{\beta}_{1sq} = \arg \min_{\beta_1} \sum_{i=1}^{n} \left[\tilde{Y}_{1sq,i} - Q_1(H_{1i}, A_{1i}; \beta_1)\right]^2. \tag{2.8}
\]

Compared with the Q-learning algorithm, the SQ algorithm replaces the step function \(1\{\cdot \geq 0\}\) in Step 2 (a) with a smooth CDF \(K(\cdot/\alpha)\) in Step 2\(^*\)(a), and as \(n\) goes to \(\infty\), \(K(\cdot/\alpha)\) approaches \(1\{\cdot \geq 0\}\). But for any finite sample size \(n\), \(\alpha > 0\) and \(K(\cdot/\alpha)\) is a smooth function, so \(\hat{\beta}_{1sq}\) is a smooth function of \(\hat{\beta}_{21}\).

3 Asymptotic Theory for the Smoothed Q-learning Algorithm

Let \(\hat{\beta}_{sq} = (\hat{\beta}_{1sq}', \hat{\beta}_{2sq}')'\). In this section, we study the asymptotic properties of \(\hat{\beta}_{sq}\). We reformulate \(\hat{\beta}_{sq}\) as a Z-estimator considered in van der Vaart (1998, ch.5).

3.1 A Method-of-Moment Formulation

Let \(B_t = (H'_{t0}, A_t H'_{t1})'\) for \(t = 1, 2\). Observe that \(\hat{\beta}_{sq}\) is a solution to the following system of estimating equations
\[
\mathbb{E}_n[m_n(\hat{\beta}_{sq})] = 0, \tag{3.1}
\]
where $\mathbb{E}_n[\cdot]$ denotes the empirical measure, $m_n(\beta) = (m_{1n}(\beta)', m_{2}(\beta)')'$,

$$m_{1n}(\beta) = m_{1n}(W; \beta) = B_1 [Y_1 + H_2' \beta_{20} + K_\alpha (H_{21}' \beta_{21}) H_{21}' \beta_{21} - B_1' \beta_1],$$

$$m_2(\beta) = m_2 (W; \beta) = B_2 (Y_2 - B_2' \beta_2),$$

(3.2)

and $W = (X_1', A_1, Y_1, X_2', A_2, Y_2)'$. So $\hat{\beta}_{sq}$ is a Method-of-Moment estimator or $Z$-estimator considered in van der Vaart (1998, ch.5). Complication arises because our moment function $m_n$ depends on $n$ through the smoothing parameter $\alpha$.

Let $M (\beta) \equiv \mathbb{E} [m (\beta)] = \mathbb{E} [m (W; \beta)]$, where $m(W; \beta) = (m_1(W; \beta)', m_2(W; \beta)')'$ in which

$$m_1(W; \beta) = B_1 [Y_1 + H_2' \beta_{20} + (H_{21}' \beta_{21})_+ - B_1' \beta_1].$$

(3.3)

Note that under Assumption A2(i), $\mathbb{E} [m(W; \beta)] = 0$ if and only if $\beta = \beta^* = (\beta_1^*, \beta_2^*)'$, where

$$\beta_1^* = (\mathbb{E} [B_1 B_1'])^{-1} \mathbb{E} [B_1 (Y_1 + H_2' \beta_{20}^* + (H_{21}' \beta_{21}^*)_+)], \quad \beta_2^* = (\mathbb{E} [B_2 B_2'])^{-1} \mathbb{E} [B_2 Y_2].$$

In contrast to the standard GMM, $\mathbb{E} [m_n (\beta^*)]$ is in general not equal to zero for finite $n$ because of smoothing $m_1(W; \beta)$ in (3.3) to $m_{1n}(W; \beta)$ in (3.2). Nonetheless, we show consistency and asymptotic normality of $\hat{\beta}_{sq}$ under assumptions below.

### 3.2 Consistency

Let $Z (\beta_{21}) = H_{21}' \beta_{21}$ and $Z = Z (\beta_{21}) = H_{21}' \beta_{21}^*$. Let $F_Z(\cdot; \beta_{21})$ denote the CDF of $Z (\beta_{21})$ and $F_Z(\cdot) = F_Z(\cdot; \beta_{21})$. Further let $\varepsilon_1 = Y_1 + H_{20}' \beta_{20}^* + (H_{21}' \beta_{21}^*)_+ - B_1' \beta_1^*$ and $\varepsilon_2 = Y_2 - B_2' \beta_2^*$. Note that $\varepsilon_1$ and $\varepsilon_2$ denote the population errors in the least squares projections of $[Y_1 + H_2' \beta_{20}^* + (H_{21}' \beta_{21}^*)_+]$ on $(A_1, H_1)$ and that of $Y_2$ on $(A_2, H_2)$ respectively.
Let $W_i = (X_{i1}, A_{i1}, Y_{i1}, X'_{i2}, A_{i2}, Y_{i2})'$ and $B_1, B_{20}, B_{21}$ denote the parameter spaces of $\beta_1$, $\beta_{20}, \beta_{21}$, respectively. Let $B$ denote the parameter space of $\beta = (\beta_1', \beta_2')'$. Let $k_1, k_2, k_{20}$ and $k_{21}$ denote the dimensions of $\beta_1, \beta_2, \beta_{20},$ and $\beta_{21}$, respectively.

To study the consistency of $\hat{\beta}_{sq}$, we make the following assumptions.

**Assumption A1.** (i) $\{W_i\}_{i=1}^n$ is a random sample.

(ii) $B, B_1, B_{20},$ and $B_{21}$ are all compact.

(iii) For each $\beta_{21} \in B_{21}$, $F_Z(\cdot; \beta_{21}) = (1 - p_{\beta_{21}}) F_C(\cdot; \beta_{21}) + p_{\beta_{21}} F_D(\cdot; \beta_{21})$, where $F_C(\cdot; \beta_{21})$ and $F_D(\cdot; \beta_{21})$ denote the continuous and discrete components of $Z(\beta_{21})$, respectively, and $p_{\beta_{21}}$ is the total probability mass of $Z(\beta_{21})$ on discrete points. $F_C(\cdot; \beta_{21})$ admits a PDF $f_C(\cdot; \beta_{21})$ that is continuously differentiable for all $\beta_{21} \in B_{21}$. $F_D(\cdot; \beta_{21})$ is a discrete CDF such that if $Z_D \sim F_D(\cdot; \beta_{21})$, then $P(Z_D = z_d) = f_D(z_d; \beta_{21})$ and $\sum_{z_d} f_D(z_d; \beta_{21}) = 1$.

(iv) Let $r_1(z; \beta_{21}) = \mathbb{E}[B_1|H_{21}\beta_{21} = z]$. For some small $\delta > 0$, there exists a constant $c_{r_1} < \infty$ such that $\sup_{\beta_{21} \in B_{21}} \sup_{|z| \leq \delta} \|r_1(z; \beta_{21}) f_C(z; \beta_{21})\| \leq c_{r_1}$.

(v) Let $r_2(z) = P[B_1 H_{21} H'_{21} \beta_{21}^* = z]$ and $g_2(z) \equiv \int_{-\infty}^{\infty} r_2(s) f_C(s) ds$. For some small $\delta > 0$, there exists a constant $c_{g_2} < \infty$ such that $\sup_{|z| \leq \delta} \|g_2(z)\| \leq c_{g_2}$. $g_2(\infty) \equiv \lim_{z \to \infty} g_2(z)$ exists and is finite.

**Assumption A2.** (i) The eigenvalues of $\mathbb{E}[B_1 B_1']$ and $\mathbb{E}[B_2 B_2']$ are all bounded and bounded away from zero.

(ii) $\mathbb{E}[\|B_1 Y_1\| + \|B_1 H_{20}\| + \|B_2 Y_2\|] < \infty$ and $\mathbb{E}[\|B_1 H_{21}'\|^2] < \infty$.

(iii) $\mathbb{E}[\|B_1 B_1'\| (Z^2 + \varepsilon_1^2)] < \infty$, $\mathbb{E}[\|B_1 B_2'\| |Z\varepsilon_2|] < \infty$, and $\mathbb{E}[\|B_2 B_2'\| \varepsilon_2^2] < \infty$.

(iv) There exists $\delta > 0$ such that $\mathbb{E}[\|B_1 \varepsilon_t\|^{2+\delta} < \infty, t = 1, 2$, $\mathbb{E}[\|B_1 H_{21}'\|^{2+\delta} < \infty$.

**Assumption A3.** (i) The kernel function $K(\cdot)$ is a CDF that admits a PDF $k(\cdot)$ with compact support $[-1, 1]$. $k(\cdot)$ is continuous, symmetric around 0, and has bounded
variation.

(ii) As $n \to \infty$, $\alpha \to 0$, $n\alpha^4 \to 0$, and $n\alpha/\ln n \to \infty$.

Assumption A1(i)-(ii) are standard in the literature on $Q$-learning. A1(iii) indicates that the distribution $F_Z(\cdot; \beta_{21})$ of $H_{21}'\beta_{21}$ is a mixed type distribution. When $\beta_{21} = \beta_{21}^*$, we will write $f_C(z; \beta_{21}^*$), $f_D(z_d; \beta_{21}^*$), and $p_{\beta_{21}}^*$ simply as $f_C(z)$, $f_D(z_d)$, and $p$, respectively. Note that we allow $p = p_n \in [0,1]$ to be $n$-dependent, in which case $\{W_i = W_{in}\}$ should be read as a triangular array process. But for notational simplicity, we suppress the dependence of all variables on $n$. A1(iv)-(v) impose conditions on the conditional moments of $B_1$ and $B_1H_{21}'$. A2(i) is an identification condition. A2(ii)-(iv) specify some moment conditions. A3(i) specifies the conditions on the kernel function $K(\cdot)$ and its derivative function $k(\cdot)$. We assume that $k(\cdot)$ has compact support $[-1,1]$ to simplify the proofs. Many kernels that are frequently used in the kernel estimation literature satisfy this condition. In particular, they include the following kernels:

$$
\begin{align*}
\text{Epanechnikov kernel: } k(z) &= \frac{3}{4} (1 - z^2) 1 \{|z| \leq 1\}; \\
\text{Triangular kernel: } k(z) &= (1 - |z|) 1 \{|z| \leq 1\}; \\
\text{Quartic kernel: } k(z) &= \frac{15}{16} (1 - z^2)^2 1 \{|z| \leq 1\}; \\
\text{Triweight kernel: } k(z) &= \frac{35}{32} (1 - z^2)^3 1 \{|z| \leq 1\}.
\end{align*}
$$

One can relax the compactness assumption at the cost of more lengthy proofs and some tail conditions on $f_C(\cdot; \beta_{21})$, $f_D(\cdot; \beta_{21})$, $r_1(\cdot; \beta_{21})$, and $r_2(\cdot)$. In particular, one can also choose $K(\cdot)$ to be the standard normal or logistic CDF. A3(ii) imposes conditions on the smoothing parameter $\alpha$. Note that $n\alpha^4 \to 0$ ensures that the bias associated with our SQ estimator $\hat{\beta}_{sq}$ is asymptotically negligible.

The following theorem studies the consistency of $\hat{\beta}_{sq}$. 

14
**Theorem 1** Suppose that Assumptions A1(i)-(iv), A2(i)-(ii), and A3 hold. Then for the SQ estimator, we have

\[ \hat{\beta}_{sq} = \beta^* + o_P(1). \]

Note that Assumptions A1(v) and A2(iii)-(iv) are not used in the proof of the above theorem.

### 3.3 Asymptotic Normality and Covariance Estimation

The following theorem studies the asymptotic normality of \( \hat{\beta}_{sq} \).

**Theorem 2** Suppose that Assumptions A1-A3 hold. Then for the SQ estimator, we have

\[ \sqrt{n}(\hat{\beta}_{sq} - \beta^*) \xrightarrow{d} N(0, \Omega), \]

where \( \Omega = \Xi^{-1} \Lambda (\Xi^{-1})' \), \( \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}' & \Lambda_{22} \end{pmatrix} \), \( \Xi = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix} = \begin{pmatrix} -\mathbb{E}[B_1B_1'] & \Xi_{12} \\ 0_{k_1 \times k_2} & -\mathbb{E}[B_2B_2'] \end{pmatrix} \).

\[ \Lambda_{11} = \mathbb{E} \left\{ B_1B_1'[Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + (H_{21}'\beta_{21}^*)_+]^2 \right\}, \]
\[ \Lambda_{12} = \mathbb{E} \left\{ B_1B_2'[Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + (H_{21}'\beta_{21}^*)_+]\left[Y_2 - B_2'\beta_2^*\right] \right\}, \]
\[ \Lambda_{22} = \mathbb{E} \left\{ B_2B_2'[Y_2 - B_2'\beta_2^*]^2 \right\}, \]

and

\[ \Xi_{12} = \left\{ \mathbb{E}[B_1H_{20}'], \left( \lim_{n \to \infty} (1 - p_n) [g_2(\infty) - g_2(0)] + p_n \frac{1}{2} r_2(0)f_D(0) + \sum_{z_d > 0} r_2(z_d)f_D(z_d) \right) \right\}. \]

**Remark 1** Noting that the eigenvalues of the block upper triangular matrix \( \Xi \) are given by those of \( \Xi_{11} \) and \( \Xi_{22} \) which are all negative under Assumption A2(i), \( \Xi \) must be non-
singular. This occurs no matter \( p \in (0, 1), p = 0, \) or \( p = 1 \) and no matter \( f_D(0) = 0 \) (regular setting) or \( f_D(0) > 0 \) (non-regular setting or exceptional law). Theorem 2 shows that our SQ estimator \( \hat{\beta}_{sq} \) is always \( \sqrt{n} \)-consistent for \( \beta^* \) and asymptotically normally distributed even under the non-regular setting. This greatly simplifies the inference procedure. In contrast, the Q-learning estimator of \( \beta_1^* \) is asymptotically normally distributed in the regular case but non-normally distributed in the non-regular case.

To make inference, we need to construct a consistent estimate of \( \Omega \). We propose to estimate \( \Lambda \) and \( \Xi \) by \( \hat{\Lambda}_n \equiv \hat{\Lambda}_n(\hat{\beta}_{sq}) \) and \( \hat{\Xi}_n \equiv \hat{\Xi}_n(\hat{\beta}_{sq}) \), where

\[
\hat{\Lambda}_n(\beta) = \begin{pmatrix}
\hat{\Lambda}_{n,11}(\beta) & \hat{\Lambda}_{n,12}(\beta) \\
\hat{\Lambda}_{n,12}'(\beta) & \hat{\Lambda}_{n,22}(\beta)
\end{pmatrix},
\hat{\Xi}_n(\beta) = \begin{pmatrix}
\hat{\Xi}_{n,11}(\beta) & [\hat{\Xi}_{n,12,0}(\beta), \hat{\Xi}_{n,12,1}(\beta)] \\
0_{k_1 \times k_2} & \hat{\Xi}_{n,22}(\beta)
\end{pmatrix},
\]

\[
\hat{\Lambda}_{n,11}(\beta) = \mathbb{E}_n \{ B_1 B_1'[Y_1 - B_1' \beta_1 + H_20 \beta_20 + K_\alpha(H_21' \beta_21)H_21' \beta_21)^2 \},
\hat{\Lambda}_{n,12}(\beta) = \mathbb{E}_n \{ B_1 B_2'[Y_1 - B_1' \beta_1 + H_20 \beta_20 + K_\alpha(H_21' \beta_21)H_21' \beta_21)Y_2 - B_2' \beta_2] \},
\hat{\Lambda}_{n,22}(\beta) = \mathbb{E}_n \{ B_2 B_2'[Y_2 - B_2' \beta_2)^2 \}.
\]

\[
\hat{\Xi}_{n,11}(\beta) = -\mathbb{E}_n[B_1 B_1'], \hat{\Xi}_{n,22}(\beta) = -\mathbb{E}_n[B_2 B_2'], \hat{\Xi}_{n,12,0}(\beta) = \mathbb{E}_n[B_1 H_20], \text{ and }
\hat{\Xi}_{n,12,1}(\beta) = \mathbb{E}_n \left\{ B_1 H_21' \left[ \frac{1}{\alpha} k_\alpha(H_21' \beta_21)H_21' \beta_21 + K_\alpha(H_21' \beta_21) \right] \right\}.
\]

in which \( k_\alpha(z) = k(\alpha/z) \). Then we can estimate \( \Omega \) by

\[
\hat{\Omega}_n = \hat{\Xi}_n^{-1} \hat{\Lambda}_n(\hat{\Xi}_n^{-1})'.
\]

The following theorem establishes the consistency of \( \hat{\Omega}_n \).

**Theorem 3** Suppose that Assumptions A1-A3 hold. Suppose that \( \mathbb{E} [ \| B_1 B_1' \| (\| Y_1 \| + \| B_1 \| + \| H_20 \|) ] < \infty \), and \( \mathbb{E} [ \| B_1 B_1' \|^2 \| H_21 \| ] < \infty \). Then \( \hat{\Omega}_n = \Omega + o_P(1) \).
3.4 Inference

Let $k = k_1 + k_2$ and $c = (c_1', c_2')' \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$. The asymptotic variance of $\sqrt{n}c' (\hat{\beta}_{sq} - \beta^*)$ is given by

$$\sigma_c^2 = c' \Xi^{-1} \Lambda (\Xi^{-1})' c.$$

We estimate $\sigma_c^2$ by $\hat{\sigma}_c^2 \equiv c' \hat{\Xi}_n^{-1} \hat{\Lambda}_n (\hat{\Xi}_n^{-1})' c$. Theorem 3 implies consistency of $\hat{\sigma}_c^2$ and the result that $T_n \overset{d}{\to} N(0, 1)$, where

$$T_n = \sqrt{n}c' (\hat{\beta}_{sq} - \beta^*)/\hat{\sigma}_c.$$  \hspace{1cm} (3.6)

It is interesting to observe that regardless of Pr ($H_{21,i}^i \beta_{21}^* = 0$) = 0 or not, the limiting distribution of the t-ratio $T_n$ is the standard normal and inference on $\beta^*$, including the first stage parameter $\beta_1^*$, is thus standard. In contrast, discontinuity of the asymptotic distribution of the $Q$-learning estimator in the underlying DGP renders inference based on it difficult.

4 Optimal Choice of the Smoothing Parameter

In this section, we study the optimal choice of the smoothing parameter $\alpha$. For this purpose, we need to conduct a higher-order asymptotic analysis.

4.1 Higher Order Expansion

Following Linton (1995), we consider the scalar standardized quantities $T_n$ and $T_0n$:

$$T_0n = \sqrt{n}c' (\hat{\beta}_{sq} - \beta^*)/\sigma_c \text{ and } S_n = \sqrt{n}(\hat{\sigma}_c - \sigma_c)/\sigma_c.$$  \hspace{1cm} (4.1)
When \( c = e_{k,j} \), a \( k \times 1 \) vector with 1 in its \( j \)th position and zeros elsewhere, \( T_0n \) and \( T_n \) can be regarded as infeasible and feasible test statistics respectively for testing the individual significance of the \( j \)th coefficient in \( \beta \).

By a Taylor expansion, we have \( T_0n = T_0^* + R_0^* \), where

\[
T_0^* = -\sigma_c^{-1}c'\Xi^{-1}\xi_N + (n\alpha)^{-1/2}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi_n^{-1}\xi_N - (n\alpha)^{-1}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N
\]

\[
+ (n\alpha)^{-3/2}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N \text{ and}
\]

\[
R_0^* = -(n\alpha)^{-2}\sigma_c^{-1}c'[\hat{\xi}_n(\bar{\beta})]^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N,
\]

in which \( \xi_N = \sqrt{n}\mathbb{E}_n[m_n(\beta^*)] \) and \( \xi_D = \sqrt{n}\mathbb{E}[\hat{\xi}_n(\bar{\beta}) - \Xi] \) with \( \bar{\beta} \) lying between \( \beta^* \) and \( \hat{\beta}_{sq} \). See Appendix A for details. We will show that the first term in \( T_0^* \) contributes to the dominant bias of \( T_0n \) and the first two terms in \( T_0^* \) contribute to the first and second order asymptotic variances of \( T_0n \), respectively.

To study the asymptotic bias and variance of \( T_n \), we also need the expansion of \( S_n \).

We show in Appendix A that

\[
S_n = \frac{-2c'\Xi^{-1}\xi_D^\dagger\Omega c - \sqrt{\alpha}c'\Xi^{-1}\xi_A\Xi^{-1'}c}{2\sigma_c^2}
\]

\[
+ \frac{2c'\Xi^{-1}\xi_D^\dagger\Xi^{-1'}\xi_D^\dagger\Omega c} {2\sqrt{n}\alpha\sigma_c^2} + \frac{c'\Xi^{-1}\xi_D^\dagger\Omega c} {2\sigma_c^2} + o_P((n\alpha)^{-1/2})
\]

\[
eq S_n^* + o_P((n\alpha)^{-1/2}), \quad (4.2)
\]

where \( \xi_A = \sqrt{n}(\hat{\Lambda}_n - \Lambda) \) and \( \xi_D^\dagger = \sqrt{n}(\hat{\xi}_n - \Xi) \). Then \( T_n = T_n^* + o_P((n\alpha)^{-1}) \), where

\[
T_n^* = -\sigma_c^{-1}c'\Xi^{-1}\xi_N + (n\alpha)^{-1/2}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi_n^{-1}\xi_N + (n\alpha)^{-1/2}\sigma_c^{-1}c'\Xi^{-1}\xi_N S_n^*
\]

\[
- (n\alpha)^{-1}\sigma_c^{-1}c'\Xi^{-1}\left\{\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N - \frac{\xi_D\Xi^{-1}\xi_N c'\Xi^{-1}\xi_D^\dagger\Omega c} {\sigma_c^2} + \frac{2\xi_N(c'\Xi^{-1}\xi_D^\dagger\Omega c)^2} {\sigma_c^4}\right\}, \quad (4.3)
\]

18
We will show that the optimal bandwidth in terms of minimizing the asymptotic mean squared error (MSE) of either $T^*_0$ or $T^*_n$ is proportional to $n^{-2/5}$.

Let $\Omega = \Xi^{-1}\Lambda\Xi^{-1'}$ with $\Omega_{ts}$ being $k_t \times k_s$ for $t, s = 1, 2$. Similarly, partition $\hat{\Omega}_n$ as
\[
\begin{pmatrix}
\hat{\Omega}_{n,11} & \hat{\Omega}_{n,12} \\
\hat{\Omega}_{n,12}' & \hat{\Omega}_{n,22}
\end{pmatrix}
\]
To state the next result, we add the following assumption.

**Assumption A4.** (i) $E\|B_t\varepsilon_t\|^8 < \infty$ and $E\|B_t B_t'\|^8 < \infty$ for $t = 1, 2$, and $E\|B_1 H_{2t}'\|^8 < \infty$ for $t = 0, 1$; (ii) $k(\cdot)$ is continuously differentiable and its first order derivative has bounded variation.

A4(i) strengthens the moment conditions on $B_t\varepsilon_t$, $B_t B_t'$, and $B_1 H_{2t}'$; A4(ii) imposes additional conditions on $k(\cdot)$. These are the cost we have to pay to derive the higher order asymptotics. A4(i) can be violated if the vectors $B_t$ and $B_t\varepsilon_t$ or matrix $B_1 H_{2t}'$ contain some element whose distribution has fat tails.

**Theorem 4** Suppose that $\alpha \propto n^{-2/5}$. Suppose that Assumptions A1-A2, A3(i), and A4 hold. Let the additional moment conditions in Theorem 3 hold. Then

(i) $T_{0n}$ and $T^*_{0n}$ have the same distribution to order $n^{-3/5}$ in the sense $\sup_u |P(T_{0n} \leq u) - P(T^*_{0n} \leq u)| = o\left(n^{-3/5}\right)$;

(ii) $E[T^*_{0n}] = -\sqrt{n\alpha^4} \sigma^{-1} c' \Xi^{-1} \kappa + o\left(n^{-3/10}\right)$ and $\text{Var}[T^*_{0n}] = 1 + (n\alpha)^{-1} \sigma^{-2} c' \Xi^{-1} \Phi_0 \Xi^{-1'} c + o\left(n^{-3/5}\right)$;

(iii) $E[T^*_{n}] = -\sqrt{n\alpha^4} \sigma^{-1} c' \Xi^{-1} \kappa + o\left(n^{-3/10}\right)$ and $\text{Var}[T^*_{n}] = 1 + (n\alpha)^{-1} \sigma^{-2} c' \Xi^{-1} \Phi \Xi^{-1'} c + o\left(n^{-3/5}\right)$, where

\[
\kappa = \begin{pmatrix}
\alpha^{-2} E \{ B_1 [K_n(H_{21}' \beta_{21}^*)H_{21}' \beta_{21}^* - (H_{21}' \beta_{21}^*)_+] \} \\
0_{k_2 \times 1}
\end{pmatrix} = O(1),
\]
\[ \Phi_0 = \begin{pmatrix} \alpha \mathbb{E} \left[ \bar{\eta}_1 (\beta_{21}^*) \Omega_{22} \bar{\eta}_1 (\beta_{21}^*)' \right] & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & 0_{k_2 \times k_2} \end{pmatrix} = O(1), \text{ and} \]

\[ \Phi = \Phi_0 + \alpha \sigma_c^{-4} \mathbb{E}[\hat{\eta}_1 (\beta_{21}^*)]^2 \Lambda + \alpha \sigma_c^{-2} \mathbb{E}[\hat{\eta}_1 (\beta_{21}^*) \eta_1^* (\beta_{21}^*) \Lambda + \Lambda \eta_1^* (\beta_{21}^*)' \eta_1^* (\beta_{21}^*)] = O(1), \]

in which

\[ \bar{\eta}_1 (\beta_{21}) \equiv \bar{\eta}_1 (H_{21}; \beta_{21}) = \eta_1 (\beta_{21}) - \mathbb{E} [\eta_1 (\beta_{21})], \]

\[ \eta_1 (\beta_{21}) \equiv \eta_1 (H_{21}; \beta_{21}) = \left[ 0_{k_1 \times k_2}, B_1 H_{21}^\prime \times \frac{1}{\alpha} k_{\alpha} (H_{21}^\prime \beta_{21}) H_{21}^\prime \beta_{21} \right], \]

\[ \eta_1^* (\beta_{21}) = c' \Xi^{-1} \begin{pmatrix} 0_{k_1 \times k_1} & \bar{\eta}_1 (\beta_{21}) \\ 0_{k_2 \times k_1} & 0_{k_2 \times k_2} \end{pmatrix} \Omega_c = c' \Xi^{-1} \bar{\eta}_1 (\beta_{21}) [\Omega_{12} c_1 + \Omega_{22} c_2], \text{ and} \]

\[ \eta_1^{**} (\beta_{21}) = \begin{pmatrix} 0_{k_1 \times k_1} & \bar{\eta}_1 (\beta_{21}) \\ 0_{k_2 \times k_1} & 0_{k_2 \times k_2} \end{pmatrix} \Xi^{-1} = \begin{pmatrix} 0_{k_1 \times k_1} & \bar{\eta}_1 (\beta_{21}) \Xi^{-1}, \beta_{21} \\ 0_{k_2 \times k_1} & 0_{k_2 \times k_2} \end{pmatrix}. \]

**Remark 2** Theorem 4 implies that \( T_n \) and \( T_{on} \) share the same dominant asymptotic bias term that is \( O(\sqrt{n\alpha^4}) = O(n^{-3/10}) \) while they have different asymptotic variances up to the order \( (n\alpha)^{-1} = O(n^{-3/5}) \). From Lemma A.1 in Appendix A, we know that \( \text{Var}(\xi_D \Xi^{-1} \xi_N) = \Phi_0 + o(1) \) and \( \text{Var}(\xi_D \Xi^{-1} \xi_N + \sigma_c^{-2} c' \Xi^{-1} \xi_D \Omega \xi_N) = \Phi + o(1) \). The second term in the last variance operator reflects the cost for estimating the asymptotic variance \( \sigma_c^2 \).

To sum up, the mean squared errors of \( T_{on} \) and \( T_n \) are approximately given by

\[ \text{MSE}_0 (\alpha) = 1 + n\alpha^4 \sigma_c^{-2} (c' \Xi^{-1} \kappa)^2 + (n\alpha)^{-1} \sigma_c^{-2} c' \Xi^{-1} \Phi_0 \Xi^{-1} c \text{ and} \]

\[ \text{MSE} (\alpha) = 1 + n\alpha^4 \sigma_c^{-2} (c' \Xi^{-1} \kappa)^2 + (n\alpha)^{-1} \sigma_c^{-2} c' \Xi^{-1} \Phi \Xi^{-1} c. \]
4.2 Optimal Bandwidths

The approximate mean squared errors of \( T_0n \) and \( T_n \) are given in (4.4) and (4.5) respectively. They depend on the smoothing parameter through the approximate bias of order \( n\alpha^4 \) and approximate variance of order \( (n\alpha)^{-1} \). The optimal bandwidth for either \( T_0n \) or \( T_n \) balances the corresponding two terms in (4.4) or (4.5) under an additional assumption stated below.

**Assumption A5.** For \( c \) in (3.6) and (4.1), it holds that \( c'\Xi^{-1}\kappa \neq 0 \).

Assumption A5 ensures that the approximate bias for \( T_0n \) and for \( T_n \) are non-zero. We can then choose \( \alpha \) to minimize either \( \text{MSE}_0(\alpha) \) or \( \text{MSE}(\alpha) \) to obtain the respective solution

\[
\alpha_0^* = \left[ \frac{c'\Xi^{-1}\Phi_0\Xi^{-1}c}{4(c'\Xi^{-1}\kappa)^2} \right]^{1/5} n^{-2/5} \quad \text{or} \quad \alpha^* = \left[ \frac{c'\Xi^{-1}\Phi\Xi^{-1}c}{4(c'\Xi^{-1}\kappa)^2} \right]^{1/5} n^{-2/5}.
\]

Since \( \alpha_0^* \) and \( \alpha^* \) depend on some unknown quantities, namely, \( \Xi, \Phi_0, \Phi, \) and \( \kappa \), they are not feasible. To obtain a feasible version of \( \alpha_0^* \) or \( \alpha^* \), we need to estimate these quantities.

Let \( \hat{\beta}_{21} \) be a consistent estimator of \( \beta_{21}^* \) that is not affected by the smoothing procedure. A rule of thumb (RoT) choice of the bandwidth is to set \( \alpha = sn^{-2/5} \), where \( s \) denotes the sample standard deviation of \( H'_{21}\hat{\beta}_{21} \). We can use this bandwidth as the pilot bandwidth and obtain the SQ estimator \( \hat{\beta}_{sq} = (\hat{\beta}_{1sq}, \hat{\beta}_{2sq})' \). Let \( \hat{\Xi}_n \) and \( \hat{\Lambda}_n \) be defined as in the previous subsection. We propose to estimate \( \Phi_0, \Phi, \) and \( \kappa \) consistently by

\[
\hat{\Phi}_0n = \begin{pmatrix}
\alpha\mathbb{E}_n[\hat{\eta}_1(\hat{\beta}_{21})\hat{\Omega}_{n,22}\hat{\eta}_1(\hat{\beta}_{21})'] & 0_{k_1 \times k_2} \\
0_{k_2 \times k_1} & 0_{k_2 \times k_2}
\end{pmatrix},
\]

\[
\hat{\Phi}_n = \hat{\Phi}_0n + \alpha\sigma_c^{-4}\mathbb{E}_n[\hat{\eta}_1^*(\hat{\beta}_{21})]^2\hat{\Lambda}_n + \alpha\sigma_c^{-2}\mathbb{E}_n[\hat{\eta}_1^*(\hat{\beta}_{21})\hat{\eta}_1^*(\hat{\beta}_{21})\hat{\Lambda}_n + \hat{\Lambda}_n\hat{\eta}_1^*(\hat{\beta}_{21})\hat{\eta}_1^*(\hat{\beta}_{21})],
\]

and

\[
\hat{\kappa}_n = \begin{pmatrix}
\alpha^{-2}\mathbb{E}_n[B_1\{K_\alpha(H'_{21}\hat{\beta}_{21})H'_{21}\hat{\beta}_{21} - (H'_{21}\hat{\beta}_{21})_+}\] & 0_{k_2 \times 1}
\end{pmatrix},
\]

21
respectively, where

\[
\hat{\eta}_1(\hat{\beta}_{21}) = \eta_1(\hat{\beta}_{21}) - E_n[\eta_1(\hat{\beta}_{21})],
\]

\[
\hat{\eta}^*_1(\hat{\beta}_{21}) \equiv \hat{\eta}^*_1(H_{21}; \hat{\beta}_{21}) = c_1' \hat{\Xi}^{-1} \hat{\eta}_1(\hat{\beta}_{21}) [\hat{\Omega}'_{n,12} c_1 + \hat{\Omega}_{n,22} c_2],
\]

and

\[
\hat{\eta}^{**}_1(\hat{\beta}_{21}) \equiv \hat{\eta}^{**}_1(H_{21}; \hat{\beta}_{21}) = \begin{bmatrix}
gerot{0}{k_1 \times k_1} & \hat{\eta}_1(\hat{\beta}_{21}) \hat{\Xi}^{-1} 
gerot{0}{k_2 \times k_1} & \hat{\eta}_1(\hat{\beta}_{21}) \hat{\Xi}^{-1} 
\end{bmatrix}.
\]

Notice that \(\hat{\eta}_1(\hat{\beta}_{21})\) can be replaced by \(\eta_1(\hat{\beta}_{21})\) without affecting the consistency of \(\hat{\Phi}_0n, \hat{\Phi}_n,\) and \(\hat{\kappa}_n\). Then a feasible optimal bandwidth is given by

\[
\hat{\alpha}^*_0 = \left[ \frac{c' \hat{\Xi}^{-1} \hat{\Phi}_0n \hat{\Xi}^{-1} c}{4 \left(c' \hat{\Xi}^{-1} \hat{\kappa}_n\right)^2} \right]^{1/5} n^{-2/5} \text{ or } \hat{\alpha}^* = \left[ \frac{c' \hat{\Xi}^{-1} \hat{\Phi}_n \hat{\Xi}^{-1} c}{4 \left(c' \hat{\Xi}^{-1} \hat{\kappa}_n\right)^2} \right]^{1/5} n^{-2/5}.
\]

We will examine the finite sample performance of the SQ estimators based on the RoT bandwidth, \(\hat{\alpha}^*_0\), and \(\hat{\alpha}^*\) in the next section. \(\hat{\alpha}^*_0\) and \(\hat{\alpha}^*\) are data-driven bandwidths based on two slightly different criteria. Applied researchers are suggested to use both for robustness in practice.

**Remark 3** The optimal bandwidths and their data-driven choices in (4.6) depend critically on Assumption A5 which may fail for some DGPs, e.g., when \(f_D(0) = 1\). To accommodate such cases, we propose the following data-driven bandwidth choices. If \(\hat{\alpha}^*_0\) or \(\hat{\alpha}^*\) exceeds a large positive constant, we use RoT bandwidth; otherwise we use \(\hat{\alpha}^*_0\) or \(\hat{\alpha}^*\). Simulation results in the next section indicate that they work well.
5 A Simulation Study

In this section, we present some simulation results on the finite sample performances of several existing estimators and CIs, and compare them with our SQ estimators and the associated Wald-type CIs.

5.1 The Data Generating Process

The simulation design we use is adapted from Chakraborty, Murphy, and Strecher (2010). For notational simplicity, we follow the notation in Chakraborty, Murphy, and Strecher (2010) in this section. Without loss of generality, we assume $Y_1 = 0$ and $Y_2$ is generated as

$$Y_2 = \gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 + \gamma_5 A_2 + \gamma_6 X_2 A_2 + \gamma_7 A_1 A_2 + v_2,$$

where $v_2 \sim N(0,1)$ and is independent of all the covariates in the above model. We consider binary treatments randomized with probability $1/2$, i.e., $P(A_t = 0) = P(A_t = 1) = 1/2$, $t = 1, 2$. The binary covariates $X_i$'s are generated as $P(X_1 = -1) = P(X_1 = 1) = 1/2$ and $P(X_2 = 1|X_1, A_1) = 1 - P(X_2 = -1|X_1, A_1) = e^{\dagger}(\delta_1 X_1 + \delta_2 A_1)$, where $e^{\dagger}(x) = e^x / (1 + e^x)$.

Recall that $H_1 = X_1$ and $H_2 = (X_1, A_1, X_2)$. Under the above DGP, we have

$$Q_2(H_2, A_2) = \gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 + (\gamma_5 + \gamma_6 X_2 + \gamma_7 A_1) A_2,$$

$$Q_1(H_1, A_1) = \mathbb{E}[Y_1 + \max_{a_2 \in \{0, 1\}} Q_2(H_2, a_2)|H_1, A_1]$$

$$= \gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 + \mathbb{E}\{[\gamma_5 + \gamma_6 X_2 + \gamma_7 A_1]_+|X_1, A_1\}.$$

It is shown in Appendix B that $Q_1(H_1, A_1)$ for the above DGP is equivalent to the following
linear function:

\[ Q_1(H_1, A_1) = \psi_1 + \psi_2 X_1 + (\psi_3 + \psi_4 X_1) A_1, \]

where

\[
\psi_1 = \gamma_1 + \frac{\rho_1 + \rho_3}{2}, \quad \psi_2 = \gamma_2 + \frac{\rho_3 - \rho_1}{2} \left(2e^\dagger(\delta_1) - 1\right), \\
\psi_3 = \gamma_3 - \frac{\rho_1 - \rho_2 + \rho_3 - \rho_4}{2} + \frac{\rho_4 - \rho_2}{2} \left(e^\dagger(-\delta_1 + \delta_2) + e^\dagger(\delta_1 + \delta_2) - 1\right), \text{ and} \\
\psi_4 = \gamma_4 + \frac{\rho_3 - \rho_1}{2} \left(1 - 2e^\dagger(\delta_1)\right) + \frac{\rho_4 - \rho_2}{2} \left(-e^\dagger(-\delta_1 + \delta_2) + e^\dagger(\delta_1 + \delta_2)\right),
\]

in which \( \rho_1 = [\gamma_5 - \gamma_6]_+, \rho_2 = [\gamma_5 - \gamma_6 + \gamma_7]_+, \rho_3 = [\gamma_5 + \gamma_6]_+, \) and \( \rho_4 = [\gamma_5 + \gamma_6 + \gamma_7]_+. \)

In the above DGP, the stage 2 treatment effect is represented by \( \gamma_5 + \gamma_6 X_2 + \gamma_7 A_1, \) and the stage 1 treatment effect is represented by \( \psi_3 + \psi_4 X_1. \) We focus on \( \psi_3 \) and \( \psi_4. \)

Recall that \( Z = \gamma_5 + \gamma_6 X_2 + \gamma_7 A_1. \) Using the notation in Section 3.2, we have \( f_D(z) = P(Z = z). \) Define the “standardized effect size” as \( \phi = |\mu_Z|/\sigma_Z, \) where \( \mu_Z = \mathbb{E}[Z] \) and \( \sigma_Z = \sqrt{\text{Var}(Z)}. \) We consider the following six designs.

**Design 1** \((f_D(0) = 0, \phi = 0.959): \) Consider a completely regular setting where there is a reasonably large stage 2 treatment effect for every subject in the population. This is achieved by setting \( \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1, \gamma_5 = 0.25, \gamma_6 = \gamma_7 = 0.5, \) and \( \delta_1 = \delta_2 = 0.5. \) In this case, \( \psi_3 = 1.491 \) and \( \psi_4 = 1.024. \)

**Design 2** \((f_D(0) = 0, \phi \text{ undefined}): \) Consider a setting where there is a very weak positive stage 2 treatment effect for every subject in the population. This is achieved by setting \( \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1, \gamma_5 = 0.01, \gamma_6 = \gamma_7 = 0, \) and \( \delta_1 = \delta_2 = 0.5. \) In this case, \( \psi_3 = \psi_4 = 1. \) This is a regular but close to non-regularity setting.

**Design 3** \((f_D(0) = 0, \phi = 1.04): \) Consider a setting where there is a very weak positive stage 2 treatment effect for half of the population, but a reasonably large positive effect
for the other half of the population. This is achieved by setting $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1$, $\gamma_5 = 0.01$, $\gamma_6 = 0$, $\gamma_7 = 0.5$, and $\delta_1 = \delta_2 = 0.5$. In this case, $\psi_3 = 1.5$ and $\psi_4 = 1$. This is a regular but close to non-regular setting.

**Design 4** ($f_D(0) = 1/2, \phi = 1$): Consider a setting where there is no stage 2 treatment effect for half of the population, but a reasonably large positive effect for the other half of the population. This is achieved by setting $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1$, $\gamma_5 = \gamma_6 = 0$, $\gamma_7 = 1$, and $\delta_1 = \delta_2 = 0.5$. In this case, $\psi_3 = 2$ and $\psi_4 = 1$. This is a non-regular setting.

**Design 5** ($f_D(0) = 1/2, \phi = 1$): Consider a setting where there is no stage 2 treatment effect for half of the population, but a reasonably large negative effect for the other half of the population. This is achieved by setting $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1$, $\gamma_5 = \gamma_6 = 0$, $\gamma_7 = -1$, and $\delta_1 = \delta_2 = 0.5$. In this case, $\psi_3 = \psi_4 = 1$. This is a non-regular setting.

**Design 6** ($f_D(0) = 1, \phi$ undefined): Consider a setting where there is no stage 2 treatment effect for any subject. This is achieved by setting $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1$, $\gamma_5 = \gamma_6 = \gamma_7 = 0$, and $\delta_1 = \delta_2 = 0.5$. In this case, $\psi_3 = \psi_4 = 1$. This is a fully non-regular scenario.

### 5.2 Estimators, CIs, and Simulation Results

We consider seven estimators of $(\psi_3, \psi_4)'$: the SQ estimator with three different bandwidth choices, the $Q$-learning estimator, the hard-thresholding (HT) estimator of Moodie and Richardson (2010), and the soft-thresholding (ST) estimator of Chakraborty et al. (2010). In the hard- and soft-thresholding methods, the predicted future rewards following the optimal policy are defined as

$$
\hat{Y}_{1i}^{HT} = Y_{1i} + H_{20,i}'\hat{\beta}_{20} + [H_{21,i}'\hat{\beta}_{21}]_+ 1 \left\{ \frac{\sqrt{n}[H_{21,i}'\hat{\beta}_{21}]_+}{\sqrt{H_{21,i}'\hat{\Sigma}_{21}H_{21,i}}} \geq z_{1-\frac{\alpha}{2}} \right\},
$$
and

\[ \tilde{Y}_{1i}^{ST} = Y_{1i} + H'_{20,i} \hat{\beta}_{20} + \left[ H'_{21,i} \hat{\beta}_{21} \right]_+ \left[ 1 - \frac{3H'_{21,i} \hat{\Sigma}_{21} H_{21,i}}{n(H'_{21,i} \hat{\beta}_{21})^2} \right]_+ , \]

respectively, where \( \hat{\Sigma}_{21} \) is the estimated covariance matrix of \( \hat{\beta}_{21} \), \( z_{1-a/2} \) is the \( 1 - a/2 \) quantile of the standard normal distribution. In our simulation, we follow Chakraborty et al. (2010) to consider two choices of the tuning parameter \( a \), 0.08 and 0.20, and denote the corresponding hard-thresholding methods as HT_{0.08} and HT_{0.20}.

Since both hard- and soft-thresholding estimators still involve non-smooth functions of \( \hat{\beta}_{21} \), they suffer from the same problem as the original Q-learning estimator as far as statistical inference is concerned. We compare eight CIs: the first three are Wald-type CIs constructed from our SQ estimators; the fourth one is the FACI in Laber et al. (2014); and the last four are centered bootstrap percentile CIs based on the last four estimators respectively. For each estimator, we report its bias and MSE. For each CI, we report its coverage probability (CP) and interval length (IL).

In our simulation, we consider two sample sizes, 200 and 1000, with 10,000 replications and 500 bootstrap replications. The CDF and PDF associated with the Epanechnikov kernel function is used. For the case when either \( \hat{\alpha}_0^* \) or \( \hat{\alpha}^* \) are large, we experimented different threshold values and found a value equal to or greater than 10 to be a reasonable choice. We set it to be 10 in both the simulation and empirical application. All estimators and CIs perform reasonably well for \( \psi_1, \psi_2 \) and for all designs, so we omit these results for the sake of brevity. Tables 1 and 2 present results for \( \psi_3 \), while Tables 3 and 4 present results for \( \psi_4 \). Like Chakraborty et al. (2010), we find that all estimators and CIs perform reasonably well for \( \psi_4 \) and for all designs, and thus focus our discussion below on Tables 1 and 2 for \( \psi_3 \). We first summarize the performances of the estimators and CIs for each
design and then offer some general observations from Tables 1 and 2.

Design 1 \((p_0 = 0, \phi = 0.959\) is a completely regular setting, so we expect the \(Q\)-learning and our SQ algorithms to perform well. (i) In terms of the bias and MSE, the \(Q\)-learning and our SQ estimators perform better than the threshold estimators for both sample sizes except when \(n = 200\) in which case the soft-thresholding estimator has the smallest MSE. (ii) In terms of the CP, all the CIs are comparable, except that the ST CI under-covers when \(n = 1000\). (iii) In terms of the IL, the \(Q\)-learning and our SQ CIs are tighter. Considering all the performance measures and the cost of computation, the three SQ estimators emerged as the best. For example, on a computer with Intel 2.00 GHz processor, when \(n = 1,000\), our smoothed \(Q\)-learning method takes less than 4 minutes for 10,000 repetitions with 10 parallel R workers. In the same setup, the HT\(_{0.20}\) method takes about 11.7 hours. The computation time for all other methods using bootstrap are similar to that of the HT\(_{0.20}\) method.

Design 2 \((p_0 = 0, \phi \text{ undefined})\) is a regular setting with a very weak positive stage 2 treatment effect for every subject in the population. This setting is close to the fully non-regular setting. Several observations emerge from results in Tables 1 and 2. (i) The ST estimator dominates the rest in terms of bias for both sample sizes. However, the \(Q\)-learning and our SQ estimators dominate in terms of MSE for both sample sizes. (ii) In terms of the CP, although all CIs over-cover, the soft-thresholding CI produces coverage rate closer to the nominal rate, followed by the \(Q\)-learning and our SQ CIs. (iii) In terms of the IL, the \(Q\)-learning CI is the tightest, followed by the soft-thresholding and our SQ CIs. Overall, the soft-thresholding, the \(Q\)-learning and our SQ methods are comparable,
but the SQ methods are less costly computationally.

**Design 3** \((p_0 = 0, \phi = 1.04)\) is a regular setting with a very weak positive stage 2 treatment effect for half of the population, but a reasonably large positive effect for the other half of the population. Qualitatively the performances of the estimators and CIs are similar to those in Design 1. (i) In terms of the bias, all estimators except the ST estimator are comparable for the smaller sample, and the two hard-thresholding estimators have the smallest biases for the larger sample size. (ii) In terms of the MSE, the \(Q\)-learning and our SQ estimators dominate for the smaller sample size, and the hard-thresholding estimators improve greatly as the sample size increases. (iii) In terms of the CP, the FACI and HT\(_{0.08}\) CIs have better coverage rates for the smaller sample. The SQ CI with bandwidth \(\hat{\alpha}^*\) and the HT\(_{0.20}\) CI have better coverage rates for the larger sample. The CIs based on the \(Q\)-learning method and the soft-thresholding method severely under-cover for both sample sizes. (iv) In terms of the IL, our SQ CIs are in general tighter. In sum, our SQ algorithm with bandwidth \(\hat{\alpha}^*\) and the HT\(_{0.20}\) method perform the best.

**Design 4** \((p_0 = 1/2, \phi = 1)\) is a non-regular setting with no stage 2 treatment effect for half of the population, but a reasonably large positive effect for the other half of the population. Qualitatively the performances of the estimators and CIs are similar to Design 3 with the SQ method using bandwidth \(\hat{\alpha}^*\) and the HT\(_{0.08}\) method emerging as the winners.

**Design 5** \((p_0 = 1/2, \phi = 1)\) is a non-regular setting with no stage 2 treatment effect for half of the population, but a reasonably large negative effect for the other half of the population. (i) The soft-thresholding estimator has the smallest bias and MSE for both sample sizes. (ii) In terms of the CP and IL, our SQ methods and the threshold methods are in general comparable. The \(Q\)-learning CI severely under-covers. In this design, the soft-thresholding method appears to be the best, followed by our SQ method.
with bandwidth $\hat{\alpha}^*$ and the HT$_{0.20}$ method.

**Design 6** ($p_0 = 1, \phi$ undefined) is a fully non-regular setting with no treatment effect for any subject in stage 2. Qualitatively the performances of the estimators and CIs are similar to those in Design 2. (i) The soft-thresholding estimator dominates in terms of the bias for both sample sizes. The $Q$-learning and our SQ estimators dominate in terms of the MSE for both sample sizes. (ii) In terms of the CP, all the CIs over-cover, but the soft-thresholding CI, the $Q$-learning, and our SQ CIs have better CPs than the hard-thresholding and FACI CIs. The $Q$-learning CI has the shortest length. Consider all the performance measures, the $Q$-learning method, our SQ method, and the soft-thresholding method are the winners.

To sum up, the $Q$-learning method has the shortest CIs for all the designs, but severely under-covers in Designs 3, 4, and 5. The soft-thresholding method also severely under-covers in Designs 3 and 4. The FACI is in general valid but can be very conservative. In terms of the MSE, the hard-thresholding estimators and our SQ estimators are comparable for all the designs. In terms of the CP, the hard-thresholding CIs over-cover more than our SQ CIs in Designs 2 and 6. Our SQ method, especially the one using the bandwidth $\hat{\alpha}^*$, and the associated Wald CI perform the best for the designs considered in terms of MSE, CP, and IL. Moreover it is the least costly computationally.

### 5.3 Expected Outcome for Different Methods

In the above section, we focused on the estimation and inference of parameters of the first-stage Q-function. The objective of a Dynamic Treatment Regime is to find the best rule $\pi = (\pi_1, \pi_2)$ to maximize $\mathbb{E}(Y_1^{\pi} + Y_2^{\pi})$. Luedtke and van der Laan (2016) develop an inference procedure for the value under the optimal treatment regime. In this section, we investigate the optimality (in terms of $\mathbb{E}(Y_1^{\pi} + Y_2^{\pi})$) of different estimation methods.
We present the estimated $\mathbb{E}(Y_{1}^{\pi} + Y_{2}^{\pi})$ under different learning algorithms, and compare them with that under the infeasible algorithm for which we know the true parameters of Q-functions.

Specifically, the simulation designs are the same as in the previous simulations. Given one design, the number of repetitions is set to be $R = 1000$. For each repetition $r$, we first estimate the Q-functions using different methods, and then generate $M = 1000$ new observations from the same population. We then use the estimated Q-functions to make a decision on the choice of $A_1, A_2$ for the new observations $m = 1, \cdots, M$, and record the outcome $Y_{jr,m}, j = 1, 2$, where $Y_{1r,m} = 0$ by the design. We compute and use the average

$$\bar{Y}_2 = \frac{1}{R} \sum_{r=1}^{R} \left( \frac{1}{M} \sum_{m=1}^{M} Y_{2r,m} \right)$$

to compare the performances of different methods. The average $\bar{Y}_2$ is first across all the new observations and then across all the repetitions. Averaging across all the new observations is to estimate the expectation in $\mathbb{E}(Y_{1}^{\pi} + Y_{2}^{\pi})$ under one specific rule $\pi$, and averaging across all the repetitions is because the estimated decision rules are themselves random.

We summarize the results in Table 5. First, the decision rule using infeasible true value of $\psi_3, \psi_4$ lead to the highest expected outcome, regardless of the sample size $n$. This is expected since in theory, it is the optimal decision rule. Second, in the first three designs, the original Q-learning and our smoothed Q-learning methods tend to perform better than the thresholding methods. Third, in the last three designs, all methods (including the optimal one) lead to exactly the same estimated expectation. A further investigation reveals that it is due to the details in these three simulation designs.

**Remark 4** Take Design 5 for an example, where $Y_2 = 1 + X_1 + A_1 + X_1A_1 - A_1A_2 + v_2$. 

<--- Table 5 --->
In this case, \( \psi_3 = \psi_4 = 1 \), so that \( \hat{\psi}_3 > 0, \hat{\psi}_4 > 0 \) given the consistency. The decision rules are \( A_1 = 1(\hat{\psi}_3 + \hat{\psi}_4 X_1 \geq 0) \), \( A_2 = 1(-A_1 \geq 0) \). For any method, given \( \hat{\psi}_3 \geq \hat{\psi}_4 \), if \( X_1 = 1 \), then \( A_1 = 1, A_2 = 0 \), and \( Y_2 = 4 + v_2 \); if \( X_1 = -1 \), then \( A_1 = 1, A_2 = 0 \), and \( Y_2 = 4 + v_2 \); if \( X_1 = -1 \), then \( A_1 = 0, A_2 = 1 \), and \( Y_2 = v_2 \). In other words, the value of \( Y_2 \) does not depend on whether \( \hat{\psi}_3 \geq \hat{\psi}_4 \) or \( \hat{\psi}_3 < \hat{\psi}_4 \) (or equivalently, the method). This implies that \( Y_{2r,m} \) does not depend on the method, and so is \( \hat{Y}_2 \). Similar discussions apply to Designs 4 and 6.

6 Summary of the Algorithm and An Empirical Application

6.1 Summary of the Algorithm

We provide a summary of the smoothed Q-learning algorithm for applied researchers.

**Smoothed Q-Learning Algorithm**

**Step 1**: regress \( Y_2 \) on \( (H'_{20}, H'_{21} A_2)' \) to obtain \( \hat{\beta}_2 = (\hat{\beta}_{20}', \hat{\beta}_{21}')' \);

**Step 2**: define \( \hat{Y}_{1sq} = Y_1 + H_{20}\hat{\beta}_{20} + (H'_{21}\hat{\beta}_{21})K_\alpha(H'_{21}\hat{\beta}_{21}) \), \( K(\cdot) \) is any CDF with a density function in (3.4). \( \alpha \) is estimated either by the rule of thumb (RoT) bandwidth \( \hat{\alpha}_{\text{RoT}} = \text{sd}(H'_{21}\hat{\beta}_{21})n^{-2/5} \), or by \( \hat{\alpha}^* \) or \( \hat{\alpha}^* \) in (4.6);

**Step 3**: calculate \( \hat{\Omega}_n \) by (3.5) and construct Wald-type confidence intervals.

6.2 An Empirical Analysis of the CATIE-AD Data

In this section, we study the optimal dynamic treatment regime for the Alzheimer’s Disease, using data from the Clinical Antipsychotic Trials of Intervention Effectiveness-Alzheimer’s Disease (CATIE-AD) study. This NIMH funded study aims to assess the effectiveness of the second-generation antipsychotic drugs (olanzapine, quetiapine, and risperidone) in treating delusions, hallucinations, aggression, and agitation for patients.
with Alzheimer’s disease.

There are four phases in this 36 weeks’ study, with double-blind Phases I and II, an open-label Phase III, and an open-choice Phase IV. In Phase I, patients were randomly assigned to receive either one of the three drugs or placebo. In Week 12, the investigator decided whether the current medication was optimal or if it would be more beneficial to try another randomized medication. If the original medication was continued, the patient remained in Phase I for the rest 24 weeks, otherwise the patient proceeded to Phase II. The patient could further proceed to Phases III and IV later on at the investigator’s discretion. A detailed description of the study design and primary analysis can be found in Schneider et al. (2001) and Schneider et al. (2006).

For the purpose of estimating the optimal dynamic treatment regime, we focus on two-stage treatment and define the two stages for a patient as follows: if the patient takes the same medication throughout the study (i.e. having only Phase I), then his/her Stage 1 is defined as the first 12 weeks, and Stage 2 is defined as the remaining 24 weeks. If the patient has both Phases I and II, then his/her Stages 1 and 2 correspond to Phases I and II in the study. There are 407 observations in the original data set. Some patients miss either Phase I or Phase II, and this leaves 287 observations in our data set.

In the empirical analysis, first, we include three covariates: (i) the Neuropsychiatric Inventory (NPI), which measures the level of delusion and hallucination, with a higher score indicating a worse condition of the patient; (ii) age; (iii) gender, with female to be 0 and male to be 1. Second, we use the Clinical Global Impression of Change (CGIC, on a 1-7 scale) as the dependent variable, and negate the original CGIC score so that a higher score represents a better outcome. The covariates are measured at the beginning of each phase, and CGIC is measure at the end of each phase. Third, we choose two different definitions of the treatment variables for illustration purpose, detailed in the
following two subsections. In either definition, we are primarily interested in estimating the optimal dynamic treatment regime so that proper guidance is available to provide personalized treatment decision. The $Q$-functions are specified as

$$Q_t(H_t, A_t) = \beta_{t0,0} + \beta_{t0,1}NPI_t + \beta_{t0,2}\text{Age} + \beta_{t0,3}\text{Gender}$$

$$+ (\beta_{t1,0} + \beta_{t1,1}NPI_t + \beta_{t1,2}\text{Age} + \beta_{t1,3}\text{Gender})A_t, \ t = 1, 2,$$

and we used the CDF and PDF of Epanechnikov kernel in the following empirical analysis.

### 6.2.1 Treatment Definition I

In this definition, for both stages, the treatment variable is defined as $A_t = 1$ if the patient takes olanzapine and 0 otherwise. This definition is due to two reasons. First, in the design of the study, patients who take drugs in the Stage 1 do not have the choice of switching to placebo in Stage 2. This prevents us from defining treatment variable as 1 if the patient takes one of the drugs and 0 for placebo, which violates the sequential randomization assumption. Second, previous studies found that olanzapine is more effective in treating Schizophrenia (see Tran et al. (1997)), we are interested if olanzapine is also more effective than other drugs in treating the Alzheimer’s Disease. Using this definition, there are 70 patients taking olanzapine and 217 otherwise in Stage 1, and 59 patients taking olanzapine and 228 otherwise in Stage 2. The estimation results are summarized in Table 6. Stage 1 results include those from the smoothed Q-learning algorithms, and Stage 2 is a standard linear regression.

Table 6

For Stage 1, due to the similarity among our methods with different bandwidths, we focus on the results with bandwidth $\hat{\alpha}^*$ for interpretation. First, NPI has a statistically
significant negative main effect, indicating that a higher score of NPI leads to a lower CGIC score. From the estimate of the interaction term between NPI and treatment, we find that olanzapine is more effective in reducing the symptoms of Alzheimer’s Disease for a patient with a higher NPI score. Second, we do not find a statistically significant effect of age, both for the main effect and its interaction effect with the treatment. Third, male tends to have a lower CGIC score. From the estimate of the interaction term between gender and treatment, olanzapine tends to be more effective for female patients than male patients. However, these two findings are not statistically significant. Fourth, the treatment tends to have a positive main effect, although it is not statistically significant. The Stage 2 results are qualitatively similar to those of Stage 1 and follow a similar discussion.

The estimated optimal DTR for the both stages (using $\hat{\alpha}^*$ for Stage 1) are given by

$$\pi_{dp}(h_1) = 1 \{2.819 + 0.027NPI_1 - 0.049Age - 0.125Gender \geq 0\},$$

$$\pi_{dp}(h_2) = 1 \{3.972 + 0.005NPI_2 - 0.052Age - 0.866Gender \geq 0\}.$$

### 6.2.2 Treatment Definition II

In this definition, we are interested in comparing olanzapine and risperidone in the first stage, and to decide whether one should change the drug choice in the second stage. Specifically, $A_1 = 1$ if the patient takes olanzapine and 0 if he/she takes risperidone, $A_2 = 1$ if they switch to another drug and 0 otherwise. This can be applied to comparing any two types of drugs, and we compare olanzapine and risperidone here for illustration purpose. This criterion gives us a subset of data with 129 observations. In Stage 1, there are 70 patients taking olanzapine and 59 taking risperidone. In Stage 2, 91 patients switch
and 38 do not. The estimation results are summarized in Table 7.

For Stage 1, we focus on the results with bandwidth $\hat{\alpha}^*$ for interpretation. First, NPI again has a negative main effect on CGIC score, but not statistically significant. Its interactive effect with treatment is positive and statistically significant, indicating that olanzapine is more effective than risperidone for a patient with a higher NPI score. Second, we do not find any statistically significant effect of age. Third, male still tends to have a lower CGIC score, but not statistically significant. Fourth, the treatment tends to have a positive but statistically insignificant main effect, providing weak evidence that olanzapine is better than risperidone. This is in line with the previous finding in Tran et al. (1997) that olanzapine is more effective than risperidone in treating Schizophrenia. Stage 2 in general follows a similar discussion, and the main effect of treatment variable suggests that switching medicine at Stage 2 might be a good choice.

The estimated optimal DTR for the both stages (using $\hat{\alpha}^*$ for Stage 1) are given by

$$
\pi_{d_1}^{dp}(h_1) = 1 \{ 1.540 + 0.036 \text{NPI}_1 - 0.043 \text{Age} + 0.115 \text{Gender} \geq 0 \},
$$

$$
\pi_{d_2}^{dp}(h_2) = 1 \{ 2.561 + 0.024 \text{NPI}_2 - 0.054 \text{Age} - 0.276 \text{Gender} \geq 0 \}.
$$

7 Concluding Remarks

In this paper, we propose a SQ algorithm for estimating optimal dynamic treatment regimes. Under assumptions in Section 3, we have established asymptotic properties of the SQ estimator and developed valid Wald-type inference procedures for parameters in the
unknown $Q$-functions. In a simulation study investigating the finite sample performance of several estimators and CIs, the SQ estimators and the associated Wald-type CIs perform favorably in terms of accuracy, computational simplicity, and robustness to the degree of non-regularity. We illustrate the new method by analyzing the data from the CATIE-AD study. The estimated optimal dynamic treatment regime allows the clinician to recommend treatment based on the NPI score, age, and gender. We find that olanzapine is in general more effective than other medications in reducing symptoms associated with the Alzheimer’s Disease. In particular, olanzapine tends to be more effective for a patient with a higher NPI score, and more effective for female patients than for male patients. We do not find any statistically significant age effect.

Like other methods, the SQ algorithm has straightforward extensions to more complex dynamic treatment regimes such as dynamic treatments involving multiple stages and multi-valued treatments. One important and also challenging extension is statistical inference for the value of the optimal treatment, see Chakraborty and Moodie (2013) for a discussion on the technical challenges involved. We hope that the technique of smoothing an otherwise non-smooth function in the parameter of interest may simplify the task of developing valid inference for the value function. This is beyond the scope of this paper, but currently under investigation by the authors.

Acknowledgements

We thank Han Hong and participants in the Causal Inference Workshop in Department of Biostatistics at University of Washington, and in Department of Economics at University of Colorado Boulder for helpful discussions. We are grateful to Lon Schneider for sharing the CATIE-AD data with us. Su acknowledges the funding support provided by the Lee Kong Chian Fund for Excellence. This manuscript reflects the views of the authors and may not reflect the opinions or views of the CATIE-AD study investigators.
Supporting Information

Additional information for this article is available online including the proofs in Appendix A, details of the generative model for simulation in Appendix B, and a guide for applied researchers to use our algorithm in Appendix C.

A Proof of The Main Results

We adopt the standard notation in the empirical process literature (e.g., van der Vaart, 1998): For any $K$-dimensional random function $g : W \to \mathbb{R}^K$, we define

$$
\mathbb{E}_n [g] = \mathbb{E}_n [g (W)] = \frac{1}{n} \sum_{i=1}^{n} g (W_i), \text{ and}
$$

$$
\mathbb{G}_n [g] = \mathbb{G}_n [g (W)] = n^{-1/2} \sum_{i=1}^{n} \{g (W_i) - \mathbb{E} [g (W_i)]\}.
$$

Let $C$ denote a generic large positive constant that may vary across lines. We also use $C_1, C_2, \ldots$ to denote finite positive constants.

Proof of Theorem 1 Let $M_n (\beta) = \mathbb{E}_n [m_n (\beta)] = \frac{1}{n} \sum_{i=1}^{n} m_n (W_i; \beta)$, where $m_n (W; \beta) = (m_{1n}(W; \beta)' , m_{2n}(W; \beta)')'$ with

$$
m_{1n}(W; \beta) = B_1 [Y_1 + H'_{20}\beta_{20} + K_{a}(H'_{21}\beta_{21})H'_{21}\beta_{21} - B'_1\beta_1],
$$

$$
m_{2n}(W; \beta) = B_2(Y_2 - B'_2\beta_2).
$$

Note that $M_n (\hat{\beta}_{sq}) = 0$, where $\hat{\beta}_{sq} = (\hat{\beta}_{1, sq}, \hat{\beta}_{2})'$.

Let $M (\beta) \equiv \mathbb{E} [m (\beta)] = \mathbb{E} [m (W; \beta)]$, where $m(W; \beta) = (m_1(W; \beta)', m_2(W; \beta)')'$ with $m_1(W; \beta) = B_1 [Y_1 + H'_{20}\beta_{20} + (H'_{21}\beta_{21})_+ - B'_1\beta_1]$. Note that under Assumption
A2(i), $M(\beta) = 0$ if and only if $\beta = \beta^* = (\beta_1^*, \beta_2^*)'$, where

$$
\beta_1^* = (\mathbb{E}[B_1 B_1'])^{-1} \mathbb{E}[B_1 (Y_1 + H_20 \beta_2^* + (H_21 \beta_2^*)_1)],
$$
$$
\beta_2^* = (\mathbb{E}[B_2 B_2'])^{-1} \mathbb{E}[B_2 Y_2].
$$
\tag{A.1}

By Theorem 5.9 in van der Vaart (1998), the claim in the theorem follows if we can show that

$$
\sup_{\beta \in \mathcal{B}} \|M_n(\beta) - M(\beta)\| = o_P(1). \tag{A.2}
$$

We prove (A.2) by showing that (i) $\sup_{\beta \in \mathcal{B}} \|M_n(\beta) - \mathbb{E}[M_n(\beta)]\| = o_P(1)$ and (ii) $\sup_{\beta \in \mathcal{B}} \|\mathbb{E}[M_n(\beta)] - M(\beta)\| = o(1)$.

To show (i), we consider a class of functions

$$
\mathcal{M}_1 = \left\{ m_{\beta} : m_{\beta}(w) = \begin{pmatrix} m_{1\beta}(y_1, b_1, h_{20}, h_{21}) \\ m_{2\beta}(y_2, b_2) \end{pmatrix}, \beta \in \mathcal{B} \right\},
$$

where $m_{1\beta}(y_1, b_1, h_{20}, h_{21}) = b_1 [y_1 + h_{20}' \beta_{20} + K(h_{21}' \beta_{21}/\alpha)h_{21}' \beta_{21} - b_1' \beta_1]$, and $m_{2\beta}(y_2, b_2) = b_2(y_2 - b_2' \beta_2)$. Note that the functions in $\mathcal{M}$ are composed of the classes of functions

$$
\begin{align*}
\mathcal{M}_{1,1} &= \{ b_1 [y_1 + h_{20}' \beta_{20} - b_1' \beta_1] : \beta_1, \beta_{20} \in \mathcal{B}_{20} \}, \\
\mathcal{M}_{1,2} &= \{ h_{21}' \beta_{21} : \beta_{21} \in \mathcal{B}_{21} \}, \\
\mathcal{M}_{1,3} &= \{ K(h_{21}' \beta_{21}/\alpha) : \beta_{21} \in \mathcal{B}_{21} \}, \\
\mathcal{M}_{1,4} &= \{ b_2 [y_2 - b_2' \beta_2] : \beta_2 \in \mathcal{B}_2 \}.
\end{align*}
$$

By Lemma 2.13 in Pakes and Pollard (1989), $\mathcal{M}_{1,1}$, $\mathcal{M}_{1,2}$ and $\mathcal{M}_{1,4}$ are all Euclidean class of functions with envelopes respectively given by $C(\|b_1 y_1\| + \|b_1 h_{20}\| + \|b_1'\|), C \|h_{21}\|,$
and \( C(\| b_2 y_2 \| + \| b_2 b'_2 \|) \). By Lemma 22 in Nolan and Pollard (1987), \( \mathcal{M}_{1,3} \) is also Euclidean for a constant envelope. As a result, \( \mathcal{M} \) is Euclidean by Lemma 2.14 in Pakes and Pollard (1989). In addition, one can readily check that \( m_\beta (W) \) has an envelope function that can be written as the summation of two terms: one is not associated with \( \mathcal{M}_{1,3} \) and has finite first moment, and the other is associated with \( \mathcal{M}_{1,3} \) and has finite second moment. Then by Theorem 37 in Pollard (1984, p.34) which allows the function class to be \( n \)-dependent\(^1\) and Lemma 2.8 in Pakes and Pollard (1989) which does not consider \( n \)-dependent function class, we have \( \sup_{\beta \in \mathcal{B}} \| M_n (\beta) - \mathbb{E}[M_n (\beta)] \| = o (1) \) almost surely, implying (i).

Next, we show that

\[
\sup_{\beta \in \mathcal{B}} \| \mathbb{E}[M_n (\beta) - M (\beta)] \| = \sup_{\beta \in \mathcal{B}} \| \mathbb{E}[B_1 \left[ H'_{21} \beta_{21} K_\alpha (H'_{21} \beta_{21}) - (H'_{21} \beta_{21})_+ \right] ] \| = o (1) .
\]

Let \( g_1 (z; \beta_{21}) = \int_{-\infty}^{z} r_1 (s; \beta_{21}) s f_C (s; \beta_{21}) ds \) and \( g_1^{(1)} (z; \beta_{21}) = d g_1 (z; \beta_{21}) / dz = r_1 (z; \beta_{21}) z f_C (z; \beta_{21}) \). By Assumption A1(iii),

\[
\mathbb{E} \left\{ B_1 \left[ H'_{21} \beta_{21} K_\alpha (H'_{21} \beta_{21}) - (H'_{21} \beta_{21})_+ \right] \right\} \\
= \mathbb{E} \left\{ r_1 (H'_{21} \beta_{21}; \beta_{21}) \left[ H'_{21} \beta_{21} K_\alpha (H'_{21} \beta_{21}) - (H'_{21} \beta_{21})_+ \right] \right\} \\
= (1 - p_{\beta_{21}}) \int_{-\infty}^{\infty} r_1 (z; \beta_{21}) \left[ z K (\frac{z}{\alpha}) - z_+ \right] f_C (z; \beta_{21}) dz \\
+ p_{\beta_{21}} \sum_{z_d} r_1 (z_d; \beta_{21}) \left[ z_d K (\frac{z_d}{\alpha}) - (z_d)_+ \right] f_D (z_d; \beta_{21}) \\
\equiv (1 - p_{\beta_{21}}) I_C (\beta_{21}) + p_{\beta_{21}} I_D (\beta_{21}) , \text{ say}.
\]

\(^1\mathcal{M}_{1,3} \) here is \( n \)-dependent through the bandwidth parameter \( \alpha \).
For the first term, we have

\[
I_C (\beta_{21}) = \int_{-\infty}^{\infty} r_1 (z; \beta_{21}) z f_C (z; \beta_{21}) \left[ K \left( \frac{\tilde{z}}{\alpha} \right) - 1 \{ z \geq 0 \} \right] dz
\]

\[
= \int_{-\infty}^{\infty} \left[ K \left( \frac{\tilde{z}}{\alpha} \right) - 1 \{ z \geq 0 \} \right] d g_1 (z; \beta_{21})
\]

\[
= - \frac{1}{\alpha} \int_{-\infty}^{\infty} g_1 (z; \beta_{21}) k \left( \frac{\tilde{z}}{\alpha} \right) d z + g_1 (0; \beta_{21})
\]

\[
= - \int_{-1}^{1} [g_1 (u\alpha; \beta_{21}) - g_1 (0; \beta_{21})] k (u) du
\]

\[
= - \alpha \int_{0}^{1} \int_{-1}^{1} g_1^{(1)} (tu\alpha; \beta_{21}) u k (u) dudt,
\]

where the second line follows from change of variables, the third line follows from integration by parts, and the last line follows from a first order Taylor series expansion with an integral remainder and the fact that \( \int_{-1}^{1} u k (u) du = 0 \) under Assumption A3(i).

Next, we note that

\[
I_D (\beta_{21}) = \sum_{z_d > 0} r_1 (z_d; \beta_{21}) z_d \left[ K \left( \frac{z_d}{\alpha} \right) - 1 \right] f_D (z; \beta_{21}) + \sum_{z_d < 0} r_1 (z_d; \beta_{21}) z_d K \left( \frac{z_d}{\alpha} \right) f_D (z; \beta_{21}).
\]

For any fixed \( z_d > 0 \), \( K \left( \frac{z_d}{\alpha} \right) = 1 \) for sufficiently small \( \alpha \) when \( n \) is large enough as \( k (\cdot) \) has compact support \([-1, 1]\). This ensures the first term in the last expression to vanish in large samples. Similarly, for any fixed \( z_d < 0 \), \( K \left( \frac{z_d}{\alpha} \right) = 0 \) for sufficiently small \( \alpha \) when \( n \) is large enough, ensuring the second term to vanish in large samples. As a result, \( I_D (\beta_{21}) = 0 \) uniformly in \( \beta_{21} \) when \( n \) is sufficiently large. It follows that for sufficiently
large $n$,

$$
\sup_{\beta_{21} \in B_{21}} \left\| \mathbb{E} \left\{ B_1 \left[ H_{21}' \beta_{21} K_h (H_{21}' \beta_{21}) - (H_{21}' \beta_{21})_+ \right] \right\} \right\|
$$

$$
= \sup_{\beta_{21} \in B_{21}} (1 - p_{\beta_{21}}) \| I_C (\beta_{21}) \| \leq \alpha \sup_{\beta_{21} \in B_{21}} \left\| \int_0^1 \int_{-1}^1 g^{(1)}_1 (tu \alpha; \beta_{21}) u k(u) du dt \right\|
$$

$$
= \alpha^2 \sup_{\beta_{21} \in B_{21}} \left\| \int_0^1 t \int_{-1}^1 r_1 (tu \alpha; \beta_{21}) f_C (tu \alpha; \beta_{21}) u^2 k(u) du dt \right\|
$$

$$
\rightarrow \alpha^2 \sup_{\beta_{21} \in B_{21}} \| r_1 (0; \beta_{21}) f_C (0; \beta_{21}) \| \int_{-1}^1 u^2 k(u) du = O (\alpha^2),
$$

where the convergence follows from the dominated convergence theorem (DCT), Assumptions A1(iv) and A3. ■

**Proof of Theorem 2** Expanding $\mathbb{E}_n[m_n(\hat{\beta}_{sq})]$ around $\beta^*$ yields

$$
0 = \mathbb{E}_n[m_n(\hat{\beta}_{sq})] = \mathbb{E}_n[m_n(\beta^*)] + \mathbb{E}_n[\nabla_{\beta} m_n(\bar{\beta})] (\hat{\beta}_{sq} - \beta^*), \quad (A.3)
$$

where $\bar{\beta}$ lies between $\hat{\beta}_{sq}$ and $\beta^*$ element by element. We prove the theorem by showing that (i) $\sqrt{n}\mathbb{E}_n[m_n(\beta^*)] \xrightarrow{d} N (0, \Lambda)$, (ii) $\Xi_n \equiv \mathbb{E}[\nabla_{\beta} m_n(\beta^*)] = \Xi + o(1)$, and (iii) $\mathbb{E}_n[\nabla_{\beta} m_n(\bar{\beta})] = \mathbb{E}[\nabla_{\beta} m_n(\beta^*)] + o_P (1)$. Then we have

$$
\sqrt{n}(\hat{\beta}_{sq} - \beta^*) = - \left[ \mathbb{E}_n[\nabla_{\beta} m_n(\bar{\beta})] \right]^{-1} \sqrt{n}\mathbb{E}_n[m_n(\beta^*)] \xrightarrow{d} N (0, \Xi^{-1} \Lambda \Xi^{-1'}).
$$

as $\Xi$ is nonsingular under Assumption A2(i).

**Step 1.** We first prove (i) $\sqrt{n}\mathbb{E}_n[m_n(\beta^*)] \xrightarrow{d} N (0, \Lambda)$. Note that $\sqrt{n}\mathbb{E}_n[m_n(\beta^*)] =
\( \mathbb{G}_n[m_n(\beta^*)] + \sqrt{n} \mathbb{E}[m_n(\beta^*)] \). Observe that

\[
\sqrt{n} \mathbb{E}[m_n(\beta^*)] = \begin{pmatrix}
\sqrt{n} \mathbb{E}[m_{1n}(\beta^*)] \\
\sqrt{n} \mathbb{E}[m_{2}(\beta^*)]
\end{pmatrix} = \begin{pmatrix}
\sqrt{n} \mathbb{E}[m_{1n}(\beta^*)] \\
0
\end{pmatrix}.
\]

In view of the fact that \( Q_1(H_1, A_1; \beta_1^*) = B_1' \beta_1^* \) is the population least squares projection of \( Y_1 + H_{20}' \beta_{20}^* + (H_{21}' \beta_{21}^*)^+ \) on \( (H_1, A_1) \) and \( B_1 = (H_{10}', A_1 H_{11}')' \), we have

\[
\mathbb{E} \{ B_1 [Y_1 + H_{20}' \beta_{20}^* + (H_{21}' \beta_{21}^*)^+ - B_1' \beta_1^*] \} = 0.
\]

It follows that

\[
\sqrt{n} \mathbb{E}[m_{1n}(\beta^*)] = \sqrt{n} \mathbb{E} \{ B_1 [Y_1 + H_{20}' \beta_{20}^* + K_\alpha (H_{21}' \beta_{21}^*) H_{21}' \beta_{21}^* - B_1' \beta_1^*] \}
= \sqrt{n} \mathbb{E} \{ B_1 [K_\alpha (H_{21}' \beta_{21}^*) H_{21}' \beta_{21}^* - (H_{21}' \beta_{21}^*)^+] \}.
\]

Following the analysis of \( \mathbb{E} \{ B_1 [K_\alpha (H_{21}' \beta_{21}^*) H_{21}' \beta_{21}^* - (H_{21}' \beta_{21}^*)^+] \} \) in the proof of Theorem 1, for sufficiently large \( n \),

\[
\sqrt{n} \mathbb{E}[m_{1n}(\beta^*)] = \sqrt{n} \mathbb{E} \{ B_1 [K_\alpha (H_{21}' \beta_{21}^*) H_{21}' \beta_{21}^* - (H_{21}' \beta_{21}^*)^+] \}
= \sqrt{n}(1 - p) I_C (\beta_{21}^*) = \sqrt{n}(1 - p) O (\alpha^2) = o(1), \quad (A.4)
\]

where the last step follows from Assumption A3(ii).

Next, we calculate the second moment of \( m_n(\beta^*) \). Observe that

\[
\Lambda_n \equiv \mathbb{E} [m_n(\beta^*) m_n(\beta^*)'] = \begin{pmatrix}
\mathbb{E}[m_{1n}(\beta^*) m_{1n}(\beta^*)'] & \mathbb{E}[m_{1n}(\beta^*) m_{2}(\beta^*)'] \\
\mathbb{E}[m_{2}(\beta^*) m_{1n}(\beta^*)'] & \mathbb{E}[m_{2}(\beta^*) m_{2}(\beta^*)']
\end{pmatrix} \equiv \begin{pmatrix}
\Lambda_{n,11} & \Lambda_{n,12} \\
\Lambda_{n,12}' & \Lambda_{22}
\end{pmatrix},
\]

42
where

\[
\Lambda_{n,11} = \mathbb{E} \left\{ B_1 B'_1 [Y_1 - B'_1 \beta_1^* + H'_{20} \beta_2^{*} + K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^*]^2 \right\},
\]

\[
\Lambda_{n,12} = \mathbb{E} \left\{ B_1 B'_2 [Y_1 - B'_1 \beta_1^* + H'_{20} \beta_2^{*} + K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^*] [Y_2 - B'_2 \beta_2^*] \right\},
\]

\[
\Lambda_{22} = \mathbb{E} \left\{ B_2 B'_2 [Y_2 - B'_2 \beta_2^*]^2 \right\}.
\]

Let \(\Lambda_{11}\) and \(\Lambda_{12}\) be defined analogously as \(\Lambda_{n,11}\) and \(\Lambda_{n,12}\) with \(K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^*\) replaced by \((H'_{21} \beta_{21}^*)_+\); see Theorem 2. We can readily show that \(\Lambda_{n,11} = \Lambda_{11} + o(1)\) and \(\Lambda_{n,12} = \Lambda_{12} + o(1)\). For example, using

\[
Y_1 - B'_1 \beta_1^* + H'_{20} \beta_2^{*} + K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^* = [Y_1 - B'_1 \beta_1^* + H'_{20} \beta_2^{*} + (H'_{21} \beta_{21}^*)_+] + [K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^* - (H'_{21} \beta_{21}^*)_+]
\]

\[
= \varepsilon_1 + [K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^* - (H'_{21} \beta_{21}^*)_+],
\]

we have

\[
\Lambda_{n,11} - \Lambda_{11} = \mathbb{E} \left\{ B_1 B'_1 [\varepsilon_1 + K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^* - (H'_{21} \beta_{21}^*)_+]^2 \right\} - \mathbb{E} \left\{ B_1 B'_1 \varepsilon_1^2 \right\} = \mathbb{E} \left\{ B_1 B'_1 [K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^* - (H'_{21} \beta_{21}^*)_+]^2 \right\}
\]

\[
+ 2 \cdot \mathbb{E} \left\{ B_1 B'_1 [K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^* - (H'_{21} \beta_{21}^*)_+] \varepsilon_1 \right\}.
\]

For any \(z > 0\), \(K_\alpha(z) z - (z)_+ \rightarrow z - z = 0\) as \(\alpha \rightarrow 0\), and for any \(z < 0\), \(K_\alpha(z) z - (z)_+ \rightarrow 0 - 0 = 0\) as \(\alpha \rightarrow 0\). In addition, noting that \(|K_\alpha(z) z - (z)_+| \leq |z|\) for all \(z \in \mathbb{R}\), we have

\[
\|B_1 B'_1 [K_\alpha (H'_{21} \beta_{21}^*) H'_{21} \beta_{21}^* - (H'_{21} \beta_{21}^*)_+]^2\| \leq \|B_1 B'_1 (H'_{21} \beta_{21}^*)_+^2\|
\]

and the right hand side object in the last expression has finite first moment by Assumption 43.
A2(iii). Then \( \mathbb{E} \{ B_1 B'_1[K_\alpha(H'_{21}\beta_{21})H'_{21}\beta_{21} - (H'_{21}\beta_{21}^*)_+]^2 \} = o(1) \) by the DCT. Similarly, \( \mathbb{E} \{ B_1 B'_1 \left[ K_\alpha(H'_{21}\beta_{21}^*)H'_{21}\beta_{21} - (H'_{21}\beta_{21}^*)_+ \right] \epsilon_1 \} = o(1) \) by the DCT and \( \Lambda_{n,11} = \Lambda_{11} + o(1) \).

Analogously, we have \( \Lambda_{n,12} = \Lambda_{12} + o(1) \) by Assumption A2(iii) and the DCT. It follows that \( \Lambda_n = \Lambda + o(1) \), where \( \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda'_{12} & \Lambda_{22} \end{pmatrix} \).

In addition,

\[
\text{Var} (m_n(\beta^*)) = \mathbb{E} \left[ m_n(\beta^*) m_n(\beta^*)' \right] - \mathbb{E} \left[ m_n(\beta^*) \right] \mathbb{E} \left[ m_n(\beta^*)' \right] = \Lambda_n - O(\alpha^2) O(\alpha^2) = \Lambda + o(1) .
\]

Now, we show that \( \mathbb{E} \| m_n(\beta^*) \|^{2+\delta} < \infty \). Let \( \omega = (\omega'_1, \omega'_2)' \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \) be nonrandom such that \( \| \omega \| = 1 \). Let \( \delta > 0 \) be given as in Assumption A2(iv). Then by the repeated use of the \( C_r \) inequality, we obtain

\[
\mathbb{E} |\omega' m_n(\beta^*)|^{2+\delta} \\
\leq 2^{1+\delta} \mathbb{E} |\omega'_1 m_{1n}(\beta^*)|^{2+\delta} + 2^{1+\delta} \mathbb{E} |\omega'_2 m_{2n}(\beta^*)|^{2+\delta} \\
\leq 2^{2+2\delta} \mathbb{E} |\omega'_1 [m_{1n}(\beta^*) - m_1(\beta^*)]|^{2+\delta} + 2^{2+2\delta} \mathbb{E} |\omega'_1 m_1(\beta^*)|^{2+\delta} + 2^{1+\delta} \mathbb{E} |\omega'_2 m_2(\beta^*)|^{2+\delta} .
\]

By Assumption A2(iii), \( \mathbb{E} |\omega'_l m_l(\beta^*)|^{2+\delta} \leq \| \omega_l \|^{2+\delta} \mathbb{E} \| m_l(\beta^*) \|^{2+\delta} < \infty \) for \( l = 1, 2 \). By the DCT (as used in the proof that \( \Lambda_{n,11} = \Lambda_{11} + o(1) \)) and Assumption A2(iv),

\[
\mathbb{E} |\omega'_1 [m_{1n}(\beta^*) - m_1(\beta^*)]|^{2+\delta} \\
\leq \| \omega_l \|^{2+\delta} \mathbb{E} \| m_{1n}(\beta^*) - m_1(\beta^*) \|^{2+\delta} \\
= \| \omega_l \|^{2+\delta} \mathbb{E} \left[ B_1 \left[ K_\alpha(H'_{21}\beta_{21}^*)H'_{21}\beta_{21}^* - (H'_{21}\beta_{21}^*)_+ \right] \right]^{2+\delta} = o(1) .
\]

It follows that \( \mathbb{E} \| m_n(\beta^*) \|^{2+\delta} < \infty \). Then by the Liapounov central limit theorem (CLT, e.g., Davidson 1994, ch23.2), \( \sqrt{n} \mathbb{E} \left[ m_n(\beta^*) \right] = \mathbb{G}_n[m_n(\beta^*)] = o(1) \xrightarrow{d} N(0, \Lambda) \).
Step 2. We prove (ii) \( \Xi_n \equiv \mathbb{E}[\nabla \beta^* m_n (\beta^*)] = \Xi + o(1) \). Observe that

\[
\Xi_n = \begin{pmatrix}
\mathbb{E}[\nabla \beta_1^* m_{1n} (\beta^*)] & \mathbb{E}[\nabla \beta_2^* m_{1n} (\beta^*)] \\
\mathbb{E}[\nabla \beta_1^* m_{2n} (\beta^*)] & \mathbb{E}[\nabla \beta_2^* m_{2n} (\beta^*)]
\end{pmatrix} = \begin{pmatrix}
\Xi_{n,11} & \Xi_{n,12} \\
\Xi_{n,21} & \Xi_{n,22}
\end{pmatrix}.
\]

Noting that \( \nabla \beta_1^* m_{1n} (\beta^*) = -B_1 B_1', \nabla \beta_2^* m_{1n} (\beta^*) = 0, \) and \( \nabla \beta_2^* m_{2n} (\beta^*) = -B_2 B_2' \), we have \( \Xi_{n,11} = -\mathbb{E}[B_1 B_1'], \Xi_{n,21} = 0, \) and \( \Xi_{n,22} = -\mathbb{E}[B_2 B_2'] \). For \( \Xi_{n,12} \), we have

\[
\nabla \beta_2^* m_{1n} (\beta^*) = [\nabla \beta_2^* m_{1n} (\beta^*), \nabla \beta_2^* m_{1n} (\beta^*)],
\]

where \( \nabla \beta_2^* m_{1n} (\beta^*) = B_1 H_1' \) and \( \nabla \beta_2^* m_{1n} (\beta^*) = B_1 H_1' \left[ \frac{1}{\alpha} k_\alpha (H_2' \beta_2^*) H_1' \beta_2^* + K_\alpha (H_2' \beta_2^* \beta_2^*) \right] \).

Let \( \Xi_{n,12,t} \equiv \mathbb{E}[\nabla \beta_2^* m_{1n} (\beta^*)] \) for \( t = 0, 1 \). It is easy to see that \( \Xi_{n,12,0} = \mathbb{E}[B_1 H_1'] \). Let \( r_2 (z) = \mathbb{E}[B_1 H_1' | H_2' \beta_2^* = z] \) and \( g_2 (z) = \int_{-\infty}^{z} r_2 (t) f_C (t) dt \). Then

\[
\Xi_{n,12,1} = \mathbb{E} \left\{ r_2 (H_2' \beta_2^*) \left[ \frac{1}{\alpha} k_\alpha (H_2' \beta_2^*) H_1' \beta_2^* + K_\alpha (H_2' \beta_2^* \beta_2^*) \right] \right\}
\]

\[
= (1 - p) \int_{-\infty}^{\infty} r_2 (z) \left[ \frac{1}{\alpha} k_\alpha (z) z + K_\alpha (z) \right] f_C (z) dz + p \sum_{z_d} r_2 (z_d) \left[ \frac{1}{\alpha} k_\alpha (z_d) z_d + K_\alpha (z_d) \right] f_D (z_d)
\]

\[
\equiv (1 - p) II_C + p II_D, \text{ say},
\]

where \( k_\alpha (z) = k(z/\alpha) \). For the first term \( II_C \), we have by the change of variables:

\[
II_C = \frac{1}{\alpha} \int_{-\infty}^{\infty} r_2 (z) z f_C (z) k(z/\alpha) dz + \int_{-\infty}^{\infty} r_2 (z) f_C (z) K(z/\alpha) dz
\]

\[
= \alpha \int_{-1}^{1} r_2 (\alpha z) u f_C (\alpha u) k(u) du + \int_{-\infty}^{\infty} K(z/\alpha) g_2 (z).
\]

By the DCT and Assumption A3, \( \int_{-1}^{1} r_2 (\alpha z) z f_C (\alpha z) k(z) dz \to r_2 (0) f_C (0) \int_{-1}^{1} z k(z) dz = 0 \) as \( \alpha \to 0 \). By the integration by parts, change of variables, DCT, and Assumptions A1(v)

45
and A3, we obtain
\[
\int_{-\infty}^{\infty} K(z/\alpha) dg_2(z) = g_2(\infty) - \frac{1}{\alpha} \int_{-\infty}^{\infty} k(z/\alpha) g_2(z) dz = g_2(\infty) - \int_{-1}^{1} k(z) g_2(\alpha z) dz
\]
\[
\rightarrow g_2(\infty) - g_2(0) \int_{-1}^{1} k(z) dz = g_2(\infty) - g_2(0).
\]

It follows that \( II_C = g_2(\infty) - g_2(0) + o(1) \). Noting that for any fixed \(|z_d| \neq 0\), \( k_\alpha(z_d) = k(z_d/\alpha) = 0 \) for sufficiently small \( \alpha \) by the fact that \( k(\cdot) \) has compact support \([-1, 1]\), we have for large enough \( n \)

\[
\frac{1}{\alpha} \sum_{z_d} r_2(z_d) k_\alpha(z_d) z_d f_D(z_d)
\]
\[
= \frac{1}{\alpha} \sum_{z_d<0} r_2(z_d) k_\alpha(z_d) z_d f_D(z_d) + \frac{1}{\alpha} \sum_{z_d>0} r_2(z_d) k_\alpha(z_d) z_d f_D(z_d) = 0 + 0 = 0.
\]

Similarly, for any fixed \( z_d > 0 \), \( K_\alpha(z_d) = K(z_d/\alpha) = 1 \) and \( K_\alpha(-z_d) = K(-z_d/\alpha) = 0 \) for sufficiently small \( \alpha \). Then for sufficiently large \( n \),

\[
\sum_{z_d} r_2(z_d) K_\alpha(z_d) f_D(z_d)
\]
\[
= \frac{1}{2} r_2(0) f_D(0) + \sum_{z_d<0} r_2(z_d) K_\alpha(z_d) f_D(z_d) + \sum_{z_d>0} r_2(z_d) K_\alpha(z_d) f_D(z_d)
\]
\[
= \frac{1}{2} r_2(0) f_D(0) + \sum_{z_d>0} r_2(z_d) f_D(z_d),
\]

where we use the fact that \( K(0) = 1/2 \) by the symmetry of \( k(\cdot) \) under Assumption A3(i).

It follows that for sufficiently large \( n \),

\[
II_D = \frac{1}{2} r_2(0) f_D(0) + \sum_{z_d>0} r_2(z_d) f_D(z_d).
\]
Consequently, \( \Xi_{n,12,2} = (1 - p) \left[ g_2 (\infty) - g_2 (0) \right] + p \left[ \frac{1}{2} r_2 (0) f_P (0) + \sum_{z_d > 0} r_2 (z_d) f_P (z_d) \right] + o (1) \) and \( \Xi_{n,12,2} = \Xi_{12} + o (1) \), where \( \Xi_{12} \) is defined in (3.4) in the main text. In sum, we have shown that \( \Xi_n = \Xi + o (1) \).

**Step 3.** We show that \( E_n [\nabla_{\beta'} m_n (\beta)] = E [\nabla_{\beta'} m_n (\beta^*)] + o_P (1) \). Note that \( \nabla_{\beta'} m_n (\beta) = \begin{pmatrix} -B_1 B'_1 & \nabla_{\beta'_2} m_{1n} (\beta) \\ 0 & -B_2 B'_2 \end{pmatrix} \), where \( \nabla_{\beta'_2} m_{1n} (\beta) = [\nabla_{\beta'_20} m_{1n} (\beta), \nabla_{\beta'_21} m_{1n} (\beta)] \).

\( \nabla_{\beta'_20} m_{1n} (\beta) = B_1 H'_{20} \), and \( \nabla_{\beta'_21} m_{1n} (\beta) = B_1 H'_{21} \left[ \frac{1}{\alpha} k_{\alpha} (H'_{21} \beta_{21}) H'_{21} \beta_{21} + K_{\alpha} (H'_{21} \beta_{21}) \right] \).

We first show that \( \sup_{\beta_{21} \in N_{\epsilon} (\beta^*_{21})} \| n^{-1/2} G_n [\nabla_{\beta'} m_n (\beta)] \| = o_P (1) \) as \( \nabla_{\beta'} m_n (\beta) \) depends on \( \beta \) only through \( \beta_{21} \). Here \( N_{\epsilon} (\beta^*_{21}) \) denotes the \( \epsilon \)-neighborhood of \( \beta^*_{21} \) with respect to the Euclidean norm on \( \mathbb{R}^{k_2} \). Since \( -B_1 B'_1, -B_2 B'_2 \), and \( B_1 H'_{20} \) are not \( \beta_{21} \)-dependent and it is standard to show that \( n^{-1/2} G_n [A_n (W)] = o_P (1) \) for \( A_n (W) = -B_1 B'_1, -B_2 B'_2 \), and \( B_1 H'_{20} \). It suffices to prove the last claim by showing that

\[
\sup_{\beta_{21} \in N_{\epsilon} (\beta^*_{21})} \| n^{-1/2} G_n [\nabla_{\beta'_21} m_{1n} (\beta)] \| = o_P (1).
\]

Consider a class of functions,

\[
\mathcal{M}_2 = \left\{ m_{\beta_{21}} : m_{\beta_{21}} (w) = b_1 h'_{21} \left[ \frac{1}{\alpha} k (h'_{21} \beta_{21} / \alpha) h'_{21} \beta_{21} + K (h'_{21} \beta_{21} / \alpha) \right], \beta_{21} \in \mathcal{B}_{21} \right\}.
\]
Note that the functions in $\mathcal{M}_2$ are composed of the following classes of functions:

\[
\mathcal{M}_{2,1} = \{h_{21}' \beta_{21} : \beta_{21} \in \mathcal{B}_{21}\},
\]
\[
\mathcal{M}_{2,2} = \{b_1 h_{21}' k(h_{21}' \beta_{21}/\alpha) : \beta_{21} \in \mathcal{B}_{21}\},
\]
\[
\mathcal{M}_{2,3} = \{b_1 h_{21}' K(h_{21}' \beta_{21}/\alpha) : \beta_{21} \in \mathcal{B}_{21}\}.
\]

By Lemma 2.13 in Pakes and Pollard (1989), $\mathcal{M}_1$ is an Euclidean class of functions with an envelope given by $C \|h_{21}\|$. By Lemma 22 in Nolan and Pollard (1987), $\mathcal{M}_{2,2}$ and $\mathcal{M}_{2,3}$ are also Euclidean with envelopes respectively given by $\|b_1 h_{21}'\| \max_{|u|\leq 1} k(u)$ and $\|b_1 h_{21}'\|$ as $k(\cdot)$ has bounded variation and $K(\cdot)$ is monotone with range $[0, 1]$. As a result, $\mathcal{M}_2$ is Euclidean by Lemma 2.14 in Pakes and Pollard (1989). Then by standard arguments used in the kernel estimation literature (e.g., Theorem 1 of Einmahl and Mason (2000)), we can show that

\[
\sup_{\beta_{21} \in N_\epsilon(\hat{\beta}_{21}^*)} \left\| n^{-1/2} \mathbb{G}_n \nabla_{\beta_{21}} m_{1n}(\beta) \right\| = O_P((n\alpha/\ln n)^{-1/2}) = o_P(1).
\]

This, in conjunction with the fact that $\mathbb{E}[\nabla_{\beta_{21}} m_{1n}(\beta)]$ is continuous in $\beta_{21}$ and that $\bar{\beta} = \beta^* + o_P(1)$, implies that

\[
\mathbb{E}_n \left[ \nabla_{\beta_{21}} m_{1n}(\bar{\beta}) \right] = \mathbb{E}_n \left[ \nabla_{\beta_{21}} m_{1n}(\beta) \right]_{\beta = \bar{\beta}} + o_P(1) = \mathbb{E}_n \left[ \nabla_{\beta_{21}} m_{1n}(\beta^*) \right] + o_P(1).
\]

This completes the proof of the theorem. ■

**Proof of Theorem 3** To prove that $\hat{\Omega}_n - \Omega = o_P(1)$, it suffices to show that (i) $\hat{\Lambda}_n(\hat{\beta}_{sq}) - \Lambda_n = o_P(1)$ and (ii) $\hat{\Xi}_n(\hat{\beta}_{sq}) - \Xi_n = o_P(1)$ as we have shown that $\Lambda_n = \Lambda + o(1)$ and $\Xi_n = \Xi + o(1)$ in the proof of Theorem 2.
To show (i) $\hat{\Lambda}_n(\hat{\beta}_{sq}) - \Lambda_n = o_p(1)$, it suffices to show that $\hat{\Lambda}_{n,11}(\hat{\beta}_{sq}) - \Lambda_{n,11} = o_p(1)$, $\hat{\Lambda}_{n,12}(\hat{\beta}_{sq}) - \Lambda_{n,12} = o_p(1)$, and $\hat{\Lambda}_{n,22}(\hat{\beta}_{sq}) - \Lambda_{n,22} = o_p(1)$. We only prove the first claim as the proofs of the other two are similar and simpler. Write

$$\hat{\Lambda}_{n,11}(\hat{\beta}_{sq}) - \Lambda_{n,11} = [\hat{\Lambda}_{n,11}(\hat{\beta}_{sq}) - \Lambda_{n,11} (\beta^*)] + [\Lambda_{n,11} (\beta^*) - \mathbb{E}[\Lambda_{n,11} (\beta^*)]].$$

The second term is $o_p(1)$ by the WLLN. Consider a class of functions,

$$\mathcal{M}_3 = \{\bar{m}_\beta : \bar{m}_\beta (w) = b_1 b'_1 [y_1 - b'_1 \beta_1 + h'_2 \beta_2 + K (h'_2 \beta_2 / \alpha) h'_2 \beta_2, \beta_1 \in \mathcal{B}_1, \beta_2 \in \mathcal{B}_2, \beta_2 \in \mathcal{B}_2\}.$$

Noting that the functions in $\mathcal{M}_3$ are composed of the following classes of functions:

$$\mathcal{M}_{3,1} = \{b_1 b'_1 [y_1 - b'_1 \beta_1 + h'_2 \beta_2] : \beta_1 \in \mathcal{B}_1, \beta_2 \in \mathcal{B}_2\},$$

$$\mathcal{M}_{3,2} = \{b_1 b'_1 h'_2 \beta_2 : \beta_2 \in \mathcal{B}_2\},$$

$$\mathcal{M}_{3,3} = \{b_1 b'_1 K (h'_2 \beta_2 / \alpha) : \beta_2 \in \mathcal{B}_2\},$$

we can readily argue that $\mathcal{M}_3$ is an Euclidean class of functions with a well-behaved envelope function. Consequently, $\bar{m}_\beta (w)$ is $P$-Glivenko-Cantelli by Theorem 37 in Pollard (1984, p. 34) and Lemma 2.8 in Pakes and Pollard (1989) under our conditions, and we have

$$\hat{\Lambda}_{n,11}(\hat{\beta}_{sq}) = \mathbb{E}[\hat{\Lambda}_{n,11}(\hat{\beta}_{sq})] + o_P (1) = \mathbb{E}[\hat{\Lambda}_{n,11}(\beta^*)] + o_P (1) = \Lambda_{n,11} + o_P (1),$$

where the second equality follows from the fact that $\mathbb{E}[\hat{\Lambda}_{n,11}(\beta)]$ is continuous in $\beta$ and that $\bar{\beta} = \beta^* + o_P (1)$.

To show (ii) $\hat{\Xi}_n(\hat{\beta}_{sq}) - \Xi_n = o_P(1)$, it suffices to show that $\hat{\Xi}_{n,11}(\hat{\beta}_{sq}) = \Xi_{n,11} + o_P (1)$,
\[ \hat{\Xi}_{n,22}(\hat{\beta}_{sq}) = \Xi_{n,22} + o_P(1), \hat{\Xi}_{n,12,0}(\hat{\beta}_{sq}) = \Xi_{n,12,0} + o_P(1), \] and \[ \hat{\Xi}_{n,12,1}(\hat{\beta}_{sq}) = \Xi_{n,12,1} + o_P(1). \]

Noting that \[ \hat{\Xi} \] is consistent, we can express \[ \Theta_n \] as follows:

\[ \Theta_n = \sum_{i=1}^{n} \left[ \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} \left( \beta_{ij} - \hat{\beta}_{ij} \right)^2 \right]. \]

Noting that \[ \hat{\Xi} \] is consistent, we can express \[ \Theta_n \] as follows:

\[ \Theta_n = \sum_{i=1}^{n} \left[ \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} \left( \beta_{ij} - \hat{\beta}_{ij} \right)^2 \right]. \]

The last term is \[ O(\sqrt{n}) \] and \[ \hat{\Xi} \] is \[ O(1) \].

**Proof of Theorem 4 (i)** Our proof is analogous to that of Theorem 1 in Linton (1995).

Note that \[ \sqrt{n}(\hat{\beta}_{sq} - \beta^*) = -[\hat{\Xi}_n(\beta)]^{-1} \sqrt{n} \mathbb{E}_n [m_n(\beta^*)]. \] Using \[ [\hat{\Xi}_n(\beta)]^{-1} = \Xi^{-1} + \{[\hat{\Xi}_n(\beta)]^{-1} - \Xi^{-1}\} \] repeatedly, we have the following expansion for \( T_{0n} \):

\[ T_{0n} = -\sigma_c^{-1} c'[\hat{\Xi}_n(\beta)]^{-1} \xi_N = -\sigma_c^{-1} c' \Xi^{-1} \xi_N - \sigma_c^{-1} c' \{[\hat{\Xi}_n(\beta)]^{-1} - \Xi^{-1}\} \xi_N \]

\[ = -\sigma_c^{-1} c' \Xi^{-1} \xi_N + \sigma_c^{-1} c' [\hat{\Xi}_n(\beta)]^{-1} [\hat{\Xi}_n(\beta) - \Xi] \Xi^{-1} \xi_N \]

\[ = -\sigma_c^{-1} c' \Xi^{-1} \xi_N + \sigma_c^{-1} c' [\hat{\Xi}_n(\beta)]^{-1} \Xi^{-1} \xi_N + \sigma_c^{-1} c' \{[\hat{\Xi}_n(\beta)]^{-1} - \Xi^{-1}\} [\hat{\Xi}_n(\beta) - \Xi] \Xi^{-1} \xi_N \]

\[ = \ldots \]

\[ = \{ -\sigma_c^{-1} c' \Xi^{-1} \xi_N + (n\alpha)^{-1/2} \sigma_c^{-1} c' \Xi^{-1} \xi_D \Xi^{-1} \xi_N - (n\alpha)^{-1} \sigma_c^{-1} c' \Xi^{-1} \xi_D \Xi^{-1} \xi_D \xi^{-1} \xi_N \}

\[ + (n\alpha)^{-3/2} \sigma_c^{-1} c' \Xi^{-1} \xi_D \Xi^{-1} \xi_D \Xi^{-1} \xi_D \xi^{-1} \xi_N \}

\[ - (n\alpha)^{-2} \sigma_c^{-1} c' [\hat{\Xi}_n(\beta)]^{-1} \xi_D \Xi^{-1} \xi_D \Xi^{-1} \xi_D \xi^{-1} \xi_N \]

\[ = T_{0n}^* + R_{0n}^*, \]

where \( \xi_N = \sqrt{n} \mathbb{E}_n [m_n(\beta^*)] \) and \( \xi_D = \sqrt{n} \mathbb{E}_n [\hat{\Xi}_n(\beta) - \Xi] \). To proceed, we apply the following result of Sargan and Mikhail (1971): for all \( u \in \mathbb{R} \) and \( \zeta > 0 \),

\[ |P(T_{0n} \leq u) - P(T_{0n}^* \leq u)| \leq P(|R_{0n}^*| > \zeta) + P(|T_{0n}^* - u| < \zeta). \]

The last term is \( O(\zeta) \) provided \( T_{0n}^* \) has a bounded density. Therefore we choose \( \zeta = \)
\[ O \left( n^{3/5} \ln n \right)^{-1} \]. We want to show that for any positive constant \( C_1 \),

\[
P \left( |R_{0n}^*| > C_1 [n^{3/5} \ln n]^{-1} \right) = o \left( n^{-3/5} \right).
\]

Let \( A_1 = \{ |R_{0n}^*| > C_1 [n^{3/5} \ln n]^{-1} \} \) and \( A_2 = \{ \| \hat{\Xi}_n(\bar{\beta})^{-1} \|_{sp} > C_2 \} \), where \( \| A \|_{sp} \equiv \lambda_{\max} \left( AA' \right)^{1/2} \) denotes the spectral norm of a matrix \( A \), with \( \lambda_{\max} (\cdot) \) (similarly \( \lambda_{\min} (\cdot) \)) to be the maximum (minimum) eigenvalue of a matrix, and \( C_2 = [\lambda_{\min}(\Xi\Xi')]^{-1/2}/2 \). Note that \( \| \hat{\Xi}_n(\bar{\beta})^{-1} \|_{sp} = [\lambda_{\min}(\hat{\Xi}_n(\bar{\beta})\hat{\Xi}_n(\bar{\beta}'))^{-1}]^{-1/2} \). Observe that

\[
P \left( A_1 \right) \leq P \left( A_1 \cap A_2 \right) + P \left( A_2^c \right).
\]

When \( \| \hat{\Xi}_n(\bar{\beta})^{-1} \|_{sp} < C_2 \), we have

\[
\| \xi_D \|_{sp} = n^{1/2} \| \hat{\Xi}_n(\bar{\beta}) - \Xi \|_{sp} = n^{1/2} \| \hat{\Xi}_n(\bar{\beta})[\Xi^{-1} - \hat{\Xi}_n(\bar{\beta})^{-1}]\Xi \|_{sp} \\
\geq n^{1/2} \{ \lambda_{\min}(\hat{\Xi}_n(\bar{\beta})\hat{\Xi}_n(\bar{\beta}')(\Xi\Xi')) \}^{1/2} \| \Xi^{-1} - \hat{\Xi}_n(\bar{\beta})^{-1} \|_{sp} \\
\geq n^{1/2} C_2^{-1} (2C_2)^{-1} \left( \| \Xi^{-1} \|_{sp} - \| \hat{\Xi}_n(\bar{\beta})^{-1} \|_{sp} \right) \\
\geq \frac{1}{2} n^{1/2} C_2^{-2} (2C_2 - C_2) = \frac{1}{2} n^{1/2} C_2^{-1},
\]

where the first inequality follows from the fact that

\[
\| ABC \|_{sp}^2 = \lambda_{\max} (ABCC'B'A') \geq \lambda_{\max} (ABB'A') \lambda_{\min} (CC') \\
= \lambda_{\max} (B'A'AB) \lambda_{\min} (CC') \geq \lambda_{\max} (B'B) \lambda_{\min} (A'A) \lambda_{\min} (CC'),
\]

and the second inequality follows from the triangle inequality. In conjunction with Cheby-
shev inequality, this implies that

\[ P(A^c_2) \leq P \left( \| \xi_D \|_{sp} \geq \frac{1}{2} n^{1/2} C_2^{-1} \right) \leq \frac{\mathbb{E}[\| \xi_D \|_{sp}^2]}{4 n C_2^{-2}} = O(n^{-1}) = o(n^{-3/5}) \]

as \( \mathbb{E}[\| \xi_D \|_{sp}^2] = O(1) \) by Lemma 1 below and Jensen inequality.

Conditional on \( A_2 \),

\[ \| R_{0n}^* \| \leq (n \alpha)^{-2} \sigma_c^{-1} \| c \| \left\| \hat{\Xi}_n(\beta) \right\|^{-1}_{sp} \| \Xi^{-1} \|_{sp}^4 \| \xi_D \|_{sp}^4 \| \xi_N \| \leq C_3 (n \alpha)^{-2} \| \xi_D \|_{sp}^4 \| \xi_N \| \]

for some \( C_3 > 0 \). Let \( A_2 = \{ c_3 (n \alpha)^{-2} \| \xi_D \|_{sp}^2 \| \xi_N \| \geq C_4 \lfloor n^{3/5} \ln n \rfloor^{-1} \} \) for some \( C_4 > 0 \).

Then \( A_1 \cap A_2 \subset \bar{A}_2 \) and

\[ P(A_1 \cap A_2) \leq P(\bar{A}_2) = P \left( \| \xi_D \|_{sp}^2 \| \xi_N \| \geq (C_4 / C_3) (n \alpha)^2 n^{-3/5} (\ln n)^{-1} \right) \]

\[ \leq (C_3 / C_4)^{8/5} \left( (n \alpha)^{-2} n^{3/5} \ln n \right)^{8/5} \cdot \mathbb{E}[\| \xi_D \|_{sp}^{22/5} \| \xi_N \|^{8/5}] \]

\[ \leq (C_3 / C_4) \alpha^{-16/5} n^{-56/25} (\ln n)^{8/5} \cdot \left\{ \mathbb{E}[\| \xi_D \|_{sp}^8] \right\}^{4/5} \left\{ \mathbb{E}[\| \xi_N \|_{sp}^8] \right\}^{1/5} \]

\[ = O(n^{-24/25} (\ln n)^{8/5}) = o(n^{-3/5}), \]

by Markov inequality, Hölder inequality, Lemma 1 below and the fact that \( \alpha \propto n^{-2/5} \).

Consequently, we have \( P(A_1) = P(|R_{0n}^*| > c_1 [n^{3/5} \ln n]^{-1}) = o(n^{-3/5}) \) and \( |P(T_{0n} \leq u) - P(T_{0n}^* \leq u)| \leq P(|R_{0n}^*| > \zeta) + P(|T_{0n}^* - u| < \zeta) = o(n^{-3/5}) \) uniformly in \( u \).

\( \textbf{(ii)} \) By Lemma 1 and direct moment calculations, we have

\[ \mathbb{E}[T_{0n}] = -\sigma_c^{-1} \zeta \mathbb{E}[\xi_N] + O(n^{-1/2}) = -\sqrt{n \alpha^4 \sigma_c^{-1} \zeta} \mathbb{E}[\xi_N] + O(n^{-3/10}), \]
and

\[
\text{Var}[T_{0n}^*] = \text{Var}[-\sigma_c^{-1}c'\Xi^{-1}\xi_N] + \text{Var}[(n\alpha)^{-1/2} \sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_N] + o(n^{-3/5})
\]

\[
= \sigma_c^{-2}c'^{-1}[\Lambda + O(\alpha^2)\Xi^{-1}c] + (n\alpha)^{-1} \sigma_c^{-1}c'\Xi^{-1}\text{Var}[\xi_D\Xi^{-1}\xi_N\Xi^{-1}c] + o(n^{-3/5})
\]

\[
= 1 + (n\alpha)^{-1} \sigma_c^{-1}c'\Xi^{-1}\Phi_0\Xi^{-1}c + o(n^{-3/5}),
\]

where we use the fact that the other variance and covariance terms in the expansion of \(\text{Var}[T_{0n}^*]\) are all \(o(n^{-3/5})\) under Assumption A4, and \(\kappa\) and \(\Phi_0\) are defined in Theorem 4.

(iii) We first consider the Taylor expansion of \(S_n\) with remainder term \(O_P((n\alpha)^{-1})\):

\[
S_n = \frac{\sqrt{n\alpha}(\hat{\sigma}_c - \sigma_c)}{\sigma_c} = \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{\sigma_c^2} \left[ \frac{1}{2 + (\hat{\sigma}_c/\sigma_c - 1)} \right]
\]

\[
= \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{\sigma_c^2} \left\{ \frac{1}{2} - \frac{1}{4} \left( \frac{\hat{\sigma}_c}{\sigma_c} - 1 \right) \right\} + O_P((n\alpha)^{-1})
\]

\[
= \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{2\sigma_c^2} - \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{4\sigma_c^2} \frac{\hat{\sigma}_c - \sigma_c}{\sigma_c} + O_P((n\alpha)^{-1})
\]

\[
= \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{2\sigma_c^2} - \frac{[\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)]^2}{8\sqrt{n\alpha}\sigma_c^4} + O_P((n\alpha)^{-1}).
\]
Next,

\[
\sqrt{n\alpha(\hat{\sigma}_c^2 - \sigma_c^2)}
\]

\[
= \sqrt{n\alpha c'} \left[ \hat{\Xi}_n^{-1} \hat{\Lambda}_n (\hat{\Xi}_n^{-1})' - \Xi^{-1} \Lambda (\Xi^{-1})' \right] c
\]

\[
= \sqrt{n\alpha c'} \left[ ((\hat{\Xi}_n^{-1} - \Xi^{-1}) + \Xi^{-1}) \hat{\Lambda}_n [(\hat{\Xi}_n^{-1} - \Xi^{-1}) + \Xi^{-1}]' - \Xi^{-1} \Lambda (\Xi^{-1})' \right] c
\]

\[
= \sqrt{n\alpha c'} \{ 2(\hat{\Xi}_n^{-1} - \Xi^{-1}) \hat{\Lambda}_n \Xi^{-1} + \Xi^{-1}(\hat{\Lambda}_n - \Lambda) \Xi^{-1} + (\hat{\Xi}_n^{-1} - \Xi^{-1}) \hat{\Lambda}_n (\hat{\Xi}_n^{-1} - \Xi^{-1})' \} c
\]

\[
= -2c' \hat{\Xi}_n^{-1} \xi_D^\dagger \Xi^{-1} \Lambda \Xi^{-1} c + \sqrt{\alpha} c' \Xi^{-1} \xi_{\Lambda} (\Xi^{-1})' c + c' \hat{\Xi}_n^{-1} \xi_D^\dagger \Xi^{-1} \Lambda_n \Xi^{-1} \xi_D^\dagger \hat{\Xi}_n^{-1} c / \sqrt{n \alpha}
\]

\[
= -2c' \Xi^{-1} \xi_D^\dagger \Xi^{-1} \Lambda \Xi^{-1} c + \sqrt{\alpha} c' \Xi^{-1} \xi_{\Lambda} (\Xi^{-1})' c
\]

\[
+ 2[ c' \Xi^{-1} \xi_D^\dagger \Xi^{-1} \xi_D^\dagger \Xi^{-1} \Lambda \Xi^{-1} c + c' \Xi^{-1} \xi_D^\dagger \Xi^{-1} \Lambda \Xi^{-1} \xi_D^\dagger \Xi^{-1} c] / \sqrt{n \alpha}
\]

\[
- 2c' \Xi^{-1} \xi_D^\dagger \Xi^{-1} \xi_{\Lambda} \Xi^{-1} c / \sqrt{n} + o_P ((n \alpha)^{-1}),
\]

where \( \xi_{\Lambda} = \sqrt{n}(\hat{\Lambda}_n - \Lambda) \) and \( \xi_D^\dagger = \sqrt{n\alpha}(\hat{\Xi}_n - \Xi) \). Plugging the last expansion into that of \( S_n \) and simplifying yields \( S_n = S_n^* + o_P ((n \alpha)^{-1/2}) \) where \( S_n^* \) is defined in (4.2) in the main text. It follows that

\[
T_n = \frac{\sqrt{n\alpha c'}(\hat{\beta}_c - \beta^*) \sigma_c}{\sigma_c} \cdot \frac{T_{0n}}{1 + (n \alpha)^{-1/2} S_n}
\]

\[
= T_{0n} \left\{ 1 - (n \alpha)^{-1/2} S_n + 2(n \alpha)^{-1} S_n^2 \right\} + o_P ((n \alpha)^{-1})
\]

\[
= T_{0n}^* + (n \alpha)^{-1/2} \left[ \sigma_c^{-1} c' \Xi^{-1} \xi_N - (n \alpha)^{-1/2} \sigma_c^{-1} c' \Xi^{-1} \xi_D \Xi^{-1} \xi_N \right] S_n^*
\]

\[
- 2(n \alpha)^{-1} \sigma_c^{-1} c' \Xi^{-1} \xi_N S_n^2 + o_P ((n \alpha)^{-1})
\]

\[
= T_{0n}^* + o_P ((n \alpha)^{-1}),
\]

where \( T_{0n}^* \) is defined in (4.3) in the main text. It is easy to verify that only the first term
in the definition of \( T_n^* \) contributes to the \( \mathbb{E}[T_n^*] \) up to the order \((n\alpha)^{-1/2} = O\left(n^{-3/10}\right)\):

\[
\mathbb{E}[T_n^*] = -\sqrt{n\alpha^2}\sigma_1 c'\Xi^{-1}\kappa + o(n^{-3/10}),
\]

and only the first three terms contribute to \( \text{Var}[T_n^*] \) up to order \((n\alpha)^{-1} = O\left(n^{-3/5}\right)\):

\[
\text{Var}[T_n^*] = \text{Var}[-(\sigma_1 c'\Xi^{-1}\xi_N)] + \text{Var}[(n\alpha)^{-1/2}\sigma_1 c'\Xi^{-1}(\xi_D\Xi^{-1}\xi_N + \sigma_1^2 c'\Xi^{-1}\xi_D^\dagger \Omega \xi_N)] + o(n^{-3/5})
\]

\[
= \sigma_1^{-2} c'\Xi^{-1}[\Lambda + O(\alpha^2)]\Xi^{-1}c + (n\alpha)^{-1}\sigma_1^{-2} c'\Xi^{-1}\text{Var}[\xi_D\Xi^{-1}\xi_N + \sigma_1^{-2} c'\Xi^{-1}\xi_D^\dagger \Omega \xi_N]\Xi^{-1}c
\]

\[
+ o\left(n^{-3/5}\right)
\]

\[
= 1 + (n\alpha)^{-1}\sigma_1^{-2} c'\Xi^{-1}\Phi\Xi^{-1}c + o\left(n^{-3/5}\right),
\]

where \( \Phi \) is defined in Theorem 4. ■

**Lemma 1** Recall that \( \xi_N = \sqrt{n}\mathbb{E}_n[m_n(\beta^*)] \) and \( \xi_D = \sqrt{n}[\hat{\xi}_n(\beta) - \Xi_n] \). Suppose that the conditions in Theorem 4 hold. Then (i) \( \mathbb{E}\|\xi_N\|^8 = O\left(1\right) \), (ii) \( \mathbb{E}\|\xi_D\|^8 = O\left(1\right) \), (iii) \( \text{Var}[\xi_D\Xi^{-1}\xi_N] = \Phi + o\left(1\right) \), and (iv) \( \text{Var}[\xi_D\Xi^{-1}\xi_N + \sigma_1^{-2} c'\Xi^{-1}\xi_D^\dagger \Omega \xi_N] = \Phi + o\left(1\right) \).

**Proof.** (i) Let \( \xi_{N,1} = \sqrt{n}\mathbb{E}_n[m(\beta^*)] = \sqrt{n}\mathbb{E}_n\left(B_1\varepsilon_1\right) \) and \( \xi_{N,2} = \xi_N - \xi_{N,1} \).

Let \( \mathcal{F}_i = \sigma\{W_i, W_{i-1}, \ldots, W_1\} \), the sigma-field generated by \( W_i, W_{i-1}, \ldots, W_1 \). Noting that \( \mathbb{E}[m(W_i; \beta^*)|\mathcal{F}_{i-1}] = \mathbb{E}[m(W_i; \beta^*)] = 0 \), we can apply Rosenthal’s inequality (e.g., Hall and Heyde 1980, p. 23) to obtain

\[
\mathbb{E}\|\xi_{N,1}\|^8 \leq C \left\{ \mathbb{E}\left[ n^{-1}\sum_{i=1}^{n}\mathbb{E}[\|m(W_i; \beta^*)\|^2|\mathcal{F}_{i-1}] \right]^4 + n^{-4}\sum_{i=1}^{n}\mathbb{E}\|m(W_i; \beta^*)\|^8 \right\} = O\left(1\right)
\]

\[
= C \left\{ [\mathbb{E}\|m(W_i; \beta^*)\|^2]^4 + n^{-3}\mathbb{E}\|m(W_i; \beta^*)\|^8 \right\} = O\left(1\right)
\]
under Assumptions A1(i) and A4(i). Similarly, we can show that $\mathbb{E} \| \xi_{N,2} - \mathbb{E}(\xi_{N,2}) \|^8 = O(1)$. In addition, noting that

$$m_n(\beta^*) - m(\beta^*) = \begin{pmatrix} B_1 \left[ K_\alpha(H'_{21} \beta^*_{21})H'_{21} \beta^*_{21} - (H'_{21} \beta^*_{21})_+ \right] \\
0 \end{pmatrix},$$

we have $\| \mathbb{E}(\xi_{N,2}) \| = O(n^{1/2} \alpha^2) = o(1)$ by the proof of Theorem 2. Then by the $C_r$ inequality,

$$\mathbb{E} \| \xi_{N,2} \|^8 \leq 128 \mathbb{E} \| \xi_{N,2} - \mathbb{E}(\xi_{N,2}) \|^8 + 128 \| \mathbb{E}(\xi_{N,2}) \|^8 = O(1) + o(1) = O(1).$$

Consequently, $\mathbb{E} \| \xi_N \|^8 \leq 128(\mathbb{E} \| \xi_{N,1} \|^8 + \mathbb{E} \| \xi_{N,2} \|^8) = O(1)$.

(ii) Let $\eta(\beta_{21}) = \eta(H_{21}; \beta_{21}) = B_1 H'_{21} [\alpha k_\alpha(H_{21} \beta_{21}) H'_{21} \beta_{21} + K_\alpha(H_{21} \beta_{21})]$. We make the following decomposition

$$\xi_D = \sqrt{\alpha} \begin{pmatrix} 0 & (\bar{G}_n[\eta(\beta^*_{21})]) \\
0 & 0 \end{pmatrix} + \sqrt{\alpha} \begin{pmatrix} 0 & (0, n^{1/2} \mathbb{E}_n[\eta(\bar{\beta}_{21}) - \eta(\beta^*_{21})]) \\
0 & 0 \end{pmatrix}$$

$$+ \sqrt{\alpha} \bar{G}_n \begin{pmatrix} -B_1 B'_1 & (B_1 H'_{20}, 0) \\
0 & -B_2 B'_2 \end{pmatrix}$$

$$\equiv \xi_{D,1} + \xi_{D,2} + \xi_{D,3}, \text{ say.}$$

As in part (i), we can readily show that $\mathbb{E} \| \xi_{D,3} \|^8 = O(\alpha^4) = o(1)$. By Theorem 2 and the fact that $\bar{\beta}_{21}$ lies between $\beta_{21}$ and $\beta^*_{21}$ element by element, we have $\bar{\beta}_{21} = \beta^*_{21} + O_P(n^{-1/2})$. Then, we can apply Sargan and Mikhail’s (1971) inequality as in the proof of Theorem 4 and show that $\mathbb{E} \| \xi_{D,3} \|^8 = o(1)$. In addition, one can readily show that $\alpha^4 \mathbb{E} \| \eta(\beta^*_{21}) \|^8 = O(1)$. It follows that $\mathbb{E} \| \xi_D \|^8 = \mathbb{E} \| \xi_{D,1} \|^8 + o(1) = \alpha^4 \mathbb{E} \| \eta(\beta^*_{21}) \|^8 = O(1).$
(iii) Note that \( \text{Var}(\xi_D^{-1}\xi_N) = \mathbb{E}(\xi_D^{-1}\xi_N\xi_N'\xi_D^{-1}) - \mathbb{E}(\xi_D^{-1}\xi_N)\mathbb{E}(\xi_N^{-1}\xi_D') \). It is easy to show that \( \mathbb{E}(\xi_D^{-1}\xi_N) = O(\alpha^{1/2}) \). Using the decomposition of \( \xi_N \) and \( \xi_D \) in (i) and (ii), we can readily show that

\[
\mathbb{E}(\xi_D^{-1}\xi_N'\xi_D') = \frac{\alpha}{n^2} \sum_{i,j,k,l=1}^{n} \mathbb{E}\left( \begin{pmatrix} 0 & \bar{\eta}_1(H_{21},i;\beta_{21}^*) \\ 0 & 0 \end{pmatrix} \Xi^{-1} \begin{pmatrix} B_{1j}B_{1k}^\varepsilon_{1j}^1\varepsilon_{1k} & B_{1j}B_{2k}^\varepsilon_{1j}2\varepsilon_{2k} \\ B_{2j}B_{1k}^\varepsilon_{2j}1\varepsilon_{1k} & B_{2j}B_{2k}^\varepsilon_{2j}2\varepsilon_{2k} \end{pmatrix} \Xi^{-1} \right) \Xi^{-1} \\
\times \begin{pmatrix} 0 & \bar{\eta}_1(H_{21},i;\beta_{21}^*) \\ 0 & 0 \end{pmatrix} \right) + o(1) \\
= \alpha \mathbb{E} \left[ \begin{pmatrix} \bar{\eta}_1(\beta_{21}^*) \Omega_{22} \bar{\eta}_1(\beta_{21}^*)' \\ 0 \end{pmatrix} \right] + o(1) = \mathbb{F}_0 + o(1),
\]

where \( \bar{\eta}_1(\beta_{21}) = \bar{\eta}_1(H_{21};\beta_{21}) = \eta_1(\beta_{21}) - \mathbb{E}[\eta_1(\beta_{21})] \), and \( \eta_1(\beta_{21}) = \eta_1(H_{21};\beta_{21}) = [0_{k_1 \times k_2}, B_1 H_{21}' \times \frac{1}{\alpha} k_{2\alpha}(H_{21}' \beta_{21})H_{21}' \beta_{21}] \). Then (iii) follows.

(iv) \( \text{Var}(\xi_D^{-1}\xi_N + \sigma_c^{-2}\xi_D^{-1}\xi_D^\dagger \Omega c\xi_N) = \text{Var}(\xi_D^{-1}\xi_N) + \text{Var}(\sigma_c^{-2}\xi_D^{-1}\xi_N^\dagger \Omega c\xi_N) + \text{Cov}(\xi_D^{-1}\xi_N, \sigma_c^{-2}\xi_D^{-1}\xi_N^\dagger \Omega c\xi_N) \). Following the analysis in (iii) and
noting that $\xi_D^\dagger$ is defined as $\xi_D$ with $\hat{\beta}$ replaced by $\hat{\beta}$, we can readily show that

$$\text{Var}\left(\sigma_c^{-2}c^\prime \Xi^{-1} \xi_D^\dagger \Omega c \xi_N\right)$$

$$= \sigma_c^{-4}c^\prime \Xi^{-1} \mathbb{E}\left(\xi_D^\dagger \Omega c \xi_N^\prime c^\prime \Omega c \xi_N^\dagger\right) \Xi^{-1} \sigma - o(1)$$

$$= \sigma_c^{-4}c^\prime \Xi^{-1} \mathbb{E}\left(\xi_D,1 \Omega c \xi_N^\prime c^\prime \Omega c \xi_N^\prime\right) \Xi^{-1} \sigma + o(1)$$

$$= \sigma_c^{-4}c^\prime \Xi^{-1} \mathbb{E}\left(\xi_D,1 \Lambda c \xi_D,1 \Omega c \xi_D,1\right) \Xi^{-1} \sigma + o(1)$$

$$= \alpha \sigma_c^{-4}c^\prime \Xi^{-1} \mathbb{E}\left[\begin{pmatrix} 0 & \bar{\eta}_1 (\beta^*_2) \\ \bar{\eta}_1 (\beta^*_2) & 0 \end{pmatrix} \Omega c \Lambda c^\prime \Omega \begin{pmatrix} 0 & \bar{\eta}_1 (\beta^*_2) \\ \bar{\eta}_1 (\beta^*_2) & 0 \end{pmatrix}^\prime\right] \Xi^{-1} \sigma + o(1)$$

and

$$\text{Cov}\left(\xi_D \Xi^{-1} \xi_N, \sigma_c^{-2}c^\prime \Xi^{-1} \xi_D^\dagger \Omega c \xi_N\right)$$

$$= \sigma_c^{-2} \mathbb{E}\left[\xi_D \Xi^{-1} \xi_N^\prime c^\prime \Omega c \xi_N^\dagger\right] \Xi^{-1} \sigma + o(1)$$

$$= \sigma_c^{-2} \mathbb{E}\left[\xi_D,1 \Xi^{-1} \xi_N^\prime c^\prime \Omega c \xi_N^\prime\right] \Xi^{-1} \sigma + o(1)$$

$$= \sigma_c^{-2} \mathbb{E}\left[\xi_D,1 \Xi^{-1} \Lambda c^\prime \Omega c \xi_D,1\right] \Xi^{-1} \sigma + o(1)$$

$$= \alpha \sigma_c^{-2} \mathbb{E}\left[\eta^*_1 (\beta_2) \eta^*_1 (\beta_2) \right] \Xi^{-1} \sigma + o(1)$$

where $\eta^*_1 (\beta_2)$ and $\eta^*_2 (\beta_2)$ are defined in Theorem 4. It follows that $\text{Var}[\xi_D \Xi^{-1} \xi_N + \sigma_c^{-2}c^\prime \Xi^{-1} \xi_D^\dagger \Omega c \xi_N] = \Phi_0 + \alpha \sigma_c^{-4} \mathbb{E}[\eta^*_1 (\beta_2)]^2 \Lambda + \alpha \sigma_c^{-2} \mathbb{E}[\eta^*_1 (\beta_2) \eta^*_1 (\beta_2) \Lambda + \Lambda \eta^*_1 (\beta_2) \eta^*_1 (\beta_2)] + o(1) = \Phi + o(1) \quad \blacksquare$
B The Generative Model

Let $\gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 = M$, then

$$
\max_{\alpha_2} Q_2(H_2, A_2)
= M + [\gamma_5 + \gamma_6 X_2 + \gamma_7 A_1] + \\
= M + \frac{\rho_1}{2}(1 - X_2)(1 - A_1) + \frac{\rho_2}{2}(1 - X_2)A_1 + \frac{\rho_3}{2}(1 + X_2)(1 - A_1) + \frac{\rho_4}{2}(1 + X_2)A_1,
$$

where $\rho_1 = [\gamma_5 - \gamma_6]$, $\rho_2 = [\gamma_5 - \gamma_6 + \gamma_7]$, $\rho_3 = [\gamma_5 + \gamma_6]$, and $\rho_4 = [\gamma_5 + \gamma_6 + \gamma_7]$. Therefore, for $e^\dagger(x) = e^x/(1 + e^x)$, we have

$$
Q_1(H_1, A_1) = \mathbb{E}[Y_1 + \max_{\alpha_2} Q_2(H_2, A_2)|H_1, A_1]
= M + \frac{\rho_1}{2}(1 - A_1)\mathbb{E}[1 - X_2|X_1, A_1] + \frac{\rho_2}{2}A_1\mathbb{E}[1 - X_2|X_1, A_1]
+ \frac{\rho_3}{2}(1 - A_1)\mathbb{E}[1 + X_2|X_1, A_1] + \frac{\rho_4}{2}A_1\mathbb{E}[1 + X_2|X_1, A_1]
= M + \frac{\rho_1 + \rho_3}{2}(1 - A_1) + \frac{\rho_2 + \rho_4}{2}A_1 + \frac{1}{2}(\rho_3 - \rho_1)(1 - A_1)\mathbb{E}[X_2|X_1, A_1]
+ \frac{1}{2}(\rho_4 - \rho_2)A_1\mathbb{E}[X_2|X_1, A_1]
\overset{(4)}{=} M + \frac{\rho_1 + \rho_3}{2}(1 - A_1) + \frac{\rho_2 + \rho_4}{2}A_1
+ \frac{1}{2}(\rho_3 - \rho_1)\left(e^\dagger(-\delta_1)(1 - X_1)(1 - A_1) + e^\dagger(\delta_1)(1 + X_1)(1 - A_1) - (1 - A_1)\right)
+ \frac{1}{2}(\rho_4 - \rho_2)\left(e^\dagger(-\delta_1 + \delta_2)(1 - X_1)A_1 + e^\dagger(\delta_1 + \delta_2)(1 + X_1)A_1 - A_1\right)
\overset{(5)}{=} \psi_1 + \psi_2 X_1 + (\psi_3 + \psi_4 X_1) A_1,
$$

where

$$
\psi_1 = \gamma_1 + \frac{\rho_1 + \rho_3}{2} + \frac{1}{2}(\rho_3 - \rho_1)\left(e^\dagger(-\delta_1) + e^\dagger(\delta_1) - 1\right) = \gamma_1 + \frac{\rho_1 + \rho_3}{2},
$$
\[
\psi_2 = \gamma_2 + \frac{\rho_3 - \rho_1}{2} \left( -e^\dagger(-\delta_1) + e^\dagger(\delta_1) \right) = \gamma_2 + \frac{\rho_3 - \rho_1}{2} \left( 2e^\dagger(\delta_1) - 1 \right),
\]
\[
\psi_3 = \gamma_3 - \frac{\rho_1 - \rho_2 + \rho_3 - \rho_4}{2} + \frac{\rho_3 - \rho_1}{2} \left( -e^\dagger(-\delta_1) - \expit(\delta_1) + 1 \right) + \frac{\rho_4 - \rho_2}{2} \left( e^\dagger(-\delta_1 + \delta_2) + e^\dagger(\delta_1 + \delta_2) - 1 \right) = \gamma_3 - \frac{\rho_1 - \rho_2 + \rho_3 - \rho_4}{2} + \frac{\rho_4 - \rho_2}{2} \left( e^\dagger(-\delta_1 + \delta_2) + e^\dagger(\delta_1 + \delta_2) - 1 \right),
\]
\[
\psi_4 = \gamma_4 + \frac{\rho_3 - \rho_1}{2} \left( 1 - 2e^\dagger(\delta_1) \right) + \frac{\rho_4 - \rho_2}{2} \left( -e^\dagger(-\delta_1 + \delta_2) + e^\dagger(\delta_1 + \delta_2) \right).
\]

In Step (4) of derivation of \(Q_1(H_1, A_1)\), we used the fact that
\[
\mathbb{E}[X_2|X_1, A_1] = 2e^\dagger(\delta_1X_1 + \delta_2A_1) - 1
\]
\[
= e^\dagger(-\delta_1)(1 - X_1)(1 - A_1) + e^\dagger(-\delta_1 + \delta_2)(1 - X_1A_1)
\]
\[
+ e^\dagger(\delta_1)(1 + X_1)(1 - A_1) + e^\dagger(\delta_1 + \delta_2)(1 + X_1A_1) - 1,
\]
and thus
\[
(1 - A_1)\mathbb{E}[X_2|X_1, A_1] = e^\dagger(-\delta_1)(1 - X_1)(1 - A_1) + e^\dagger(\delta_1)(1 + X_1)(1 - A_1) - (1 - A_1),
\]
\[
A_1\mathbb{E}[X_2|X_1, A_1] = e^\dagger(-\delta_1 + \delta_2)(1 - X_1)A_1 + e^\dagger(\delta_1 + \delta_2)(1 + X_1)A_1 - A_1,
\]
which is in turn due the facts that \(A_1^2 = A_1\), \(A_1(1 - A_1) = 0\) and \((1 - A_1)^2 = 1 - A_1\) since \(A_1 \in \{0, 1\}\). In Step (5), we used the fact that \(e^\dagger(-x) + e^\dagger(x) = 1\).

**C  An Example to Guide Applied Researchers**

In this section, we provide detailed steps of using our R code of the smoothed Q-learning algorithm for the empirical example in the main text. It is performed on data from the Clinical Antipsychotic Trials of Intervention Effectiveness-Alzheimer’s Disease (CATIE-
AD) study, and we focus on treatment definition I of the empirical section.

First, download the file “Q-Learning-All-Methods.R”, in which the following methods are implemented: (1) Smoothed Q-learning algorithm in Fan et al. (2018); (2) Standard Q-learning algorithm with percentile bootstrapped confidence intervals; (3) Adaptive method in Laber et al. (2014); (4) Hard-thresholding method in Moodie and Richardson (2010); (5) Soft-thresholding method in Chakraborty, Murphy, and Strecher (2010).

The main function in “Q-Learning-All-Methods.R” is

\[
Q.Learning.All.Methods = \text{function}(H_{10}, H_{11}, A_1, Y_1, \\
H_{20}, H_{21}, A_2, Y_2, \\
\text{kernel} = \text{Gaussian'}, alpha = 0.05, NB = 500),
\]

where \( H_{tj}, t = 1, 2, j = 0, 1 \) is part \( j \) of history in period \( t \), of dimension \( n \times (k_{tj} + 1) \), where 1 represents the intercept, \( k_{tj} \) is the number of covariates, and \( n \) is the sample size; \( A_t, Y_t, t = 1, 2 \) are the vectors of treatment variable and dependent variable in period \( t \), each of dimension \( n \times 1 \); \( \text{kernel} \) specifies the choice of kernel function in the smoothed Q-learning algorithm, with default value ‘Gaussian’, and a user can change it to one of ‘Epanechikov’, ‘Triangular’, ‘Quartic’, or ‘Triweight’; \( \text{alpha} \) is the significance level with default value 0.05, and \( \text{NB} \) is the number of bootstrap draws for other methods such as Q-learning, thresholding methods, with default value 500.

In the empirical example in Fan et al. (2018), \( H_{10} \) includes a constant, Neuropsychiatric Inventory (NPI), age, and gender in period 1, \( H_{11} = H_{10} \), \( H_{20} \) includes a constant, Neuropsychiatric Inventory (NPI), age, and gender in period 2. The treatment variable \( A_t = 1, t = 1, 2 \) if a patient uses olanzapine and 0 otherwise. \( Y_{t}, t = 1, 2 \) are the
negated score of Clinical Global Impression of Change (CGIC). We import the data and define the variables in a separate R file ‘xxx.R’, and then type ‘source(‘Q-Learning-All-Methods.R’’)’ in it to import the code for Q-learning methods. Next, we use the function ‘Q.Learning.All.Methods’, choose kernel = ‘Epanechikov’, and use the default values for alpha and NB to obtain the first-stage estimation results for all methods. We present the results from the smoothed Q-learning algorithms below, where Stage 1 results are from the function ‘Q.Learning.All.Methods’, and Stage 2 is a standard ordinary least square regression that a user can easily obtain from the ‘lm’ command in R.

<table>
<thead>
<tr>
<th></th>
<th>Stage 1</th>
<th>Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SQ(RoT)</td>
<td>SQ(α₀)</td>
</tr>
<tr>
<td>NPIₜ</td>
<td>-0.011</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>[-0.023, 0.000]</td>
<td>[-0.023, 0.000]</td>
</tr>
<tr>
<td>Age</td>
<td>0.033</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>[-0.008, 0.076]</td>
<td>[-0.008, 0.076]</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.392</td>
<td>-0.391</td>
</tr>
<tr>
<td></td>
<td>[-1.002, 0.218]</td>
<td>[-1.001, 0.218]</td>
</tr>
<tr>
<td>Aₜ</td>
<td>2.819</td>
<td>2.820</td>
</tr>
<tr>
<td>NPIₜ·Aₜ</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>[0.007, 0.047]</td>
<td>[0.006, 0.047]</td>
</tr>
<tr>
<td>Age·Aₜ</td>
<td>-0.049</td>
<td>-0.049</td>
</tr>
<tr>
<td></td>
<td>[-0.102, 0.003]</td>
<td>[-0.102, 0.003]</td>
</tr>
<tr>
<td>Gender·Aₜ</td>
<td>-0.125</td>
<td>-0.125</td>
</tr>
<tr>
<td></td>
<td>[-0.920, 0.668]</td>
<td>[-0.919, 0.668]</td>
</tr>
</tbody>
</table>

Regression Results for Treatment Definition I (Line 1: point estimate; Line 2: 95% CIs)
References


**Corresponding author:** Yanqin Fan, Department of Economics, University of Washington, Box 353330, Seattle, WA, 98195. E-mail: fany88@uw.edu
<table>
<thead>
<tr>
<th>Design</th>
<th>Sample Size: 200</th>
<th></th>
<th>Sample Size: 1000</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (×10²)</td>
<td>MSE (×10²)</td>
<td>CP (%)</td>
<td>IL</td>
</tr>
<tr>
<td>1</td>
<td>SQ(RoT) -0.0949 3.4930 94.37 0.7335 -0.0449 0.7359 94.79 0.3336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SQ((\hat{\alpha}_0)) -0.0914 3.4750 94.46 0.7339 -0.0469 0.7360 94.82 0.3333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SQ((\hat{\alpha}^*)) -0.0648 3.5150 94.34 0.7353 -0.1354 0.7480 94.86 0.3376</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FACI - - 95.74 0.7971 - - 95.13 0.3448</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q -0.1003 3.4660 94.69 0.7163 -0.0420 0.7338 94.61 0.3275</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HT_{0.08} -3.2920 4.1490 95.89 0.7725 -1.3190 0.9445 94.85 0.3509</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HT_{0.20} -1.9280 3.8920 95.37 0.7497 -0.4750 0.8157 94.75 0.3402</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ST -3.6590 3.3750 94.79 0.7167 -3.9560 0.9658 92.79 0.3362</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SQ((\hat{\alpha}^*)) -0.1417 2.5730 97.12 0.7203 -0.0701 0.5170 97.53 0.3273</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SQ((\hat{\alpha}^*)) -0.1458 2.6090 97.10 0.7307 -0.0709 0.5226 97.48 0.3281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FACI - - 98.33 0.8063 - - 98.47 0.3625</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q -0.1423 2.5650 97.07 0.6664 -0.0569 0.5169 97.14 0.2963</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HT_{0.08} -0.1209 3.6680 98.37 0.7926 -0.0170 0.7221 98.41 0.3543</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HT_{0.20} -0.2410 3.3550 98.33 0.7594 -0.0666 0.6782 98.31 0.3393</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ST -0.0027 3.6290 96.41 0.7342 0.0015 0.7099 96.66 0.3263</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SQ((\alpha^{*})) -8.5060 4.0560 93.52 0.7550 -3.5630 0.8023 94.19 0.3414</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SQ((\hat{\alpha}_0)) -8.5420 4.0500 93.63 0.7568 -3.6000 0.8024 94.23 0.3414</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SQ((\hat{\alpha}^*)) -8.1870 4.0330 93.80 0.7611 -3.3890 0.7930 94.53 0.3446</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FACI - - 95.57 0.8252 - - 96.43 0.3602</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q -8.6080 4.0510 90.69 0.7130 -3.6150 0.8026 90.55 0.3242</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HT_{0.08} -8.6900 6.9100 94.41 0.8946 -0.1508 0.7776 96.11 0.3586</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HT_{0.20} -6.3550 5.1070 94.08 0.8207 -1.2150 0.7710 94.84 0.3467</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ST -24.710 10.5700 80.54 0.7917 -5.5940 1.2070 89.01 0.3664</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Performances of Different Estimators for \(\psi_3\)
### Table 2: Performances of Different Estimators for $\psi_3$

<table>
<thead>
<tr>
<th>Design</th>
<th>Sample Size: 200</th>
<th>Sample Size: 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (×10²)</td>
<td>MSE (×10²)</td>
</tr>
<tr>
<td>4</td>
<td>SQ(RoT)</td>
<td>-8.8710</td>
</tr>
<tr>
<td></td>
<td>SQ($\hat{\alpha}_0^*$)</td>
<td>-9.0770</td>
</tr>
<tr>
<td></td>
<td>SQ($\hat{\alpha}^*$)</td>
<td>-8.4760</td>
</tr>
<tr>
<td></td>
<td>FACI</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>-9.1970</td>
</tr>
<tr>
<td></td>
<td>HT_{0.08}</td>
<td>-2.3470</td>
</tr>
<tr>
<td></td>
<td>HT_{0.20}</td>
<td>-4.3260</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>-16.5100</td>
</tr>
<tr>
<td>5</td>
<td>SQ(RoT)</td>
<td>-8.8110</td>
</tr>
<tr>
<td></td>
<td>SQ($\hat{\alpha}_0^*$)</td>
<td>-9.0180</td>
</tr>
<tr>
<td></td>
<td>SQ($\hat{\alpha}^*$)</td>
<td>-8.4130</td>
</tr>
<tr>
<td></td>
<td>FACI</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>-9.1390</td>
</tr>
<tr>
<td></td>
<td>HT_{0.08}</td>
<td>-2.1350</td>
</tr>
<tr>
<td></td>
<td>HT_{0.20}</td>
<td>-4.2460</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>-0.6916</td>
</tr>
<tr>
<td>6</td>
<td>SQ(RoT)</td>
<td>-0.1440</td>
</tr>
<tr>
<td></td>
<td>SQ($\hat{\alpha}_0^*$)</td>
<td>-0.1427</td>
</tr>
<tr>
<td></td>
<td>SQ($\hat{\alpha}^*$)</td>
<td>-0.1447</td>
</tr>
<tr>
<td></td>
<td>FACI</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>-0.1431</td>
</tr>
<tr>
<td></td>
<td>HT_{0.08}</td>
<td>-0.0949</td>
</tr>
<tr>
<td></td>
<td>HT_{0.20}</td>
<td>-0.1858</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>-0.0032</td>
</tr>
</tbody>
</table>
Table 3: Performances of Different Estimators for $\psi_4$

<table>
<thead>
<tr>
<th>Design</th>
<th>Sample Size: 200</th>
<th>Sample Size: 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias ($\times 10^2$)</td>
<td>MSE ($\times 10^2$)</td>
</tr>
<tr>
<td>1 SQ(RoT)</td>
<td>−0.2118</td>
<td>2.4430</td>
</tr>
<tr>
<td></td>
<td>2 SQ(\hat{\alpha}_0^2)</td>
<td>2.4460</td>
</tr>
<tr>
<td></td>
<td>3 SQ(\hat{\alpha}^2)</td>
<td>2.4560</td>
</tr>
<tr>
<td></td>
<td>FACI</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>−0.2174</td>
</tr>
<tr>
<td></td>
<td>HT_{0.08}</td>
<td>1.3120</td>
</tr>
<tr>
<td></td>
<td>HT_{0.20}</td>
<td>0.4963</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>3.2640</td>
</tr>
<tr>
<td>2 SQ(RoT)</td>
<td>−0.2971</td>
<td>2.0950</td>
</tr>
<tr>
<td></td>
<td>2 SQ(\hat{\alpha}_0^2)</td>
<td>2.0960</td>
</tr>
<tr>
<td></td>
<td>3 SQ(\hat{\alpha}^2)</td>
<td>2.0970</td>
</tr>
<tr>
<td></td>
<td>FACI</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>−0.2964</td>
</tr>
<tr>
<td></td>
<td>HT_{0.08}</td>
<td>−0.2233</td>
</tr>
<tr>
<td></td>
<td>HT_{0.20}</td>
<td>−0.2475</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>−0.1996</td>
</tr>
</tbody>
</table>

3 SQ(RoT) | −0.1980          | 2.1030            | 94.58  | 0.5597 | −0.0031 | 0.4028            | 95.19  | 0.2506 |
<p>|          | 2 SQ(\hat{\alpha}<em>0^2) | 2.1030            | 94.54  | 0.5599 | −0.0032 | 0.4028            | 95.19  | 0.2506 |
|          | 3 SQ(\hat{\alpha}^2) | 2.1050            | 94.53  | 0.5593 | −0.0035 | 0.4030            | 95.21  | 0.2504 |
|          | FACI              | -                 | 95.76  | 0.5927 | -         | -                 | 95.62  | 0.2572 |
|          | Q                 | −0.1982           | 2.1020            | 94.37  | 0.5614 | −0.0032 | 0.4027            | 94.83  | 0.2482 |
|          | HT</em>{0.08}         | 0.0348            | 2.2120            | 94.90  | 0.5757 | 0.0038 | 0.4108            | 94.96  | 0.2507 |
|          | HT_{0.20}         | −0.1147           | 2.1530            | 94.67  | 0.5691 | 0.0025 | 0.4080            | 94.86  | 0.2497 |
|          | ST                | 0.0818            | 2.1520            | 94.49  | 0.5673 | 0.1457 | 0.4125            | 94.73  | 0.2507 |</p>
<table>
<thead>
<tr>
<th>Design</th>
<th>Sample Size: 200</th>
<th>Sample Size: 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ(RoT)</td>
<td>-0.1906</td>
<td>2.1040</td>
</tr>
<tr>
<td>SQ(\hat{\alpha}_0^*)</td>
<td>-0.1908</td>
<td>2.1030</td>
</tr>
<tr>
<td>SQ(\hat{\alpha}_0^*)</td>
<td>-0.1898</td>
<td>2.1040</td>
</tr>
<tr>
<td>FACI</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Q</td>
<td>-0.1909</td>
<td>2.1020</td>
</tr>
<tr>
<td>HT_{0.08}</td>
<td>-0.1740</td>
<td>2.1410</td>
</tr>
<tr>
<td>HT_{0.20}</td>
<td>-0.2119</td>
<td>2.1320</td>
</tr>
<tr>
<td>ST</td>
<td>0.1676</td>
<td>2.1590</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ(RoT)</td>
<td>-0.1675</td>
<td>2.1030</td>
</tr>
<tr>
<td>SQ(\hat{\alpha}_0^*)</td>
<td>-0.1668</td>
<td>2.1030</td>
</tr>
<tr>
<td>SQ(\hat{\alpha}_0^*)</td>
<td>-0.1663</td>
<td>2.1040</td>
</tr>
<tr>
<td>FACI</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Q</td>
<td>-0.1677</td>
<td>2.1020</td>
</tr>
<tr>
<td>HT_{0.08}</td>
<td>-0.1732</td>
<td>2.0970</td>
</tr>
<tr>
<td>HT_{0.20}</td>
<td>-0.1911</td>
<td>2.1110</td>
</tr>
<tr>
<td>ST</td>
<td>-0.1856</td>
<td>2.0790</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ(RoT)</td>
<td>-0.2983</td>
<td>2.0950</td>
</tr>
<tr>
<td>SQ(\hat{\alpha}_0^*)</td>
<td>-0.3010</td>
<td>2.0960</td>
</tr>
<tr>
<td>SQ(\hat{\alpha}_0^*)</td>
<td>-0.3035</td>
<td>2.0970</td>
</tr>
<tr>
<td>FACI</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Q</td>
<td>-0.2975</td>
<td>2.0940</td>
</tr>
<tr>
<td>HT_{0.08}</td>
<td>-0.2086</td>
<td>2.1100</td>
</tr>
<tr>
<td>HT_{0.20}</td>
<td>-0.2535</td>
<td>2.1200</td>
</tr>
<tr>
<td>ST</td>
<td>-0.1995</td>
<td>2.0810</td>
</tr>
</tbody>
</table>
Table 5: Comparison of Optimality for Different Methods

<table>
<thead>
<tr>
<th>Design</th>
<th>$n$</th>
<th>$Y_2$</th>
<th>SQ(RoT)</th>
<th>SQ($\hat{\alpha}_0$)</th>
<th>SQ($\hat{\alpha}^*$)</th>
<th>Q</th>
<th>HT$_{0.08}$</th>
<th>HT$_{0.20}$</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>2.8685</td>
<td>2.8502</td>
<td>2.8500</td>
<td>2.8508</td>
<td>2.8509</td>
<td>2.8481</td>
<td>2.8475</td>
<td>2.8480</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.8622</td>
<td>2.8625</td>
<td>2.8624</td>
<td>2.8618</td>
<td>2.8618</td>
<td>2.8616</td>
<td>2.8622</td>
<td>2.8618</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>2.0123</td>
<td>2.0078</td>
<td>2.0078</td>
<td>2.0078</td>
<td>2.0078</td>
<td>2.0074</td>
<td>2.0075</td>
<td>2.0073</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.0070</td>
<td>2.0025</td>
<td>2.0025</td>
<td>2.0025</td>
<td>2.0025</td>
<td>2.0021</td>
<td>2.0022</td>
<td>2.0020</td>
</tr>
<tr>
<td>3</td>
<td>200</td>
<td>2.5123</td>
<td>2.4979</td>
<td>2.4979</td>
<td>2.4979</td>
<td>2.4980</td>
<td>2.4898</td>
<td>2.4947</td>
<td>2.4670</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.5070</td>
<td>2.5070</td>
<td>2.5070</td>
<td>2.5070</td>
<td>2.5070</td>
<td>2.5070</td>
<td>2.5070</td>
<td>2.5070</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.9970</td>
<td>2.9970</td>
<td>2.9970</td>
<td>2.9970</td>
<td>2.9970</td>
<td>2.9970</td>
<td>2.9970</td>
<td>2.9970</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
<td>2.0023</td>
<td>2.0023</td>
<td>2.0023</td>
<td>2.0023</td>
<td>2.0023</td>
<td>2.0023</td>
<td>2.0023</td>
<td>2.0023</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.9970</td>
<td>1.9970</td>
<td>1.9970</td>
<td>1.9970</td>
<td>1.9970</td>
<td>1.9970</td>
<td>1.9970</td>
<td>1.9970</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>1.9986</td>
<td>1.9986</td>
<td>1.9986</td>
<td>1.9986</td>
<td>1.9986</td>
<td>1.9986</td>
<td>1.9986</td>
<td>1.9986</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.0009</td>
<td>2.0009</td>
<td>2.0009</td>
<td>2.0009</td>
<td>2.0009</td>
<td>2.0009</td>
<td>2.0009</td>
<td>2.0009</td>
</tr>
</tbody>
</table>
Table 6: Regression Results for Treatment Definition I (Line 1: point estimate; Line 2: 95% CIs)

<table>
<thead>
<tr>
<th></th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 2</th>
<th>Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SQ(RoT)</td>
<td>SQ($\hat{\alpha}_0$)</td>
<td>SQ($\hat{\alpha}^*$)</td>
<td>OLS</td>
</tr>
<tr>
<td><strong>NPI_t</strong></td>
<td>-0.011</td>
<td>-0.011</td>
<td>-0.011</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>[-0.023, 0.000]</td>
<td>[-0.023, 0.000]</td>
<td>[-0.023, 0.000]</td>
<td>[-0.027, -0.005]</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>0.033</td>
<td>0.033</td>
<td>0.033</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>[-0.008, 0.076]</td>
<td>[-0.008, 0.076]</td>
<td>[-0.008, 0.076]</td>
<td>[-0.015, 0.046]</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>-0.392</td>
<td>-0.391</td>
<td>-0.390</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td>[-1.002, 0.218]</td>
<td>[-1.001, 0.218]</td>
<td>[-1.001, 0.219]</td>
<td>[-0.529, 0.341]</td>
</tr>
<tr>
<td><strong>A_t</strong></td>
<td>2.819</td>
<td>2.820</td>
<td>2.819</td>
<td>3.972</td>
</tr>
<tr>
<td><strong>NPI_t·A_t</strong></td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>[0.007, 0.047]</td>
<td>[0.006, 0.047]</td>
<td>[0.007, 0.047]</td>
<td>[-0.019, 0.030]</td>
</tr>
<tr>
<td><strong>Age·A_t</strong></td>
<td>-0.049</td>
<td>-0.049</td>
<td>-0.049</td>
<td>-0.052</td>
</tr>
<tr>
<td></td>
<td>[-0.102, 0.003]</td>
<td>[-0.102, 0.003]</td>
<td>[-0.102, 0.003]</td>
<td>[-0.111, 0.006]</td>
</tr>
<tr>
<td><strong>Gender·A_t</strong></td>
<td>-0.125</td>
<td>-0.125</td>
<td>-0.125</td>
<td>-0.866</td>
</tr>
<tr>
<td></td>
<td>[-0.920, 0.668]</td>
<td>[-0.919, 0.668]</td>
<td>[-0.919, 0.667]</td>
<td>[-1.945, 0.212]</td>
</tr>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SQ(RoT)</td>
<td>SQ((\hat{\alpha}_0^*))</td>
<td>SQ((\hat{\alpha}^*))</td>
<td>OLS</td>
</tr>
<tr>
<td>NPI(_t)</td>
<td>-0.023</td>
<td>-0.023</td>
<td>-0.023</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>[-0.050, 0.004]</td>
<td>[-0.051, 0.004]</td>
<td>[-0.050, 0.004]</td>
<td>[-0.064, 0.018]</td>
</tr>
<tr>
<td>Age</td>
<td>0.042</td>
<td>0.043</td>
<td>0.042</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>[-0.046, 0.132]</td>
<td>[-0.046, 0.132]</td>
<td>[-0.046, 0.132]</td>
<td>[-0.053, 0.087]</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.315</td>
<td>-0.315</td>
<td>-0.315</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>[-1.891, 1.261]</td>
<td>[-1.910, 1.279]</td>
<td>[-1.891, 1.261]</td>
<td>[-1.238, 1.274]</td>
</tr>
<tr>
<td>(A_t)</td>
<td>1.541</td>
<td>1.540</td>
<td>1.540</td>
<td>2.561</td>
</tr>
<tr>
<td>NPI(_t)·(A_t)</td>
<td>0.036</td>
<td>0.036</td>
<td>0.036</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>[0.008, 0.064]</td>
<td>[0.008, 0.064]</td>
<td>[0.008, 0.064]</td>
<td>[-0.020, 0.070]</td>
</tr>
<tr>
<td>Age·(A_t)</td>
<td>-0.043</td>
<td>-0.043</td>
<td>-0.043</td>
<td>-0.054</td>
</tr>
<tr>
<td></td>
<td>[-0.131, 0.044]</td>
<td>[-0.131, 0.044]</td>
<td>[-0.131, 0.044]</td>
<td>[-0.141, 0.032]</td>
</tr>
<tr>
<td>Gender·(A_t)</td>
<td>0.112</td>
<td>0.114</td>
<td>0.115</td>
<td>-0.276</td>
</tr>
<tr>
<td></td>
<td>[-1.002, 1.228]</td>
<td>[-1.009, 1.238]</td>
<td>[-1.004, 1.235]</td>
<td>[-1.697, 1.145]</td>
</tr>
</tbody>
</table>