6-2019

A smoothed Q-learning algorithm for estimating optimal dynamic treatment regime

Yanqin FAN
University of Washington

Ming HE
University of Washington

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Xiao-Hua ZHOU
University of Washington

Citation

FAN, Yanqin; HE, Ming; SU, Liangjun; and ZHOU, Xiao-Hua. A smoothed Q-learning algorithm for estimating optimal dynamic treatment regime. (2019). *Scandinavian Journal of Statistics*. 46, (2), 446-469.

Available at: https://ink.library.smu.edu.sg/soe_research/2044

Published in Scandinavian Journal of Statistics, December 2018. Advance online
https://doi.org/10.1111/sjos.12359

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

Yanqin Fan[†] Ming He[‡] Liangjun Su[§] Xiao-Hua Zhou[¶]

This version: March 24, 2016

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

[*] We thank Han Hong, Eric Laber, and participants in the Causal Inference Workshop in Department of Biostatistics at the University of Washington 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.

[†] Department of Economics, University of Washington, Box 353330, Seattle, WA, 98195.

[‡] Department of Economics, University of Washington, Box 353330, Seattle, WA, 98195.

[§] School of Economics, Singapore Management University, Stamford Road 90, Singapore 178903.

[¶] Department of Biostatistics, University of Washington, Seattle, WA, 98195.

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[1] 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 the pros and cons 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 so can be done using any statistical software. 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. (2013) for the exact expression. As shown in Robins (2004) for g-estimation[2] and Laber et al. (2013) 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. (2013) 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.[3]

[1] In general, the mean outcome can be replaced by the mean of any known function of the outcome.

[2] We refer interested readers to Robins (2004) for g-estimation and the implications of the max operation on its asymptotic properties.

[3] We note that none of these bootstrap CIs is asymptotically valid under exceptional laws.

Using local sequences, Laber et al. (2013) 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).

This paper develops a new algorithm, a smoothed Q-learning algorithm, for estimating optimal DTRs. It involves the same two estimation steps as the Q-learning algorithm except that in the smoothed Q-learning 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 smoothed Q-learning 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 smoothed Q-learning estimators possess standard asymptotic properties, namely, they are asymptotically unbiased and normally distributed even under exceptional laws.[4] As a result, statistical inference based on the smoothed Q-learning estimator is standard. Indeed we construct a consistent estimator of the asymptotic variance of the smoothed Q-learning estimator based on which Wald-type inference can be carried out. Further by deriving higher order mean square error (MSE) expansions for the smoothed Q-learning estimator, we derive the optimal smoothing parameter and propose a data-driven method for estimating it.

To shed further light on the finite sample performances of some existing statistical estimation and inference methods and our smoothed Q-learning algorithm, we conduct an extensive simulation study following a similar design in Chakraborty et al. (2010). First, we consider seven estimators: the smoothed Q-learning estimator with three different bandwidth choices, the Q-learning estimator, the hard-thresholding estimator of Moodie and Richardson (2010) with two choices of the tuning parameter, and the soft-thresholding (ST) estimator of Chakraborty et al. (2010). Second, we compare eight CIs: the first three are Wald-type CIs constructed from our smoothed Q-learning estimators; the fourth one is the FACI in Laber et al. (2013); 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). We construct six designs ranging from fully regular to close to non-regular to fully non-regular. We find that (i) all methods perform well in fully regular designs; (ii) the

[4] In Hirano and Porter (2012) and Robins (2004), it is shown that when the true parameter is a non-differentiable functional of the underlying DGP, there is no regular or locally asymptotically unbiased estimator. However this does not imply non-existence of pointwise asymptotically unbiased estimator. The Q-learning estimator is point-wise asymptotically biased and non-normally distributed under exceptional laws, but our smoothed Q-learning estimator is pointwise asymptotically unbiased and asymptotically normally distributed under assumptions adopted in this paper.

Q-learning and soft-thresholding CIs may severely under-cover in non-regular and close to non-regular designs; (iii) the FACI is in general valid but can be very conservative; (iv) in term of the MSE, the hard-thresholding estimators and our smoothed Q-learning estimators are comparable for all the designs, while in terms of the CP, the hard-thresholding CIs over-cover more than our smoothed Q-learning CI in fully non-regular and close to fully non-regular designs. Overall, our smoothed Q-learning method and the associated Wald CI perform the best for the designs considered in terms of MSE, CP, and IL.[5] Moreover it is the least costly computationally.

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 smoothed Q-learning 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 follow. Section 2 first reviews the basic set-up and the Q-learning algorithm, then presents the smoothed Q-learning algorithm. Section 3 establishes the asymptotic theory for the smoothed Q-learning algorithm. The optimal smoothing parameter is derived in Section 4. Section 4 also develops a data-driven method for estimating the optimal smoothing parameter. 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 offered in Section 7. All the technical proofs are relegated to Appendix A. 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

[5] We expect the penalized Q-learning in Song et al. (2015), the adaptive Q-learning in Goldberg, Song, and Kosorok (2013), and the m-out-of-n bootstrap in Chakraborty, Laber, and Zhao (2013) to perform similarly to our smoothed Q-learning algorithm. All three methods are shown to be asymptotically valid under regularity conditions. They are however computationally more costly than our smoothed Q-learning algorithm.

transpose as A', its Frobenius norm as $\|A\| \left(\equiv [\mathrm{tr}(AA')]^{1/2}\right)$, and its spectral norm as $\|A\|_{\mathrm{sp}} \left(\equiv [\lambda_{\max}(AA')]^{1/2}\right)$, where $\lambda_{\max}(\cdot)$ (similarly $\lambda_{\min}(\cdot)$) denotes the maximum (minimum) eigenvalue of certain matrix. We will use $0_{a\times b}$ to denote the $a \times b$ matrix of zeros. Let $\overset{d}{\to}$ and $\overset{p}{\to}$ denote convergence in distribution and 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 n i.i.d. observations denoted as $\{X_{1i}, A_{1i}, Y_{1i}, X_{2i}, A_{2i}, Y_{2i}\}_{i=1}^{n}$, 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 π which maximizes the expected outcome $E\left(Y_1^{\pi} + Y_2^{\pi}\right)$, where Y_t^{π} denotes the potential outcome in period t associated with treatment regime π. 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 smoothed Q-learning 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 $(a_1, a_2) \in \{0, 1\}^2$,

$$\begin{aligned} A_1 &\perp \left(Y_1(a_1), X_2(a_1), Y_2(a_1, a_2)\right) | H_1 \text{ and} \\ A_2 &\perp Y_2(a_1, a_2) | H_2, \end{aligned}$$

  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[6] 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.

[6] We note that $X_2(a_1)$ could include $Y_1(a_1)$ as a component in which case $Y_1(a_1)$ in the first condition in NUC is redundant.

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:

$$\begin{aligned} Q_2\left(h_2, a_2\right) &= \mathbb{E}\left[Y_2 | H_2 = h_2, A_2 = a_2\right] \text{ and} \\ Q_1\left(h_1, a_1\right) &= \mathbb{E}\Big[Y_1 + \max_{a_2 \in \{0,1\}} Q_2\left(H_2, a_2\right) | H_1 = h_1, A_1 = a_1\Big], \end{aligned} \tag{2.1}$$

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

$$\pi_t^{dp}\left(h_t\right) \in \arg \max_{a_t \in \{0,1\}} Q_t\left(h_t, a_t\right), \ t = 1, 2, \tag{2.2}$$

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

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

$$Q_t\left(H_{ti}, A_{ti}; \beta_t\right) = H'_{t0,i}\beta_{t0} + (H'_{t1,i}\beta_{t1})A_{ti}, \ t = 1, 2, \tag{2.3}$$

where $\beta_t = (\beta'_{t0}, \beta'_{t1})'$, $H_{t0,i}$ and $H_{t1,i}$ are column vectors of features depending on individual i's history at stage t. They can be subvectors or functions of H_{ti} 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, where $H_{t1,i}$ which also includes an intercept term denotes the treatment effects of history. Under the above linear specification, an optimal DTR is given by

$$\pi_t^{dp}\left(h_t\right) \in \arg \max_{a_t \in \{0,1\}} \left[(h'_{t1}\beta^*_{t1})a_t\right], t = 0, 1, \tag{2.4}$$

where $h'_t = (h'_{t0}, h'_{t1})$ and β^*_{t1} denotes the true value of β_{t1}.

Remark. 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 $h'_{t1}\beta^*_{t1} = 0$.

Estimating the optimal DTR is equivalent to estimating the unknown parameters in the Q functions. Let β^*_t denote the true value of β_t with $\beta^*_t = (\beta^{*\prime}_{t0}, \beta^{*\prime}_{t1})'$. The Q-learning algorithm proceeds to estimate β^*_1 and β^*_2 in two steps.

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

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

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

$$\begin{aligned}\tilde{Y}_{1q,i} &= Y_{1i} + \max_{A_{2i}\in\{0,1\}} Q_2\left(H_{2i}, A_{2i}; \hat{\beta}_2\right) \\ &= Y_{1i} + H'_{20,i}\hat{\beta}_{20} + \left(H'_{21,i}\hat{\beta}_{21}\right) 1\{H'_{21,i}\hat{\beta}_{21} \geq 0\},\end{aligned}$$

where $1\{\cdot\}$ is the indicator function.

(b) Regress $\tilde{Y}_{1q,i}$ on H_{1i} and A_{1i}, $i = 1, \cdots, n$, to obtain

$$\hat{\beta}_{1q} = \arg\min_{\beta_1} \sum_{i=1}^{n} \left[\tilde{Y}_{1q,i} - Q_1\left(H_{1i}, A_{1i}; \beta_1\right)\right]^2. \tag{2.6}$$

The Q-learning algorithm is popular in applied work because of its computational simplicity. However statistical inference for the first stage parameter β_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 $\Pr\left(H'_{21,i}\beta_{21}^* = 0\right) = 0$ to non-normal when $\Pr\left(H'_{21,i}\beta_{21}^* = 0\right) > 0$ (exceptional law).[7] This was first observed by Robins (2004) for g-estimation. Laber et al. (2013) 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 smoothed Q-learning algorithm. In sharp contrast to $\hat{\beta}_{1q}$, we show that under conditions adopted in Section 3 in this paper, the asymptotic distribution of the smoothed Q-learning estimator is normal whether $\Pr\left(H'_{21,i}\beta_{21}^* = 0\right) = 0$ or $\Pr\left(H'_{21,i}\beta_{21}^* = 0\right) > 0$.

The smoothed Q-learning 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 above Q-learning algorithm with Step 2^* below.

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

$$\tilde{Y}_{1sq,i} = Y_{1i} + H'_{20,i}\hat{\beta}_{20} + \left(H'_{21,i}\hat{\beta}_{21}\right) K_\alpha(H'_{21,i}\hat{\beta}_{21}), \tag{2.7}$$

where $K_\alpha(x) = K(x/\alpha)$, $K(\cdot)$ is a cumulative distribution function (CDF) that admits a probability density function (PDF) denoted as $k(\cdot)$, and $\alpha = \alpha_n > 0$ is a smoothing parameter (bandwidth) that converges to zero as $n \to \infty$.

[7] The case where the underlying DGP satisfies $\Pr\left(H'_{21,i}\beta_{21}^* = 0\right) = 0$ is referred to as the regular setting, while the case where the underlying DGP is that $\Pr\left(H'_{21,i}\beta_{21}^* = 0\right) > 0$ is referred to as non-regular setting and the corresponding DGP is called an exceptional law. In this paper we use non-regular setting and exceptional law interchangeably.

(b) Obtain the smoothed estimator of β_1 as

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

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

3 Asymptotic Theory for the Smoothed Q-learning Algorithm

Let $\hat{\beta}_{sq} = (\hat{\beta}'_{1,sq}, \hat{\beta}'_2)'$. In this section, we study the asymptotic properties of $\hat{\beta}_{sq}$. To simplify the technical analysis, we first reformulate $\hat{\beta}_{sq}$ as a method-of-moment estimator or 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, $m_n(\beta) = (m_{1n}(\beta)', m_2(\beta)')'$,

$$\begin{aligned} m_{1n}(\beta) = m_{1n}(W;\beta) &= B_1\left[Y_1 + H'_{20}\beta_{20} + K_\alpha(H'_{21}\beta_{21})H'_{21}\beta_{21} - B'_1\beta_1\right], \\ m_2(\beta) = m_2\left(W;\beta\right) &= B_2(Y_2 - B'_2\beta_2), \end{aligned} \tag{3.2}$$

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 α.

Let $\beta^* = (\beta_1^{*\prime}, \beta_2^{*\prime})'$, where for $t = 1, 2$, $\beta_t^* = (\beta_{t0}^{*\prime}, \beta_{t1}^{*\prime})'$ denotes the true value of $\beta_t = (\beta'_{t0}, \beta'_{t1})'$. Let

$$M\left(\beta\right) \equiv \mathbb{E}\left[m\left(\beta\right)\right] = \mathbb{E}\left[m\left(W;\beta\right)\right],$$

where $m(W;\beta) = (m_1(W;\beta)', m_2(W;\beta)')'$ in which

$$m_1(W;\beta) = B_1\left[Y_1 + H'_{20}\beta_{20} + (H'_{21}\beta_{21})_+ - B'_1\beta_1\right]. \tag{3.3}$$

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

$$\begin{aligned} \beta_1^* &= \left(\mathbb{E}\left[B_1 B'_1\right]\right)^{-1} \mathbb{E}[B_1\left(Y_1 + H'_{20}\beta^*_{20} + (H'_{21}\beta^*_{21})_+\right)] \text{ and} \\ \beta_2^* &= \left(\mathbb{E}\left[B_2 B'_2\right]\right)^{-1} \mathbb{E}[B_2 Y_2]. \end{aligned}$$

In contrast to the standard GMM, $\mathbb{E}\left[m_n\left(\beta^*\right)\right]$ 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 $Z\left(\beta_{21}\right) = H_{21}'\beta_{21}$ and $Z = Z\left(\beta_{21}^*\right) = H_{21}'\beta_{21}^*$. Let $F_Z(\cdot;\beta_{21})$ denote the CDF of $Z\left(\beta_{21}\right)$ and $F_Z(\cdot) = F_Z(\cdot;\beta_{21}^*)$. Further let

$$\varepsilon_1 = Y_1 + H_{20}'\beta_{20}^* + (H_{21}'\beta_{21}^*)_+ - B_1'\beta_1^* \text{ and } \varepsilon_2 = Y_2 - B_2'\beta_2^*.$$

Note that ε_1 and ε_2 denote the population errors in the least squares projections of $[Y_1 + H_{20}'\beta_{20}^* + (H_{21}'\beta_{21}^*)_+]$ on (A_1, H_1) and that of Y_2 on (A_2, H_2) respectively. Let $W_i = (X_{1i}', A_{1i}, Y_{1i}, X_{2i}', A_{2i}, Y_{2i})'$ and $\mathcal{B}_1$, $\mathcal{B}_{20}$, $\mathcal{B}_{21}$ denote the parameter spaces of β_1, β_{20}, β_{21}, respectively. Let $\mathcal{B}$ denote the parameter space of $\beta = \left(\beta_1', \beta_2'\right)'$. Let k_1, k_2, k_{20} and k_{21} denote the dimensions of β_1, β_2, β_{20}, and β_{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) $\mathcal{B}$, $\mathcal{B}_1$, $\mathcal{B}_{20}$, and $\mathcal{B}_{21}$ are all compact.

(iii) For each $\beta_{21} \in \mathcal{B}_{21}$, the CDF $F_Z(\cdot;\beta_{21})$ can be written as $F_Z(\cdot;\beta_{21}) = \left(1 - p_{\beta_{21}}\right) F_C\left(\cdot;\beta_{21}\right) + p_{\beta_{21}} F_D\left(\cdot;\beta_{21}\right)$, where $F_C\left(\cdot;\beta_{21}\right)$ and $F_D\left(\cdot;\beta_{21}\right)$ denote the continuous and discrete components of $Z\left(\beta_{21}\right)$, respectively. $F_C\left(\cdot;\beta_{21}\right)$ admits a PDF $f_C\left(\cdot;\beta_{21}\right)$ that is continuously differentiable for all $\beta_{21} \in \mathcal{B}_{21}$. $F_D\left(\cdot;\beta_{21}\right)$ is a discrete CDF such that if $Z_D \backsim F_D\left(\cdot;\beta_{21}\right)$, then $P\left(Z_D = z_d\right) = f_D\left(z_d;\beta_{21}\right)$ and $\sum_{z_d} f_D\left(z_d;\beta_{21}\right) = 1$.

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

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

Assumption A2. (i) The eigenvalues of $\mathbb{E}\left[B_1 B_1'\right]$ and $\mathbb{E}\left[B_2 B_2'\right]$ 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) $\mathbb{E}[\,\|B_1 B_1'\| \left(Z^2 + \varepsilon_1^2\right)] < \infty$, $\mathbb{E}[\,\|B_1 B_2'\|\, \|Z\varepsilon_2\|] < \infty$, and $\mathbb{E}[\,\|B_2 B_2'\|\, \varepsilon_2^2] < \infty$.

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

Assumption A3. (i) The kernel function $K\left(\cdot\right)$ is a CDF that admits a PDF $k\left(\cdot\right)$ with compact support $\left[-1, 1\right]$. $k\left(\cdot\right)$ 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 $f_C\left(z;\beta_{21}^*\right), f_D\left(z_d;\beta_{21}^*\right)$, and $p_{\beta_{21}^*}$ simply as $f_C\left(z\right), f_D\left(z_d\right)$, 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 $K\left(\cdot\right)$ and its derivative function $k\left(\cdot\right)$. We assume that $k\left(\cdot\right)$ 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:

1. Epanechnikov kernel: $k\left(z\right)=\frac{3}{4}\left(1-z^2\right)1\left\{|z|\leq 1\right\}$;
2. Triangular kernel: $k\left(z\right)=\left(1-|z|\right)1\left\{|z|\leq 1\right\}$;
3. Quartic kernel: $k\left(z\right)=\frac{15}{16}\left(1-z^2\right)^2 1\left\{|z|\leq 1\right\}$;
4. Triweight kernel: $k\left(z\right)=\frac{35}{32}\left(1-z^2\right)^3 1\left\{|z|\leq 1\right\}$.

One can relax the compactness assumption at the cost of more lengthy proofs and some tail conditions on $f_C\left(\cdot;\beta_{21}\right), f_D\left(\cdot;\beta_{21}\right), r_1\left(\cdot;\beta_{21}\right)$, and $r_2(\cdot)$. A3(ii) imposes conditions on the smoothing parameter α. Note that $n\alpha^4\rightarrow 0$ ensures that the bias associated with our smoothed Q-learning estimator $\hat{\beta}_{sq}$ is asymptotically negligible.

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

Theorem 3.1 *Suppose that Assumptions A1(i)-(iv), A2(i)-(ii), and A3 hold. Then for the smoothed Q-learning estimator, we have*

$$\hat{\beta}_{sq}=\beta^*+o_P\left(1\right).$$

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 3.2 *Suppose that Assumptions A1-A3 hold. Then for the smoothed Q-learning estimator, we have*

$$\sqrt{n}(\hat{\beta}_{sq}-\beta^*)\overset{d}{\rightarrow}N(0,\ \Omega),$$

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

$$\begin{aligned}
\Lambda_{11} &= \mathbb{E}\left\{B_1B_1'[Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + \left(H_{21}'\beta_{21}^*\right)_+]^2\right\}, \\
\Lambda_{12} &= \mathbb{E}\left\{B_1B_2'[Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + \left(H_{21}'\beta_{21}^*\right)_+][Y_2 - B_2'\beta_2^*]\right\}, \\
\Lambda_{22} &= \mathbb{E}\left\{B_2B_2'[Y_2 - B_2'\beta_2^*]^2\right\}, \ \textit{and} \\
\Xi_{12} &= \lim_{n\to\infty}(\mathbb{E}[B_1H_{20}'], (1-p_n)\left[g_2(\infty) - g_2(0)\right] + p_n[\frac{1}{2}r_2(0)f_D(0) + \sum_{z_d>0} r_2(z_d)f_D(z_d)]). \qquad (3.4)
\end{aligned}$$

Remark. Noting that the eigenvalues of the block upper triangular matrix Ξ are given by those of Ξ_{11} and Ξ_{22} which are all negative under Assumption A2(i), Ξ must be nonsingular. 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 3.2 shows that our smoothed Q-learning estimator $\hat{\beta}_{sq}$ is always$\sqrt{n}$-consistent for β^* and asymptotically normally distributed even under the non-regular setting. This greatly simplifies the inference procedure. In contrast, the Q-learning estimator of β_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 Ω. We propose to estimate Λ and Ξ by $\hat{\Lambda}_n \equiv \hat{\Lambda}_n(\hat{\beta}_{sq})$ and $\hat{\Xi}_n \equiv \hat{\Xi}_n(\hat{\beta}_{sq})$, where

$$\begin{aligned}
\hat{\Lambda}_n(\beta) &= \begin{pmatrix} \hat{\Lambda}_{n,11}(\beta) & \hat{\Lambda}_{n,12}(\beta) \\ \hat{\Lambda}_{n,12}'(\beta) & \hat{\Lambda}_{n,22}(\beta) \end{pmatrix}, \ \hat{\Xi}_n(\beta) = \begin{pmatrix} \hat{\Xi}_{n,11}(\beta) & [\hat{\Xi}_{n,12,0}(\beta), \hat{\Xi}_{n,12,1}(\beta)] \\ 0_{k_1\times k_2} & \hat{\Xi}_{n,22}(\beta) \end{pmatrix}, \\
\hat{\Lambda}_{n,11}(\beta) &= \mathbb{E}_n\left\{B_1B_1'[Y_1 - B_1'\beta_1 + H_{20}'\beta_{20} + K_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21}]^2\right\}, \\
\hat{\Lambda}_{n,12}(\beta) &= \mathbb{E}_n\left\{B_1B_2'[Y_1 - B_1'\beta_1 + H_{20}'\beta_{20} + K_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21}](Y_2 - B_2'\beta_2)\right\}, \\
\hat{\Lambda}_{n,22}(\beta) &= \mathbb{E}_n\left\{B_2B_2'(Y_2 - B_2'\beta_2)^2\right\}. \\
\hat{\Xi}_{n,11}(\beta) &= -\mathbb{E}_n[B_1B_1'], \ \hat{\Xi}_{n,22}(\beta) = -\mathbb{E}_n[B_2B_2'], \ \hat{\Xi}_{n,12,0}(\beta) = \mathbb{E}_n[B_1H_{20}'], \ \text{and} \\
\hat{\Xi}_{n,12,1}(\beta) &= \mathbb{E}_n\left\{B_1H_{21}'\left[\frac{1}{\alpha}k_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21} + K_\alpha(H_{21}'\beta_{21})\right]\right\}.
\end{aligned}$$

in which $k_\alpha(z) = k(z/\alpha)$. Then we can estimate Ω 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 3.3 *Suppose that Assumptions A1-A3 hold. Suppose that* $\mathbb{E}[\,\|B_1B_1'\|\,(\|Y_1\| + \|B_1\| + \|H_{20}\|)] < \infty$, *and* $\mathbb{E}[\,\|B_1B_1'\|^2\,\|H_{21}\|^2] < \infty$. *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

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

We estimate σ_c^2 by $\hat{\sigma}_c^2 \equiv c'\hat{\Xi}_n^{-1}\hat{\Lambda}_n(\hat{\Xi}_n^{-1})'c$. Theorem 3.3 implies consistency of $\hat{\sigma}_c^2$ and the result that $T_n \xrightarrow{d} N(0, 1)$, where

$$T_n = \sqrt{n}c'(\hat{\beta}_{sq} - \beta^*)/\hat{\sigma}_c. \tag{3.5}$$

It is interesting to observe that regardless of $\Pr\left(H_{21,i}'\beta_{21}^* = 0\right) = 0$ or not, the limiting distribution of the t-ratio T_n is the standard normal and inference on β^*, including the first stage parameter β_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 α. 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} :

$$T_{0n} = \sqrt{n}c'(\hat{\beta}_{sq} - \beta^*)/\sigma_c \text{ and } S_n = \sqrt{n}(\hat{\sigma}_c - \sigma_c)/\sigma_c. \tag{4.1}$$

When $c = e_{k,j}$, a $k \times 1$ vector with 1 in its jth 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 jth coefficient in β.

By a Taylor expansion, we have $T_{0n} = T_{0n}^* + R_{0n}^*$, where

$$\begin{aligned}
T_{0n}^* &= -\sigma_c^{-1}c'\Xi^{-1}\xi_N + (n\alpha)^{-1/2}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi_n^{-1}\xi_N - (n\alpha)^{-1}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N \\
&\quad +(n\alpha)^{-3/2}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N \text{ and} \\
R_{0n}^* &= -(n\alpha)^{-2}\sigma_c^{-1}c'[\hat{\Xi}_n(\bar{\beta})]^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N,
\end{aligned}$$

in which $\xi_N = \sqrt{n}\mathbb{E}_n[m_n(\beta^*)]$ and $\xi_D = \sqrt{n\alpha}[\hat{\Xi}_n(\bar{\beta}) - \Xi]$ with $\overline{\beta}$ lies between β^* and $\hat{\beta}_{sq}$. See Appendix A for details. We will show that the first term in T_{0n}^* contributes to the dominant bias of T_{0n} and the first two terms in T_{0n}^* 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

$$\begin{aligned} S_n &= \frac{-2c'\Xi^{-1}\xi_D^{\dagger}\Omega c - \sqrt{\alpha}c'\Xi^{-1}\xi_{\Lambda}\Xi^{-1\prime}c}{2\sigma_c^2} \\ &\quad + \frac{2c'\Xi^{-1}\xi_D^{\dagger}\Xi^{-1}\xi_D^{\dagger}\Omega c + c'\Xi^{-1}\xi_D^{\dagger}\Omega\xi_D^{\dagger}\Xi^{-1\prime}c + \sigma_c^{-2}(c'\Xi^{-1}\xi_D^{\dagger}\Omega c)^2}{2\sqrt{n\alpha}\sigma_c^2} + o_P((n\alpha)^{-1/2}) \\ &\equiv S_n^* + o_P((n\alpha)^{-1/2}), \end{aligned} \tag{4.2}$$

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

$$\begin{aligned} T_n^* &= -\sigma_c^{-1}c'\Xi^{-1}\xi_N + (n\alpha)^{-1/2}\,\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi_n^{-1}\xi_N + (n\alpha)^{-1/2}\sigma_c^{-1}c'\Xi^{-1}\xi_N S_n^* \\ &\quad - (n\alpha)^{-1}\,\sigma_c^{-1}c'\Xi^{-1}\left\{\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N - \frac{\xi_D\Xi^{-1}\xi_N c'\Xi^{-1}\xi_D^{\dagger}\Omega c}{\sigma_c^2} + \frac{2\xi_N(c'\Xi^{-1}\xi_D^{\dagger}\Omega c)^2}{\sigma_c^4}\right\}. \end{aligned} \tag{4.3}$$

We will show that the optimal bandwidth in terms of minimizing the asymptotic mean squared error (MSE) of either T_{0n}^* or T_n^* is proportional to $n^{-2/5}$.

Let $\Omega = \Xi^{-1}\Lambda\Xi^{-1\prime} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix}$ with Ω_{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) $\mathbb{E}\left\|B_t\varepsilon_t\right\|^8 < \infty$ and $\mathbb{E}\left\|B_tB_t'\right\|^8 < \infty$ for $t = 1, 2$, and $\mathbb{E}\left\|B_1H_{2t}\right\|^8 < \infty$ 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_tB_t', and B_1H_{2t}. A4(ii) imposes additional conditions on $k(\cdot)$.

Theorem 4.1 *Suppose that $\alpha \propto n^{-2/5}$. Suppose that Assumptions A1-A2, A3(i), and A4 hold. Let the additional moment conditions in Theorem 3.3 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) $\mathbb{E}[T_{0n}^*] = -\sqrt{n\alpha^4}\sigma_c^{-1}c'\Xi^{-1}\mathcal{B} + o\left(n^{-3/10}\right)$ *and* $Var[T_{0n}^*] = 1 + (n\alpha)^{-1}\sigma_c^{-2}c'\Xi^{-1}\Phi_0\Xi^{-1\prime}c + o\left(n^{-3/5}\right)$;

(iii) $\mathbb{E}[T_n^*] = -\sqrt{n\alpha^4}\sigma_c^{-1}c'\Xi^{-1}\mathcal{B} + o\left(n^{-3/10}\right)$ *and* $Var[T_n^*] = 1 + (n\alpha)^{-1}\sigma_c^{-2}c'\Xi^{-1}\Phi\Xi^{-1\prime}c + o\left(n^{-3/5}\right)$,

where

$$
\begin{aligned}
\mathcal{B} &= \begin{pmatrix} \alpha^{-2}\mathbb{E}\left\{B_1\left[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - (H_{21}'\beta_{21}^*)_+\right]\right\} \\ 0_{k_2\times 1} \end{pmatrix} = O\left(1\right), \\
\Phi_0 &= \begin{pmatrix} \alpha\mathbb{E}\left[\bar{\eta}_1\left(\beta_{21}^*\right)\Omega_{22}\bar{\eta}_1\left(\beta_{21}^*\right)'\right] & 0_{k_1\times k_2} \\ 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{pmatrix} = O\left(1\right), \;\; \textit{and} \\
\Phi &= \Phi_0 + \alpha\sigma_c^{-4}\mathbb{E}[\eta_1^*\left(\beta_{21}^*\right)]^2\Lambda + \alpha\sigma_c^{-2}\mathbb{E}[\eta_1^*\left(\beta_{21}^*\right)\eta_1^{**}\left(\beta_{21}^*\right)\Lambda + \Lambda\eta_1^{**}\left(\beta_{21}^*\right)'\eta_1^*\left(\beta_{21}^*\right)] = O\left(1\right),
\end{aligned}
$$

in which

$$
\begin{aligned}
\bar{\eta}_1\left(\beta_{21}\right) &\equiv \bar{\eta}_1\left(H_{21};\beta_{21}\right) = \eta_1\left(\beta_{21}\right) - \mathbb{E}\left[\eta_1\left(\beta_{21}\right)\right], \\
\eta_1\left(\beta_{21}\right) &\equiv \eta_1(H_{21};\beta_{21}) = [0_{k_1\times k_{20}}, B_1H_{21}' \times \frac{1}{\alpha}k_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21}], \\
\eta_1^*\left(\beta_{21}\right) &= c'\Xi^{-1}\begin{pmatrix} 0_{k_1\times k_1} & \bar{\eta}_1(\beta_{21}) \\ 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{pmatrix}\Omega c = c_1'\Xi_{11}^{-1}\bar{\eta}_1\left(\beta_{21}\right)\left[\Omega_{12}'c_1 + \Omega_{22}c_2\right], \; \textit{and} \\
\eta_1^{**}\left(\beta_{21}\right) &= \begin{pmatrix} 0_{k_1\times k_1} & \bar{\eta}_1(\beta_{21}) \\ 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{pmatrix}\Xi^{-1} = \begin{pmatrix} 0_{k_1\times k_1} & \bar{\eta}_1\left(\beta_{21}\right)\Xi_{22}^{-1} \\ 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{pmatrix}.
\end{aligned}
$$

Remark. Theorem 4.1 implies that T_n and T_{0n} share the same dominant asymptotic bias term that is $O(\sqrt{n\alpha^4}) = O\left(n^{-3/10}\right)$ while they have different asymptotic variances up to the order $(n\alpha)^{-1} = O\left(n^{-3/5}\right)$. From Lemma A.1 in Appendix A, we know that $\mathrm{Var}\left(\xi_D\Xi^{-1}\xi_N\right) = \Phi_0 + o\left(1\right)$ and $\mathrm{Var}\left(\xi_D\Xi^{-1}\xi_N + \sigma_c^{-2}c'\Xi^{-1}\xi_D^\dagger\Omega c\xi_N\right) = \Phi + o\left(1\right)$. The second term in the last variance operator reflects the cost for estimating the asymptotic variance σ_c^2.

To sum up, we have shown that the mean squared errors of T_{0n} and T_n are approximately given by

$$
\begin{aligned}
\mathrm{MSE}_0\left(\alpha\right) &= 1 + n\alpha^4\sigma_c^{-2}\left(c'\Xi^{-1}\mathcal{B}\right)^2 + \left(n\alpha\right)^{-1}\sigma_c^{-2}c'\Xi^{-1}\Phi_0\Xi^{-1\prime}c \text{ and} && (4.4) \\
\mathrm{MSE}\left(\alpha\right) &= 1 + n\alpha^4\sigma_c^{-2}\left(c'\Xi^{-1}\mathcal{B}\right)^2 + \left(n\alpha\right)^{-1}\sigma_c^{-2}c'\Xi^{-1}\Phi\Xi^{-1\prime}c. && (4.5)
\end{aligned}
$$

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 (3.5) and (4.1), it holds that $c'\Xi^{-1}\mathcal{B} \neq 0$.

Assumption A5 ensures that the approximate bias for T_{0n} and for T_n are non-zero. We can then choose α to minimize either $\mathrm{MSE}_0\left(\alpha\right)$ or $\mathrm{MSE}(\alpha)$ to obtain the respective solution

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

Since α_0^* and α^* depend on some unknown quantities, namely, Ξ, Φ_0, Φ, and $\mathcal{B}$, they are not feasible. To obtain a feasible version of α_0^* or α^*, we need to estimate these quantities.

Let $\hat{\beta}_{21}$ be a consistent estimator of β_{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 smoothed Q-learning estimator $\hat{\beta}_{sq} = (\hat{\beta}_{1,sq}', \hat{\beta}_2')'$. Let $\hat{\Xi}_n$ and $\hat{\Lambda}_n$ be defined as in the previous subsection. We propose to estimate Φ_0, Φ, and $\mathcal{B}$ consistently by

$$\begin{aligned}
\hat{\Phi}_{0n} &= \begin{pmatrix} \alpha\mathbb{E}_n[\hat{\eta}_1(\hat{\beta}_{21})\hat{\Omega}_{n,22}\hat{\eta}_1(\hat{\beta}_{21})'] & 0_{k_1\times k_2} \\ 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{pmatrix}, \\
\hat{\Phi}_n &= \hat{\Phi}_{0n} + \alpha\sigma_c^{-4}\mathbb{E}_n[\hat{\eta}_1^*(\hat{\beta}_{21})]^2\hat{\Lambda}_n + \alpha\sigma_c^{-2}\mathbb{E}_n[\hat{\eta}_1^*(\hat{\beta}_{21})\hat{\eta}_1^{**}(\hat{\beta}_{21})\hat{\Lambda}_n + \hat{\Lambda}_n\hat{\eta}_1^{**}(\hat{\beta}_{21})'\hat{\eta}_1^*(\hat{\beta}_{21})], \text{ and} \\
\hat{\mathcal{B}}_n &= \begin{pmatrix} \alpha^{-2}\mathbb{E}_n[B_1\{K_\alpha(H_{21}'\hat{\beta}_{21})H_{21}'\hat{\beta}_{21} - (H_{21}'\hat{\beta}_{21})_+\}] \\ 0_{k_2\times 1} \end{pmatrix},
\end{aligned}$$

respectively, where

$$\begin{aligned}
\hat{\eta}_1(\hat{\beta}_{21}) &= \eta_1(\hat{\beta}_{21}) - \mathbb{E}_n[\eta_1(\hat{\beta}_{21})], \\
\hat{\eta}_1^*(\hat{\beta}_{21}) &\equiv \hat{\eta}_1^*(H_{21};\hat{\beta}_{21}) = c_1'\hat{\Xi}_{n,11}^{-1}\hat{\eta}_1(\hat{\beta}_{21})[\hat{\Omega}_{n,12}'c_1 + \hat{\Omega}_{n,22}c_2], \text{ and} \\
\hat{\eta}_1^{**}(\hat{\beta}_{21}) &\equiv \hat{\eta}_1^{**}(H_{21};\hat{\beta}_{21}) = \begin{bmatrix} 0_{k_1\times k_1} & \hat{\eta}_1(\hat{\beta}_{21})\hat{\Xi}_{n,22}^{-1} \\ 0_{k_2\times k_1} & 0_{k_2\times k_2} \end{bmatrix}.
\end{aligned}$$

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 $\hat{\mathcal{B}}_n$. Then a feasible optimal bandwidth is given by

$$\hat{\alpha}_0^* = \left[\frac{c'\hat{\Xi}_n^{-1}\hat{\Phi}_{0n}\hat{\Xi}_n^{-1\prime}c}{4\left(c'\hat{\Xi}_n^{-1}\hat{\mathcal{B}}_n\right)^2}\right]^{1/5} n^{-2/5} \text{ or } \hat{\alpha}^* = \left[\frac{c'\hat{\Xi}_n^{-1}\hat{\Phi}_n\hat{\Xi}_n^{-1\prime}c}{4\left(c'\hat{\Xi}_n^{-1}\hat{\mathcal{B}}_n\right)^2}\right]^{1/5} n^{-2/5}. \tag{4.6}$$

We will examine the finite sample performance of the smoothed Q-learning estimators based on the RoT bandwidth, $\hat{\alpha}_0^*$, and $\hat{\alpha}^*$ in the next section. Note that $\hat{\alpha}_0^*$ is also optimal in terms of minimizing the MSE of $c'\hat{\beta}_{sq}$.

Remark. 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*; 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 smoothed Q-learning 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

$$Y_2 = \gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 + \gamma_5 A_2 + \gamma_6 X_2 A_2 + \gamma_7 A_1 A_2 + \epsilon_2,$$

where $\epsilon_2 \sim N(0,1)$ 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_t's are generated as $P(X_1 = -1) = P(X_1 = 1) = 1/2$ and $P(X_2 = 1|X_1, A_1) = 1 - P(X_2 = -1|X_1, A_1) = \text{expit}(\delta_1 X_1 + \delta_2 A_1)$, where $\text{expit}(x) = e^x/(1 + e^x)$.

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

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

and

$$\begin{aligned} Q_1(H_1, A_1) &= \mathbb{E}[Y_1 + \max_{a_2 \in \{0,1\}} Q_2(H_2, a_2) | H_1, A_1] \\ &= \gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 + \mathbb{E}\{[\gamma_5 + \gamma_6 X_2 + \gamma_7 A_1]_+ | X_1, A_1\}. \end{aligned}$$

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

$$\begin{aligned} &\psi_1 = \gamma_1 + \frac{a_1 + a_3}{2}, \ \psi_2 = \gamma_2 + \frac{a_3 - a_1}{2}\Big(2\text{expit}(\delta_1) - 1\Big), \\ &\psi_3 = \gamma_3 - \frac{a_1 - a_2 + a_3 - a_4}{2} + \frac{a_4 - a_2}{2}\Big(\text{expit}(-\delta_1 + \delta_2) + \text{expit}(\delta_1 + \delta_2) - 1\Big), \text{ and} \\ &\psi_4 = \gamma_4 + \frac{a_3 - a_1}{2}\Big(1 - 2\text{expit}(\delta_1)\Big) + \frac{a_4 - a_2}{2}\Big(- \text{expit}(-\delta_1 + \delta_2) + \text{expit}(\delta_1 + \delta_2)\Big), \end{aligned}$$

in which $a_1 = [\gamma_5 - \gamma_6]_+$, $a_2 = [\gamma_5 - \gamma_6 + \gamma_7]_+$, $a_3 = [\gamma_5 + \gamma_6]_+$, and $a_4 = [\gamma_5 + \gamma_6 + \gamma_7]_+$. 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 ψ_3 and ψ_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 $\sigma_Z = \sqrt{\mathrm{Var}(Z)}$. 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 ($f_D(0) = 0$, ϕ 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 ($f_D(0) = 1$, ϕ 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 smoothed Q-learning estimator with three different bandwidth choices and with the Gaussian kernel, 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

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

and

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

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 Moodie and Richardson (2010) to consider two choices of the tuning parameter a, 0.08 and 0.20, and denote the corresponding hard-thresholding methods as $\text{HT}_{0.08}$ and $\text{HT}_{0.20}$.

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 smoothed Q-learning estimators; the fourth one is the FACI in Laber et al. (2013); 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. Tables 1 and 2 present results for ψ_3, while Tables 3 and 4 present results for ψ_4. Like Chakraborty et al. (2010), we find that all estimators and CIs perform reasonably well for ψ_4 and for all designs. So we focus our discussion below on Tables 1 and 2 for ψ_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.

< − − − − − − − Table 1 and Table 2 Here − − − − − − − >

Design 1 ($p_0 = 0$, $\phi = 0.959$) is a completely regular setting, so we expect the Q-learning and our smoothed Q-learning algorithms to perform well. (i) In terms of the bias and MSE, the Q-learning and our smoothed Q-learning 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 smoothed Q-learning CIs are tighter. Considering all the performance measures and the cost of computation, the three smoothed Q-learning estimators emerged as the best.[8]

[8] For example, on a computer with Intel 2.00 GHz processor, when $n = 1,000$, our smoothed Q-learning method

Design 2 ($p_0 = 0$, ϕ 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 smoothed Q-learning 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 smoothed Q-learning CIs. (iii) In terms of the IL, the Q-learning CI is the tightest, followed by the soft-thresholding and our smoothed Q-learning CIs. Overall, the soft-thresholding, the Q-learning and our smoothed Q-learning methods are comparable, but the smoothed Q-learning 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 smoothed Q-learning 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 $\mathrm{HT}_{0.08}$ CIs have better coverage rates for the smaller sample. The smoothed Q-learning CI with bandwidth $\hat{\alpha}^*$ and the $\mathrm{HT}_{0.20}$ 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 smoothed Q-learning CIs are in general tighter. In sum, our smoothed Q-learning algorithm with bandwidth $\hat{\alpha}^*$ and the $\mathrm{HT}_{0.20}$ 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 smoothed Q-learning method using bandwidth $\hat{\alpha}^*$ and the $\mathrm{HT}_{0.08}$ 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 smoothed Q-learning 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 smoothed Q-learning method with bandwidth $\hat{\alpha}^*$ and the $\mathrm{HT}_{0.20}$ method.

takes less than 4 minutes for $10,000$ repetitions with 10 parallel R workers. In the same setup, the $\mathrm{HT}_{0.20}$ method takes about 11.7 hours. The computation time for all other methods using bootstrap are similar to that of the $\mathrm{HT}_{0.20}$ method.

Design 6 ($p_0 = 1$, ϕ 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 smoothed Q-learning 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 smoothed Q-learning 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 smoothed Q-learning 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 term of the MSE, the hard-thresholding estimators and our smoothed Q-learning estimators are comparable for all the designs. In terms of the CP, the hard-thresholding CIs over-cover more than our smoothed Q-learning CIs in Designs 2 and 6. Our smoothed Q-learning 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.

6 An Empirical Application: 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. At 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, 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, 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 ever changes the type of medication, then his/her Stages 1 and 2 correspond to Phases I and II in the study. 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, and there are 125 male and 171 female out of 296 patients. Second, we choose the Clinical Global Impression of Change (CGIC, on a 1-7 scale) as the dependent variable. We negate the original CGIC score so that a higher score represents a better outcome. Third, the treatment variable is defined as $A_t = 1,\ t = 1, 2$ if the patient takes olanzapine and 0 otherwise.[9] In Stage 1, there are 70 patients taking olanzapine and 226 otherwise. In Stage 2, there are 60 patients taking olanzapine and 236 otherwise.

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

$$\begin{aligned} Q_t(H_t, A_t) = &\ \beta_{t0,0} + \beta_{t0,1}\mathrm{NPI}_t + \beta_{t0,2}\mathrm{Age} + \beta_{t0,3}\mathrm{Gender} \\ &+ (\beta_{t1,0} + \beta_{t1,1}\mathrm{NPI}_t + \beta_{t1,2}\mathrm{Age} + \beta_{t1,3}\mathrm{Gender})A_t,\ t = 1, 2. \end{aligned}$$

The estimation results are summarized in Tables 5 and 6. Stage 2 is a standard linear regression, and its estimated parameters are qualitatively similar to those of Stage 1 for the same covariates. We thus focus on Stage 1 for discussion.

From Table 6, it is clear that the point estimates from different methods are very close. In terms of lengths of confidence intervals, our smoothed Q CIs dominate all others in 4 out of 8 parameters of interest, and the soft-threshold CI dominates all others in the rest 4 out of 8 parameters of interest. As in the simulation results, the length of the FACI confidence interval is relatively conservative, with the longest interval lengths for all parameters of interest.

< − − − − − − − Tables 5 and 6 Here − − − − − − − >

Next, given the similarity of estimated parameters across different methods, we focus on those of the smoothed Q-learning method with bandwidth $\hat{\alpha}^*$ for interpretation. First, NPI has a statistically significant negative main effect, which indicates that given any treatment status, a higher score of NPI leads to a lower CGIC score. From the estimate of the interaction term

[9] This definition of treatment variable is due to two reasons. First, in the design of the study, patients who take drugs in the Stage 1 does not have the choice of switching to placebo in Stage 2. This does not allow us to define treatment variable as 1 if the patient takes one of the drugs and 0 for placebo, as this will violate 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.

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 main effect of age, but the estimate of the interaction term between age and treatment suggests that olanzapine tends to be more effective for a younger patient. Third, male tends to have a lower CGIC score given that he suffers from the Alzheimer's Disease. From the estimate of the interaction term between gender and treatment, we find that olanzapine tends to be more effective for female patients than male patients. Fourth, the treatment tends to have a positive main effect, although it is not statistically significant. This is consistent with the previous finding in Tran et al. (1997) that olanzapine is more effective than risperidone in treating Schizophrenia.

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

$$\begin{aligned}\pi_2^{dp}(h_2) &= 1\left\{3.904 + 0.004\text{NPI}_2 - 0.051\text{Age} - 0.874\text{Gender} \geq 0\right\},\\ \pi_1^{dp}(h_1) &= 1\left\{2.539 + 0.026\text{NPI}_1 - 0.046\text{Age} - 0.041\text{Gender} \geq 0\right\}.\end{aligned}$$

As an illustration, we consider the treatment prescriptions for the average patient[10] in each gender group, which is summarized in Table 7. For the average female patient in our sample, it is suggested not to use olanzapine in Stage 1 and to use olanzapine in Stage 2. For the average male patient in our sample, it is suggested not to use olanzapine in both stages. In general, the above estimated optimal dynamic treatment regime allows the clinician to prescribe a treatment based on NPI score, age, and gender in each stage.

< − − − − − − − − − Table 7 Here − − − − − − − − − >

7 Concluding Remarks

In this paper, we have proposed a smoothed Q-learning algorithm for estimating optimal dynamic treatment regimes. Under assumptions in Section 3, we have established asymptotic properties of the smoothed Q-learning 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 smoothed Q-learning 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

[10] An average patient refers to a hypothetical one with each covariate to be at the sample mean.

effective for female patients than for male patients. We do not find any statistically significant age effect.

Like other methods, the smoothed Q-learning 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.

APPENDIX

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 $g : \mathcal{W} \to \mathbb{R}^K$, we define

$$\begin{aligned} \mathbb{E}_n [g] &= \mathbb{E}_n [g (W)] = \frac{1}{n} \sum_{i=1}^{n} g (W_i) , \text{ and} \\ \mathbb{G}_n [g] &= \mathbb{G}_n [g (W)] = n^{-1/2} \sum_{i=1}^{n} \{g (W_i) - \mathbb{E} [g (W_i)]\}. \end{aligned}$$

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

Proof of Theorem 3.1. Let $M_n (\beta) = \mathbb{E}_n[m_n(\beta)] = \frac{1}{n} \sum_{i=1}^{n} m_n (W_i; \beta)$, where $m_n(W; \beta) = (m_{1n}(W; \beta)', m_2(W; \beta)')'$, $m_{1n}(W; \beta) = B_1 [Y_1 + H'_{20}\beta_{20} + K_\alpha(H'_{21}\beta_{21})H'_{21}\beta_{21} - B'_1\beta_1]$, and $m_2(W; \beta) = B_2(Y_2 - B'_2\beta_2)$. 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 $m(W; \beta) = (m_1(W; \beta)', m_2(W; \beta)')'$ and $m_1(W; \beta) = B_1 [Y_1 + H'_{20}\beta_{20} + (H'_{21}\beta_{21})_+ - B'_1\beta_1]$.

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

$$\beta_1^* = \left(\mathbb{E} \left[B_1 B'_1\right]\right)^{-1} \mathbb{E}[B_1 \left(Y_1 + H'_{20}\beta^*_{20} + (H'_{21}\beta^*_{21})_+\right)], \text{ and } \beta_2^* = \left(\mathbb{E} \left[B_2 B'_2\right]\right)^{-1} \mathbb{E}[B_2 Y_2]. \quad \text{(A.1)}$$

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

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

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

$$\begin{aligned} \mathcal{M}_{1,1} &= \left\{b_1 \left[y_1 + h'_{20}\beta_{20} - b'_1\beta_1\right] : \beta_1 \in \mathcal{B}_1, \beta_{20} \in \mathcal{B}_{20}\right\} , \\ \mathcal{M}_{1,2} &= \left\{h'_{21}\beta_{21} : \beta_{21} \in \mathcal{B}_{21}\right\} , \\ \mathcal{M}_{1,3} &= \left\{K(h'_{21}\beta_{21}/\alpha) : \beta_{21} \in \mathcal{B}_{21}\right\} , \\ \mathcal{M}_{1,4} &= \left\{b_2 \left[y_2 - b'_2\beta_2\right] : \beta_2 \in \mathcal{B}_2\right\} . \end{aligned}$$

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 $C(\|b_1 y_1\| + \|b_1 h_{20}\| + \|b_1 b_1'\|)$, $C\|h_{21}\|$, 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[11] 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

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

Let $g_1(z;\beta_{21}) = \int_{-\infty}^{z} r_1(s;\beta_{21})\, s f_C(s;\beta_{21})\, ds$ and $g_1^{(1)}(z;\beta_{21}) = dg_1(z;\beta_{21})/dz = r_1(z;\beta_{21})\, z f_C(z;\beta_{21})$. By Assumption A1(iii),

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

[11] $\mathcal{M}_{1,3}$ here is n-dependent through the bandwidth parameter α.

For the first term, we have

$$\begin{aligned}
I_C(\beta_{21}) &= \int_{-\infty}^{\infty} r_1(z;\beta_{21})\, z f_C(z;\beta_{21}) \left[K(\frac{z}{\alpha}) - 1\{z \geq 0\}\right] dz \\
&= \int_{-\infty}^{\infty} \left[K(\frac{z}{\alpha}) - 1\{z \geq 0\}\right] dg_1(z;\beta_{21}) \\
&= -\int_{-\infty}^{\infty} g_1(z;\beta_{21})\, d\left[K(\frac{z}{\alpha}) - 1\{z \geq 0\}\right] \\
&= -\frac{1}{\alpha}\int_{-\infty}^{\infty} g_1(z;\beta_{21})\, k(\frac{z}{\alpha})dz + g_1(0;\beta_{21}) \\
&= -\int_{-1}^{1} \left[g_1(u\alpha;\beta_{21}) - g_1(0;\beta_{21})\right] k(u)du \\
&= -\alpha\int_{0}^{1}\int_{-1}^{1} g_1^{(1)}(tu\alpha;\beta_{21})uk(u)dudt,
\end{aligned}$$

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

Next, we note that

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

For any fixed $z_d > 0$, $K(\frac{z_d}{\alpha}) = 1$ for sufficiently small α 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 $z_d < 0$, $K(\frac{z_d}{\alpha}) = 0$ for sufficiently small α when n is large enough, ensuring the second term to vanish in large samples. As a result, $I_D(\beta_{21}) = 0$ uniformly in β_{21} when n is sufficiently large. It follows that for sufficiently large n,

$$\begin{aligned}
& \sup_{\beta_{21}\in\mathcal{B}_{21}} \left\| \mathbb{E}\left\{B_1\left[H_{21}'\beta_{21}K_h(H_{21}'\beta_{21}) - \left(H_{21}'\beta_{21}\right)_+\right]\right\}\right\| \\
= \; & \sup_{\beta_{21}\in\mathcal{B}_{21}} (1-p_{\beta_{21}}) \left\|I_C(\beta_{21})\right\| \leq \alpha \sup_{\beta_{21}\in\mathcal{B}_{21}} \left\|\int_0^1\int_{-1}^1 g_1^{(1)}(tu\alpha;\beta_{21})uk(u)dudt\right\| \\
= \; & \alpha^2 \sup_{\beta_{21}\in\mathcal{B}_{21}} \left\|\int_0^1 t\int_{-1}^1 r_1(tu\alpha;\beta_{21})\, f_C(tu\alpha;\beta_{21})\, u^2k(u)dudt\right\| \\
\to \; & \alpha^2 \sup_{\beta_{21}\in\mathcal{B}_{21}} \left\|r_1(0;\beta_{21})\, f_C(0;\beta_{21})\right\| \int_{-1}^1 u^2k(u)du = O\left(\alpha^2\right),
\end{aligned}$$

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

Proof of Theorem 3.2. Expanding $\mathbb{E}_n[m_n(\hat{\beta}_{sq})]$ around β^* yields

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

where $\bar{\beta}$ lies between $\hat{\beta}_{sq}$ and β^* element by element. We prove the theorem by showing that (i) $\sqrt{n}\mathbb{E}_n[m_n(\beta^*)] \overset{d}{\to} N(0, \Lambda)$, (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

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

as Ξ is nonsingular under Assumption A2(i).

Step 1. We first prove (i) $\sqrt{n}\mathbb{E}_n[m_n(\beta^*)] \overset{d}{\to} N(0, \Lambda)$. 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

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

It follows that

$$\begin{aligned} \sqrt{n}\mathbb{E}[m_{1n}(\beta^*)] &= \sqrt{n}\mathbb{E}\left\{B_1\left[Y_1 + H_{20}'\beta_{20}^* + K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - B_1'\beta_1^*\right]\right\} \\ &= \sqrt{n}\mathbb{E}\left\{B_1\left[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - (H_{21}'\beta_{21}^*)_+\right]\right\}. \end{aligned}$$

Following the analysis of $\mathbb{E}\left\{B_1\left[K_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21} - (H_{21}'\beta_{21})_+\right]\right\}$ in the proof of Theorem 3.1, for sufficiently large n,

$$\begin{aligned} \sqrt{n}\mathbb{E}[m_{1n}(\beta^*)] &= \sqrt{n}\mathbb{E}\left\{B_1\left[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - (H_{21}'\beta_{21}^*)_+\right]\right\} \\ &= \sqrt{n}(1-p)I_C(\beta_{21}^*) = \sqrt{n}(1-p)O\left(\alpha^2\right) = o(1) \text{ by Assumption A3(ii).} \end{aligned} \tag{A.4}$$

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

$$\Lambda_n \equiv \mathbb{E}\left[m_n(\beta^*)m_n(\beta^*)'\right] = \begin{pmatrix} \mathbb{E}\left[m_{1n}(\beta^*)m_{1n}(\beta^*)'\right] & \mathbb{E}\left[m_{1n}(\beta^*)m_2(\beta^*)'\right] \\ \mathbb{E}\left[m_2(\beta^*)m_{1n}(\beta^*)'\right] & \mathbb{E}\left[m_2(\beta^*)m_2(\beta^*)'\right] \end{pmatrix} \equiv \begin{pmatrix} \Lambda_{n,11} & \Lambda_{n,12} \\ \Lambda_{n,12}' & \Lambda_{22} \end{pmatrix},$$

where

$$\begin{aligned} \Lambda_{n,11} &= \mathbb{E}\left\{B_1B_1'[Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^*]^2\right\}, \\ \Lambda_{n,12} &= \mathbb{E}\left\{B_1B_2'[Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^*][Y_2 - B_2'\beta_2^*]\right\}, \\ \Lambda_{22} &= \mathbb{E}\left\{B_2B_2'[Y_2 - B_2'\beta_2^*]^2\right\}. \end{aligned}$$

Let Λ_{11} and Λ_{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 3.2. We can readily show that $\Lambda_{n,11} = \Lambda_{11} + o\,(1)$ and $\Lambda_{n,12} = \Lambda_{12} + o\,(1)$. For example, using

$$\begin{aligned} & Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* \\ = & \left[Y_1 - B_1'\beta_1^* + H_{20}'\beta_{20}^* + \left(H_{21}'\beta_{21}^*\right)_+\right] + \left[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - \left(H_{21}'\beta_{21}^*\right)_+\right] \\ = & \ \varepsilon_1 + \left[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - \left(H_{21}'\beta_{21}^*\right)_+\right], \end{aligned}$$

we have

$$\begin{aligned} \Lambda_{n,11} - \Lambda_{11} &= \mathbb{E}\left\{B_1B_1'[\varepsilon_1 + K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - \left(H_{21}'\beta_{21}^*\right)_+]^2\right\} - \mathbb{E}\left(B_1B_1'\varepsilon_1^2\right) \\ &= \mathbb{E}\left\{B_1B_1'[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - \left(H_{21}'\beta_{21}^*\right)_+]^2\right\} \\ &\quad +2\cdot\mathbb{E}\left\{B_1B_1'[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - \left(H_{21}'\beta_{21}^*\right)_+]\varepsilon_1\right\}. \end{aligned}$$

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

$$\left\|B_1B_1'[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - \left(H_{21}'\beta_{21}^*\right)_+]^2\right\| \le \left\|B_1B_1'\right\|\left(H_{21}'\beta_{21}^*\right)^2,$$

and the right hand side object in the last expression has finite first moment by Assumption A2(iii). Then $\mathbb{E}\left\{B_1B_1'[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - (H_{21}'\beta_{21}^*)_+]^2\right\} = o\,(1)$ by the DCT. Similarly, $\mathbb{E}\{B_1B_1'$ $[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - (H_{21}'\beta_{21}^*)_+]\varepsilon_1\} = o\,(1)$ 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,

$$\begin{aligned} \mathrm{Var}\left(m_n(\beta^*)\right) &= \mathbb{E}\left[m_n(\beta^*)m_n(\beta^*)'\right] - \mathbb{E}\left[m_n(\beta^*)\right]\mathbb{E}\left[m_n(\beta^*)'\right] \\ &= \Lambda_n - O\left(\alpha^2\right)O\left(\alpha^2\right) = \Lambda + o\,(1). \end{aligned}$$

Now, we show that $\mathbb{E}\left\|m_n(\beta^*)\right\|^{2+\delta} < \infty$. 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

$$\begin{aligned} \mathbb{E}\left|\omega' m_n(\beta^*)\right|^{2+\delta} &\le 2^{1+\delta}\mathbb{E}\left|\omega_1' m_{1n}(\beta^*)\right|^{2+\delta} + 2^{1+\delta}\mathbb{E}\left|\omega_2' m_2(\beta^*)\right|^{2+\delta} \\ &\le 2^{2+2\delta}\mathbb{E}\left|\omega_1'\left[m_{1n}(\beta^*) - m_1(\beta^*)\right]\right|^{2+\delta} + 2^{2+2\delta}\mathbb{E}\left|\omega_1' m_1(\beta^*)\right|^{2+\delta} + 2^{1+\delta}\mathbb{E}\left|\omega_2' m_2(\beta^*)\right|^{2+\delta}. \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}\left|\omega_1'\left[m_{1n}(\beta^*) - m_1(\beta^*)\right]\right|^{2+\delta} &\le \|\omega_1\|^{2+\delta}\,\mathbb{E}\left\|m_{1n}(\beta^*) - m_1(\beta^*)\right\|^{2+\delta} \\ &= \|\omega_1\|^{2+\delta}\,\mathbb{E}\left\|B_1\left[K_\alpha(H_{21}'\beta_{21}^*)H_{21}'\beta_{21}^* - \left(H_{21}'\beta_{21}^*\right)_+\right]\right\|^{2+\delta} = o\,(1). \end{aligned}$$

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

Step 2. We prove (ii) $\Xi_n \equiv \mathbb{E}[\nabla_{\beta'} m_n(\beta^*)] = \Xi + o(1)$. Observe that

$$\Xi_n = \begin{pmatrix} \mathbb{E}[\nabla_{\beta_1'} m_{1n}(\beta^*)] & \mathbb{E}[\nabla_{\beta_2'} m_{1n}(\beta^*)] \\ \mathbb{E}[\nabla_{\beta_1'} m_2(\beta^*)] & \mathbb{E}[\nabla_{\beta_2'} m_2(\beta^*)] \end{pmatrix} \equiv \begin{pmatrix} \Xi_{n,11} & \Xi_{n,12} \\ \Xi_{n,21} & \Xi_{n,22} \end{pmatrix}.$$

Noting that $\nabla_{\beta_1'} m_{1n}(\beta^*) = -B_1 B_1'$, $\nabla_{\beta_1'} m_2(\beta^*) = 0$, and $\nabla_{\beta_2'} m_2(\beta^*) = -B_2 B_2'$, 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

$$\nabla_{\beta_2'} m_{1n}(\beta^*) = \left[\nabla_{\beta_{20}'} m_{1n}(\beta^*), \nabla_{\beta_{21}'} m_{1n}(\beta^*)\right],$$

where $\nabla_{\beta_{20}'} m_{1n}(\beta^*) = B_1 H_{20}'$ and $\nabla_{\beta_{21}'} m_{1n}(\beta^*) = B_1 H_{21}'[\frac{1}{\alpha} k_\alpha(H_{21}'\beta_{21}^*) H_{21}'\beta_{21}^* + K_\alpha(H_{21}'\beta_{21}^*)]$. Let $\Xi_{n,12,t} \equiv \mathbb{E}[\nabla_{\beta_{2t}'} m_{1n}(\beta^*)]$ for $t = 0, 1$. It is easy to see that $\Xi_{n,12,0} = \mathbb{E}[B_1 H_{20}']$. Let $r_2(z) = \mathbb{E}[B_1 H_{21}' | H_{21}'\beta_{21}^* = z]$ and $g_2(z) = \int_{-\infty}^{z} r_2(t) f_C(t)\, dt$. Then

$$\begin{aligned} \Xi_{n,12,1} &= \mathbb{E}\left\{ r_2(H_{21}'\beta_{21}^*)[\frac{1}{\alpha} k_\alpha(H_{21}'\beta_{21}^*) H_{21}'\beta_{21}^* + K_\alpha(H_{21}'\beta_{21}^*)] \right\} \\ &= (1-p)\int_{-\infty}^{\infty} r_2(z)[\frac{1}{\alpha} k_\alpha(z) z + K_\alpha(z)] f_C(z)\, dz + p \sum_{z_d} r_2(z_d)[\frac{1}{\alpha} k_\alpha(z_d) z_d + K_\alpha(z_d)] f_D(z_d) \\ &\equiv (1-p) II_C + p II_D, \text{ say,} \end{aligned}$$

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

$$\begin{aligned} II_C &= \frac{1}{\alpha}\int_{-\infty}^{\infty} r_2(z) z f_C(z) k(z/\alpha) dz + \int_{-\infty}^{\infty} r_2(z) f_C(z) K(z/\alpha) dz \\ &= \alpha \int_{-1}^{1} r_2(\alpha u) u f_C(\alpha u) k(u) du + \int_{-\infty}^{\infty} K(z/\alpha) dg_2(z). \end{aligned}$$

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) and A3, we obtain

$$\begin{aligned} \int_{-\infty}^{\infty} K(z/\alpha) dg_2(z) &= g_2(\infty) - \frac{1}{\alpha}\int_{-\infty}^{\infty} k(z/\alpha) g_2(z) dz = g_2(\infty) - \int_{-1}^{1} k(z) g_2(\alpha z) dz \\ &\to g_2(\infty) - g_2(0)\int_{-1}^{1} k(z) dz = g_2(\infty) - g_2(0). \end{aligned}$$

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 α by the fact that $k(\cdot)$ has compact support $[-1, 1]$, we have for large enough

n

$$\begin{aligned}
&\frac{1}{\alpha}\sum_{z_d} r_2(z_d)k_\alpha(z_d)z_d f_D(z_d) \\
= &\frac{1}{\alpha}\sum_{z_d<0} r_2(z_d)k_\alpha(z_d)z_d f_D(z_d) + \frac{1}{\alpha}\sum_{z_d>0} r_2(z_d)k_\alpha(z_d)z_d f_D(z_d) = 0+0=0.
\end{aligned}$$

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 α. Then for sufficiently large n,

$$\begin{aligned}
&\sum_{z_d} r_2(z_d)K_\alpha(z_d)f_D(z_d) \\
= &\frac{1}{2}r_2(0)f_D(0) + \sum_{z_d<0} r_2(z_d)K_\alpha(z_d)f_D(z_d) + \sum_{z_d>0} r_2(z_d)K_\alpha(z_d)f_D(z_d) \\
= &\frac{1}{2}r_2(0)f_D(0) + \sum_{z_d>0} r_2(z_d)f_D(z_d),
\end{aligned}$$

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

Step 3. We show that $\mathbb{E}_n[\nabla_{\beta'}m_n(\bar{\beta})] = \mathbb{E}[\nabla_{\beta'}m_n(\beta^*)] + o_P(1)$. Note that $\nabla_{\beta'}m_n(\beta) = \begin{pmatrix} -B_1B_1' & \nabla_{\beta_2'}m_{1n}(\beta) \\ 0 & -B_2B_2' \end{pmatrix}$, where $\nabla_{\beta_2'}m_{1n}(\beta) = \left[\nabla_{\beta_{20}'}m_{1n}(\beta),\ \nabla_{\beta_{21}'}m_{1n}(\beta)\right]$,

$$\nabla_{\beta_{20}'}m_{1n}(\beta) = B_1H_{20}', \text{ and } \nabla_{\beta_{21}'}m_{1n}(\beta) = B_1H_{21}'[\frac{1}{\alpha}k_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21} + K_\alpha(H_{21}'\beta_{21})].$$

We first show that $\sup_{\beta_{21}\in N_\epsilon(\beta_{21}^*)}\left\|n^{-1/2}\mathbb{G}_n[\nabla_{\beta'}m_n(\beta)]\right\| = o_P(1)$ as $\nabla_{\beta'}m_n(\beta)$ depends on β only through β_{21}. Here $N_\epsilon(\beta_{21}^*)$ denotes the ϵ-neighborhood of β_{21}^* with respect to the Euclidean norm on $\mathbb{R}^{k_{21}}$. Since $-B_1B_1'$, $-B_2B_2'$, and B_1H_{20}' are not β_{21}-dependent and it is standard to show that $n^{-1/2}\mathbb{G}_n[A_n(W)] = o_P(1)$ for $A_n(W) = -B_1B_1'$, $-B_2B_2'$, and B_1H_{20}'. It suffices to prove the last claim by showing that

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

Consider a class of functions,

$$\mathcal{M}_2 = \left\{m_{\beta_{21}} : m_{\beta_{21}}(w) = b_1h_{21}'[\frac{1}{\alpha}k(h_{21}'\beta_{21}/\alpha)h_{21}'\beta_{21} + K(h_{21}'\beta_{21}/\alpha)], \beta_{21}\in\mathcal{B}_{21}\right\}.$$

Note that the functions in $\mathcal{M}_2$ are composed of the following classes of functions:

$$\begin{aligned}\mathcal{M}_{2,1} &= \left\{h_{21}'\beta_{21} : \beta_{21} \in \mathcal{B}_{21}\right\},\\ \mathcal{M}_{2,2} &= \left\{b_1 h_{21}' k(h_{21}'\beta_{21}/\alpha) : \beta_{21} \in \mathcal{B}_{21}\right\},\\ \mathcal{M}_{2,3} &= \left\{b_1 h_{21}' K(h_{21}'\beta_{21}/\alpha) : \beta_{21} \in \mathcal{B}_{21}\right\}.\end{aligned}$$

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 $\|b_1 h_{21}'\| \max_{|u|\le 1} k(u)$ 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

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

This, in conjunction with the fact that $\mathbb{E}[\nabla_{\beta_2'} m_{1n}(\beta)]$ is continuous in β_{21} and that $\bar{\beta} = \beta^* + o_P(1)$, implies that

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

This completes the proof of the theorem. ■

Proof of Theorem 3.3. 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 3.2.

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,

$$\mathcal{M}_3 = \{\bar{m}_\beta : \bar{m}_\beta(w) = b_1 b_1'[y_1 - b_1'\beta_1 + h_{20}'\beta_{20} + K(h_{21}'\beta_{21}/\alpha)h_{21}'\beta_{21}, \beta_1 \in \mathcal{B}_1, \beta_{20} \in \mathcal{B}_{20}, \beta_{21} \in \mathcal{B}_{21}\}.$$

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

$$\begin{aligned}\mathcal{M}_{3,1} &= \left\{b_1 b_1'\left[y_1 - b_1'\beta_1 + h_{20}'\beta_{20}\right] : \beta_1 \in \mathcal{B}_1, \beta_{20} \in \mathcal{B}_{20}\right\},\\ \mathcal{M}_{3,2} &= \left\{b_1 b_1' h_{21}'\beta_{21} : \beta_{21} \in \mathcal{B}_{21}\right\},\\ \mathcal{M}_{3,3} &= \left\{b_1 b_1' K(h_{21}'\beta_{21}/\alpha) : \beta_{21} \in \mathcal{B}_{21}\right\},\end{aligned}$$

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 (1987, 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 β 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 $\hat{\Xi}_{n,11}(\beta)$, $\hat{\Xi}_{n,22}(\beta)$, and $\hat{\Xi}_{n,12,0}(\beta)$ are not β-dependent, it is trivial to show the first three claims. The last claim can be shown as in the analysis of $\hat{\Lambda}_{n,11}(\hat{\beta}_{sq})$. This completes the proof of the theorem. ■

Proof of Theorem 4.1. (i) Our proof is analogous to that of Theorem 1 in Linton (1995). Note that $\sqrt{n}(\hat{\beta}_{sq} - \beta^*) = -[\hat{\Xi}_n(\bar{\beta})]^{-1}\sqrt{n}\mathbb{E}_n[m_n(\beta^*)]$. Using $[\hat{\Xi}_n(\bar{\beta})]^{-1} = \Xi^{-1} + \{[\hat{\Xi}_n(\bar{\beta})]^{-1} - \Xi^{-1}\}$ and $[\hat{\Xi}_n(\bar{\beta})]^{-1} - \Xi^{-1} = -[\hat{\Xi}_n(\bar{\beta})]^{-1}[\hat{\Xi}_n(\bar{\beta}) - \Xi]\Xi^{-1}$ repeatedly, we have the following expansion for T_{0n} :

$$\begin{aligned}
T_{0n} &= -\sigma_c^{-1}c'[\hat{\Xi}_n(\bar{\beta})]^{-1}\xi_N = -\sigma_c^{-1}c'\Xi^{-1}\xi_N - \sigma_c^{-1}c'\{[\hat{\Xi}_n(\bar{\beta})]^{-1} - \Xi^{-1}\}\xi_N \\
&= -\sigma_c^{-1}c'\Xi^{-1}\xi_N + \sigma_c^{-1}c'[\hat{\Xi}_n(\bar{\beta})]^{-1}[\hat{\Xi}_n(\bar{\beta}) - \Xi]\Xi^{-1}\xi_N \\
&= -\sigma_c^{-1}c'\Xi^{-1}\xi_N + \sigma_c^{-1}c'\Xi^{-1}[\hat{\Xi}_n(\bar{\beta}) - \Xi]\Xi^{-1}\xi_N + \sigma_c^{-1}c'\{[\hat{\Xi}_n(\bar{\beta})]^{-1} - \Xi^{-1}\}[\hat{\Xi}_n(\bar{\beta}) - \Xi]\Xi^{-1}\xi_N \\
&= \ldots \\
&= \{-\sigma_c^{-1}c'\Xi^{-1}\xi_N + (n\alpha)^{-1/2}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_N - (n\alpha)^{-1}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N \\
&\quad + (n\alpha)^{-3/2}\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N\} \\
&\quad - (n\alpha)^{-2}\sigma_c^{-1}c'[\hat{\Xi}_n(\bar{\beta})]^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_D\Xi^{-1}\xi_N \\
&= T_{0n}^* + R_{0n}^*,
\end{aligned}$$

where $\xi_N = \sqrt{n}\mathbb{E}_n[m_n(\beta^*)]$ and $\xi_D = \sqrt{n\alpha}[\hat{\Xi}_n(\bar{\beta}) - \Xi]$. 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 $\zeta = O\left([n^{3/5}\ln n]^{-1}\right)$. 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 $A_1 = \left\{|R_{0n}^*| > C_1[n^{3/5}\ln n]^{-1}\right\}$ and $A_2 = \{\left\|[\hat{\Xi}_n(\bar{\beta})]^{-1}\right\|_{\text{sp}} > C_2\}$, where $C_2 = [\lambda_{\min}(\Xi\Xi')]^{-1/2}/2$. Note that $\left\|[\hat{\Xi}_n(\bar{\beta})]^{-1}\right\|_{\text{sp}} = [\lambda_{\min}(\hat{\Xi}_n(\bar{\beta})\hat{\Xi}_n(\bar{\beta})')]^{-1/2}$. Observe that

$$P(A_1) \leq P(A_1 \cap A_2) + P(A_2^c).$$

When $\left\|[\hat{\Xi}_n(\bar{\beta})]^{-1}\right\|_{\mathrm{sp}} < C_2$, we have

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

where the first inequality follows from the fact that

$$\begin{aligned}
\|ABC\|_{\mathrm{sp}}^2 &= \lambda_{\max}\left(ABCC'B'A'\right) \geq \lambda_{\max}\left(ABB'A'\right)\lambda_{\min}\left(CC'\right) \\
&= \lambda_{\max}\left(B'A'AB\right)\lambda_{\min}\left(CC'\right) \geq \lambda_{\max}\left(B'B\right)\lambda_{\min}\left(A'A\right)\lambda_{\min}\left(CC'\right),
\end{aligned}$$

and the second inequality follows from the triangle inequality. In conjunction with Chebyshev inequality, this implies that

$$P\left(A_2^c\right) \leq P\left(\|\xi_D\|_{\mathrm{sp}} \geq \frac{1}{2}n^{1/2}C_2^{-1}\right) \leq \frac{\mathbb{E}[\,\|\xi_D\|_{\mathrm{sp}}^2]}{4nC_2^{-2}} = O\left(n^{-1}\right) = o(n^{-3/5})$$

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

Conditional on A_2,

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

for some $C_3 > 0$. Let $\bar{A}_2 = \{c_3(n\alpha)^{-2}\|\xi_D\|_{\mathrm{sp}}^2\|\xi_N\| \geq C_4[n^{3/5}\ln n]^{-1}\}$ for some $C_4 > 0$. Then $A_1 \cap A_2 \subset \bar{A}_2$ and

$$\begin{aligned}
P\left(A_1 \cap A_2\right) &\leq P\left(\bar{A}_2\right) = P\left(\|\xi_D\|_{\mathrm{sp}}^2\|\xi_N\| \geq (C_4/C_3)(n\alpha)^2 n^{-3/5}(\ln n)^{-1}\right) \\
&\leq (C_3/C_4)^{8/5}\left[(n\alpha)^{-2}n^{3/5}\ln n\right]^{8/5}\cdot\mathbb{E}[\|\xi_D\|_{\mathrm{sp}}^{32/5}\|\xi_N\|^{8/5}] \\
&\leq (C_3/C_4)\alpha^{-16/5}n^{-56/25}(\ln n)^{8/5}\cdot\left\{\mathbb{E}\|\xi_D\|_{\mathrm{sp}}^8\right\}^{4/5}\left\{\mathbb{E}\|\xi_N\|^8\right\}^{1/5} \\
&= O\left(n^{-24/25}(\ln n)^{8/5}\right) = o(n^{-3/5}),
\end{aligned}$$

by Markov inequality, Hölder inequality, Lemma A.1 below and the fact that $\alpha \propto n^{-2/5}$. Consequently, we have $P\left(A_1\right) = P\left(|R_{0n}^*| > C_1[n^{3/5}\ln n]^{-1}\right) = o(n^{-3/5})$ and $|P\left(T_{0n} \leq u\right) - P\left(T_{0n}^* \leq u\right)| \leq P\left(|R_{0n}^*| > \zeta\right) + P\left(|T_{0n}^* - u| < \zeta\right) = o(n^{-3/5})$ uniformly in u.

(ii) By Lemma A.1 and direct moment calculations, we have

$$\mathbb{E}[T_{0n}^*] = -\sigma_c^{-1}c'\Xi^{-1}\mathbb{E}[\xi_N] + O(n^{-1/2}) = -\sqrt{n\alpha^4}\sigma_c^{-1}c'\Xi^{-1}\mathcal{B} + o(n^{-3/10}),$$

and

$$
\begin{aligned}
\mathrm{Var}[T_{0n}^*] &= \mathrm{Var}[-\sigma_c^{-1} c'\Xi^{-1}\xi_N] + \mathrm{Var}[(n\alpha)^{-1/2}\,\sigma_c^{-1} c'\Xi^{-1}\xi_D\Xi^{-1}\xi_N] + o(n^{-3/5}) \\
&= \sigma_c^{-2} c'\Xi^{-1}[\Lambda + O\left(\alpha^2\right)]\Xi^{-1\prime} c + (n\alpha)^{-1}\,\sigma_c^{-1} c'\Xi^{-1}\mathrm{Var}[\xi_D\Xi^{-1}\xi_N\Xi^{-1\prime}c] + o(n^{-3/5}) \\
&= 1 + (n\alpha)^{-1}\,\sigma_c^{-1} c'\Xi^{-1}\Phi_0\Xi^{-1\prime} c + o(n^{-3/5}),
\end{aligned}
$$

where we use the fact that the other variance and covariance terms in the expansion of $\mathrm{Var}[T_{0n}^*]$ are all $o(n^{-3/5})$ under Assumption A4, and $\mathcal{B}$ and Φ_0 are defined in Theorem 4.1.

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

$$
\begin{aligned}
S_n &= \frac{\sqrt{n\alpha}(\hat{\sigma}_c - \sigma_c)}{\sigma_c} = \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{\sigma_c^2}\left[\frac{1}{2 + (\hat{\sigma}_c/\sigma_c - 1)}\right] \\
&= \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{\sigma_c^2}\left\{\frac{1}{2} - \frac{1}{4}\left(\frac{\hat{\sigma}_c}{\sigma_c} - 1\right)\right\} + O_P\left((n\alpha)^{-1}\right) \\
&= \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{2\sigma_c^2} - \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{4\sigma_c^2}\frac{\hat{\sigma}_c - \sigma_c}{\sigma_c} + O_P\left((n\alpha)^{-1}\right) \\
&= \frac{\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)}{2\sigma_c^2} - \frac{\left[\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2)\right]^2}{8\sqrt{n\alpha}\sigma_c^4} + O_P\left((n\alpha)^{-1}\right).
\end{aligned}
$$

Next,

$$
\begin{aligned}
\sqrt{n\alpha}(\hat{\sigma}_c^2 - \sigma_c^2) &= \sqrt{n\alpha}c'\left[\hat{\Xi}_n^{-1}\hat{\Lambda}_n(\hat{\Xi}_n^{-1})' - \Xi^{-1}\Lambda\left(\Xi^{-1}\right)'\right]c \\
&= \sqrt{n\alpha}c'\left[[(\hat{\Xi}_n^{-1} - \Xi^{-1}) + \Xi^{-1}]\hat{\Lambda}_n[(\hat{\Xi}_n^{-1} - \Xi^{-1}) + \Xi^{-1}]' - \Xi^{-1}\Lambda\left(\Xi^{-1}\right)'\right]c \\
&= \sqrt{n\alpha}c'\{2(\hat{\Xi}_n^{-1} - \Xi^{-1})\hat{\Lambda}_n\Xi^{-1\prime} + \Xi^{-1}(\hat{\Lambda}_n - \Lambda)\Xi^{-1\prime} + (\hat{\Xi}_n^{-1} - \Xi^{-1})\hat{\Lambda}_n(\hat{\Xi}_n^{-1} - \Xi^{-1})'\}c \\
&= -2c'\hat{\Xi}_n^{-1}\xi_D^{\dagger}\Xi^{-1}\hat{\Lambda}_n\Xi^{-1\prime}c + \sqrt{\alpha}c'\Xi^{-1}\xi_\Lambda\left(\Xi^{-1}\right)'c + c'\hat{\Xi}_n^{-1}\xi_D^{\dagger}\Xi^{-1}\hat{\Lambda}_n\Xi^{-1\prime}\xi_D^{\dagger}\hat{\Xi}_n^{-1\prime}c/\sqrt{n\alpha} \\
&= -2c'\Xi^{-1}\xi_D^{\dagger}\Xi^{-1}\Lambda\Xi^{-1\prime}c + \sqrt{\alpha}c'\Xi^{-1}\xi_\Lambda\left(\Xi^{-1}\right)'c \\
&\quad +2[c'\Xi^{-1}\xi_D^{\dagger}\Xi^{-1}\xi_D^{\dagger}\Xi^{-1}\Lambda\Xi^{-1\prime}c + c'\Xi^{-1}\xi_D^{\dagger}\Xi^{-1}\Lambda\Xi^{-1\prime}\xi_D^{\dagger}\Xi^{-1\prime}c]/\sqrt{n\alpha} \\
&\quad -2c'\Xi^{-1}\xi_D^{\dagger}\Xi^{-1}\xi_\Lambda\Xi^{-1\prime}c/\sqrt{n} + O_P\left((n\alpha)^{-1}\right)
\end{aligned}
$$

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 $S_n = S_n^* + o_P\left((n\alpha)^{-1/2}\right)$ where S_n^* is defined in (4.2). It follows that

$$
\begin{aligned}
T_n &= \frac{\sqrt{n}c'(\hat{\beta}_{sq} - \beta^*)}{\sigma_c}\cdot\frac{\sigma_c}{\hat{\sigma}_c} = \frac{T_{0n}}{1 + (n\alpha)^{-1/2}S_n} \\
&= T_{0n}\left\{1 - (n\alpha)^{-1/2}S_n + 2(n\alpha)^{-1}S_n^2\right\} + o_P((n\alpha)^{-1}) \\
&= T_{0n}^* + (n\alpha)^{-1/2}\left[\sigma_c^{-1}c'\Xi^{-1}\xi_N - (n\alpha)^{-1/2}\,\sigma_c^{-1}c'\Xi^{-1}\xi_D\Xi_n^{-1}\xi_N\right]S_n^* - 2(n\alpha)^{-1}\sigma_c^{-1}c'\Xi^{-1}\xi_N S_n^{*2} \\
&\quad +o_P((n\alpha)^{-1}) \\
&= T_n^* + o_P((n\alpha)^{-1}),
\end{aligned}
$$

where T_n^* is defined in (4.3). 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 $(n\alpha)^{-1/2} = O\left(n^{-3/10}\right)$:

$$\mathbb{E}[T_n^*] = -\sqrt{n\alpha^4}\sigma_c^{-1}c'\Xi^{-1}\mathcal{B} + o(n^{-3/10}),$$

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

$$\begin{aligned}
\mathrm{Var}[T_n^*] &= \mathrm{Var}[-\sigma_c^{-1}c'\Xi^{-1}\xi_N] + \mathrm{Var}[(n\alpha)^{-1/2}\,\sigma_c^{-1}c'\Xi^{-1}(\xi_D\Xi^{-1}\xi_N + \sigma_c^{-2}c'\Xi^{-1}\xi_D^{\dagger}\Omega c\xi_N)] + o(n^{-3/5}) \\
&= \sigma_c^{-2}c'\Xi^{-1}[\Lambda + O(\alpha^2)]\Xi^{-1\prime}c + (n\alpha)^{-1}\,\sigma_c^{-1}c'\Xi^{-1}\mathrm{Var}[\xi_D\Xi^{-1}\xi_N + \sigma_c^{-2}c'\Xi^{-1}\xi_D^{\dagger}\Omega c\xi_N]\Xi^{-1\prime}c \\
&\quad +o\left(n^{-3/5}\right) \\
&= 1 + (n\alpha)^{-1}\,\sigma_c^{-1}c'\Xi^{-1}\Phi\Xi^{-1\prime}c + o\left(n^{-3/5}\right),
\end{aligned}$$

where Φ is defined in Theorem 4.1. ■

Lemma A.1 *Recall that* $\xi_N = \sqrt{n}\mathbb{E}_n[m_n(\beta^*)]$ *and* $\xi_D = \sqrt{n}[\hat{\Xi}_n(\bar{\beta}) - \Xi_n]$. *Suppose that the conditions in Theorem 4.1 hold. Then (i)* $\mathbb{E}\left\|\xi_N\right\|^8 = O\left(1\right)$, *(ii)* $\mathbb{E}\left\|\xi_D\right\|^8 = O\left(1\right)$, *(iii) Var*$[\xi_D\Xi^{-1}\xi_N] = \Phi_0 + o\left(1\right)$, *and (iv) Var*$[\xi_D\Xi^{-1}\xi_N + \sigma_c^{-2}c'\Xi^{-1}\xi_D^{\dagger}\Omega c\xi_N] = \Phi + o\left(1\right)$.

Proof. (i) Let $\xi_{N,1} = \sqrt{n}\mathbb{E}_n[m(\beta^*)] = \sqrt{n}\mathbb{E}_n\begin{pmatrix} B_1\varepsilon_1 \\ B_2\varepsilon_2 \end{pmatrix}$ and $\xi_{N,2} = \xi_N - \xi_{N,1}$. Let $\mathcal{F}_i = \sigma\left\{W_i, W_{i-1}, ...W_1\right\}$, the sigma-field generated by $W_i, W_{i-1}, ...W_1$. Noting that $\mathbb{E}\left[m(W_i;\beta^*)|\mathcal{F}_{i-1}\right] = \mathbb{E}\left[m(W_i;\beta^*)\right] = 0$, we can apply Rosenthal's inequality (e.g., Hall and Heyde 1980, p. 23) to obtain

$$\begin{aligned}
\mathbb{E}\left\|\xi_{N,1}\right\|^8 &\leq C\left\{\mathbb{E}\left[n^{-1}\sum_{i=1}^{n}\mathbb{E}[\,\left\|m(W_i;\beta^*)\right\|^2|\mathcal{F}_{i-1}]\right]^4 + n^{-4}\sum_{i=1}^{n}\mathbb{E}\left\|m(W_i;\beta^*)\right\|^8\right\} \\
&= C\left\{\left[\mathbb{E}\left\|m(W_i;\beta^*)\right\|^2\right]^4 + n^{-3}\mathbb{E}\left\|m(W_i;\beta^*)\right\|^8\right\} = O\left(1\right)
\end{aligned}$$

under Assumptions A1(i) and A4(i). Similarly, we can show that $\mathbb{E}\left\|\xi_{N,2} - \mathbb{E}(\xi_{N,2})\right\|^8 = O\left(1\right)$. 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 3.2. Then by the C_r inequality,

$$\mathbb{E}\left\|\xi_{N,2}\right\|^8 \leq 128\mathbb{E}\left\|\xi_{N2} - \mathbb{E}(\xi_{N,2})\right\|^8 + 128\left\|\mathbb{E}(\xi_{N,2})\right\|^8 = O\left(1\right) + o\left(1\right) = O\left(1\right).$$

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

(ii) Let $\eta(\beta_{21}) = \eta(H_{21};\beta_{21}) = B_1 H_{21}'[\frac{1}{\alpha}k_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21} + K_\alpha(H_{21}'\beta_{21})]$. We make the following decomposition

$$\begin{aligned}
\xi_D &= \sqrt{\alpha}\begin{pmatrix} 0 & (0, \mathbb{G}_n[\eta(\beta_{21}^*)]) \\ 0 & 0 \end{pmatrix} + \sqrt{\alpha}\begin{pmatrix} 0 & \left(0, n^{1/2}\mathbb{E}_n\left[\eta\left(\bar{\beta}_{21}\right) - \eta\left(\beta_{21}^*\right)\right]\right) \\ 0 & 0 \end{pmatrix} \\
&\quad + \sqrt{\alpha}\mathbb{G}_n\begin{pmatrix} -B_1B_1' & (B_1H_{20}', 0) \\ 0 & -B_2B_2' \end{pmatrix} \\
&\equiv \xi_{D,1} + \xi_{D,2} + \xi_{D,3}, \text{ say.}
\end{aligned}$$

As in part (i), we can readily show that $\mathbb{E}\left\|\xi_{D,3}\right\|^8 = O\left(\alpha^4\right) = o(1)$. By Theorem 3.2 and the fact that $\bar{\beta}_{21}$ lies between $\hat{\beta}_{21}$ and β_{21}^* element by element, we have $\bar{\beta}_{21} = \beta_{21}^* + O_P\left(n^{-1/2}\right)$. Then, we can apply Sargan and Mikhail's (1971) inequality as in the proof of Theorem 4.1 and show that $\mathbb{E}\left\|\xi_{D,3}\right\|^8 = o(1)$. In addition, one can readily show that $\alpha^4\mathbb{E}\left\|\eta\left(\beta_{21}^*\right)\right\|^8 = O(1)$. It follows that $\mathbb{E}\left\|\xi_D\right\|^8 = \mathbb{E}\left\|\xi_{D,1}\right\|^8 + o(1) = \alpha^4\mathbb{E}\left\|\eta\left(\beta_{21}^*\right)\right\|^8 = O(1)$.

(iii) Note that $\mathrm{Var}\left(\xi_D\Xi^{-1}\xi_N\right) = \mathbb{E}\left(\xi_D\Xi^{-1}\xi_N\xi_N'\Xi^{-1\prime}\xi_D'\right) - \mathbb{E}\left(\xi_D\Xi^{-1}\xi_N\right)\mathbb{E}\left(\xi_N'\Xi^{-1\prime}\xi_D'\right)$. It is easy to show that $\mathbb{E}\left(\xi_D\Xi^{-1}\xi_N\right) = O\left(\alpha^{1/2}\right)$. Using the decomposition of ξ_N and ξ_D in (i) and (ii), we can readily show that

$$\begin{aligned}
&\mathbb{E}\left(\xi_D\Xi^{-1}\xi_N\xi_N'\Xi^{-1\prime}\xi_D'\right) \\
=\ &\mathbb{E}\left(\xi_{D,1}\Xi^{-1}\xi_{N,1}\xi_{N,1}'\Xi^{-1\prime}\xi_{D,1}'\right) + o(1) \\
=\ &\frac{\alpha}{n^2}\sum_{i,j,k,l=1}^{n}\mathbb{E}\left\{\begin{pmatrix} 0 & \bar{\eta}_1\left(H_{21,i};\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}\Xi^{-1}\begin{pmatrix} B_{1j}B_{1k}'\varepsilon_{1j}\varepsilon_{1k} & B_{1j}B_{2k}'\varepsilon_{1j}\varepsilon_{2k} \\ B_{2j}B_{1k}'\varepsilon_{2j}\varepsilon_{1k} & B_{2j}B_{2k}'\varepsilon_{2j}\varepsilon_{2k} \end{pmatrix}\Xi^{-1\prime}\right. \\
&\left.\times\begin{pmatrix} 0 & \bar{\eta}_1\left(H_{21,l};\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}'\right\} + o(1) \\
=\ &\alpha\mathbb{E}\left\{\begin{pmatrix} 0 & \bar{\eta}_1\left(\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}\Omega\begin{pmatrix} 0 & \bar{\eta}_1\left(\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}'\right\} + o(1) \\
=\ &\begin{pmatrix} \alpha\mathbb{E}\left[\bar{\eta}_1\left(\beta_{21}^*\right)\Omega_{22}\bar{\eta}_1\left(\beta_{21}^*\right)'\right] & 0 \\ 0 & 0 \end{pmatrix} + o(1) = \Phi_0 + o(1),
\end{aligned}$$

where $\bar{\eta}_1(\beta_{21}) \equiv \bar{\eta}_1(H_{21};\beta_{21}) = \eta_1(\beta_{21}) - \mathbb{E}[\eta_1(\beta_{21})]$, and $\eta_1(\beta_{21}) \equiv \eta_1(H_{21};\beta_{21}) = [0_{k_1\times k_{20}}, B_1H_{21}' \times\frac{1}{\alpha}k_\alpha(H_{21}'\beta_{21})H_{21}'\beta_{21}]$. Then (iii) follows.

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

$$\begin{aligned}
& \operatorname{Var}\left(\sigma_c^{-2}c'\Xi^{-1}\xi_D^{\dagger}\Omega c\xi_N\right) \\
= \;& \sigma_c^{-4}c'\Xi^{-1}\mathbb{E}\left(\xi_D^{\dagger}\Omega c\xi_N\xi_N'c'\Omega\xi_D^{\dagger\prime}\right)\Xi^{-1\prime}c - o\left(1\right) \\
= \;& \sigma_c^{-4}c'\Xi^{-1}\mathbb{E}\left(\xi_{D,1}\Omega c\xi_{N,1}\xi_{N,1}'c'\Omega\xi_{D,1}'\right)\Xi^{-1\prime}c + o\left(1\right) \\
= \;& \sigma_c^{-4}c'\Xi^{-1}\mathbb{E}\left(\xi_{D,1}\Omega c\Lambda c'\Omega\xi_{D,1}'\right)\Xi^{-1\prime}c + o\left(1\right) \\
= \;& \alpha\ \sigma_c^{-4}c'\Xi^{-1}\mathbb{E}\left\{\begin{pmatrix} 0 & \bar{\eta}_1\left(\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}\Omega c\Lambda c'\Omega\begin{pmatrix} 0 & \bar{\eta}_1\left(\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}'\right\}\Xi^{-1\prime}c + o\left(1\right) \\
= \;& \alpha\sigma_c^{-4}\mathbb{E}[\eta_1^*\left(\beta_{21}^*\right)]^2\Lambda + o\left(1\right)
\end{aligned}$$

and

$$\begin{aligned}
& \operatorname{Cov}\left(\xi_D\Xi^{-1}\xi_N, \sigma_c^{-2}c'\Xi^{-1}\xi_D^{\dagger}\Omega c\xi_N\right) \\
= \;& \sigma_c^{-2}\mathbb{E}\left[\xi_D\Xi^{-1}\xi_N\xi_N'c'\Omega\xi_D^{\dagger\prime}\right]\Xi^{-1\prime}c - o\left(1\right) \\
= \;& \sigma_c^{-2}\mathbb{E}\left[\xi_{D,1}\Xi^{-1}\xi_{N,1}\xi_{N,1}'c'\Omega\xi_{D,1}'\right]\Xi^{-1\prime}c + o\left(1\right) \\
= \;& \sigma_c^{-2}\mathbb{E}\left[\xi_{D,1}\Xi^{-1}\Lambda c'\Omega\xi_{D,1}'\right]\Xi^{-1\prime}c + o\left(1\right) \\
= \;& \alpha\sigma_c^{-2}\mathbb{E}\left[\begin{pmatrix} 0 & \bar{\eta}_1\left(\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}\Xi^{-1}\Lambda c'\Omega\begin{pmatrix} 0 & \bar{\eta}_1\left(\beta_{21}^*\right) \\ 0 & 0 \end{pmatrix}'\right]\Xi^{-1\prime}c + o\left(1\right) \\
= \;& \alpha\sigma_c^{-2}\mathbb{E}[\eta_1^*\left(\beta_{21}^*\right)\eta_1^{**}\left(\beta_{21}^*\right)]\Lambda,
\end{aligned}$$

where $\eta_1^*\left(\beta_{21}\right)$ and $\eta_1^{**}\left(\beta_{21}\right)$ are defined in Theorem 4.1. It follows that $\operatorname{Var}[\xi_D\Xi^{-1}\xi_N + \sigma_c^{-2}c'\Xi^{-1}\xi_D^{\dagger}\Omega c\xi_N]$ $= \Phi_0 + \alpha\sigma_c^{-4}\mathbb{E}[\eta_1^*\left(\beta_{21}^*\right)]^2\Lambda + \alpha\sigma_c^{-2}\mathbb{E}[\eta_1^*\left(\beta_{21}^*\right)\eta_1^{**}\left(\beta_{21}^*\right)\Lambda + \Lambda\eta_1^{**}\left(\beta_{21}^*\right)'\eta_1^*\left(\beta_{21}^*\right)] + o\left(1\right) = \Phi + o\left(1\right)$ ■

B The Generative Model

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

$$\begin{aligned}
\max_{\alpha_2} Q_2(H_2, A_2) &= M + [\gamma_5 + \gamma_6 X_2 + \gamma_7 A_1]_+ \\
&= M + \frac{\alpha_1}{2}(1 - X_2)(1 - A_1) + \frac{\alpha_2}{2}(1 - X_2)A_1 + \frac{\alpha_3}{2}(1 + X_2)(1 - A_1) + \frac{\alpha_4}{2}(1 + X_2)A_1,
\end{aligned}$$

where $\alpha_1 = [\gamma_5 - \gamma_6]_+$, $\alpha_2 = [\gamma_5 - \gamma_6 + \gamma_7]_+$, $\alpha_3 = [\gamma_5 + \gamma_6]_+$, and $\alpha_4 = [\gamma_5 + \gamma_6 + \gamma_7]_+$. Therefore,

$$
\begin{aligned}
&Q_1(H_1, A_1)\\
&= \mathbb{E}[Y_1 + \max_{\alpha_2} Q_2(H_2, \alpha_2)|H_1, A_1]\\
&= M + \frac{\alpha_1}{2}(1 - A_1)\mathbb{E}[1 - X_2|X_1, A_1] + \frac{\alpha_2}{2}A_1\mathbb{E}[1 - X_2|X_1, A_1]\\
&+ \frac{\alpha_3}{2}(1 - A_1)\mathbb{E}[1 + X_2|X_1, A_1] + \frac{\alpha_4}{2}A_1\mathbb{E}[1 + X_2|X_1, A_1]\\
&= M + \frac{\alpha_1 + \alpha_3}{2}(1 - A_1) + \frac{\alpha_2 + \alpha_4}{2}A_1 + \frac{1}{2}(\alpha_3 - \alpha_1)(1 - A_1)\mathbb{E}[X_2|X_1, A_1] + \frac{1}{2}(\alpha_4 - \alpha_2)A_1\mathbb{E}[X_2|X_1, A_1]\\
&\overset{(4)}{=} M + \frac{\alpha_1 + \alpha_3}{2}(1 - A_1) + \frac{\alpha_2 + \alpha_4}{2}A_1\\
&+ \frac{1}{2}(\alpha_3 - \alpha_1)\Big(\text{expit}(-\delta_1)(1 - X_1)(1 - A_1) + \text{expit}(\delta_1)(1 + X_1)(1 - A_1) - (1 - A_1)\Big)\\
&+ \frac{1}{2}(\alpha_4 - \alpha_2)\Big(\text{expit}(-\delta_1 + \delta_2)(1 - X_1)A_1 + \text{expit}(\delta_1 + \delta_2)(1 + X_1)A_1 - A_1\Big)\\
&\overset{(5)}{=} \psi_1 + \psi_2 X_1 + (\psi_3 + \psi_4 X_1)\, A_1,
\end{aligned}
$$

where

$$
\begin{aligned}
\psi_1 &= \gamma_1 + \frac{\alpha_1 + \alpha_3}{2} + \frac{1}{2}(\alpha_3 - \alpha_1)\Big(\text{expit}(-\delta_1) + \text{expit}(\delta_1) - 1\Big) = \gamma_1 + \frac{\alpha_1 + \alpha_3}{2},\\
\psi_2 &= \gamma_2 + \frac{\alpha_3 - \alpha_1}{2}\Big(- \text{expit}(-\delta_1) + \text{expit}(\delta_1)\Big) = \gamma_2 + \frac{\alpha_3 - \alpha_1}{2}\Big(2\text{expit}(\delta_1) - 1\Big),\\
\psi_3 &= \gamma_3 - \frac{\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4}{2} + \frac{\alpha_3 - \alpha_1}{2}\Big(- \text{expit}(-\delta_1) - \text{expit}(\delta_1) + 1\Big)\\
&\quad + \frac{\alpha_4 - \alpha_2}{2}\Big(\text{expit}(-\delta_1 + \delta_2) + \text{expit}(\delta_1 + \delta_2) - 1\Big)\\
&= \gamma_3 - \frac{\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4}{2} + \frac{\alpha_4 - \alpha_2}{2}\Big(\text{expit}(-\delta_1 + \delta_2) + \text{expit}(\delta_1 + \delta_2) - 1\Big),\\
\psi_4 &= \gamma_4 + \frac{\alpha_3 - \alpha_1}{2}\Big(1 - 2\text{expit}(\delta_1)\Big) + \frac{\alpha_4 - \alpha_2}{2}\Big(- \text{expit}(-\delta_1 + \delta_2) + \text{expit}(\delta_1 + \delta_2)\Big).
\end{aligned}
$$

In Step (4) of derivation of $Q_1(H_1, A_1)$, we used the fact that

$$
\begin{aligned}
\mathbb{E}[X_2|X_1, A_1] &= 2\text{expit}(\delta_1 X_1 + \delta_2 A_1) - 1\\
&= \text{expit}(-\delta_1)(1 - X_1)(1 - A_1) + \text{expit}(-\delta_1 + \delta_2)(1 - X_1)A_1\\
&\quad + \text{expit}(\delta_1)(1 + X_1)(1 - A_1) + \text{expit}(\delta_1 + \delta_2)(1 + X_1)A_1 - 1,
\end{aligned}
$$

and thus

$$
\begin{aligned}
(1 - A_1)\mathbb{E}[X_2|X_1, A_1] &= \text{expit}(-\delta_1)(1 - X_1)(1 - A_1) + \text{expit}(\delta_1)(1 + X_1)(1 - A_1) - (1 - A_1),\\
A_1\mathbb{E}[X_2|X_1, A_1] &= \text{expit}(-\delta_1 + \delta_2)(1 - X_1)A_1 + \text{expit}(\delta_1 + \delta_2)(1 + X_1)A_1 - A_1,
\end{aligned}
$$

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 $\text{expit}(-x) + \text{expit}(x) = 1$.

References

[1] Chakraborty, B, E.B., Laber, and Y. Zhao (2013), "Inference for Optimal Dynamic Treatment Regimes using an Adaptive m-out-of-n Bootstrap Scheme," *Biometrics* 69, 714-723.

[2] Chakraborty, B., S.A. Murphy and V. Strecher (2010), "Inference for Non-regular Parameters in Optimal Dynamic Treatment Regimes," *Statistical Methods in Medical Research* 19, 317-343.

[3] Chakraborty, B. and E.E.M. Moodie (2013), *Statistical Methods for Dynamic Treatment Regimes*. Springer, New York.

[4] Davidson, J. (1994), *Stochastic Limit Theory*. Oxford University Press, Oxford.

[5] Einmahl, U., and D. M. Mason (2000), "An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators," *Journal of Theoretical Probability* 13, 1-37.

[6] Goldberg, Y. R. Song, and M.R. Kosorok (2013), "Adaptive Q-learning," *Institute of Mathematical Statistics* 9, 150-162.

[7] Hall, P., and C. Heyde (1980), *Martingale Limit Theory and Its Application*. Academic Press, New York.

[8] Hirano, K. and J.R. Porter (2012), "Impossibility Results for Nondifferentiable Functionals," *Econometrica* 80, 1769-1790.

[9] Horowitz, J.L. (1992), "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica* 60, 505-531.

[10] Laber, E.B., D.J. Lizotte, M. Qian, W.E. Pelham, and S.A. Murphy (2013), "Dynamic Treatment Regimes: Technical Challenges and Applications," Working Paper.

[11] Linton, O. (1995), "Second Order Approximation in the Partially Linear Regression Models," *Econometrica* 63, 1079-1112.

[12] Manski, C. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics* 3, 205-228.

[13] Manski, C. (1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics* 27, 313-334.

[14] Moodie, E.E.M., T.S. Richardson, and D.A. Stephens (2007), "Demystifying Dynamic Treatment Regimes," *Biometrics* 63, 447-455.

[15] Moodie, E.E.M. and T.S. Richardson (2010), "Estimating Optimal Dynamic Regimes: Correcting Bias under the Null," *Scandinavian Journal of Statistics* 37, 126-146.

[16] Nolan, D. and D. Pollard (1987), "U-processes: Rates of Convergence," *Annals of Statistics* 15, 780-799.

[17] Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica* 57, 1027-1057.

[18] Pollard, D. (1984), *Convergence of Stochastic Processes.* Springer-Verlag, New York,

[19] Robins, J. (2004), "Optimal Structural Nested Models for Optimal Sequential Decisions," In Lin D.Y., Heagerty P., eds. *Proceedings of the Second Seattle Symposium in Biostatistics.* Springer, New York, 189-326.

[20] Rubin, D.B. (1980), "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment," *Journal of the American Statistical Association* 75, 591-593.

[21] Sargan, J.D. and W.M. Mikhail (1991), "A General Approximation to the Distribution of Instrumental Variables Estimates," *Econometrica* 39, 131-169.

[22] Schneider L.S., P.N. Tariot, C.G. Lyketsos, K.S. Dagerman, K.L. Davis, S. Davis, J.K. Hsiao, D.V. Jeste, I.R. Katz, J.T. Olin, B.G. Pollock, P.V. Rabins, R.A. Rosenheck, G.W. Small, B. Lebowitz, and J.A. Lieberman (2001), "National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer Disease Trial Methodology," *American Journal of Geriatric Psychiatry* 9(4), 346-360.

[23] Schneider, L.S., P.N. Tariot, K.S. Dagerman, S.M. Davis, J.K. Hsiao, M. S. Ismail, B.D. Lebowitz, C.G. Lyketsos, J.M. Ryan, T.S. Stroup, D.L. Sultzer, D. Weintraub, and J.A. Lieberman (2006), "Effectiveness of Atypical Antipsychotic Drugs in Patients with Alzheimer's Disease," *New England Journal of Medicine* 355, 1525-1538.

[24] Song, R., W. Wang, D. Zheng, and M.R. Kosorok (2015), "Penalized Q-learning for Dynamic Treatment Regimens," *Statistica Sinica* 25, 901-920.

[25] Tran, P.V., S.H. Hamilton, A.J. Kuntz, J.H. Potvin, S.W. Andersen, C. Beasley, and G.D.Tollefson (1997), "Double-Blind Comparison of Olanzapine versus Risperidone in the Treatment of Schizophrenia and other Psychotic Disorders," *Journal of Clinical Psychopharmacology* 17(5), 407-418.

[26] van der Vaart, A.W. (1998), *Asymptotic Statistics*. Cambridge University Press, Cambridge.

		Sample Size: 200				Sample Size: 1000			
Design		Bias	MSE	CP	IL	Bias	MSE	CP	IL
		$(\times 10^2)$	$(\times 10^2)$	(%)		$(\times 10^2)$	$(\times 10^2)$	(%)	
1	SQ(RoT)	−0.0949	3.4930	94.37	0.7335	−0.0449	0.7359	94.79	0.3336
	SQ($\hat{\alpha}_0^*$)	−0.0914	3.4780	94.46	0.7339	−0.0469	0.7360	94.82	0.3333
	SQ($\hat{\alpha}^*$)	−0.0648	3.5150	94.34	0.7353	−0.1354	0.7480	94.86	0.3376
	FACI	-	-	95.74	0.7971	-	-	95.13	0.3448
	Q	−0.1003	3.4660	94.69	0.7163	−0.0420	0.7338	94.61	0.3275
	$\text{HT}_{0.08}$	−3.2920	4.1490	95.89	0.7725	−1.3190	0.9445	94.85	0.3509
	$\text{HT}_{0.20}$	−1.9280	3.8920	95.37	0.7497	−0.4750	0.8157	94.75	0.3402
	ST	−3.6590	3.3750	94.79	0.7167	−3.9560	0.9658	92.79	0.3362
2	SQ(RoT)	−0.1429	2.5730	97.07	0.7261	−0.0661	0.5173	97.50	0.3271
	SQ($\hat{\alpha}_0^*$)	−0.1417	2.5730	97.12	0.7293	−0.0701	0.5170	97.53	0.3273
	SQ($\hat{\alpha}^*$)	−0.1458	2.6090	97.10	0.7307	−0.0709	0.5226	97.48	0.3281
	FACI	-	-	98.33	0.8063	-	-	98.47	0.3625
	Q	−0.1423	2.5650	97.07	0.6664	−0.0659	0.5169	97.14	0.2963
	$\text{HT}_{0.08}$	−0.1209	3.6680	98.37	0.7926	−0.0170	0.7221	98.41	0.3543
	$\text{HT}_{0.20}$	−0.2410	3.3550	98.33	0.7594	−0.0666	0.6782	98.31	0.3393
	ST	−0.0027	3.6290	96.41	0.7342	0.0015	0.7099	96.66	0.3263
3	SQ(RoT)	−8.5060	4.0560	93.52	0.7550	−3.5630	0.8023	94.19	0.3414
	SQ($\hat{\alpha}_0^*$)	−8.5420	4.0500	93.63	0.7568	−3.6000	0.8024	94.23	0.3414
	SQ($\hat{\alpha}^*$)	−8.1870	4.0330	93.80	0.7611	−3.3890	0.7930	94.53	0.3446
	FACI	-	-	95.57	0.8252	-	-	96.43	0.3602
	Q	−8.6080	4.0510	90.69	0.7130	−3.6150	0.8026	90.55	0.3242
	$\text{HT}_{0.08}$	−8.6900	6.9100	94.41	0.8946	−0.1508	0.7776	96.11	0.3586
	$\text{HT}_{0.20}$	−6.3550	5.1070	94.08	0.8207	−1.2150	0.7710	94.84	0.3467
	ST	−24.710	10.5700	80.54	0.7917	−5.5940	1.2070	89.01	0.3664

Table 1: Performances of Different Estimators for ψ_3

		Sample Size: 200				Sample Size: 1000			
Design		Bias	MSE	CP	IL	Bias	MSE	CP	IL
		($\times 10^2$)	($\times 10^2$)	(%)		($\times 10^2$)	($\times 10^2$)	(%)	
4	SQ(RoT)	-8.8710	4.1900	93.46	0.7592	-3.9100	0.8320	93.82	0.3421
	SQ($\hat{\alpha}_0^*$)	-9.0770	4.1940	93.51	0.7599	-4.0640	0.8370	93.69	0.3414
	SQ($\hat{\alpha}^*$)	-8.4760	4.1090	94.01	0.7648	-3.7410	0.8155	94.33	0.3449
	FACI	-	-	96.03	0.8074	-	-	96.23	0.3598
	Q	-9.1970	4.2040	89.82	0.7298	-4.0940	0.8381	89.60	0.3242
	HT$_{0.08}$	-2.3470	4.0680	95.34	0.8246	-0.9354	0.7808	95.38	0.3510
	HT$_{0.20}$	-4.3260	3.9740	94.07	0.7856	-1.8560	0.7850	94.06	0.3441
	ST	-16.5100	7.5520	85.65	0.8372	-3.3960	0.9158	92.45	0.3446
5	SQ(RoT)	-8.8110	4.1070	94.11	0.7592	-3.8250	0.8116	94.00	0.3419
	SQ($\hat{\alpha}_0^*$)	-9.0180	4.1120	94.12	0.7600	-3.9800	0.8165	93.88	0.3413
	SQ($\hat{\alpha}^*$)	-8.4130	4.0240	94.54	0.7653	-3.6570	0.7950	94.58	0.3448
	FACI	-	-	96.16	0.8035	-	-	96.16	0.3596
	Q	-9.1390	4.1210	90.00	0.7285	-4.0090	0.8175	90.31	0.3242
	HT$_{0.08}$	-2.1350	3.9090	95.68	0.7891	-0.8507	0.7658	95.61	0.3509
	HT$_{0.20}$	-4.2460	3.8680	94.47	0.7732	-1.7710	0.7669	94.59	0.3441
	ST	-0.6916	3.8530	95.20	0.7606	-0.2797	0.7574	95.39	0.3384
6	SQ(RoT)	-0.1440	2.5720	97.12	0.7262	-0.0635	0.5152	97.44	0.3270
	SQ($\hat{\alpha}_0^*$)	-0.1427	2.5720	97.19	0.7293	-0.0622	0.5149	97.45	0.3273
	SQ($\hat{\alpha}^*$)	-0.1447	2.6080	97.11	0.7309	-0.0633	0.5201	97.43	0.3279
	FACI	-	-	98.28	0.8054	-	-	98.45	0.3616
	Q	-0.1431	2.5640	97.03	0.6663	-0.0635	0.5148	97.02	0.2961
	HT$_{0.08}$	-0.0949	3.6590	98.33	0.7901	-0.0210	0.7229	98.23	0.3516
	HT$_{0.20}$	-0.1858	3.3530	98.22	0.7577	-0.0710	0.6710	98.21	0.3375
	ST	0.0032	3.6590	96.37	0.7353	0.0034	0.7222	96.54	0.3274

Table 2: Performances of Different Estimators for ψ_3

		Sample Size: 200				Sample Size: 1000			
Design		Bias	MSE	CP	IL	Bias	MSE	CP	IL
		($\times 10^2$)	($\times 10^2$)	(%)		($\times 10^2$)	($\times 10^2$)	(%)	
1	SQ(RoT)	-0.2118	2.4430	94.75	0.6091	-0.0204	0.4809	95.24	0.2735
	SQ($\hat{\alpha}_0^*$)	-0.2128	2.4460	94.79	0.6100	-0.0233	0.4807	95.24	0.2735
	SQ($\hat{\alpha}^*$)	-0.2079	2.4560	94.76	0.6105	-0.0482	0.4819	95.30	0.2741
	FACI	-	-	95.15	0.6278	-	-	95.18	0.2756
	Q	-0.2174	2.4400	94.74	0.6085	-0.0216	0.4808	94.67	0.2714
	HT$_{0.08}$	1.3120	2.5900	95.00	0.6258	0.3630	0.4947	94.78	0.2741
	HT$_{0.20}$	0.4963	2.5210	94.90	0.6181	0.1091	0.4834	94.72	0.2727
	ST	3.2640	2.6150	94.50	0.6174	1.6110	0.5187	94.45	0.2741
2	SQ(RoT)	-0.2971	2.0950	94.53	0.5609	-0.0551	0.4024	95.34	0.2513
	SQ($\hat{\alpha}_0^*$)	-0.3004	2.0960	94.57	0.5614	-0.0546	0.4023	95.36	0.2515
	SQ($\hat{\alpha}^*$)	-0.3028	2.0970	94.51	0.5608	-0.0553	0.4026	95.36	0.2513
	FACI	-	-	95.75	0.5958	-	-	96.20	0.2638
	Q	-0.2964	2.0940	94.36	0.5606	-0.0551	0.4023	94.90	0.2481
	HT$_{0.08}$	-0.2233	2.1130	94.69	0.5690	-0.0241	0.4075	95.06	0.2513
	HT$_{0.20}$	-0.2475	2.1210	94.64	0.5671	-0.0361	0.4073	95.02	0.2506
	ST	-0.1996	2.0820	94.37	0.5602	-0.0121	0.4001	94.85	0.2480
3	SQ(RoT)	-0.1980	2.1030	94.58	0.5597	-0.0031	0.4028	95.19	0.2506
	SQ($\hat{\alpha}_0^*$)	-0.1962	2.1030	94.54	0.5599	-0.0032	0.4028	95.19	0.2506
	SQ($\hat{\alpha}^*$)	-0.1941	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$_{0.08}$	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

Table 3: Performances of Different Estimators for ψ_4

		Sample Size: 200				Sample Size: 1000			
Design		Bias	MSE	CP	IL	Bias	MSE	CP	IL
		$(\times 10^2)$	$(\times 10^2)$	(%)		$(\times 10^2)$	$(\times 10^2)$	(%)	
4	SQ(RoT)	−0.1906	2.1040	94.60	0.5597	−0.0023	0.4035	95.26	0.2507
	SQ($\hat{\alpha}_0^*$)	−0.1908	2.1030	94.56	0.5594	−0.0023	0.4034	95.26	0.2506
	SQ($\hat{\alpha}^*$)	−0.1898	2.1040	94.53	0.5586	−0.0028	0.4035	95.23	0.2503
	FACI	-	-	95.39	0.5826	-	-	95.59	0.2570
	Q	−0.1909	2.1020	94.40	0.5617	−0.0022	0.4032	94.83	0.2483
	HT$_{0.08}$	−0.1740	2.1410	94.68	0.5695	−0.0015	0.4099	94.92	0.2504
	HT$_{0.20}$	−0.2119	2.1320	94.54	0.5663	0.0035	0.4088	94.87	0.2497
	ST	0.1676	2.1590	94.50	0.5688	0.0685	0.4101	94.73	0.2499
5	SQ(RoT)	−0.1675	2.1030	94.40	0.5582	−0.0094	0.4034	95.07	0.2506
	SQ($\hat{\alpha}_0^*$)	−0.1668	2.1030	94.35	0.5579	−0.0096	0.4033	95.04	0.2505
	SQ($\hat{\alpha}^*$)	−0.1663	2.1040	94.34	0.5571	−0.0096	0.4034	94.99	0.2501
	FACI	-	-	95.07	0.5809	-	-	95.40	0.2569
	Q	−0.1677	2.1020	94.35	0.5601	−0.0094	0.4032	94.69	0.2481
	HT$_{0.08}$	−0.1732	2.0970	94.55	0.5634	−0.0086	0.4025	94.97	0.2491
	HT$_{0.20}$	−0.1911	2.1110	94.53	0.5629	−0.0037	0.4046	94.85	0.2490
	ST	−0.1856	2.0790	94.30	0.5582	−0.0082	0.3991	94.77	0.2472
6	SQ(RoT)	−0.2983	2.0950	94.51	0.5608	−0.0551	0.4024	95.34	0.2513
	SQ($\hat{\alpha}_0^*$)	−0.3010	2.0960	94.53	0.5614	−0.0549	0.4025	95.33	0.2515
	SQ($\hat{\alpha}^*$)	−0.3035	2.0970	94.49	0.5608	−0.0551	0.4026	95.32	0.2513
	FACI	-	-	95.71	0.5956	-	-	96.20	0.2637
	Q	−0.2975	2.0940	94.38	0.5606	−0.0551	0.4024	94.92	0.2481
	HT$_{0.08}$	−0.2086	2.1100	94.62	0.5685	−0.0230	0.4056	95.05	0.2509
	HT$_{0.20}$	−0.2535	2.1200	94.58	0.5668	−0.0320	0.4074	95.07	0.2504
	ST	−0.1995	2.0810	94.37	0.5600	−0.0116	0.3998	94.84	0.2478

Table 4: Performances of Different Estimators for ψ_4

Const.	NPI_2	Age	Gender	A_2	$NPI_2 * A_2$	$Age * A_2$	$Gender * A_2$
-3.938^{***}	-0.016^{***}	0.014	-0.120	3.904	0.004	-0.051^{*}	-0.874
(1.222)	(0.005)	(0.015)	(0.216)	(2.405)	(0.012)	(0.030)	(0.531)
Adj. R^2: 0.814 F statistic: 162.5^{***} (df=8; 288) $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$							

Table 5: Stage 2 Regression Result

	SQ(G, ROT)	SQ(G, $\hat{\alpha}_0^*$)	SQ(G, $\hat{\alpha}^*$)	FACI	Q	$HT_{0.08}$	$HT_{0.20}$	ST
Const.	-8.831	-8.842	-8.871	-8.83	-8.83	-10.504	-9.501	-10.763
	[-12.249,-5.413]	[-12.256,-5.429]	[-12.305,-5.437]	[-12.391,-4.733]	[-12.141,-5.616]	[-13.173,-6.416]	[-12.878,-6.007]	[-13.655,-7.358]
	6.836	6.827	6.868	7.658	6.525	6.757	6.871	6.298
NPI_1	-0.011	-0.011	-0.011	-0.011	-0.011	-0.012	-0.011	-0.012
	[-0.022,0]	[-0.022,0]	[-0.022,0]	[-0.022,0.003]	[-0.022,-0.001]	[-0.023,-0.002]	[-0.022,-0.001]	[-0.023,-0.002]
	0.022	0.022	0.022	0.025	0.021	0.021	0.021	0.021
Age	0.03	0.03	0.03	0.03	0.03	0.05	0.038	0.053
	[-0.012,0.072]	[-0.012,0.072]	[-0.012,0.073]	[-0.022,0.075]	[-0.01,0.072]	[-0.001,0.084]	[-0.006,0.078]	[0.01,0.088]
	0.084	0.085	0.085	0.097	0.081	0.085	0.084	0.078
Gender	-0.49	-0.49	-0.488	-0.492	-0.492	-0.304	-0.387	-0.287
	[-1.091,0.112]	[-1.091,0.112]	[-1.09,0.115]	[-1.476,0.204]	[-1.104,0.103]	[-1.036,0.246]	[-1.068,0.157]	[-0.992,0.331]
	1.203	1.203	1.205	1.680	1.207	1.282	1.224	1.323
A_1	2.538	2.53	2.529	2.541	2.541	2.956	2.234	3.216
	[-1.744,6.82]	[-1.754,6.815]	[-1.755,6.814]	[-2.468,7.677]	[-1.691,6.871]	[-1.652,7.199]	[-1.984,7.21]	[-1.16,7.279]
	8.564	8.569	8.569	10.145	8.562	8.851	9.194	8.438
$NPI_1 * A_1$	0.027	0.026	0.026	0.027	0.027	0.028	0.028	0.028
	[0.006,0.047]	[0.006,0.047]	[0.006,0.047]	[0.004,0.051]	[0.005,0.047]	[0.005,0.049]	[0.005,0.048]	[0.005,0.049]
	0.041	0.041	0.041	0.047	0.043	0.044	0.043	0.044
$Age * A_1$	-0.046	-0.046	-0.046	-0.046	-0.046	-0.052	-0.043	-0.056
	[-0.1,0.007]	[-0.099,0.007]	[-0.099,0.007]	[-0.11,0.013]	[-0.105,0.007]	[-0.11,0.008]	[-0.108,0.008]	[-0.111,0.001]
	0.107	0.107	0.107	0.123	0.112	0.118	0.116	0.111
$Gender * A_1$	-0.042	-0.042	-0.041	-0.042	-0.042	-0.042	0.011	-0.059
	[-0.836,0.752]	[-0.836,0.751]	[-0.836,0.754]	[-1.041,0.879]	[-0.843,0.911]	[-0.809,0.907]	[-0.894,0.916]	[-0.863,0.872]
	1.588	1.587	1.59	1.919	1.754	1.716	1.809	1.736

Table 6: Stage 1 Regression Result (for each covariate, line 1: point estimate; line 2: 95% CIs; line 3: lengths of CIs)

	Mean of NPI_1	Mean of NPI_2	Mean of Age	Stage 1	Stage 2
Female	38.7	31.5	78.6	0	1
Male	34.4	27.3	76.5	0	0

Table 7: Example of Treatment Prescription from Estimated Optimal DTR